International Journal of Statistics and Probability; Vol. 13, No. 1; March 2024
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Comparisons of the Satterthwaite Approaches for Fixed Effects in Linear Mixed Models

Waseem Alnosaier
Correspondence: Waseem Alnosaier, Department of Statistics, Institute of Public Administration, Riyadh, Saudi Arabia

Received: December 20, 2023 Accepted: February 23, 2024 Online Published: February 28, 2024
doi:10.5539/ijsp.v13n1p22 URL: https://doi.org/10.5539/ijsp.v13n1p22

Abstract

Four approximate F- tests derived by Fai and Cornelious in 1996 to make inference for fixed effects in mixed linear models of rank greater than one. Two of these approaches derived by introducing a Wald-type statistic distributed approximately as an F distribution, and the denominator degrees of freedom computed by matching the approximated one moment of the Wald-type statistic with the exact one moment of the F distribution. The other two approaches were derived by introducing a scaled Wald-type statistic to be distributed approximately as an F distribution, and the denominator degrees of freedom and the scale factor computed by matching the two moments of the statistic with the moments of the F distribution. This paper proposes two more approximate F-tests analogous to the four approaches where an adjusted estimator of the variance of the estimate of fixed effects used. In addition, the paper evaluates and compares the performance of the six approaches analytically, and some useful results are presented. Also, a simulation study for block designs was run to assess and compare the performance of the approaches based on their observed test levels. The simulation study shows that the approaches usually perform reasonably based on their test levels, and in some cases some approaches found to more adequately than other approaches.

Keywords: mixed linear model, Satterthwaite test, restricted maximum likelihood estimator, linear hypothesis, Variance components

1. Introduction

Data analysts and practicing statisticians usually encounter problems in making inference for the fixed effects in normal linear mixed models where the fixed effects are of interest and the random effects are a source of errors. When some observations are lost or not available for some reasons, the data become not balanced and the analysis becomes more complicated, where the conventional Anova table does not provide an exact test, and approximation is needed. For a data vector $\mathbf{y}$ distributed as multivariate normal distribution with mean $\mathbf{X\beta}$, and $\boldsymbol{\beta}$ is the vector 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 fixed and has ℓ independent columns. A Wald-type statistic of the form

$$Q = \hat{\beta}' L[L'\hat{V}ar(\hat{\beta})L]^{-1} L'\hat{\beta},$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{V}ar(\hat{\boldsymbol{\beta}})$ are estimators of $\boldsymbol{\beta}$ and $Var(\hat{\boldsymbol{\beta}})$ respectively, usually follows a chi square distribution with ℓ degrees of freedom for large samples. However, for small samples, this approximation might not be appropriate. Giesbrecht and Burns (1985) suggested a method to determine the degrees of freedom for an approximate t-test (GB test). Also, Jeske and Harville (1988) suggested a procedure to determine the degrees of freedom for an approximate t-test with adjusting the variance estimate of the fixed effects estimator $\hat{V}ar(\hat{\boldsymbol{\beta}})$ (JH test). Both procedures are limited for hypotheses of rank 1, and the degrees of freedom is estimated in a way analogous to Satterthwaite's approximation (Satterthwaite, 1946). Based on these two procedures, Fai and Cornelious (1996) proposed four approximate *F* tests for the hypotheses of rank greater than 1. In fact, they extended the GB and JH procedures by matching the first and second moments of the *F* distribution with the test statistic for each procedure. For procedures obtained by matching the first moment, a Wald-type *F* statistic used, and the denominator degrees of freedom is estimated based on the data. On the other hand, for procedures obtained by matching the second moment, a scaled *F* statistic used, and the denominator degrees of freedom and scale factor are estimated based on the data. The extended procedures of the GB test by matching the first and second moments will be called the GB1 and GB2 approaches respectively, and similarly, the extended procedures of the JH test will be called the JH1 and JH2 approaches. Fai and Cornelious compared the performance of the four proposed procedures through a simulation study for split plot designs. They concluded that the GB1 and JH1 approaches performed similarly and adequately more than other approaches which found to be more liberal. The popular procedure called the Satterthwaite procedure to approximate the denominator degrees of freedom, and embedded in most statistical software packages is actually the GB1 procedure. This test might be executed in SAS by using the option DDFM=SATTERTHWAITE in the model statement of the PROC MIXED procedure (SAS Institute Inc, 2021).

Many researchers evaluated the performance of the GB1 approach and compared it to other known approaches (e.g., Luke, 2017; Monar & Zucker, 2004; Kuznetsova et al., 2017; Spilke, 2005; Schaalje et al., 2002). However, the other forms of tests derived by Fai and Cornelious, except the GB1 procedure, have not received enough attention in the research literature neither analytically nor through simulation studies. In this paper, we investigated the inner working of these four procedures in order to compare the performance of these procedures analytically and through a simulation study for block designs different from the design studied by Fai and Cornelious (1996). In addition, we proposed two more procedures analogous to procedures JH1 and JH2 with using the adjusted variance-covariance of the estimate of fixed effects parameters derived by Harville and Jeske (1992).

In Section 2, and in addition to the model and notation, we review the four approaches developed by Fai and Corenelius (1996). Also, we present two more approaches based on the adjustment of the variance-covariance matrix of the fixed effects proposed by Harville and Jeske (1992). These two approaches will be called HJ1 and HJ2 approaches. Comparisons of the six approaches (i.e., the GB1, GB2, JH1, JH2, HJ1, and HJ2 approaches) based on their test statistics, computed denominator degrees of freedom and scales is discussed analytically in Section 3. In Section 4, we present a simulation study for three block designs to evaluate and compare the performance of the six approaches based on the observed test levels. The computed denominator degrees of freedom and scales are presented and compared for each approach.

2. The Satterthwaite Approaches to Test the Fixed Effects in Linear Mixed Models

Consider n observations y follow a multivariate normal distribution with mean $\mathbf{X\beta}$, and variance covariance $\mathbf{\Sigma}$. The matrix $\mathbf{X}$ is a known $n\times p$ matrix, and $\boldsymbol{\beta}$ is a $p\times 1$ vector of the fixed effects parameters, and $\mathbf{\Sigma}$ is a positive definite $n\times n$ matrix, and linear of r unknown variance components $\mathbf{\Theta}$, that is $\mathbf{\Sigma}=\theta_1\mathbf{G}_1+\cdots+\theta_r\mathbf{G}_r$ for $\mathbf{G}_i\,'s$ are known $n\times n$ symmetric matrices. 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.

Fai and Cornelious (1996) proposed four approximate F-tests. The first two tests are extensions for the GB test (Giesbrecht & Burns, 1985) to accommodate the hypotheses of fixed effects for rank greater than 1. The first test, called GB1, derived by introducing a Wald-type statistic follows approximately an F distribution, and the denominator degrees of freedom of the F distribution is computed by matching the exact first moment of the F distribution and the approximated first moment of the Wald-type test statistic. For the second test, called GB2, a scaled Wald-type statistic follows approximately an F distribution, and the denominator degrees of freedom and scale factor computed by matching the exact two moments of the F distribution with the approximated two moments of the scaled Wald-type test statistic. Similarly, the third and fourth tests, called JH1 and JH2, are extensions of the JH test (Jeske & Harville, 1988), and the denominator degrees of freedom and scale factor computed by using the one moment and two moments approximations respectively.

2.1 Review of the GB1 and GB2 Approaches

Fai and Cornelious (1996) extended the GB method for testing hypotheses of rank greater than one. They started with the sum of squares for H_0 as:

$$T=\hat{\beta}'L[L'\hat{V}ar(\hat{\beta})L]^{-1}L'\hat{\beta} \tag{2.1}$$

where $\hat{\boldsymbol{\beta}}$ is the estimated generalized least squares estimator of the fixed effects $\boldsymbol{\beta}$, which is

$$\hat{\beta}=(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y \tag{2.2}$$

and $\hat{\Sigma}=\hat{\theta}_1G_1+\cdots+\hat{\theta}_rG_r$, where $\hat{\theta}_i\,'s$ are the REML estimates of the variance components $\theta_i\,'s$ (Corbiel & Searle, 1976). In the GB method, the matrix $\hat{\Phi}=X'\hat{\Sigma}^{-1}X$ is used as the estimator of $Var(\hat{\boldsymbol{\beta}})$.

That is $\hat{V}ar(\hat{\beta})=\hat{\Phi}=X'\hat{\Sigma}^{-1}X$,

and

$$T_{GB}=\hat{\beta}'L[L'\hat{\Phi}L]^{-1}L'\hat{\beta} \tag{2.3}$$

Under H_0, $F_{GB1}=T_{GB}/\ell\sim F(\ell,\nu_{GB1})$ approximately, and we can choose ν_{GB1} by using the one-moment approximation, where the approximate first moment of F_{GB1} matched with the exact first moment of $F(\ell,\nu_{GB1})$ to obtain:

$$E(T_{GB}) = \frac{\ell \nu_{GB1}}{\nu_{GB1} - 2}, \text{ provided } \nu_{GB1} > 2 \tag{2.4a}$$

and

$$\nu_{GB1} = \frac{2E(T_{GB})}{E(T_{GB}) - \ell}, \text{ provided } E(T_{GB}) > \ell. \tag{2.4b}$$

The approximation of $E(T_{GB})$ is obtained by performing the spectral decomposition of $\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L}$, where $\mathbf{L}'\hat{\mathbf{\Phi}}\mathbf{L} = \mathbf{P}'\mathbf{D}\mathbf{P}$ for $\mathbf{P}$ an orthogonal matrix, $\mathbf{D}$ is a diagonal matrix of eigenvalues d_m's and $d_m > 0$ for $m = 1, \cdots, \ell$. The quantity T_{GB} in (2.3) above can be expressed as

$$T_{GB} = \hat{\beta}' LPP'[L'\hat{\Phi}L]^{-1} PP'L'\hat{\beta}$$

$$= \hat{\beta}' LP[^{P} P'L'\hat{\beta}$$

$$= \hat{\beta}' LPD^{-1} P'L'\hat{\beta}$$

$$= \sum_{m=1}^{\ell} \frac{[(\mathbf{P}'\mathbf{L}'\hat{\boldsymbol{\beta}})_m]^2}{d_m} = \sum_{m=1}^{\ell} t_{\nu_m}^2 \tag{2.5}$$

Supposing that $t_{\nu_m}^2$'s are distributed approximately independently as t-distribution with degrees of freedom ν_m for each, we obtain:

$$E(T_{GB}) = \sum_{m=1}^{\ell} E(t_{\nu_m}^2) = \sum_{m=1}^{\ell} E(F_{1,\nu_m}) = \sum_{m=1}^{\ell} \frac{\nu_m}{\nu_m - 2}, \text{ provided } \nu_m > 2 \tag{2.6}$$

and

$$\nu_m \approx \frac{2(d_m)^2}{\mathbf{g}'_m \mathbf{W} \mathbf{g}_m} \tag{2.7}$$

where $\mathbf{g}_m$ is the gradient of $\mathbf{a}_m \hat{\mathbf{\Phi}} \mathbf{a}'_m$ with respect to $\mathbf{\theta}$, $\mathbf{a}_m$ is the m^{th} row of $\mathbf{PL}'$, and $\mathbf{W} = [w_{ij}]_{r\times r}$ is the variance-covariance matrix of the REML estimators of the variance components $\hat{\theta}$ which can be approximated by the inverse of the expected information matrix $\tilde{\mathbf{W}}$.

For extending the GB method by using the two-moment approximation, two quantities need to be computed: the scale λ_{GB2} and the denominator degrees of freedom ν_{GB2} such that a scaled F statistic $F_{GB2} = \lambda_{GB2} T_{GB} / \ell \sim F(\ell, \nu_{GB2})$ approximately under H_0. By matching the approximated two moments of F_{GB2} and the exact two moments of $F(\ell, \nu_{GB2})$ we obtain:

$$\lambda_{GB2} E(T_{GB}) = \frac{\ell \nu_{GB2}}{\nu_{GB2} - 2},$$

and

$$\frac{\lambda_{GB2}^2}{\ell^2} Var(T_{GB}) = \frac{2\nu_{GB2}^2 (\nu_{GB2} + \ell - 2)}{\ell (\nu_{GB2} - 2)^2 (\nu_{GB2} - 4)}.$$

Hence, we have

$$\nu_{GB2} = 4 + \frac{2[E(T_{GB})]^2 (\ell + 2)}{\ell Var(T_{GB}) - 2[E(T_{GB})]^2}, \text{ provided } \ell Var(T_{GB}) - 2[E(T_{GB})]^2 > 0 \tag{2.8a}$$

and

$$\lambda_{GB2} = \frac{\ell \nu_{GB2}}{E(T_{GB})(\nu_{GB2} - 2)}, \tag{2.8b}$$

where $Var(T_{GB})$ is obtained from expression (2.5):

$$Var(T_{GB}) = \sum_{m=1}^{\ell} Var(t_{\nu_m}^2) = \sum_{m=1}^{\ell} Var(F_{1,\nu_m}) = \sum_{m=1}^{\ell} \frac{2\nu_m^2(\nu_m - 1)}{(\nu_m - 2)^2(\nu_m - 4)}, \text{ provided } \nu_m > 4.$$

$E(T_{GB})$ is as in (2.6), and ν_m is approximated as in (2.7).

2.2 Review of the JH1 and JH2 Approaches

Since the traditional estimator of $Var(\hat{\boldsymbol{\beta}})$, matrix $\hat{\boldsymbol{\Phi}}$, tends to underestimate, Jeske and Harville (1988) adjusted the estimator using the adjustment suggested by Kackar and Harville (1984), which is

$$Var(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Phi} + \boldsymbol{\Lambda},$$

where

$$\boldsymbol{\Lambda} \approx \tilde{\boldsymbol{\Lambda}} = \boldsymbol{\Phi}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j)\right\}\boldsymbol{\Phi} \tag{2.9}$$

and

$$P_i = X'\frac{\partial\Sigma^{-1}}{\partial\theta_i}X, \text{ and } Q_{ij} = X'\frac{\partial\Sigma^{-1}}{\partial\theta_i}\Sigma\frac{\partial\Sigma^{-1}}{\partial\theta_j}X. \tag{2.10}$$

That is

$$\hat{V}ar(\hat{\beta}) = \hat{\Phi}_{JH}(\hat{\theta}) = \hat{\Phi}_{JH} = \hat{\Phi} + \hat{\Lambda}, \tag{2.11}$$

where

$$\hat{\Lambda} = \hat{\Phi}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r}\hat{w}_{ij}(\hat{Q}_{ij} - \hat{P}_i\hat{\Phi}\hat{P}_j)\right\}\hat{\Phi} \tag{2.12}$$

These quantities to be estimated by substituting the REML estimates of the variance components $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$. Similar to the extension of the GB approach, Fai and Cornelious (1996) started the extension of JH approach with Wald-type statistic:

$$T_{JH} = \hat{\beta}'L[L'\hat{\Phi}_{JH}L]^{-1}L'\hat{\beta} \tag{2.13}$$

In the first extension, that is the JH1 approach, we have $F_{JH1} = T_{JH}/\ell \sim F(\ell, \nu_{JH1})$ approximately under H_0, and we compute ν_{JH1} by using the one-moment approximation where the approximate first moment of F_{JH1} matched with the exact moment of $F(\ell, \nu_{JH1})$. That is

$$E(T_{JH}) = \frac{\ell\,\nu_{JH1}}{\nu_{JH1} - 2}, \text{ provided } \nu_{JH1} > 2 \tag{2.14a}$$

and

$$\nu_{JH1} = \frac{2E(T_{JH})}{E(T_{JH}) - \ell}. \tag{2.14b}$$

Similar to the extension of the GB approach, $E(T_{JH})$ is approximated by performing the spectral decomposition $\mathbf{L}'\hat{\boldsymbol{\Phi}}_{JH}\mathbf{L} = \mathbf{P}'\mathbf{D}\mathbf{P}$ for $\mathbf{P}$ an orthogonal matrix, $\mathbf{D}$ is a diagonal matrix of eigenvalues d_s's and $d_s > 0$ for $s = 1, \cdots, \ell$.

$$T_{jH} = \sum_{s=1}^{\ell}\frac{[(\mathbf{P}'\mathbf{L}'\hat{\boldsymbol{\beta}})_s]^2}{d_s} = \sum_{s=1}^{\ell} t_{\nu_s}^2, \tag{2.15}$$

$$E(T_{JH}) = \sum_{s=1}^{\ell}\frac{\nu_s}{\nu_s - 2}, \text{ provided } \nu_s > 2, \tag{2.16}$$

and $\nu_s \approx \dfrac{2(d_s)^2}{\mathbf{g}_s'\mathbf{W}\mathbf{g}_s}$ where $\mathbf{g}_s$ is the gradient of $\mathbf{a}_s\hat{\boldsymbol{\Phi}}_{JH}\mathbf{a}_s'$ with respect to $\boldsymbol{\theta}$, and $\mathbf{a}_s$ is the s^th^ row of $\mathbf{PL}'$.

For the two-moment approximation, and analogous to the GB2 approach, the scale λ_{JH2} and ν_{JH2} are computed such that $F_{JH2} = \lambda_{JH2}T_{JH}/\ell \sim F(\ell, \nu_{JH2})$ approximately under H_0.

$$\nu_{JH2} = 4 + \frac{2[E(T_{JH})]^2(\ell + 2)}{\ell Var(T_{JH}) - 2[E(T_{JH})]^2}, \tag{2.17a}$$

and

$$\lambda_{JH2} = \frac{\ell \nu_{JH2}}{E(T_{JH})(\nu_{JH2}-2)}, \tag{2.17b}$$

where $Var(T_{JH}) = \sum_{j=1}^{\ell} \frac{2\nu_j^2(\nu_j - 1)}{(\nu_j - 2)^2(\nu_j - 4)}$, provided $\nu_j > 4.$

2.3 Two Proposed Approximation Approaches

The estimator for $Var(\hat{\boldsymbol{\beta}})$ used in JH approximations, $\hat{\boldsymbol{\Phi}}_{JH}$, is a better estimator than $\hat{\boldsymbol{\Phi}}$ which tends to underestimate. However, $\hat{\boldsymbol{\Phi}}_{JH}$ still tends to underestimate. In these proposed approaches, we use a better estimator for $Var(\hat{\boldsymbol{\beta}})$ suggested by Harville and Jeske (1992), and used by Kenward and Roger in their approach known as the KR estimation (Kenward & Roger, 1997) (See remark 1 below). That is:

$$Var(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Phi} + 2\boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda}$ is approximated as in expression (2.9).

Hence

$$\hat{V}ar(\hat{\beta}) = \hat{\Phi}_{HJ}(\hat{\theta}) = \hat{\Phi}_{HJ} = \hat{\Phi} + 2\hat{\Lambda} \tag{2.18}$$

where $\hat{\boldsymbol{\Lambda}}$ is as in expression (2.12).

Analogous to the JH1, and JH2 approaches, consider:

$$T_{HJ} = \hat{\beta}' L[L'\hat{\Phi}_{HJ} L]^{-1} L'\hat{\beta}, \tag{2.19}$$

and two approaches developed by using the one-moment and two-moment approximations. The first approximation, which will be called JH1, is obtained by using the one-moment approximation, where we have $F_{HJ1} = T_{HJ}/\ell \sim F(\ell, \nu_{HJ1})$ approximately under H_0, and ν_{HJ1} is obtained by matching the approximate first moment of F_{HJ1} with the exact moment of $F(\ell, \nu_{HJ1})$.

$$E(T_{HJ}) = \frac{\ell \nu_{HJ1}}{\nu_{HJ1} - 2}, \text{ provided } \nu_{HJ1} > 2 \tag{2.20}$$

and

$$\nu_{HJ1} = \frac{2E(T_{HJ})}{E(T_{HJ}) - \ell}. \tag{2.21}$$

Also, similar to the extensions of the GB and JH approaches, the approximation of $E(T_{HJ})$ is obtained by using the spectral decomposition $\mathbf{L}'\hat{\boldsymbol{\Phi}}_{HJ}\mathbf{L} = \mathbf{P}'\mathbf{D}\mathbf{P}$ for $\mathbf{P}$ an orthogonal matrix, $\mathbf{D}$ is a diagonal matrix of eigenvalues d_k's and $d_k > 0$ for $k = 1, \cdots, \ell$.

$$E(T_{HJ}) = \sum_{k=1}^{\ell} \frac{\nu_k}{\nu_k - 2}, \text{ provided } \nu_k > 2, \tag{2.22}$$

and $\nu_k \approx \frac{2(d_k)^2}{\mathbf{g}_k' \mathbf{W} \mathbf{g}_k}$, where $\mathbf{g}_k$ is the gradient of $\mathbf{a}_k \hat{\boldsymbol{\Phi}}_{HJ} \mathbf{a}_k'$ with respect to $\boldsymbol{\theta}$, and $\mathbf{a}_k$ is the k^{th} row of $\mathbf{PL}'$.

For the two-moment approximation, $F_{HJ2} = \lambda_{HJ2} T_{HJ}/\ell \sim F(\ell, \nu_{HJ2})$ approximately under H_0, and ν_{HJ2} and λ_{HJ2} are calculated as

$$\nu_{HJ2} = 4 + \frac{2[E(T_{HJ})]^2(\ell+2)}{\ell Var(T_{HJ}) - 2[E(T_{HJ})]^2}, \tag{2.23a}$$

$$\lambda_{HJ2} = \frac{\ell \nu_{HJ2}}{E(T_{HJ})(\nu_{HJ2} - 2)}, \tag{2.23b}$$

and

$$Var(T_{HJ}) = \sum_{k=1}^{\ell} \frac{2\nu_k^2(\nu_k - 1)}{(\nu_k - 2)^2(\nu_k - 4)}, \text{ provided } \nu_k > 4.$$

Remark 1 The variance matrix of the fixed effects estimates used in the proposed Wald-type statistic is the same as in the Wald-type Kenward-Roger test. In fact, in some special cases, both tests might deliver same values for the denominator degrees of freedom and scale factor. However, in general, the tests are different regarding the obtained denominator degrees of freedom and scale. The Kenward-Roger test uses the Taylor series expansion method to obtain an approximate moment for the test statistic, whereas the spectral decomposition method is used in the Satterthwaite based approaches.

3. Comparisons of the Satterthwaite Approaches

All approaches presented in section 2 use similar Wald-type statistics, and the differences among them may be described based on two sources: first, the estimator of $Var(\hat{\boldsymbol{\beta}})$ used in the statistics, where it is $\widehat{\Phi}, \widehat{\Phi}_{JH}$, and $\widehat{\Phi}_{HJ}$ for the extension of the GB, JH, and HJ approaches respectively. Second, the GB1, JH1, and HJ1 approaches were developed based on using the one-moment approximation where the denominator degrees of freedom are computed, whereas the GB2, JH2, and HJ2 approaches developed based on the two-moment approximation, where a scaled test statistic used, and both the denominator degrees of freedom and the scale factor are computed. The performance of these approaches typically depends on the test statistic (including the obtained scale) and the denominator degrees of freedom. Accordingly, besides evaluating and comparing the performance of the approaches through simulation studies, it is reasonable and useful to compare their test statistics, computed denominator degrees of freedom, and scale analytically. In this section, we present some useful results about the approaches' statistics, computed denominator degrees of freedom and cale.

3.1 General Formulas for the Satterthwaite Approaches

It can be seen that the denominator degrees of freedom and scales produced by approaches presented in Section 2 follow general formulas. For the test statistic T as in (2.1), and for the one-moment approximation, we have $F = \lambda T/\ell \sim F(\ell, \nu_{one})$ approximately under H_0. To find ν_{one}, the approximate first moment of F statistic is matched with the exact first moment of $F(\ell, \nu_{one})$, and we have:

$$E(T) = \frac{\ell\,\nu_{one}}{\nu_{one} - 2}, \text{ provided } \nu_{one} > 2 \tag{3.1a}$$

and

$$\nu_{one} = \frac{2E(T)}{E(T) - \ell}, \quad \text{provided} \quad E(T) > \ell \tag{3.1b}$$

$E(T)$ is approximated by performing the spectral decomposition $\mathbf{L}'\hat{V}ar(\boldsymbol{\beta})\mathbf{L} = \mathbf{P}'\mathbf{D}\mathbf{P}$ for $\mathbf{P}$ an orthogonal matrix, $\mathbf{D}$ is a diagonal matrix of eigenvalues d_q's and $d_q > 0$ for $q = 1, \cdots, \ell,$ where

$$E(T) = \sum_{q=1}^{\ell} \frac{\nu_q}{\nu_q - 2}, \text{ provided } \nu_q > 2, \tag{3.2}$$

and

$$\nu_q \approx \frac{2(d_q)^2}{\mathbf{g}'_q\, \mathbf{W} \mathbf{g}_q}. \tag{3.3}$$

The quantity $\mathbf{g}_q$ is the gradient of $\mathbf{a}_q Var(\hat{\boldsymbol{\beta}})\mathbf{a}'_q$ with respect to $\boldsymbol{\theta}$, and $\mathbf{a}_q$ is the q^{th} row of $\mathbf{PL}'$. Note that $T = T_{GB}, T_{JH}$, and T_{HJ} for the extension of the GB, JH, and HJ approaches respectively.

For the two-moment approximation, the scale λ and the denominator degrees of freedom ν_{two} are

$$\nu_{two} = 4 + \frac{2[E(T)]^2(\ell + 2)}{\ell Var(T) - 2[E(T)]^2}, \text{ provided } \ell Var(T) - 2[E(T)]^2 > 0 \tag{3.4a}$$

and

$$\lambda = \frac{\ell\,\nu_{two}}{E(T)(\nu_{two} - 2)}, \tag{3.4b}$$

where

$$Var(T) = \sum_{q=1}^{\ell} \frac{2\nu_q^2(\nu_q - 1)}{(\nu_q - 2)^2(\nu_q - 4)}, \text{ provided } \nu_q > 4. \tag{3.4c}$$

3.2 Comparisons of the Test Statistics, Computed Denominator Degrees of Freedom, and Scales

From expressions (3.1b) and (3.4b), we obtain:

$$\lambda\left(\frac{\nu_{one}}{\nu_{one}-2}\right)=\frac{\nu_{two}}{\nu_{two}-2}, \tag{3.5}$$

where ν_{one} is the denominator degrees of freedom obtained by using the one-moment approximation, ν_{two} and λ are the computed denominator degrees of freedom and scale obtained using the two-moment approximation. Since typically $\lambda \le 1$, and the function $f(\nu)=\nu/(\nu-2)$ for $\nu > 2$ is a decreasing function, it is reasonable to expect that typically $\nu_{one} \le \nu_{two}$ (see remark 2 below). On the other hand, and provided $\lambda \le 1$, we have $F_1 = T/\ell \ge F_2 = \lambda F_1$, where F_1 and F_2 are the test statistics used in the one-moment and two-moment approximations respectively for the extensions of the same approach (i.e., the GB, JH, HJ approaches). In addition, comparing the test statistics for different approaches using the one-moment approximation is analytically possible by adopting the definition of Loewner order of symmetric matrices (Pukelshiem, 2006). Since $\mathbf{\Lambda}$ is a non-negative matrix, we have:

$$\widehat{\Phi}_{HJ} = \widehat{\Phi} + 2\hat{\Lambda} \ge \widehat{\Phi}_{JH} = \widehat{\Phi} + \hat{\Lambda} \ge \widehat{\Phi},$$

and form Abadir and Magnus (2005), we obtain

$$(L'\widehat{\Phi}_A L)^{-1} \le (L'\widehat{\Phi}_{JH} L)^{-1} \le (L'\widehat{\Phi} L)^{-1}$$

Hence

$$F_{HJ1} \le F_{JH1} \le F_{GB1}.$$

For the case $\ell = 1$, it should be true that both extensions for each approach; using the one-moment and two-moment approximations, are identical and they are equivalent to the original approach (i.e., the GB and JH approaches). The following lemma is to verify this fact.

Lemma 1 when $\ell = 1$, then $\nu_{one} = \nu_{two} = \nu_1$, and $\lambda = 1$.

Remark 2 From the expression of the scale (3.4b) above, and provided that $E(T) > \ell$, the simulation studies show that typically $\lambda \le 1$. However, the expression formula does not guarantee that the scale λ doesn't exceed 1. In fact, this occurs most likely when the computed denominator degrees of freedom using the second moment approximation, ν_2 gets closer to the value of 4. This happens because of the data character.

Consider the design matrix of the model $\mathbf{X}$ is partitioned as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ where:

$$\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X} = \begin{bmatrix} \mathbf{X}_1'\mathbf{\Sigma}^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{\Sigma}^{-1}\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{X}_2 \end{bmatrix}$$

Suppose $\mathbf{X}_1'\mathbf{\Sigma}^{-1}\mathbf{X}_2 = \mathbf{0}$, and $\mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{X}_2 = f(\boldsymbol{\theta})\mathbf{A}$, where f is a scalar function, and $\mathbf{A}$ is a fixed matrix (i.e., doesn't depend on $\boldsymbol{\theta}$). Also, suppose that $\mathbf{X}_2$ is the correspondent matrix to the fixed effects to be tested, so that $E(\mathbf{y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$, and $\mathbf{L}'\boldsymbol{\beta} = [\mathbf{0} \quad \mathbf{B}'][\boldsymbol{\beta}_2 \quad \boldsymbol{\beta}_2]' = \mathbf{B}'\boldsymbol{\beta}_2$. Some well-known designs follow this model partition (e.g., the balanced incomplete block designs), and this model partition has been noticed to have specific properties for several approximate tests that use a Wald-type statistic. Alnosaier and Birkes (2019) studied this partition model for the Kenward-Roger approximate test and other alternative tests.

Lemma 2 For models that satisfy the partition as above, the approximations of ν_q as in expression (3.3) are the same for all $q = 1, \cdots, \ell$.

Lemma 3 (a) For models that satisfy the partition above, the denominator degrees of freedom using the two-moment approximation, ν_{two} is a linear function of the denominator degrees of freedom using the one-moment approximation, ν_{one}.

(b) $\nu_{two} \ge \nu_{one}$, and the scale $\lambda \le 1$.

If a model satisfies the condition that $\mathbf{P_X\Sigma} = \mathbf{\Sigma P_X}$ for all $\mathbf{\Sigma}$, where $\mathbf{P_X}$ is an orthogonal projection operator on the range of $\mathbf{X}$, it is said that the model satisfies Zyskind's condition (Zyskind, 1967). For models that satisfy Zyskind's condition, we add an assumption that met by most models which is $\mathbf{P_X G}_i = \mathbf{G}_i\mathbf{P_X}$ for all *i*, where $\mathbf{G}_i$ as in Section 2.

Lemma 4 (a) For models that satisfy Zyskind's condition and the assumption above, the GB1, JH1 and HJ1 approaches are identical, and the GB2, JH2 and HJ2 approaches are identical.

(b) When $\ell = 1$, and the model satisfies Zyskind's condition, the GB1, GB2, JH1, JH2, HJ1, and HJ2 approaches are identical.

Detailed proofs of lemmas 1-4 are in the Appendix.

4. A Simulation Study

To evaluate and compare the performance of the Satterthwaite approaches presented in this paper, we conducted a simulation study for 15 models with three different block designs (e.g., five models for each design).

Design 1: A balanced incomplete block design with 52 observations, 13 treatments, 13 blocks, and the maximum block size is 4, obtained from (Cochran & Cox, 1992, p. 448).

Design 2: A complete block design with 36 observations, 6 treatments, 6 blocks, and deleting two observations from different blocks and different treatments.

Design 3: A partial incomplete block design with 21 observations, 9 treatments, 7 blocks, and the maximum block size is 3, (Kuehl, 2000, p. 329 with omitting the last two blocks).

The model for the block designs can be written as

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

where μ is the mean, τ_i is the treatment effects, b_j is the random block effects, e_{ijk} is the residual errors, $i = 1,...,t,\ j = 1,...,s,\ k = 1,\dots,n_{ij}$ and all n_{ij} are 0 or 1. The b_j and e_{ijk} are all independent, $b_j \sim \mathrm{N}(0,\sigma_b^2),$ and $e_{ijk} \sim \mathrm{N}(0,\sigma_e^2).$ In order to have a full column rank matrix of the fixed effects, we reparametrize the model as : $\tau_i^* = \tau_i - \tau_t$ for $i = 1,...,t-1,$ and $\mu^* = \mu + \tau_t$. The reparametrized model can be expressed as $y_{ijk} = \mu^* + \tau_i^* + b_j + e_{ijk}$, for $i = 1,...,t-1,$ and $y_{tjk} = \mu^* - \tau_1^* - \cdots \tau_{t-1}^* + b_j + e_{tjk}$ for $i = t$. The null hypothesis to test that there are no treatment effects is $\mathrm{H}_0 : \tau_1^* = \cdots = \tau_{t-1}^* = 0$.

For each design, 10,000 data sets have been simulated under the null distribution with no treatment effects, and assuming $\mu^* = 0$, for five settings of the ratio $\rho = \sigma_b / \sigma_e$: 0.25, 0.5, 1, 2, 4. The simulation was done by generating the random terms in the model (i.e., the blocks and residual errors) for each setting of ρ using Matlab 2020.

To run the simulation study for the Satterthwaite approaches, the quantities that needed to be computed have their expressions presented in the previous sections. However, some expressions need to be derived to be in computable forms (See the Appendix).

4.1 Computing the Denominator Degrees of Freedom and Scales

The iteration algorithm in expression 4.1in the Appendix to compute the REML estimates of the variance components did not converge for some data sets. Table 1 presents the percentage of the generated data sets that converge for each model under the null hypothesis (rounded to two decimal places). The percentage increased as the ratio ρ increased, and almost all of data sets for models with designs 1 and 2 converged. For models with the smallest design (i.e., design 3), it was noted that there were significant number of data sets that did not converge for smaller values of ρ.

Table 1. The percentage of data sets for which the iteration algorithm converges to compute the REML estimates of the variance components under the null hypothesis

ρ	Design 1	Design 2	Design 3
0.25	99.51	99.84	89.86
0.50	99.94	99.98	93.34
1.00	100.00	100.00	97.95
2.00	100.00	100.00	99.80
4.00	100.00	100.00	100.00

In addition, and to compute the denominator degrees of freedom for each approach, we limited the simulation to data sets which met the conditions as expressed in the general formulas (3.1b) and (3.4a) in Section 3. That is, to compute ν_{one}, we considered data sets satisfied $E(T) > \ell,$ and to compute ν_{two}, we also applied the condition $\ell Var(T) - 2[E(T)]^2 > 0.$ These conditions imply that $\nu_{one} > 2,$ and $\nu_{two} > 4.$ Table 2 presents the percentage of generated data sets satisfied these required conditions for each approach and each model under the null distribution. For models with larger designs (i.e., designs 1 and 2), it was noted that almost all data sets met the conditions for the GB1 and GB2 approaches. However, significant number of data sets did not meet the conditions with smaller values of ρ for the other approaches. It was also noted that more data sets did not meet the conditions for models with the smallest design (i.e., design 3), and especially with small values of ρ. Only 14.02% of data sets met the conditions of the HJ2 approach for the model with design 3, and $\rho = 1$. In fact, except the GB1 and GB2 approaches, most data sets did not meet the conditions for $\rho \leq 1$ to compute the denominator degrees of freedom. Unlike the PROC MIXED procedure of SAS where the denominator degrees of freedom ν_{one} set as zero when the condition is not met, for the simulation purpose of evaluating the performance of the approaches, we considered only data sets met the conditions required to compute the denominator degrees of freedom.

Table 4 in the Appendix shows the average of the computed denominator degrees of freedom for the Satterthwaite approaches. It can be clearly seen from table 4 that the average of computed denominator degrees of freedom using the one-moment approximation is less than the average of computed denominator degrees of freedom using the two-moment approximation for all six approaches. This coincides with our expectation in Section 3.2. Table 5 in the Appendix shows the average of scales of all approaches using the two-moment approximation, and they found to be less than one. However, this does not mean necessarily that the scales were less than one for all data sets. In fact, it was found that the scale for the GB2, JH2, and HJ2 approaches exceeded one for few data sets. For the smallest design, the scale values computed by the approaches using the two-moment approximation, were not very stable, where more data sets produced scales exceeded one, and other data sets produced scales significantly smaller than usual (e.g. 0.2 and 0.3). This explains why the average scales for design 3 were smaller than for the other designs. For the models with design 1, it was found that all produced scales for the GB2, JH2, and HJ2 approaches with all values of ρ did not exceed one, and this coincides with Lemma 3 (part b) in Section 3 since the balanced incomplete block design satisfies the model partition of lemma 3.

Table 2. The percentage of generated data sets that for which the iteration algorithm converged, and met the conditions to compute the denominator degrees of freedom under the null hypothesis

	ρ	Rate of data sets to compute degrees of freedom					
		GB1	GB2	JH1	JH2	HJ1	HJ2
Design 1	0.25	99.51	99.21	85.02	82.95	81.19	78.48
	0.50	99.94	99.90	89.12	87.48	86.39	84.51
	1.00	100.00	100.00	98.69	98.42	98.20	97.80
	2.00	100.00	100.00	99.99	99.99	99.99	99.99
	4.00	100.00	100.00	100.00	100.00	100.00	100.00
Design 2	0.25	99.84	99.84	87.11	84.82	84.68	81.92
	0.50	99.98	99.98	90.68	87.82	88.94	85.86
	1.00	100.00	100.00	97.94	97.22	97.43	96.64
	2.00	100.00	100.00	99.87	99.83	99.83	99.80
	4.00	100.00	100.00	99.99	99.99	99.99	99.99
Design 3	0.25	77.89	72.49	55.48	49.57	52.31	41.78
	0.50	85.08	81.15	52.21	35.04	44.57	24.29
	1.00	95.40	93.93	49.10	25.88	36.30	14.02
	2.00	99.53	99.44	92.24	83.02	86.80	71.48
	4.00	99.96	99.95	99.71	99.10	99.36	98.14

4.2 Comparisons of the Observed Test Levels

The performance of all six approaches was evaluated and compared based on the computed test levels of the approach. The observed test level was computed by the proportion of generated data sets under the null hypothesis, converged to compute the REML estimates of the variance components, and satisfied the required conditions to compute the denominator degrees of freedom of which $F_1 > F(\ell, \nu_{one})$ for approaches using the one-moment approximation, and of which $F_2 > F(\ell, \nu_{two})$ for approaches using the two-moment approximation, as these quantities described in Section 3. Typically, this proportion is desirable to be close to the nominal level which chosen to be 0.05. However, it is not expected that the observed test level to be exactly 0.05 due to the simulation error which is $22\% / \sqrt{n}$ where n is the number of generated data sets, converged to produce the REML estimates of the variance components, and met the conditions to produce the denominator degrees of freedom as explained in Section 4.1. Also, in comparing the observed test levels, the difference of 0.2% was considered unimportant.

As presented in table 3, it is notable that the test level for approaches using the two-moment approximation were more liberal than for approaches using the one-moment approximation for all models. This was expected because the computed denominator degrees of freedom tended to be larger for approaches using the two-moment approximation as seen in Section 4.1. Apparently, the effect of the computed scale factors on the test level was minimal comparing to the computed denominator degrees of freedom. For models with the largest design(i.e., design 1), the JH1 approach performed well for all values of ρ, whereas the GB1 and HJ1 approaches performed well only for larger values of ρ. The GB1 approach found to be so liberal, and the HJ1 approach found to be excessively conservative for small values of ρ, and this is due to their average of computed denominator degrees of freedom. For example, for $\rho = 0.25$, the average computed denominator degrees of freedom for the GB1, JH1, and HJ1 approaches are 34.61, 27.94, and 24.41 respectively. The approaches GB2, JH2, and HJ2 were found to be liberal for all values of ρ, however, the HJ2 approach performed more adequately and less liberal than others. For models with design 2, the GB1, JH1, and HJ1 approaches performed almost the same, and they performed adequately. They found to be slightly conservative for $\rho = 1$ and 2. The GB2, JH2, and HJ2 performed similarly and found to be liberal for all values of ρ. For models with the smallest design (i.e., design 3), and for small values of ρ, the GB1 approach was so liberal, and the HJ1 approach seemed to perform slightly conservative. The JH1 approach was found to outperformed other approaches for small values of ρ. However, the majority of generated data did not meet the conditions to compute the denominator degrees of freedom for the JH1 and HJ1 approaches for small values of ρ. For larger values of ρ, the GB1, JH1, and HJ1 approaches performed adequately for models with design 3. The GB2, JH2, and HJ2 approaches did not perform adequately, however, the JH2 and HJ2 approaches were found to be more similar and more acceptable. In general, the approaches did not perform with the smallest design as well as with the other deigns, and this was expected from Section 4.1 where the approaches were not very stable in producing the typical denominator degrees of freedom and scales.

Table 3. The observed test levels under the null hypothesis for generated data sets and met the conditions to compute the denominator degrees of freedom

	ρ	Observed test levels					
		GB1	GB2	JH1	JH2	HJ1	HJ2
Design 1	0.25	0.0741	0.1002	0.0480	0.0679	0.0387	0.0577
	0.50	0.0700	0.0965	0.0523	0.0760	0.0450	0.0676
	1.00	0.0549	0.0798	0.0487	0.0726	0.0431	0.0679
	2.00	0.0500	0.0740	0.0486	0.0724	0.0477	0.0707
	4.00	0.0523	0.0765	0.0519	0.0758	0.0513	0.0754
Design 2	0.25	0.0527	0.0622	0.0525	0.0627	0.0527	0.0627
	0.50	0.0488	0.0562	0.0489	0.0574	0.0486	0.0571
	1.00	0.0475	0.0566	0.0463	0.0567	0.0461	0.0565
	2.00	0.0471	0.0551	0.0472	0.0551	0.0472	0.0549
	4.00	0.0497	0.0560	0.0497	0.0560	0.0497	0.0559
Design 3	0.25	0.0901	0.1228	0.0530	0.0754	0.0436	0.0608
	0.50	0.0835	0.1113	0.0460	0.0759	0.0455	0.0762
	1.00	0.0687	0.0935	0.0501	0.0862	0.0601	0.1041
	2.00	0.0578	0.0760	0.0451	0.0623	0.0425	0.0621
	4.00	0.0438	0.0615	0.0421	0.0584	0.0413	0.0550

5. Conclusions

The six approximate *F*-tests considered in this paper are different based on the estimator for the variance of the estimate of fixed effects used in the Wald-type test statistic, and the type of moment method approximation used to compute the denominator degrees of freedom and scale. For a Wald-type statistic that distributed approximately as F distribution, the one-moment approximation was used to compute the denominator degrees of freedom. However, when a scaled Wald-type statistic distributed approximately as F distribution, the two-moment approximation was used to determine the denominator degrees of freedom and the scale factor. Analytically, and under specific conditions, we found that all the six approaches' performances are the same, and under other conditions, some of these approaches are identical. In general, the denominator degrees of freedom computed by approaches which used the two-moment approximation found to be greater than those for approaches which used the one-moment approximation. This typically affected the approaches which used the two-moment approximation to be significantly liberal comparing to those used the one-moment approximation. This coincided with the results of the simulation study, which consisted of running 15 models (3 block designs with 5 settings of the ratio of the variance components), where the approaches used the two-moment approximation (i.e., GB2, JH2, and HJ2) found produced larger denominator degrees of freedom, and hence these approaches performed liberally. The approaches used the one-moment approximation (i.e., GB1, JH1, and HJ1) performed similarly for some models which coincides with the findings of the simulations study of Fai and Cornelious (1996) for the similarity of the performance of the GB1 and JH1 approaches, however, these approaches performed differently for some other models. The JH1 approach which used the adjusted estimator of the variance of the estimate of fixed effects suggested by Kackar and Harville (1984), and derived by using one-moment approximation, found usually to outperform all other approaches. The GB1 approach which uses the traditional estimator of the variance of the estimate of fixed effects found to be significantly liberal for some models, whereas the HJ1 approach which uses the adjustment of the estimator of the variance of the estimate of fixed effects that suggested by Harville and Jeske (1992) found to be excessively conservative test for some models. For models with the small design and small values of ρ, the approaches usually did not perform adequately. The simulation findings are for block designs, and it is suggested to do more studies for the Satterthwaite based approaches for other models with different designs in order to evaluate their performance and ensure the adequacy of the JH1 approach, and to explore the potential of developing the approaches such that the performance becomes adequately even for models with small designs.

References

Abadir, K. M., & Magnus J. R. (2005). *Matrix algebra* (Vol. 1). Cambridge University Press.

Alnosaier, W., & Birkes, D. (2019). Inner workings of the Kenward-Roger test. *Metrika, 82*(2), 195-223.

Cochran, W. G., & Cox, G. M. (1992). *Experimental designs* (2nd ed.). John Wiley and Sons.

Corbiel, R. R., & Searl, S. R. (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics, 18*(1), 31-38.

Giesbrecht, F. G., & Burns, J. C. (1985). Two-stage analysis based on a mixed model: large sample asymptotic theory and small-sample simulation results. *Biometrics, 41*(2), 477-486.

Harville, D. A., & Jeske, D. R. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association, 87*(419), 724-731.

Hrong-Tai Fai, A., & Cornelius, P. L. (1996). Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analysis of unbalanced split-plot experiments. *Journal of statistical computation and simulation, 54*(4), 363-378.

Jeske, D. R., & Harville, D. A. (1988). Prediction-interval procedure and (fixed-effects) confidence-interval procedures for mixed linear models. *Communications in Statistics-Theory and Methods, 17*(4), 1053-1087.

Kackar, R. N., & Harville, D. A. (1984). Approximation for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association, 79*(388), 853-862.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*(3), 983-997.

Kuehl, R. O. (2000). *Design of experiments: statistical principles of research design and analysis* (2nd ed.). Duxbury Press.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R. H. (2017). ImerTest package: tests in linear mixed models. *Journal of Statistical Software, 82*(13), 1-26.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects model in R. *Behavior Research Methods, 49*, 1494-1502.

Manor, O., & Zucker, D. M. (2004). Small sample inference for fixed effects in the mixed linear model. *Computational Statistics & Data Analysis, 46*(4), 801-817.

Pase, L., & Salvan, A. (1997). *Principles of statistical inference from a Neo-Fisherian prospective* (Vol. 4). World Scientific.

Pukelshiem, F. (2006). *Optimal design of experiments* (2nd ed.). Society for Industrial and Applied Mathematics.

SAS Institute Inc. (2021). *SAS/STAT15.2 User's guide*. SAS Institute Inc.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics, 2*(6), 110-114.

Schaalje, G. B., McBride, J. B., & Fellingham, G. W. (2002). Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural, Biological, and Environmental Statistics, 7*(4), 512-524.

Searl, S. R., Casella, G., & McCulloch, C. E. (2006). *Variance components*. John Wiley &Sons.

Spilke, J., Piepho, H. P., & Hu, X. (2005). Analysis of unbalance data by mixed linear models using the mixed procedure of the SAS system. *Journal of Agronomy and Crop Science, 191*(1), 47-54.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *The Annals of Mathematical Statistics, 38*(4), 1092-1109.

Appendix

1. Proof of Lemma 1

From the general expression (3.1a), and since $\ell = 1,$ we obtain

$$E(T) = \nu_1/(\nu_1 - 2),$$

and substituting this quantity in the general expressions for the denominator degrees of freedom (3.1b), and (3.4a), and simplifying the expressions we have

$$\nu_{one} = \nu_1, \text{ and } \nu_{two} = \nu_1.$$

Also, from the general expression for the scale (3.4b), and since $\ell = 1,$ we obtain

$$\lambda = \frac{\nu_{two}}{E(T)(\nu_{two} - 2)},$$

and substituting $E(T) = \nu_1/(\nu_1 - 2)$, and since $\nu_{two} = \nu_1$, we have $\lambda = 1$.

2. Proof of Lemma 2

Without loss of generality, it suffices to show that the approximates of ν_s for the extensions of the JH approaches (e.g., the JH1 and JH2 approaches) are the same for $s = 1, \cdots, \ell$.

$$\nu_s \approx \frac{2(d_s)^2}{\mathbf{g}_s'\mathbf{W}\mathbf{g}_s} \quad \text{for} \quad s = 1, \cdots, \ell,$$

and $\mathbf{g}_s$ is the gradient of $\mathbf{a}_s \hat{\mathbf{\Phi}}_{JH} \mathbf{a}_s'$ with respect to $\boldsymbol{\theta}$, and $\mathbf{a}_j$ is the s^{th} row of $\mathbf{PL}'$. The spectral decomposition for $\mathbf{L}'\hat{\mathbf{\Phi}}_{JH}\mathbf{L} = \mathbf{P}'\mathbf{DP}$ for $\mathbf{P}$ an orthogonal matrix, $\mathbf{D}$ is a diagonal matrix of eigenvalues d_s's and $d_s > 0$ for $s = 1, \cdots, \ell$.

According to the partition of the model, we have:

$$\mathbf{\Phi} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} = \begin{bmatrix} (\mathbf{X}_1'\mathbf{\Sigma}^{-1}\mathbf{X}_1)^{-1} & \mathbf{0} \\ \mathbf{0} & [f(\boldsymbol{\theta})\mathbf{A}]^{-1} \end{bmatrix},$$

and from expression (2.11) for $\mathbf{P}_i$ and $\mathbf{Q}_{ij}$ we have

$$\mathbf{P}_i = \mathbf{X}'\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_i}\mathbf{X} = \frac{\partial(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})}{\partial \theta_i} = \begin{bmatrix} (\partial/\partial\theta_i)(\mathbf{X}_1'\mathbf{\Sigma}^{-1}\mathbf{X}_1') & \mathbf{0} \\ \mathbf{0} & (\partial/\partial\theta_i) f(\boldsymbol{\theta})\mathbf{A} \end{bmatrix},$$

$$\mathbf{Q}_{ij} = \mathbf{X}'\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_i}\mathbf{\Sigma}\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_j}\mathbf{X} = \begin{bmatrix} * & * \\ * & \mathbf{X}_2'\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_i}\mathbf{\Sigma}\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_j}\mathbf{X}_2 \end{bmatrix},$$

$$\mathbf{\Phi}\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j\mathbf{\Phi} = \begin{bmatrix} * & \mathbf{0} \\ \mathbf{0} & f_i(\boldsymbol{\theta}) f_j(\boldsymbol{\theta})[f(\boldsymbol{\theta})\mathbf{A}]^{-1} \end{bmatrix}, \text{ and } \quad \mathbf{\Phi}\mathbf{Q}_{ij}\mathbf{\Phi} = \begin{bmatrix} * & * \\ * & [f(\boldsymbol{\theta})\mathbf{A}]^{-1}\mathbf{X}_2'\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_i}\mathbf{\Sigma}\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_j}\mathbf{X}_2[f(\boldsymbol{\theta})\mathbf{A}]^{-1} \end{bmatrix},$$

where $f_i(\boldsymbol{\theta}) = (\partial/\partial\theta_i) f(\boldsymbol{\theta})$, and (*) refers to quantities in the matrices that we don't need to compute. Now, since $\mathbf{L}' = [\mathbf{0} \quad \mathbf{B}']$ from the model partition, we obtain

$$\mathbf{L}'\mathbf{\Phi}\mathbf{L} = \frac{1}{f(\boldsymbol{\theta})}\mathbf{B}'\mathbf{A}^{-1}\mathbf{B},$$

and

$$\mathbf{L}'\mathbf{\Phi}\mathbf{Q}_{ij}\mathbf{\Phi}\mathbf{L} = \frac{1}{[f(\boldsymbol{\theta})]^2}\mathbf{B}'\mathbf{A}^{-1}\mathbf{X}_2'\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_i}\mathbf{\Sigma}\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_j}\mathbf{X}_2\mathbf{A}^{-1}\mathbf{B}$$

Since $\mathbf{\Phi}_{JH} = \mathbf{\Phi} + \tilde{\mathbf{\Lambda}}$, and $\tilde{\mathbf{\Lambda}} = \mathbf{\Phi}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}(\mathbf{Q}_{ij} - \mathbf{P}_i\mathbf{\Phi}\mathbf{P}_{ij})\right\}\mathbf{\Phi}$,

then

$$\mathbf{L}'\mathbf{\Phi}_{JH}\mathbf{L} = \mathbf{L}'\mathbf{\Phi}\mathbf{L} + \mathbf{L}'\tilde{\mathbf{\Lambda}}\mathbf{L}$$
$$= \frac{1}{f(\boldsymbol{\theta})}\mathbf{B}'\mathbf{A}^{-1}\mathbf{B} + \sum_{i=1}^{r}\sum_{j=1}^{r}\frac{w_{ij}}{[f(\boldsymbol{\theta})]^2}\mathbf{B}'\mathbf{A}^{-1}\mathbf{X}_2'\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\theta_i}\mathbf{\Sigma}\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\theta_j}\mathbf{X}_2\mathbf{A}^{-1}\mathbf{B}$$
$$-\frac{1}{[f(\boldsymbol{\theta})]^3}\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}f_i(\boldsymbol{\theta})f_j(\boldsymbol{\theta})\mathbf{B}'\mathbf{A}^{-1}\mathbf{B} \tag{3.6}$$

Note that according to the model partition,

$$\mathbf{X}_2'\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\theta_i}\mathbf{X}_2 = \frac{\partial(\mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{X}_2')}{\partial\theta_i} = f_i(\boldsymbol{\theta})\mathbf{A},$$

and also, we have
$$\mathbf{X}_2'\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\theta_i}\mathbf{X}_2 = -\mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{G}_i\mathbf{\Sigma}^{-1}\mathbf{X}_2.$$

This implies that

$$-\frac{\partial(\mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{G}_j\mathbf{\Sigma}^{-1}\mathbf{X}_2)}{\partial\theta_i} = \mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{G}_i\mathbf{\Sigma}^{-1}\mathbf{G}_j\mathbf{\Sigma}^{-1}\mathbf{X}_2 + \mathbf{X}_2'\mathbf{\Sigma}^{-1}\mathbf{G}_i\mathbf{\Sigma}^{-1}\mathbf{G}_j\mathbf{\Sigma}^{-1}\mathbf{X}_2$$
$$= f_{ij}(\boldsymbol{\theta})\mathbf{A},\ \text{where } f_{ij}(\boldsymbol{\theta}) = (\partial/\partial\theta_i)f_j(\boldsymbol{\theta}).$$

Then,

$$\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}\mathbf{X}_2'\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\theta_i}\mathbf{\Sigma}\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\theta_j}\mathbf{X}_2 = \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}f_{ij}(\boldsymbol{\theta})\mathbf{A}. \tag{3.7}$$

Substituting expression (3.7) in (3.6), we obtain

$$\mathbf{L}'\mathbf{\Phi}_{JH}\mathbf{L} = h_{ij}(\boldsymbol{\theta})\mathbf{B}'\mathbf{A}^{-1}\mathbf{B},$$

where
$$h_{ij}(\boldsymbol{\theta}) = \frac{1}{f(\boldsymbol{\theta})} + \frac{1}{2[f(\boldsymbol{\theta})]^2}\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}f_{ij}(\boldsymbol{\theta}) - \frac{1}{[f(\boldsymbol{\theta})]^3}\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}f_i(\boldsymbol{\theta})f_j(\boldsymbol{\theta}).$$

So, the approximate of ν_s for $s = 1,\cdots,\ell$ will be

$$\frac{2(d_s)^2}{\mathbf{g}_s'\mathbf{W}\mathbf{g}_s} = \frac{2(d_s)^2}{[(\partial/\partial\theta_i)h_{ij}(\boldsymbol{\theta})\mathbf{b}_s\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\mathbf{b}_s']'_{r\times1}\mathbf{W}[(\partial/\partial\theta_i)h_{ij}(\boldsymbol{\theta})\mathbf{b}_s\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\mathbf{b}_j']_{r\times1}}$$

$$= \frac{2(d_s)^2}{(\mathbf{b}_s\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\mathbf{b}_s')^2[(\partial/\partial\theta_i)h_{ij}(\boldsymbol{\theta})]'_{r\times1}\mathbf{W}[(\partial/\partial\theta_i)h_{ij}(\boldsymbol{\theta})]_{r\times1}}$$

$$= \frac{2(d_s)^2}{(\mathbf{b}_s\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\mathbf{b}_s')^2\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}(\partial/\partial\theta_i)h_{ij}(\boldsymbol{\theta})(\partial/\partial\theta_j)h_{ij}(\boldsymbol{\theta})},$$

where $\mathbf{b}_s$ is the *s* row of the orthogonal matrix **P**.
Note that $d_s = \mathbf{b}_s\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\mathbf{b}_s'$, and hence we have

$$\frac{2(d_s)^2}{\mathbf{g}_s'\mathbf{W}\mathbf{g}_s} = \frac{2}{\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}(\partial/\partial\theta_i)h_{ij}(\boldsymbol{\theta})(\partial/\partial\theta_j)h_{ij}(\boldsymbol{\theta})}$$

which doesn't depend on the row *s*, and the approximation of ν_s are the same for $s = 1,\cdots,\ell$.
For the extensions of the GB approach,

$$h_{ij}(\boldsymbol{\theta}) = h(\boldsymbol{\theta}) = \frac{1}{f(\boldsymbol{\theta})},$$

and for the HJ approaches,

$$h_{ij}(\boldsymbol{\theta}) = \frac{1}{f(\boldsymbol{\theta})} + \frac{1}{[f(\boldsymbol{\theta})]^2}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij} f_{ij}(\boldsymbol{\theta}) - \frac{2}{[f(\boldsymbol{\theta})]^3}\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij} f_i(\boldsymbol{\theta}) f_j(\boldsymbol{\theta}).$$

3. Proof of Lemma 3

(a) From expressions (3.2), (3.4c), and lemma 2, we have

$$E(T) = \ell\nu_1/(\nu_1-2), \text{ provided } \nu_1 > 2 \tag{3.8}$$

Also, from expression (3.4c), and lemma 2, we have

$$Var(T) = [2\ell\nu_1^2(\nu_1-1)]/[(\nu_1-2)^2(\nu_1-4)], \text{ provided } \nu_1 > 4 \tag{3.9}$$

Substituting expressions (3.8) and (3.9) for *E*(*T*) and *Var*(*T*) in expression (3.4a), we obtain

$$\nu_{two} = 4 + \frac{2\left(\frac{\ell\nu_1}{\nu_1-2}\right)^2(\ell+2)}{\ell\frac{2\ell\nu_1^2(\nu_1-1)}{(\nu_1-2)^2(\nu_1-4)} - 2\left(\frac{\ell\nu_1}{\nu_1-2}\right)^2} = \frac{\ell+2}{3}\nu_1 - \frac{4}{3}(\ell-1)$$

Note that from expressions (3.8) and (3.1b), we obtain $\nu_{one} = \nu_1$, and hence we have

$$\nu_{two} = \frac{\ell+2}{3}\nu_{one} - \frac{4}{3}(\ell-1) \tag{3.10}$$

(b) Because of the model partition, as we have seen in part (a), $\nu_q = \nu_1 = \nu_{one}$ for all $q = 1,\cdots,\ell$. So, for data sets produce $\nu_1 = \nu_{one} \le 4$, the two-moment approximation is not applicable since we require $\nu_1 > 4$. Hence the proof will be for the data sets produce $\nu_1 = \nu_{one} > 4$.

Since $\nu_{one} > 4,$

then $\left(\frac{\ell-1}{3}\right)\nu_{one} + \nu_{one} > 4\left(\frac{\ell-1}{3}\right) + \nu_{one},$

and $\left(\frac{\ell+2}{3}\right)\nu_{one} - 4\left(\frac{\ell-1}{3}\right) > \nu_{one}$

Note that from part (a), the left side is ν_{two} as in expression (3.10).

For the scale, recall form expression (3.4b) that

$$\lambda = \frac{\ell\nu_{two}}{E(T)(\nu_{two}-2)},$$

and expressing *E*(*T*), and ν_{two} in terms of ν_{one} using expressions (3.8), and (3.10), we obtain

$$\lambda = \frac{\ell\left\{\left(\frac{\ell+2}{3}\right)\nu_{one} - 4\left(\frac{\ell-1}{3}\right)\right\}}{\frac{\ell\nu_{one}}{\nu_{one}-2}\left\{\left(\frac{\ell+2}{3}\right)\nu_{one} - 4\left(\frac{\ell-1}{3}\right) - 2\right\}}$$

$$= \frac{(\ell+2)\nu_{one}^2 - 4\ell\nu_{one} - 2[\ell\nu_{one} - 4(\ell-1)]}{(\ell+2)\nu_{one}^2 - 4\ell\nu_{one} - 2\nu_{one}}$$

To prove that $\lambda \le 1$, it suffices to show that $\ell\nu_{one} - 4(\ell-1) \ge \nu_{one}$,

and since $\nu_{one} > 4,$

then $$(\ell-1)\nu_{one}+\nu_{one}>4(\ell-1)+\nu_{one},$$

and $$\ell\,\nu_{one}-4(\ell-1)>\nu_{one}$$

Note that the equality happens for the case of $\ell=1$, where we obtain $\lambda=1$ as in lemma 1.

4. Proof of Lemma 4

(a) For models that satisfy Zyskind's condition, and the assumption above, it can be proved that $\mathbf{Q}_{ij}=\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_{ij}$ (Alnosaier & Birkes, 2019, Lemma 6).

Since $$\tilde{\mathbf{\Lambda}}=\mathbf{\Phi}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r}w_{ij}(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_{ij})\right\}\mathbf{\Phi},$$

then $\tilde{\mathbf{\Lambda}}=\mathbf{0}$, and hence $\mathbf{\Phi}_{JH}=\mathbf{\Phi}_{HJ}=\mathbf{\Phi}$

This implies that $T_{JH}=T_{HJ}=T_{GB}$, and hence the test statistics using one-moment (i.e., $F_{JH1}=F_{HJ1}=F_{GB1}$) are equal, and the denominator degrees of freedom, as in expression (3.1b) are equal as well. Note that the test statistic using one-moment is different from the test statistic using two-moment because of the scale value. However, since $T_{JH}=T_{HJ}=T_{GB}$, then the denominator degrees of freedom using two-moment, as in expression (3.4a) are equal. Also, the scales, as in expression (3.4b) are equal, and hence the test statistics are equal.

(b) Direct. From part (a), and since the scale is one (Lemma 1), then all test statistics are equal.

5. Preparing Formulas for the Simulation Study

The REML estimates of the variance components σ_b^2 and σ_e^2 were computed by the iteration algorithm on p.252 of Searl et al. (2006):

$$\left\{tr(\mathbf{G}_i\mathbf{M}^{\bullet}\mathbf{G}_j\mathbf{M}^{\bullet})\right\}_{i,j=1}^{r}\boldsymbol{\theta}^{\bullet}=\left\{y'\mathbf{M}^{\bullet}\mathbf{G}_i\mathbf{M}^{\bullet}y)\right\}_{i=1}^{r},$$

where

$$\mathbf{M}=\mathbf{\Sigma}^{-1}-\mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}, \tag{4.1}$$

$\mathbf{M}^{\bullet}=\mathbf{M}(\boldsymbol{\theta}^{\bullet})$, and $\boldsymbol{\theta}^{\bullet}$ is the solution of the equation. The variance-covariance matrix of the REML estimators of the variance components, $\mathbf{W}=[w_{ij}]_{r\times r}$ is approximated by the inverse of the expected information matrix $\tilde{\mathbf{W}}=\tilde{\mathbf{W}}(\boldsymbol{\theta})$ (Kenward and Roger, 1997). To approximate the quantity ν_q in expression (3.3),

$$\nu_q\approx\frac{2(d_q)^2}{\mathbf{g}'_q\,\mathbf{W}\mathbf{g}_q},$$

we need to compute the estimate of the quantity $\mathbf{g}_q$ which is the gradient of $\mathbf{a}_q Var(\hat{\boldsymbol{\beta}})\mathbf{a}'_q$ with respect to $\boldsymbol{\theta}$, and $\mathbf{a}_q$ is the q^{th} row of $\mathbf{PL}'$. So, we need to find expressions for $(\partial/\partial\theta_i)\mathbf{\Phi}$, and $(\partial/\partial\theta_i)\tilde{\mathbf{\Lambda}}$, for $i=1,\ 2.$

$$\frac{\partial\mathbf{\Phi}}{\partial\theta_i}=\frac{\partial(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}}{\partial\theta_i}=-(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\frac{\partial\mathbf{\Sigma}^{-1}}{\partial\theta_i}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}=-\mathbf{\Phi}\mathbf{P}_i\mathbf{\Phi} \tag{4.2}$$

Let $\theta_1=\sigma_e^2$, and $\theta_2=\sigma_b^2$, and by noting $\mathbf{G}_1=\mathbf{I}_n$, we obtain

$$(\partial/\partial\theta_1)\mathbf{\Phi}=-\mathbf{\Phi}\mathbf{P}_1\mathbf{\Phi}=-\mathbf{\Phi}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{\Phi}, \tag{4.2a}$$

and

$$(\partial/\partial\theta_2)\mathbf{\Phi}=-\mathbf{\Phi}\mathbf{P}_2\mathbf{\Phi}=-\mathbf{\Phi}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{G}_2\mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{\Phi} \tag{4.2b}$$

$$\begin{aligned}\frac{\partial\tilde{\mathbf{\Lambda}}}{\partial\theta_i}&=\frac{\partial}{\partial\theta_i}\left\{\mathbf{\Phi}\left[\sum_{i=1}^{2}\sum_{j=1}^{2}w_{ij}(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)\right]\mathbf{\Phi}\right\}\\&=\frac{\partial\mathbf{\Phi}}{\partial\theta_i}\left[\sum_{i=1}^{2}\sum_{j=1}^{2}w_{ij}(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)\right]\mathbf{\Phi}+\mathbf{\Phi}\left[\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\partial w_{ij}}{\partial\theta_i}(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)\right]\mathbf{\Phi}\\&+\mathbf{\Phi}\left[\sum_{i=1}^{2}\sum_{j=1}^{2}w_{ij}\frac{\partial(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)}{\partial\theta_i}\right]\mathbf{\Phi}+\mathbf{\Phi}\left[\sum_{i=1}^{2}\sum_{j=1}^{2}w_{ij}(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)\right]\frac{\partial\mathbf{\Phi}}{\partial\theta_i},\end{aligned} \tag{4.3}$$

The above expression consists of four terms. The first term and fourth term are respectively:

$$\frac{\partial \mathbf{\Phi}}{\partial \theta_i}\left[\sum_{i=1}^{2}\sum_{j=1}^{2} w_{ij}(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)\right]\mathbf{\Phi} = -\mathbf{\Phi}\mathbf{P}_i\mathbf{\Phi}\left[\sum_{i=1}^{2}\sum_{j=1}^{2} w_{ij}(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)\right]\mathbf{\Phi}, \quad (4.4)$$

and

$$\mathbf{\Phi}\left[\sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)\right]\frac{\partial \mathbf{\Phi}}{\partial \theta_i} = -\mathbf{\Phi}\left[\sum_{i=1}^{2}\sum_{j=1}^{2} w_{ij}(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)\right]\mathbf{\Phi}\mathbf{P}_i\mathbf{\Phi} \quad (4.5)$$

Note that

$$\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j = \mathbf{X}'\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_i}\mathbf{\Sigma}\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_j}\mathbf{X} - \mathbf{X}'\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_i}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_j}\mathbf{X}$$

$$= \mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_i}\left[\mathbf{\Sigma}^{-1}-\mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\right]\frac{\partial \mathbf{\Sigma}}{\partial \theta_j}\mathbf{\Sigma}^{-1}\mathbf{X}$$

$$= \mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_i}\mathbf{M}\frac{\partial \mathbf{\Sigma}}{\partial \theta_j}\mathbf{\Sigma}^{-1}\mathbf{X},$$

where $\mathbf{M}$ is as in expression (4.1).

So, the third term can be expressed as

$$\mathbf{\Phi}\left[\sum_{i=1}^{2}\sum_{j=1}^{2} w_{ij}\frac{\partial(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)}{\partial \theta_i}\right]\mathbf{\Phi}$$

$$= \mathbf{\Phi}\left[w_{11}\frac{\partial(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{M}\mathbf{\Sigma}^{-1}\mathbf{X})}{\partial \theta_i} + 2w_{12}\frac{\partial(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{M}\mathbf{G}_2\mathbf{\Sigma}^{-1}\mathbf{X})}{\partial \theta_i} + w_{22}\frac{\partial(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{G}_2\mathbf{M}\mathbf{G}_2\mathbf{\Sigma}^{-1}\mathbf{X})}{\partial \theta_i}\right]\mathbf{\Phi}$$

$$= \mathbf{\Phi}\mathbf{X}'\left[w_{11}\frac{\partial \mathbf{H}}{\partial \theta_i} + 2w_{12}\frac{\partial \mathbf{H}}{\partial \theta_i}\mathbf{G}_2 + w_{22}\mathbf{G}_2\frac{\partial \mathbf{H}}{\partial \theta_i}\mathbf{G}_2\right]\mathbf{X}\mathbf{\Phi}, \quad (4.6)$$

where

$$\mathbf{H} = \mathbf{\Sigma}^{-1}\mathbf{M}\mathbf{\Sigma}^{-1},$$

$$(\partial/\partial\theta_1)\mathbf{H} = -\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{-1}\mathbf{M}\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{M}\mathbf{M}\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{M}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{-1},$$

and

$$(\partial/\partial\theta_2)\mathbf{H} = -\mathbf{\Sigma}^{-1}\mathbf{G}_2\mathbf{\Sigma}^{-1}\mathbf{M}\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{M}\mathbf{G}_2\mathbf{M}\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{M}\mathbf{\Sigma}^{-1}\mathbf{G}_2\mathbf{\Sigma}^{-1}$$

Note that for the models considered in the simulation study, we have $\mathbf{G}_2\mathbf{\Sigma}^{-1} = \mathbf{\Sigma}^{-1}\mathbf{G}_2$, and this implies $\mathbf{Q}_{12} = \mathbf{Q}_{21}$, and $\mathbf{P}_1\mathbf{\Phi}\mathbf{P}_2 = \mathbf{P}_2\mathbf{\Phi}\mathbf{P}_1$.

To compute the second term, $\mathbf{\Phi}\left[\sum_{i=1}^{r}\sum_{j=1}^{r}\frac{\partial w_{ij}}{\partial \theta_i}(\mathbf{Q}_{ij}-\mathbf{P}_i\mathbf{\Phi}\mathbf{P}_j)\right]\mathbf{\Phi}$, we need to find an expression for $(\partial/\partial\theta_i)w_{ij}$, where, as mentioned above, the inverse of the expected information matrix $\tilde{\mathbf{W}}$ is used to approximate $\mathbf{W}$.

The expected information matrix $\mathbf{I}_E$ can be expressed as on p. 253 of Searl et al (2006):

$$\mathbf{I}_E = \frac{1}{2}\begin{bmatrix} tr(\mathbf{M}\mathbf{M}) & tr(\mathbf{M}\mathbf{M}\mathbf{G}_2) \\ tr(\mathbf{M}\mathbf{M}\mathbf{G}_2) & tr(\mathbf{M}\mathbf{G}_2\mathbf{M}\mathbf{G}_2) \end{bmatrix},$$

then

$$\tilde{\mathbf{W}} = \frac{2}{c}\begin{bmatrix} tr(\mathbf{M}\mathbf{G}_2\mathbf{M}\mathbf{G}_2) & -tr(\mathbf{M}\mathbf{M}\mathbf{G}_2) \\ -tr(\mathbf{M}\mathbf{M}\mathbf{G}_2) & tr(\mathbf{M}\mathbf{M}) \end{bmatrix},$$

where

$$c = tr(\mathbf{M}\mathbf{M})tr(\mathbf{M}\mathbf{G}_2\mathbf{M}\mathbf{G}_2) - [tr(\mathbf{M}\mathbf{M}\mathbf{G}_2)]^2.$$

The quantity $(\partial/\partial\theta_i)w_{ij}$ is approximated by $(\partial/\partial\theta_i)\tilde{w}_{ij}$, and a useful formula to compute this approximation can be found on p. 341 of Pace and Salvan (1997):

$$(\partial/\partial\theta_i)\tilde{w}_{ij} = -\sum_{k=1}^{2}\sum_{l=1}^{2}\tilde{w}_{ik}\tilde{w}_{lj}(\partial/\partial\theta_i)i_{kl}, \text{ where } \mathbf{I}_E = [i_{kl}]_{2\times 2}.$$

That is, we obtain:

$$(\partial/\partial\theta_i)\tilde{w}_{11} = -\frac{2}{a^2}\Big\{[tr(\mathbf{MG}_2\mathbf{MG}_2)]^2(\partial/\partial\theta_i)tr(\mathbf{MM}) \\ -2tr(\mathbf{MG}_2\mathbf{MG}_2)tr(\mathbf{MMG}_2)(\partial/\partial\theta_i)tr(\mathbf{MMG}_2) \\ +[tr(\mathbf{MMG}_2)]^2(\partial/\partial\theta_i)tr(\mathbf{MG}_2\mathbf{MG}_2)\Big\},$$

$$(\partial/\partial\theta_i)\tilde{w}_{12} = \frac{2}{a^2}\Big\{tr(\mathbf{MG}_2\mathbf{MG}_2)tr(\mathbf{MMG}_2)(\partial/\partial\theta_i)tr(\mathbf{MM}) \\ -2tr(\mathbf{MG}_2\mathbf{MG}_2)tr(\mathbf{MM})(\partial/\partial\theta_i)tr(\mathbf{MMG}_2) \\ +tr(\mathbf{MMG}_2)tr(\mathbf{MM})(\partial/\partial\theta_i)tr(\mathbf{MG}_2\mathbf{MG}_2)\Big\},$$

$$(\partial/\partial\theta_i)\tilde{w}_{22} = -\frac{2}{a^2}\Big\{[tr(\mathbf{MMG}_2]^2(\partial/\partial\theta_i)tr(\mathbf{MM}) \\ -2tr(\mathbf{MMG}_2)tr(\mathbf{MM})(\partial/\partial\theta_i)tr(\mathbf{MMG}_2) \\ +[tr(\mathbf{MM})]^2(\partial/\partial\theta_i)tr(\mathbf{MG}_2\mathbf{MG}_2)\Big\},$$

where

$$(\partial/\partial\theta_1)tr(\mathbf{MM}) = -2tr(\mathbf{MM}), \qquad (\partial/\partial\theta_1)tr(\mathbf{MMG}_2) = -2tr(\mathbf{MMMG}_2),$$
$$(\partial/\partial\theta_1)tr(\mathbf{MG}_2\mathbf{MG}_2) = -2tr(\mathbf{MG}_2\mathbf{MMG}_2),$$
$$(\partial/\partial\theta_2)tr(\mathbf{MM}) = -2tr(\mathbf{MMG}_2\mathbf{M}),$$
$$(\partial/\partial\theta_2)tr(\mathbf{MMG}_2) = -2tr(\mathbf{MMG}_2\mathbf{MG}_2),$$

and

$$(\partial/\partial\theta_2)tr(\mathbf{MG}_2\mathbf{MG}_2) = -2tr(\mathbf{MG}_2\mathbf{MG}_2\mathbf{MG}_2).$$

All quantities used in the simulation study and are functions of the variance components $\boldsymbol{\theta}$ are to be estimated by substituting the REML estimates $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$.

6. Computed Denominator Degrees of Freedom and Scale

The average of computed denominator degrees of freedom and scales for the approaches are presented in table 3 and 4 below.

Table 4. The average of computed denominator degrees of freedom for generate data sets of which the iteration algorithm to compute the REML estimates converged, and met the conditions to compute the degrees of freedom under the null hypothesis

	ρ	The average of computed denominator degrees of freedom					
		GB1	GB2	JH1	JH2	HJ1	HJ2
Design 1	0.25	34.61	147.30	27.94	118.62	24.41	102.74
	0.50	33.59	142.14	27.17	114.24	24.48	101.81
	1.00	30.07	125.65	27.08	112.03	25.27	103.67
	2.00	27.91	115.56	27.50	113.68	27.12	111.89
	4.00	27.23	112.41	27.14	111.99	27.05	111.57
Design 2	0.25	23.50	49.49	21.72	45.84	21.40	45.48
	0.50	23.32	49.08	22.14	46.99	22.20	46.29
	1.00	23.13	48.64	23.28	47.98	23.11	47.74
	2.00	23.04	48.43	23.00	48.34	22.98	48.28
	4.00	23.01	48.36	23.00	48.34	22.99	48.33
Design 3	0.25	8.06	21.16	5.49	11.21	5.37	23.64
	0.50	7.94	18.15	6.57	39.69	13.51	25.57
	1.00	7.47	15.57	16.55	31.95	13.52	29.31
	2.00	6.60	12.53	5.79	12.43	6.02	17.03
	4.00	6.17	11.19	5.83	10.45	5.67	9.90

Table 5. The average of computed scales for generate data sets of which the iteration algorithm to compute the REML estimates converged, and met the conditions to compute the denominator degrees of freedom under the null hypothesis

	ρ	The average of scales		
		GB2	JH2	HJ2
Design 1	0.25	0.9535	0.9399	0.9319
	0.50	0.9534	0.9399	0.9330
	1.00	0.9484	0.9498	0.9374
	2.00	0.9447	0.9439	0.9431
	4.00	0.9433	0.9432	0.9430
Design 2	0.25	0.9534	0.9530	0.9507
	0.50	0.9531	0.9544	0.9570
	1.00	0.9527	0.9536	0.9532
	2.00	0.9525	0.9525	0.9524
	4.00	0.9525	0.9525	0.9524
Design 3	0.25	0.8545	0.8044	0.7747
	0.50	0.8533	0.6765	0.7202
	1.00	0.8440	0.6130	0.6507
	2.00	0.8297	0.7583	0.7398
	4.00	0.8227	0.8171	0.8156