Thailand Statistician
April 2026; 24(2): 494-504
http://statassoc.or.th
Contributed paper

Conditions Under Which the Kenward-Roger Approximate Test is an Exact F Test

Waseem Alnosaier*[a] and David Birkes [b]
[a] Statistics Department, Institute of Public Administration, Riyadh, Saudi Arabia.
[b] Statistics Department, Oregon State University, Corvallis, Oregon, USA.
*Corresponding author; e-mail: alnosaierw@ipa.edu.sa

Received: 24 June 2025
Revised: 12 November 2025
Accepted: 8 January 2026

Abstract

The Kenward-Roger (KR) test is an approximate F test, which is to say that it uses a test statistic that has approximately an F distribution under the null hypothesis. The test statistic is constructed so that in two special cases the test is an exact F test, with a test statistic that has exactly an F distribution under the null hypothesis. One of the special cases is testing for the significance of a fixed group effect in a one-factor ANOVA model. We show that this testing problem can be extended to a more general class of testing problems in which the KR test is an exact F test.

Keywords: Kenward Roger test, mixed linear model, exact F test, residual maximum likelihood estimator, fixed effects.

1. Introduction

The Kenward-Roger (KR) test (Kenward and Roger 1997) is a well-known procedure for testing linear hypotheses about fixed effects in normal mixed linear models. It has been cited more than 4800 times according to Google Scholar. The KR test has been embedded in the MIXED procedure of the SAS statistical package (SAS Institute Inc 2021). It also has been incorporated in many other statistical packages (e.g., Matlab Central, R, and SPSS). Many simulation studies have investigated the performance of the test in mixed linear models and showed that the test performed well comparing to other available tests (Schaalje et al. 2002; Spike et al. 2005; Wulff and Robinson 2009; Gregory 2011; Chawla et al. 2014; Halekoh and Højsgaard 2014). Alnosaier and Birkes (2019) provided details of the inner workings of the test, and proposed two alternative similar tests.

Consider $\mathbf{y}$ a random vector distributed according to a multivariate normal distribution with mean $\mathbf{X}\boldsymbol{\beta}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\beta}$ is a vector of unknown fixed-effects parameters, and $\boldsymbol{\Sigma}$ depends on a vector $\boldsymbol{\psi}$ of variance-covariance parameters. Suppose we want to test the linear hypotheses $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ where $\mathbf{L}$ is a known matrix with ℓ linearly independent columns. Kenward and Roger (1997) started their procedure with a Wald-type statistic of the form:

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

where $\hat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$ and $\widehat{\text{Var}}(\mathbf{L}'\hat{\boldsymbol{\beta}})$ is an estimator of $\text{Var}(\mathbf{L}'\hat{\boldsymbol{\beta}})$. It is usually true that the test statistic F has approximately an F distribution under the null hypothesis, and this is the test statistic used for the Satterthwaite test developed by Fai and Cornellious (1996). The Kenward and Roger's test may be summarized as described by Alnosaier and Birkes **(**2019**)**, that is they improve on the approximation in three ways**:** First, based on the results of Kackar and Harville (1984) and Harville and Jeske (1992), Kenward and Roger improve the estimator of $\text{Var}(\hat{\boldsymbol{\beta}})$. Second, they suppose that there are two positive quantities λ and m such that a scaled form of *F*, that is λF, is distributed approximately as an F distribution with ℓ numerator degrees of freedom and m denominator degrees of freedom under the null hypothesis. Third, they modify the approach in such a way that the test is exact for two special cases.

The two special cases used to modify the approach so that the KR test becomes exact tests is testing for a fixed group effect in a one-factor ANOVA model, in which case the KR test statistic is the usual F-ratio from the ANOVA table. The other special case is testing independent identical distributed vectors from a multivariate normal population for a specific value of the mean vector, in which case the KR test is equivalent to Hotelling's one-sample T^2 test. The simulation studies of Alnosaier and Birkes (2019) have shown that this modification is worth doing and improved the performance of the KR test significantly. In this paper, we show that the modification applied on the KR test does not produce exact tests only for the two special cases. In fact, the adjustment of the KR test for two special cases suffices to achieve exact F tests for much a larger class of testing problems than the special case of testing a fixed group effect in a one-factor ANOVA model. We present more general conditions under which the KR test is an exact test. These conditions are satisfied in all fixed-effects models and many standard testing problems in balanced mixed-effects ANOVA models. The detailed work of the KR test is described in Section 2, and Section 3 presents more general testing problems than testing a fixed group effect in a one-way ANOVA model that satisfy specific assumptions and will be called focused quadratic commutative orthogonal block structure. In Section 3, some definitions, remarks and theoretical results are presented. The KR test of this type of testing problems is shown to be an exact test in Section 4.

2. Description of the KR Test

Consider the normal mixed linear model,

$$\mathbf{y} \sim \text{N}\left(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}\right), \tag{2}$$

where $\text{E}\left(\mathbf{y}\right) = \mathbf{X}\boldsymbol{\beta}$ and variance-covariance matrix $\text{Var}\left(\boldsymbol{y}\right) = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}\left(\boldsymbol{\psi}\right)$ where $\mathbf{X}$ is a known $n \times p$ matrix of full column rank, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\boldsymbol{\Sigma}$ is an $n \times n$ positive definite matrix depends on an $r \times 1$ vector $\boldsymbol{\psi}$ of unknown parameters. The variance-covariance matrix $\boldsymbol{\Sigma}$ is assumed to depend linearly on the parameters ψ_i, so that $\boldsymbol{\Sigma} = \psi_1\mathbf{G}_1 + \cdots + \psi_r\mathbf{G}_r$ for known symmetric matrices $\mathbf{G}_i$. Note that a one-factor ANOVA model satisfies this assumption.

Consider the testing problem of the linear hypothesis $\text{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{L}'\boldsymbol{\beta}_0$ where $\boldsymbol{\beta}_0$ is a $p \times 1$ vector and $\mathbf{L}$ is a known $p \times \ell$ matrix of full column rank. To describe the KR test (Kenward and Roger 1997), Kenward and Roger begin with the idea of a Wald test using a test statistic of the form

$$T = \left(\mathbf{L}'\hat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}_0\right)' \left[\widehat{\text{Var}}(\mathbf{L}'\hat{\boldsymbol{\beta}})\right]^{-1} \left(\mathbf{L}'\hat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}_0\right) = \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)' \mathbf{L}\left[\mathbf{L}'\widehat{\text{Var}}(\mathbf{L}'\hat{\boldsymbol{\beta}})\mathbf{L}\right]^{-1} \mathbf{L}'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right), \tag{3}$$

where $\hat{\boldsymbol{\beta}}$ and $\widehat{\mathrm{Var}}(\mathbf{L}'\hat{\boldsymbol{\beta}})$ are estimators of $\boldsymbol{\beta}$ and $\mathrm{Var}(\mathbf{L}'\hat{\boldsymbol{\beta}})$, respectively. They choose $\hat{\boldsymbol{\beta}}$ to be an estimated generalized least-squares estimator (EGLSE), that is,

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y}, \tag{4}$$

where $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\left(\hat{\boldsymbol{\psi}}\right)$ and $\hat{\boldsymbol{\psi}}$ is an estimator of $\boldsymbol{\psi}$, and $\hat{\boldsymbol{\psi}}$ is chosen to be residual maximum likelihood estimator (REML) of $\boldsymbol{\psi}$. It is often true that a Wild test statistic like *T* has approximately a chi-squared distribution with ℓ degrees of freedom under the null hypothesis. The approximation is improved by (a) using an estimator $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})$ developed by Kackar and Harville (1984), (b) allowing the approximating null distribution T to be a scaled F distribution, and (c) adjusting the approximating null distribution to be exact in the two special cases, namely for the one-way ANOVA fixed effects model, and Hotelling's one-sample T^2 test.

The vector $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}\left(\boldsymbol{\psi}\right) = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$ is called a generalized least-squares "estimator" (GLSE) (although it is an estimator only if either $\boldsymbol{\Sigma}$ is known or $\tilde{\boldsymbol{\beta}}\left(\boldsymbol{\psi}\right)$ is the same for all $\boldsymbol{\psi}$). Note that $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\psi}}\right)$. Kackar and Harville (1984) used the expression

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\Phi} + \boldsymbol{\Lambda} \tag{5}$$

where

$$\boldsymbol{\Phi} = \boldsymbol{\Phi}\left(\boldsymbol{\psi}\right) = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1} = \mathrm{Var}\left(\tilde{\boldsymbol{\beta}}\right),\ \boldsymbol{\Lambda} = \mathrm{Var}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) \tag{6}$$

and $\boldsymbol{\Lambda}$ is approximated by

$$\tilde{\boldsymbol{\Lambda}} = \tilde{\boldsymbol{\Lambda}}\left(\boldsymbol{\psi}\right) = \boldsymbol{\Phi}\left[\sum_{i=1}^{r}\sum_{j=1}^{r}\omega_{ij}\left(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j\right)\right]\boldsymbol{\Phi} \tag{7}$$

where $\mathbf{W} = \left[\omega_{ij}\right]_{r\times r} = \mathrm{Var}\left(\hat{\boldsymbol{\psi}}\right)$, and $\mathbf{P}_i = \mathbf{P}_i\left(\boldsymbol{\psi}\right) = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{X}$,

$\mathbf{Q}_{ij} = \mathbf{Q}_{ij}\left(\boldsymbol{\psi}\right) = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{G}_j\boldsymbol{\Sigma}^{-1}\mathbf{X}$.

The REML $\hat{\boldsymbol{\psi}}$ is derived from a residual model not involving $\boldsymbol{\beta}$. Let $\tilde{\mathbf{W}} = \mathbf{W}\left(\boldsymbol{\psi}\right)$ denote the inverse of the expected information matrix for the residual model, and $\mathbf{W}$ can be approximated by $\hat{\mathbf{W}} = \tilde{\mathbf{W}}\left(\hat{\boldsymbol{\psi}}\right)$. Also, define $\hat{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\left(\hat{\boldsymbol{\psi}}\right)$, $\hat{\boldsymbol{\Lambda}} = \tilde{\boldsymbol{\Lambda}}\left(\hat{\boldsymbol{\psi}}\right)$, $\widehat{\mathbf{P}}_i = \mathbf{P}_i\left(\hat{\boldsymbol{\psi}}\right)$, $\hat{\mathbf{Q}}_{ij} = \mathbf{Q}_{ij}\left(\hat{\boldsymbol{\psi}}\right)$.

The matrix $\hat{\boldsymbol{\Phi}}$ has traditionally been used as a convenient estimator of $\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right)$, but generally it tends to underestimate, for two reasons. First, although $\hat{\boldsymbol{\Phi}}$ is a sensible estimator for $\boldsymbol{\Phi}$, it is true that $\boldsymbol{\Phi}$ is not the same as $\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right)$ unless $\boldsymbol{\Lambda} = \mathbf{0}$, which happens only when Zyskind's condition (see Section 3) is satisfied. The formula $\boldsymbol{\Phi} + \tilde{\boldsymbol{\Lambda}}$ from Kackar and Harville (1984) is a more accurate approximation for $\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right)$ than $\hat{\boldsymbol{\Phi}}$ is, and correspondingly, $\hat{\boldsymbol{\Phi}} + \hat{\boldsymbol{\Lambda}}$ is a better approximation for $\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right)$ than $\hat{\boldsymbol{\Phi}}$ is. But $\hat{\boldsymbol{\Phi}} + \hat{\boldsymbol{\Lambda}}$ still tends to underestimate $\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right)$. The bias is reduced by using $\hat{\boldsymbol{\Phi}}_{\mathrm{A}} = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Lambda}}$, and Kenward and Roger (1997) use this quantity as the approximation for $\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right)$ in the test statistic T and rescale it by dividing by ℓ. That is

$$F_{\mathrm{KR}} = \frac{1}{\ell}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)' \mathbf{L}\left[\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L}\right]^{-1}\mathbf{L}'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right). \tag{8}$$

The distribution of F_{KR} is approximated by supposing that there are positive numbers λ and m such that the distribution of λF_{KR} is approximately an F distribution with ℓ numerator degrees of freedom and m denominator degrees of freedom, that is,

$$\lambda F_{\mathrm{KR}} \sim \mathrm{F}\left(\ell, m\right)$$

Note that m is not required to be an integer. The values of λ and m are determined by matching the first two moments of the test statistic F_{KR} and the F distribution, that is,

$$\lambda \mathrm{E}\left(F_{\mathrm{KR}}\right) = \mathrm{E}\left[\mathrm{F}\left(\ell, m\right)\right] \text{ and } \lambda^2 \mathrm{Var}\left(F_{\mathrm{KR}}\right) = \mathrm{Var}\left[\mathrm{F}\left(\ell, m\right)\right]. \tag{9}$$

Generally, there are no exact formulas for $\mathrm{E}\left(F_{KR}\right)$ and $\mathrm{Var}\left(F_{\mathrm{KR}}\right)$, but Taylor series expansion can be used to obtain approximate formulas, say $\hat{E}$ and $\hat{V}$.

$$\hat{E} = 1 + \frac{A_2}{\ell} \text{ and } \hat{V} = \frac{2}{\ell}\left(1 + B\right) \tag{10}$$

where $B = \frac{1}{2\ell}\left(A_1 + 6A_2\right)$, $A_1 = \sum_{i=1}^{r}\sum_{j=1}^{r}\hat{\omega}_{ij}\mathrm{tr}\left(\hat{\boldsymbol{\Delta}}\hat{\mathbf{P}}_i\right)\mathrm{tr}\left(\hat{\boldsymbol{\Delta}}\hat{\mathbf{P}}_j\right)$, $A_2 = \sum_{i=1}^{r}\sum_{j=1}^{r}\hat{\omega}_{ij}\mathrm{tr}\left(\hat{\boldsymbol{\Delta}}\hat{\mathbf{P}}_i\hat{\boldsymbol{\Delta}}\hat{\mathbf{P}}_j\right)$,

$\hat{\boldsymbol{\Delta}} = \hat{\boldsymbol{\Phi}}\mathbf{L}\left[\mathbf{L}'\hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{L}\right]^{-1}\mathbf{L}'\hat{\boldsymbol{\Phi}}$.

By matching the approximate moments of λF_{KR} and the exact moments of $\mathrm{F}\left(\ell, m\right)$:

$$\hat{m} = 4 + \frac{\ell + 2}{\ell\hat{\rho} - 1} \text{ and } \hat{\lambda} = \frac{\hat{m}}{\left(\hat{m} - 2\right)\hat{E}} \tag{11}$$

where $\hat{\rho} = \dfrac{\hat{V}}{2\left(\hat{E}\right)^2}$.

It can be seen that these formulas do not give the exact distributions of F_{KR} in the two special cases mentioned above, and Kenward and Roger (1997) make adjustments of the procedure. The adjustments can be described as modifications of the quantities $\hat{E}$ and $\hat{V}$, say E^* and V^* respectively so the produced estimates of the scale factor and the denominator degrees of freedom match the known exact values for the two special cases:

$$E^* = \left(1 - \frac{A_2}{\ell}\right)^{-1} \text{ and } V^* = \frac{2}{\ell}\left[\frac{1 + c_1 B}{\left(1 - c_2 B\right)^2\left(1 - c_3 B\right)}\right] \tag{12}$$

where $c_1 = \dfrac{g}{h}$, $c_2 = \dfrac{\ell - g}{h}$, $c_3 = \dfrac{\ell + 2 - g}{h}$ for $g = \dfrac{\left(\ell + 1\right)A_1}{\left(\ell + 2\right)A_2} - \dfrac{\ell + 4}{\ell + 2}$ and $h = 3\ell + 2\left(1 - g\right)$.

3. Testing Problems with Focused Quadratic Commutative Orthogonal Block Structure

Consider the mixed linear model as in expression (2). We assume, as is often true, the following:

(i) $\mathbf{G}_r = \mathbf{I}$, the identity matrix.

(ii) The set of possible parameter vectors $\boldsymbol{\psi}$ contains a nondegenrate rectangle in $\mathbb{R}^r$ containing the vector $\boldsymbol{\psi} = (0,\cdots,0,1)$ (so that $\boldsymbol{\Sigma} = \mathbf{I}$ is a possible variance-covariance matrix).

(iii) The matrices $\mathbf{G}_1,\ldots,\ \mathbf{G}_{r-1}$, $\mathbf{I}$ are linearly independent.

For a linear hypothesis $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{L}'\boldsymbol{\beta}_0$, it is convenient to re-express this testing problem by defining the random vector $\mathbf{z} = \boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}_0$ and the parameter vector $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}_0$. Then, the model can be written as $\mathrm{E}(\mathbf{z}) = \mathbf{X}\boldsymbol{\delta}$, $\mathrm{Var}(\mathbf{z}) = \boldsymbol{\Sigma}$, and the hypothesis as $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\delta} = \mathbf{0}$. Rather than changing notation from $\boldsymbol{\beta}$ to $\boldsymbol{\delta}$, it is more convenient to keep the original notation $\boldsymbol{\beta}$ while assuming, without loss of generality, that $\boldsymbol{\beta}_0 = \mathbf{0}$. That is the problem of testing the linear hypothesis can be stated as $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$.

Let $\mathbf{P}_\mathbf{X}$ denote the orthogonal projection matrix on $\mathcal{R}(\mathbf{X})$ which denotes the range of $\mathbf{X}$, Ψ denote the set of all possible parameter vectors $\boldsymbol{\psi}$, and let $\mathcal{V} = \{\boldsymbol{\Sigma}(\boldsymbol{\psi}) : \boldsymbol{\psi} \in \Psi\}$ *(*the set of all possible variance-covariance matrices*).*

Definition 1 (*a*) *The model is said to satisfy Zyskind's condition if the matrix* $\mathbf{P}_\mathbf{X}$ *commutes with the matrices in* $\mathcal{V}$*, i.e.,* $\mathbf{P}_\mathbf{X}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\mathbf{P}_\mathbf{X}$ *for all* $\boldsymbol{\Sigma} \in \mathcal{V}$ (*Zyskind 1967*).

(*b*) *The model has orthogonal block structure* (*OBS*) *if the matrices in* $\mathcal{V}$ *commute with one another, i.e.,* $\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i$ *for all* $\boldsymbol{\Sigma}_i$, $\boldsymbol{\Sigma}_j \in \mathcal{V}$.

(*c*) *The model has commutative orthogonal block structure* (*COBS*) *if it has OBS and satisfies Zyskind's condition.*

(*d*) *A linear subspace* S *of symmetric matrices is a quadratic subspace if* $\mathbf{A}^2 \in S$ *for all* $\mathbf{A} \in S$.

(*e*) *The model has quadratic orthogonal block structure* (*QOBS*) *if it has OBS and the linear span of* $\mathcal{V}$*, denoted by* $\mathrm{sp}(\mathcal{V})$*, is a quadratic subspace.*

(*f*) *The model has quadratic commutative orthogonal block structure (QCOBS) if it has QOBS and COBS.*

(*g*) *The testing problem is said to satisfy Zyskind's condition if both the full model and the hypothesized submodel satisfy Zyskind's condition.*

A few useful remarks about these definitions are presented below.

Remarks (a) The condition of Zyskind is equivalent to the condition that $\mathbf{P}_\mathbf{X}\mathbf{G}_i = \mathbf{P}_\mathbf{X}\mathbf{G}_i$ for all $i = 1,...,r-1$. Another equivalent condition is that the least-squares estimator of $\boldsymbol{\beta}$ is a uniformly best linear unbiased estimator, which implies $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Theorem 2 in Zyskind *(*1967*)* gives several other equivalent conditions.

(b) The definition of OBS considered here is a restatement of the original definition given by Nelder (1965) and differs from the definition given in VanLeeuwen et al. (1998) and (1999), where "OBS" is actually "QOBS" in the terminology used here.

(c) If the model has OBS, then the possible variance-covariance matrices have a simultaneous spectral decomposition, which means that there are nonzero pairwise orthogonal projection matrices $\mathbf{E}_h$ such that $\mathbf{G}_i = \sum_{h=1}^{s}\lambda_{ih}\mathbf{E}_h$ for coefficients λ_{ih} with $\lambda_{rh} = 1$ (so that $\sum_{h=1}^{s}\mathbf{E}_h = \mathbf{I}$). Then

$\boldsymbol{\Sigma} = \sum_{i=1}^{r} \psi_i \mathbf{G}_i = \sum_{h=1}^{s} \alpha_h(\boldsymbol{\psi}) \mathbf{E}_h$ where $\alpha_h(\boldsymbol{\psi}) = \sum_{i=1}^{r} \psi_i \lambda_{ih}$. An OBS model has QOBS if and only if $s = r$. To show that a model has QOBS, it suffices to show that $\mathbf{G}_i \mathbf{G}_j \in \mathrm{sp}\{\mathbf{G}_1, \cdots, \mathbf{G}_r\}$ for all $i, j = 1, \cdots, r$.

(d) The term "commutative" in commutative OBS refers to the commutativity of $\mathbf{P}_\mathbf{X}$ with all $\boldsymbol{\Sigma} \in \mathcal{V}$ and not to the commutativity of all $\boldsymbol{\Sigma} \in \mathcal{V}$ with one another. The latter kind of commutativity is referred to simply as OBS. COBS models have been studied in Fonseca et al. (2008).

Below and in Section 4, we present some useful theoretical results about models and testing problems that satisfy conditions described above in this section. Proofs of the results are provided in the Appendix.

Lemma 1 *Assume a testing problem satisfies Zyskind's condition. Let* $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{L}$, $\mathbf{M} = \mathbf{P}_\mathbf{X} - \mathbf{P}_\mathbf{U}$, *and* $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\psi}) = \boldsymbol{\Phi}\mathbf{L}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-}\mathbf{L}'\boldsymbol{\Phi}$ *(so that* $\hat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}(\hat{\boldsymbol{\psi}})$ *when* $\hat{\boldsymbol{\Phi}}_\mathrm{A} = \hat{\boldsymbol{\Phi}}$*), where* $\mathbf{U}$ *is a matrix such that* $\mathcal{R}(\mathbf{U}) = \{\mathbf{X}\boldsymbol{\beta} : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}\}$ *, the space of possible mean vectors in the hypothesized submodel. The*

(*a*) $\mathcal{R}(\boldsymbol{\Sigma}\mathbf{X}\mathbf{A}) = \mathcal{R}(\mathbf{X}\mathbf{A}) = \mathcal{R}(\mathbf{M})$.

(*b*) $\mathbf{X}\boldsymbol{\Delta}\mathbf{X}' = \boldsymbol{\Sigma}\mathbf{X}\mathbf{A}(\mathbf{A}'\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}\mathbf{A}')^{-}\mathbf{A}'\mathbf{X}'\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\mathbf{M}$.

(*c*) $F_\mathrm{KR} = \frac{1}{\ell}\boldsymbol{y}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{M}\boldsymbol{y}$.

Lemma 2 *Assume a model with QCOBS. Let* $\mathbf{N} = \mathbf{I} - \mathbf{P}_\mathbf{X}$ *and* $t_h = \mathrm{rank}(\mathbf{N}\mathbf{E}_h)$. *Then*

(*a*) *The REML estimator* $\hat{\boldsymbol{\psi}}_\mathrm{REML}$ *satisfies* $\alpha_h(\hat{\boldsymbol{\psi}}_\mathrm{REML}) = \boldsymbol{y}'\mathbf{N}\mathbf{E}_h\boldsymbol{y} / t_h$.

(*b*) *The inverse of the expected information matrix for the residual model is*

$$\tilde{\mathbf{W}} = 2\left(\mathbf{C}^{-1}\right)' \mathbf{D}\mathbf{C}^{-1}, \text{ where } \mathbf{C} = [\lambda_{ih}]_{r\times r} \text{ and } \mathbf{D} = \mathbf{D}(\boldsymbol{\psi}) = \mathrm{diag}\left[\alpha_h(\boldsymbol{\psi})^2 / t_h\right].$$

4. When the Approximate KR Test is an Exact Test

Consider the mixed linear model and the testing problem as described in Section 3.

Definition 2 *The testing problem has focused QCOBS (or FQCOBS) if (i) the full model has QCOBS, (ii) the hypothesized submodel satisfies Zyskind's condition, and (iii)* $\mathcal{R}(\mathbf{M}) \subset \mathcal{R}(\mathbf{E}_f)$ *for some* f *, where* $\mathbf{M}$ *is as described in Lemma 1, and* $\mathbf{E}_f$ *is as described in part c of the Remarks in Section 3.*

The models with QCOBS, and testing problems with FQCOBS have been studied by Rady (1986). The class of models with QCOBS includes all fixed-effects models which include one-way ANOVA models. For balanced mixed-effects ANOVA models in which the intersection of every pair of random effects is again a random effect, if the significance of a single fixed effect is tested (rather than simultaneously test more than one effect), then the testing problem has FQCOBS.

In a testing problem with FQCOBS, Rady (1986) showed that the generalized ANOVA test is an exact F test. That is,

$$F_{\text{ANOVA}} = \frac{1}{\ell}\boldsymbol{y}'\mathbf{M}\boldsymbol{y} \Big/ \frac{1}{t_f}\boldsymbol{y}'\mathbf{N}\mathbf{E}_f\boldsymbol{y} \tag{13}$$

where the test statistic F_{ANOVA} is the ratio of two mean squares with expected values:

$$\mathrm{E}\left(\frac{1}{\ell}\boldsymbol{y}'\mathbf{M}\boldsymbol{y}\right) = \frac{1}{\ell}\mathbf{M}\mathbf{X}\boldsymbol{\beta}^2 + \frac{1}{\ell}\mathrm{tr}\left(\mathbf{M}\boldsymbol{\Sigma}\right) = \frac{1}{\ell}\mathbf{M}\mathbf{X}\boldsymbol{\beta}^2 + \alpha_f\left(\boldsymbol{\psi}\right)$$

and

$$\mathrm{E}\left(\frac{1}{t_f}\boldsymbol{y}'\mathbf{N}\mathbf{E}_f\boldsymbol{y}\right) = \frac{1}{t_f}\mathrm{tr}\left(\mathbf{N}\mathbf{E}_f\boldsymbol{\Sigma}\right) = \alpha_f\left(\boldsymbol{\psi}\right) \tag{14}$$

where $\mathbf{M}$ is as described in Lemma 1, and $\mathbf{N}$ and t_f are as described in Lemma 2. The matrix $\mathbf{E}_f$ and $\alpha_f\left(\boldsymbol{\psi}\right)$ are as described in part c of the Remarks in Section 3. If the null hypothesis is true, then $\mathbf{M}\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$ and the ratio has an exact F distribution with ℓ numerator degrees of freedom and t_f denominator degrees of freedom. The p-value of the test is $\mathrm{P}\left\{\mathrm{F}\left(\ell, t_f\right) \geq F_{\text{ANOVA}}\right\}$. For such a testing problem, we will show below that the KR test coincides with this ANOVA test, that is the KR test is an exact F test and produces the exact and known values for the denominator degrees of freedom and the scale factor.

Lemma 3 *Assume a testing problem with FQCOBS. Then*

(*a*) $\mathrm{tr}\left(\boldsymbol{\Delta}\mathbf{P}_i\right) = -\ell\lambda_{if}\alpha_f\left(\boldsymbol{\psi}\right)^{-1}$.

(*b*) $\mathrm{tr}\left(\boldsymbol{\Delta}\mathbf{P}_i \boldsymbol{\Delta}\boldsymbol{\Delta}\mathbf{P}_j\right) = \ell\lambda_{if}\lambda_{jf}\alpha_f\left(\boldsymbol{\psi}\right)^{-2}$.

Theorem 1 *Assume a testing problem with FQCOBS. Then*

(*a*) $F_{\text{KR}} = F_{\text{ANOVA}} = \frac{1}{\ell}\boldsymbol{y}'\mathbf{M}\boldsymbol{y} \Big/ \frac{1}{t_f}\boldsymbol{y}'\mathbf{N}\mathbf{E}_f\boldsymbol{y}$.

(*b*) $m^* = t_f$, and $\lambda^* = 1$.

Therefore, we can say that in a testing problem with FQCOBS, the KR test and the generalized ANOVA test described above are identical, and they are an exact F test.

5. Conclusions

The approximation of the KR test (Kenward and Roger 1997) was developed by using an estimator of the variance of the fixed effects estimator developed by Kackar and Harville (1984), and using a scaled form of the Wald-type statistic to have an F distribution approximately. Then, they adjusted the approach so that the test coincides with the exact test for two special cases, namely for the one-way ANOVA fixed effects model, and Hotelling's one-sample T^2 test. Alnosaier and Birkes (2019) found that the adjustment applied to the KR test was worth doing and enhanced the performance of the KR test. Although the KR test was adjusted based on the special case of testing a fixed group effect in a one-way ANOVA model, it was showed that the test is exact for more general testing problems, that called FQCOBS testing problems. This might explain the significant improvement of the performance after applying the adjustment on the KR test. Also, it can be seen as the test approximates to the F distribution sufficiently when the test is not exact.

Appendix

A.1 Proof of Lemma 1

(a) Since $\mathbf{L} = \mathbf{X}'\mathbf{X}\mathbf{A}$, then $\mathcal{R}(\mathbf{U}) = \{\mathbf{X}\boldsymbol{\beta} : \mathbf{A}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0}\} = \mathcal{R}(\mathbf{X}) \cap \mathcal{R}(\mathbf{X}\mathbf{A})^{\perp}$.

Similarly, $\mathcal{R}(\boldsymbol{\Sigma}\mathbf{U}) = \mathcal{R}(\boldsymbol{\Sigma}\mathbf{X}) \cap \mathcal{R}(\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{A})^{\perp}$.

The Condition of Zyskind for the model and the submodel implies

$$\mathcal{R}(\boldsymbol{\Sigma}\mathbf{X}) = \mathcal{R}(\mathbf{X}) \text{ and } \mathcal{R}(\boldsymbol{\Sigma}\mathbf{U}) = \mathcal{R}(\mathbf{U}),$$

so $\mathcal{R}(\mathbf{X}) \cap \mathcal{R}(\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{A})^{\perp} = \mathcal{R}(\mathbf{U}) = \mathcal{R}(\mathbf{X}) \cap \mathcal{R}(\mathbf{X}\mathbf{A})^{\perp}$

Now, $\mathcal{R}(\mathbf{M}) = \mathcal{R}(\mathbf{P}_{\mathbf{X}} - \mathbf{P}_{\mathbf{U}}) = \mathcal{R}(\mathbf{X}) \cap \mathcal{R}(\mathbf{U})^{\perp} = \mathcal{R}(\mathbf{X}) \cap \left[\mathcal{R}(\mathbf{X}) \cap \mathcal{R}(\mathbf{X}\mathbf{A})^{\perp}\right]^{\perp}$.

Similarly, $\mathcal{R}(\mathbf{M}) = \mathcal{R}(\mathbf{X}) \cap \left[\mathcal{R}(\mathbf{X}) \cap \mathcal{R}(\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{A})^{\perp}\right]^{\perp} = \mathcal{R}(\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{A})$.

Thus, $\mathcal{R}(\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{A}) = \mathcal{R}(\mathbf{X}\mathbf{A})$, which implies $\mathcal{R}(\mathbf{X}\mathbf{A}) = \mathcal{R}(\boldsymbol{\Sigma}\mathbf{X}\mathbf{A})$.

(b) $\mathbf{X}\boldsymbol{\Delta}\mathbf{X}' = \mathbf{X}\boldsymbol{\Phi}\mathbf{L}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}\mathbf{L}'\boldsymbol{\Phi}\mathbf{X}'$, $\mathbf{X}\boldsymbol{\Phi}\mathbf{L} = \mathbf{X}\boldsymbol{\Phi}\mathbf{X}'\mathbf{X}\mathbf{A} = \boldsymbol{\Sigma}\mathbf{P}_{\mathbf{X}}\mathbf{X}\mathbf{A} = \boldsymbol{\Sigma}\mathbf{X}\mathbf{A}$, and $\mathbf{L}'\boldsymbol{\Phi}\mathbf{L} = \mathbf{A}'\mathbf{X}'\mathbf{X}\boldsymbol{\Phi}\mathbf{L} = \mathbf{A}'\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}\mathbf{A}$.

Note that $\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Delta}\mathbf{X}' = \mathbf{X}\mathbf{A}(\mathbf{A}'\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\boldsymbol{\Sigma}$ is idempotent with range space $\mathcal{R}(\mathbf{X}\mathbf{A}) = \mathcal{R}(\mathbf{M})$ (by part a), and null space $\mathcal{N}(\mathbf{A}'\mathbf{X}'\boldsymbol{\Sigma}) = \mathcal{R}(\boldsymbol{\Sigma}\mathbf{X}\mathbf{A})^{\perp} = \mathcal{R}(\mathbf{M})^{\perp} = \mathcal{N}(\mathbf{M})$, so it must coincide with $\mathbf{M}$.

(c) Recall that we assume without loss of generality that $\boldsymbol{\beta}_0 = \mathbf{0}$. We can write

$$\mathbf{L}'\hat{\boldsymbol{\beta}} = \mathbf{A}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = \mathbf{A}'\mathbf{X}'\boldsymbol{y}.$$

By Lemma 6 of Alnosaier and Birkes (2019), $\hat{\boldsymbol{\Phi}}_{\mathrm{A}} = \hat{\boldsymbol{\Phi}}$, and from the proof of part b,

$$\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L} = \mathbf{A}'\mathbf{X}'\hat{\boldsymbol{\Sigma}}\mathbf{X}\mathbf{A}.$$

Now, $F_{\mathrm{KR}} = \frac{1}{\ell}\hat{\boldsymbol{\beta}}'\mathbf{L}\left(\mathbf{L}'\hat{\boldsymbol{\Phi}}\mathbf{L}\right)^{-1}\mathbf{L}'\hat{\boldsymbol{\beta}} = \frac{1}{\ell}\boldsymbol{y}'\mathbf{X}\mathbf{A}\left(\mathbf{A}'\mathbf{X}'\hat{\boldsymbol{\Sigma}}\mathbf{X}\mathbf{A}\right)^{-1}\mathbf{A}'\mathbf{X}'\boldsymbol{y} = \frac{1}{\ell}\boldsymbol{y}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{M}\boldsymbol{y}$ by part b.

A.2 Proof of Lemma 2

(a) The REML is the MLE of $\boldsymbol{\psi}$ in the residual model, i.e., the model induced on the vector $\mathbf{N}\boldsymbol{y}$ of least-squares residuals, or equivalently, any vector $\boldsymbol{z} = \mathbf{K}'\boldsymbol{y}$ where $\mathbf{K}\mathbf{K}' = \mathbf{N}$ and $\mathbf{K}'\mathbf{K} = \mathbf{I}_{n-p}$. It is convenient to reparametrize using $\boldsymbol{\alpha} = (\alpha_1(\boldsymbol{\psi}), ..., \alpha_r(\boldsymbol{\psi}))' = \mathbf{C}'\boldsymbol{\psi}$, which is a one-to-one function of $\boldsymbol{\psi}$. The log-likelihood of the residual model is

$$\log L = constant - \frac{1}{2}\log\left|\mathbf{K}'\boldsymbol{\Sigma}\mathbf{K}\right| - \frac{1}{2}\mathbf{z}'(\mathbf{K}'\boldsymbol{\Sigma}\mathbf{K})^{-1}\mathbf{z}.$$

Recall that $\boldsymbol{\Sigma} = \sum_{h=1}^{r}\alpha_h\mathbf{E}_h$, so that $\frac{\partial}{\partial\alpha_h}\boldsymbol{\Sigma} = \mathbf{E}_h$ and

$$\frac{\partial}{\partial\alpha_h}\log L = -\frac{1}{2}\mathrm{tr}\left[(\mathbf{K}'\boldsymbol{\Sigma}\mathbf{K})^{-1}\mathbf{K}'\mathbf{E}_h\mathbf{K}\right] + \frac{1}{2}\mathbf{z}'(\mathbf{K}'\boldsymbol{\Sigma}\mathbf{K})^{-1}\mathbf{K}'\mathbf{E}_h\mathbf{K}(\mathbf{K}'\boldsymbol{\Sigma}\mathbf{K})^{-1}\mathbf{z}$$

$$= -\frac{1}{2}\mathrm{tr}(\mathbf{F}\mathbf{E}_h) + \frac{1}{2}\boldsymbol{y}'\mathbf{F}\mathbf{E}_h\mathbf{F}\boldsymbol{y}$$

where $\mathbf{F} = \mathbf{F}(\boldsymbol{\psi}) = \mathbf{K}\left(\mathbf{K}'\boldsymbol{\Sigma}\mathbf{K}\right)^{-1}\mathbf{K}'$. The product $\mathbf{F}\boldsymbol{\Sigma}$ is idempotent with range space $\mathcal{R}(\mathbf{K}) = \mathcal{R}(\mathbf{N})$ and null space $\mathcal{N}(\mathbf{K}'\boldsymbol{\Sigma}) = \mathcal{R}(\boldsymbol{\Sigma}\mathbf{K})^{\perp} = \mathcal{R}(\boldsymbol{\Sigma}\mathbf{N})^{\perp} = \mathcal{R}(\mathbf{N}\boldsymbol{\Sigma})^{\perp} = \mathcal{R}(\mathbf{N})^{\perp} = \mathcal{N}(\mathbf{N})$. Therefore $\mathbf{F}\boldsymbol{\Sigma} = \mathbf{N}$ and $\mathbf{F} = \mathbf{N}\boldsymbol{\Sigma}^{-1} = \mathbf{X}\mathbf{A}\sum_{h=1}^{r}\alpha_h^{-1}\mathbf{N}\mathbf{E}_h$, so $\mathbf{F}\mathbf{E}_h = \alpha_h^{-1}\mathbf{N}\mathbf{E}_h$ and $\frac{\partial}{\partial\alpha_h}\log L = -\frac{1}{2}\alpha_h^{-1}t_h + \frac{1}{2}\alpha_h^{-2}\boldsymbol{y}'\mathbf{N}\mathbf{E}_h\boldsymbol{y}$. Setting this derivative to 0, it can be seen that the REML of α_h is $\widehat{\alpha}_{h,\mathrm{REML}} = \boldsymbol{y}'\mathbf{N}\mathbf{E}_h\boldsymbol{y}/t_h$.

(b) $\frac{\partial^2}{\partial\alpha_h\partial\alpha_g}\log L = 0$ if $g \neq h$ and $\frac{\partial^2}{\partial\alpha_h^2}\log L = \frac{1}{2}\alpha_h^{-2}t_h - \alpha_h^{-3}\boldsymbol{y}'\mathbf{N}\mathbf{E}_h\boldsymbol{y}$.

Since $\mathrm{E}\left(\boldsymbol{y}'\mathbf{N}\mathbf{E}_h\boldsymbol{y}\right) = \mathrm{tr}\left(\mathbf{N}\mathbf{E}_h\boldsymbol{\Sigma}\right) = \mathrm{tr}\left(\mathbf{N}\mathbf{E}_h\boldsymbol{\Sigma}\alpha_h\right) = t_h\alpha_h$, then $-\mathrm{E}\left(\frac{\partial^2}{\partial\alpha_h^2}\log L\right) = \frac{1}{2}\alpha_h^{-2}t_h$. Thus, the expected information matrix for the residual model is $\frac{1}{2}\mathbf{D}^{-1}$ when $\boldsymbol{\alpha}$ is used as the parameter vector. The information matrix changes to $\frac{1}{2}\mathbf{C}\mathbf{D}^{-1}\mathbf{C}'$ when the parameter vector is changed to $\boldsymbol{\psi} = \left(\mathbf{C}'\right)^{-1}\boldsymbol{\alpha}$, so that $\tilde{\mathbf{W}}(\boldsymbol{\psi}) = 2\left(\mathbf{C}'\right)^{-1}\mathbf{D}\mathbf{C}^{-1}$.

A.3 Proof of Lemma 3

(a) $-\mathrm{tr}\left(\boldsymbol{\Delta}\mathbf{P}_i\right) = \mathrm{tr}\left(\boldsymbol{\Delta}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{X}\right) = \mathrm{tr}\left(\mathbf{X}\boldsymbol{\Delta}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\right)$

$= \mathrm{tr}\left(\mathbf{M}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\right) = \mathrm{tr}\left(\mathbf{M}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\right)$ (by Lemma 1b)

$= \mathrm{tr}\left[\mathbf{M}\left\{\sum_{h=1}^{r}\lambda_{ih}\mathbf{E}_h\right\}\left\{\sum_{h=1}^{r}\alpha_h\left(\boldsymbol{\psi}\right)^{-1}\mathbf{E}_h\right\}\right]$

$= \mathrm{tr}\left(\lambda_{if}\alpha_f\left(\boldsymbol{\psi}\right)^{-1}\mathbf{M}\right) = \lambda_{if}\alpha_f\left(\boldsymbol{\psi}\right)^{-1}\mathrm{tr}\left(\mathbf{M}\right)$

Next, $\mathrm{tr}\left(\mathbf{M}\right) = \mathrm{tr}\left(\mathbf{P}_{\mathbf{X}}\right) - \mathrm{tr}\left(\mathbf{P}_{\mathbf{U}}\right) = \mathrm{rank}\left(\mathbf{X}\right) - \mathrm{rank}\left(\mathbf{U}\right) = p - \left(p - \ell\right) = \ell$, so, $\mathrm{tr}\left(\boldsymbol{\Delta}\mathbf{P}_i\right) = -\ell\lambda_{if}\alpha_f\left(\boldsymbol{\psi}\right)^{-1}$.

(b) Similarly, $\mathrm{tr}\left(\boldsymbol{\Delta}\mathbf{P}_i\boldsymbol{\Delta}\mathbf{P}_j\right) = \mathrm{tr}\left(\mathbf{M}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{M}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{G}_j\boldsymbol{\Sigma}^{-1}\right) = \mathrm{tr}\left(\mathbf{M}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{G}_j\boldsymbol{\Sigma}^{-1}\right)$

$= \mathrm{tr}\left[\mathbf{M}\lambda_{if}\alpha_f\left(\boldsymbol{\psi}\right)^{-1}\lambda_{jf}\alpha_f\left(\boldsymbol{\psi}\right)^{-1}\right] = \ell\lambda_{if}\lambda_{jf}\alpha_f\left(\boldsymbol{\psi}\right)^{-2}$.

A.4 Proof of the Theorem

(a) By Lemma 1c and 2a, $F_{\mathrm{KR}} = \frac{1}{\ell}\boldsymbol{y}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{M}\boldsymbol{y} = \frac{1}{\ell}\boldsymbol{y}'\left[\sum_{h=1}^{r}\alpha_h\left(\hat{\boldsymbol{\psi}}\right)^{-1}\mathbf{E}_h\right]\mathbf{M}\boldsymbol{y}$.

$= \frac{1}{\ell}\boldsymbol{y}'\mathbf{M}\boldsymbol{y}/\alpha_h\left(\hat{\boldsymbol{\psi}}\right) = \frac{1}{\ell}\boldsymbol{y}'\mathbf{M}\boldsymbol{y}/\frac{1}{t_f}\boldsymbol{y}'\mathbf{N}\mathbf{E}_f\boldsymbol{y}$

(b) By Lemma 3a, $A_1 = \sum_{i=1}^{r}\sum_{j=1}^{r}\hat{\omega}_{ij}\ell^2\lambda_{if}\lambda_{jf}\alpha_f\left(\hat{\boldsymbol{\psi}}\right)^{-2} = \ell^2\alpha_f\left(\hat{\boldsymbol{\psi}}\right)^{-2}\sum_{i=1}^{r}\sum_{j=1}^{r}\hat{\omega}_{ij}\lambda_{if}\lambda_{jf}$.

Note that $\sum_{i=1}^{r}\sum_{j=1}^{r}\hat{\omega}_{ij}\lambda_{if}\lambda_{jf}$ is the *(f,f)* entry of the matrix $\mathbf{C}'\hat{\mathbf{W}}\mathbf{C}$ (by Lemma 3b)

$= 2\mathbf{C}'\left(\mathbf{C}'\right)^{-1}\hat{\mathbf{D}}\mathbf{C}^{-1}\mathbf{C} = 2\hat{\mathbf{D}}$, so $\sum_{i=1}^{r}\sum_{j=1}^{r}\hat{\omega}_{ij}\lambda_{if}\lambda_{jf} = 2\alpha_f\left(\hat{\boldsymbol{\psi}}\right)^2/t_f$ and $A_1 = 2\ell^2/t_f$.

By Lemma 3b, $A_2 = \sum_{i=1}^{r}\sum_{j=1}^{r}\hat{\omega}_{ij}\ell\lambda_{if}\lambda_{jf}\alpha_f\left(\hat{\boldsymbol{\psi}}\right)^{-2} = A_1/\ell = 2\ell/t_f$. By Lemma 1 of Alnosaier and Birkes (2019), $m^* = 2\ell/A_2 = t_f$, and $\lambda^* = 1$.

References

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

Chawla A, Maiti T, Sinha S. Kenward-Roger approximation for linear mixed models with missing covariates. Technical Report [No. RM-706]. Department of Statistics and Probability: Michigan State University. 2014.

Fai AHT, Cornelius PL. Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. J Stat Comput Simul. 1996; 54(4): 363-378.

Fonseca M, Mexia JT, Zmyślony R. Inference in normal models with commutative orthogonal block structure. Acta Comment Univ Tartu Math. 2008; 12: 3-16.

Gregory KB. A comparison of denominator degrees of freedom approximation methods in the unbalanced two-way factorial mixed model. Master's [thesis]. College Station (TX): Texas A&M University; 2011.

Halekoh U, Højsgaard S. A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models–the R package pbkrtest. J Stat Softw. 2014; 59: 1-32.

Harville DA, Jeske DR. Mean squared error of estimation or prediction under a general linear model. J Am Stat Assoc. 1992; 87 (419): 724-731.

Kackar RN, Harville DA. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. J Am Stat Assoc. 1984; 79(388): 853-862.

Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997; 53(3): 983-997.

Nelder JA. The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. Proc R Soc Lond A. 1965; 283(1393): 147-162.

Rady EA. Testing fixed effects in mixed linear models. PhD [dissertation]. Corvallis (OR): Oregon State University; 1986.

SAS Institute Inc. SAS/STAT15.2 User's Guide. Corey: SAS Institute Inc; 2021.

Schaalje GB, McBride JB, Fellingham GW. Adequacy of approximations to distributions of test statistics in complex mixed linear models. J Agric Biol Environ Stat. 2002; 7(4): 512-524.

Spilke J, Piepho HP, Hu X. A simulation study on tests of hypotheses and confidence intervals for fixed effects in mixed models for blocked experiments with missing data. J Agric Biol Environ Stat. 2005; 10 (3): 374-389.

VanLeeuwen DM, Seely JS, Birkes DS. Sufficient conditions for orthogonal designs in mixed linear models. J Stat Plan Inference. 1998; 73(1-2): 373-389.

VanLeeuwen DM, Birkes DS, Seely JS. Balance and orthogonality in designs for mixed classification models. Ann Stat. 1999; 27(6): 1927-1947.

Wulff SS, Robinson TJ. Assessing the uncertainty of regression estimates in a response surface model for repeated measures. Qual Technol Quant M. 2009; 6 (3): 309-324.

Zyskind G. On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. Ann Math Stat. 1967; 38(4): 1092-1109.