




Inner workings of the Kenward–Roger test

Waseem Alnosaier¹ · David Birkes^{2,3} 

Received: 12 August 2017 / Published online: 21 July 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

For testing a linear hypothesis about fixed effects in a normal mixed linear model, a popular approach is to use a Wald test, in which the test statistic is assumed to have a null distribution that is approximately chi-squared. This approximation is questionable, however, for small samples. In 1997 Kenward and Roger constructed a test that addresses this problem. They altered the Wald test in three ways: (a) adjusting the test statistic, (b) approximating the null distribution by a scaled F distribution, and (c) modifying the formulas to achieve an exact F test in two special cases. Alterations (a) and (b) lead to formulas that are somewhat complicated but can be explained by using Taylor series approximations and a few convenient assumptions. The modified formulas used in alteration (c), however, are more mysterious. Restricting attention to models with linear variance–covariance structure, we provide details of a derivation that justifies these formulas. We show that similar but different derivations lead to different formulas that also produce exact F tests in the two special cases and are equally justifiable. A simulation study was done for testing the equality of treatment effects in block-design models. Tests based on the different derivations performed very similarly. Moreover, the simulations confirm that alteration (c) is worthwhile. The Kenward–Roger test showed greater accuracy in its p values than did the unmodified version of the test.

Keywords Mixed linear model · Linear hypothesis · Small sample approximation · Kenward Roger test · Residual maximum likelihood estimator · Exact F test

✉ David Birkes
birkes@stat.oregonstate.edu

¹ Institute of Public Administration, Riyadh, Saudi Arabia

² Statistics Department, Oregon State University, Corvallis, OR 97331, USA

³ 3514 NE Alameda St, Portland, OR 97212, USA

1 Introduction

The Kenward–Roger (KR) test is widely used for testing linear hypotheses about fixed effects in normal mixed linear models. Following its introduction in 1997 (Kenward and Roger 1997), it has been cited in the literature more than 2500 times according to Google Scholar. Since 1999 (Stroup 1999) it has been incorporated in the SAS statistical package as the DDFM = KR option in the MODEL statement of its MIXED procedure (SAS Institute Inc 2015). The KR test has also been made available in the R package ‘pbkrtest’ (Halekoh and Højsgaard 2014) and can be found at Matlab Central as function ‘ddfmixed.m’ written by Witkovsky (2012). Simulation studies have shown that it performs well in a variety of mixed linear models (Schaalje et al. 2002; Guiard et al. 2003; Kowalchuk et al. 2004; Spilke et al. 2005; Chen 2006; Wimmer and Witkovsky 2007; Arnau et al. 2009; Wulff and Robinson 2009; Livacic-Rojas et al. 2010).

Consider a data vector \mathbf{y} whose distribution can be assumed to be multivariate normal with mean vector $\mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a vector of fixed-effect parameters, and with an invertible variance–covariance matrix $\boldsymbol{\Sigma}$ depending on a vector $\boldsymbol{\theta}$ of variance–covariance parameters. We will assume the variance–covariance structure is linear. Suppose we want to test a linear hypothesis $H_0: \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ where \mathbf{L} is a matrix whose ℓ columns are linearly independent. In developing their test, Kenward and Roger (1997) begin with the idea of a Wald test statistic of the form

$$T = \hat{\boldsymbol{\beta}}' \mathbf{L} [\mathbf{L}' \hat{\text{Var}}(\hat{\boldsymbol{\beta}}) \mathbf{L}]^{-1} \mathbf{L}' \hat{\boldsymbol{\beta}} \quad (1.1)$$

where $\hat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$ and $\hat{\text{Var}}(\hat{\boldsymbol{\beta}})$ is an estimator of $\text{Var}(\hat{\boldsymbol{\beta}})$. For large samples it is often true that a Wald test statistic like T has approximately a chi-squared distribution with ℓ degrees of freedom under the null hypothesis. For small samples, however, the null distribution of T may not be well approximated by this distribution. Kenward and Roger improve on the approximation in three ways:

- (a) They use an improved estimator $\hat{\text{Var}}(\hat{\boldsymbol{\beta}})$ based on results of Harville and coworkers (Kackar and Harville 1984; Harville and Jeske 1992).
- (b) They allow the approximating null distribution of T to be a scaled F distribution.
- (c) They modify the approximating null distribution to be exact in two special cases.

Improvements (a) and (b) mainly involve Taylor series approximations and a few convenient assumptions. The resulting formulas are somewhat complicated but the outline of their derivation in Kenward and Roger (1997) makes the formulas seem reasonable. The modified formulas used to achieve improvement (c), however, are more mysterious. Below we provide details of a derivation that justifies these formulas (see Sect. 7). We show that similar but different derivations lead to different formulas that also produce exact null distributions in the two special cases. The two alternative procedures in Sects. 8 and 9 have formulas and derivations that are somewhat simpler than the Kenward–Roger procedure. The simulation study reported in Sect. 11 suggests that the three procedures perform similarly, at least when testing the equality of treatment effects in a block-design model with random blocks.

Section 2 presents basic notation and assumptions; Sect. 3 contains notation and formulas for improvement (a); Sect. 4 describes improvement (b); a description and justification of improvement (c) are given in Sects. 5–7. The justification leads us to two variations on improvement (c) that are derived in Sects. 8 and 9. In Sect. 10 are some theoretical results about when the three modifications produce the same formulas. Simulation results are presented in Sect. 11. Details are provided in the “Appendix”.

2 The testing problem

Consider a random vector distributed according to a multivariate normal distribution:

$$\mathbf{y} \stackrel{d}{=} N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \tag{2.1}$$

with mean vector $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and variance–covariance matrix $\text{Var}(\mathbf{y}) = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ where \mathbf{X} is a known $n \times p$ matrix of full column rank, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown fixed-effect parameters, and $\boldsymbol{\Sigma}$ is an $n \times n$ positive-definite matrix depending on an $r \times 1$ vector $\boldsymbol{\theta}$ of unknown variance–covariance parameters. We will assume the model includes an intercept term. Let $\boldsymbol{\Omega}$ denote the set of allowable values of $\boldsymbol{\theta}$. We will assume that one of the allowable variance–covariance matrices is the identity matrix \mathbf{I}_n (which is true for most models). We will assume that the variance–covariance structure is intrinsically linear, so that, perhaps after reparameterization,

$$\boldsymbol{\Sigma} = \theta_1 \mathbf{G}_1 + \cdots + \theta_r \mathbf{G}_r$$

for known symmetric matrices \mathbf{G}_i . Types of variance–covariance structures that are linear include variance–components, random-coefficient, Toeplitz, Huynh–Feldt, and banded structures, as well as the unstructured structure. Two technical assumptions (which are satisfied for most models) are that the matrices \mathbf{G}_i are linearly independent and the set $\boldsymbol{\Omega}$ contains a nonempty open subset of \mathbb{R}^r .

Consider the problem of testing a linear hypothesis $H_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ where \mathbf{L} is a known $p \times \ell$ matrix of full column rank. A general approach to testing such a hypothesis is to form a Wald-type test statistic of the form

$$T = (\mathbf{L}'\hat{\boldsymbol{\beta}})'[\hat{\text{Var}}(\mathbf{L}'\hat{\boldsymbol{\beta}})]^{-1}(\mathbf{L}'\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}'\mathbf{L}[\mathbf{L}'\hat{\text{Var}}(\hat{\boldsymbol{\beta}})\mathbf{L}]^{-1}\mathbf{L}'\hat{\boldsymbol{\beta}}$$

where $\hat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$ and $\hat{\text{Var}}(\hat{\boldsymbol{\beta}})$ is an estimator of $\text{Var}(\hat{\boldsymbol{\beta}})$. According to asymptotic theory, if the sample size is large (and suitable assumptions are met), one can test the null hypothesis by rejecting it if T is greater than a critical value of the $\chi^2(\ell)$ distribution. This may not be a good test, however, when the sample size is small. As stated in Sect. 1, Kenward and Roger (1997) have introduced alterations (a), (b) and (c) in order to improve the performance of the test for small samples.

3 Choosing an estimator $\hat{\beta}$ and an estimator of its variance–covariance matrix

Kenward and Roger (1997) choose $\hat{\beta}$ to be an estimated generalized least-squares estimator (EGLSE), that is,

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

where $\hat{\Sigma} = \Sigma(\hat{\theta})$ and $\hat{\theta}$ is an estimator of θ . They choose $\hat{\theta}$ to be the residual maximum likelihood estimator (REMLE) of θ .

Let $\tilde{\beta} = \tilde{\beta}(\theta) = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y}$, which is often called a generalized least-squares estimator (GLSE). (See the remark below.) Note that the EGLSE can be written as $\hat{\beta} = \tilde{\beta}(\hat{\theta})$. Kackar and Harville (1984) expressed

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

where

$$\Phi = \Phi(\theta) = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} = \text{Var}(\tilde{\beta}) \quad \text{and} \quad \Lambda = \text{Var}(\hat{\beta} - \tilde{\beta})$$

and they approximated Λ by

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

where $\mathbf{W} = [w_{ij}]_{r \times r} = \text{Var}(\hat{\theta})$ and

$$\begin{aligned} \mathbf{P}_i &= \mathbf{P}_i(\theta) = -\mathbf{X}'\Sigma^{-1}\mathbf{G}_i\Sigma^{-1}\mathbf{X} \\ \mathbf{Q}_{ij} &= \mathbf{Q}_{ij}(\theta) = \mathbf{X}'\Sigma^{-1}\mathbf{G}_i\Sigma^{-1}\mathbf{G}_j\Sigma^{-1}\mathbf{X}. \end{aligned}$$

Remark In general, the GLSE depends on the value of the unknown parameter vector θ and therefore is not a true estimator because it cannot be calculated from the observed data. A true estimator of β is obtained by substituting an estimator $\hat{\theta}$ for θ , thus producing an EGLSE. For special models, the GLSE does not depend on the value of θ . This happens if and only if the model satisfies the condition that, for all allowable Σ , the column space of $\Sigma\mathbf{X}$ is contained in the column space of \mathbf{X} (Zyskind 1967, Theorem 2). For such models, the GLSE of β coincides with the LSE and is a uniformly best linear unbiased estimator. Zyskind’s condition is met by models such as balanced mixed-effects classification models that are ‘proper’ as defined by VanLeeuwen et al. (1999). Most, if not all, mixed-effects classification models used in practice are proper.

The REMLE $\hat{\theta}$ is derived from a residual model not involving β . Let $\tilde{\mathbf{W}} = \tilde{\mathbf{W}}(\theta)$ denote the inverse of the expected information matrix for the residual model. We can

approximate \mathbf{W} by $\hat{\mathbf{W}} = \tilde{\mathbf{W}}(\hat{\theta})$. Also define $\hat{\Phi} = \Phi(\hat{\theta})$, $\hat{\Lambda} = \tilde{\Lambda}(\hat{\theta})$, $\hat{\mathbf{P}}_i = \mathbf{P}_i(\hat{\theta})$, $\hat{\mathbf{Q}}_{ij} = \mathbf{Q}_{ij}(\hat{\theta})$.

The matrix $\hat{\Phi}$ has traditionally been used as a convenient estimator for $\text{Var}(\hat{\beta})$, but it tends to underestimate. First, although $\hat{\Phi}$ is a sensible estimator for Φ , Φ is not the same as $\text{Var}(\hat{\beta})$ unless $\Lambda = \mathbf{0}$ (which happens if and only if the model satisfies the condition of Zyskind (1967) mentioned in the remark above). The formula $\Phi + \tilde{\Lambda}$ from Kackar and Harville (1984) is a more accurate approximation for $\text{Var}(\hat{\beta})$ than Φ is, and correspondingly, $\hat{\Phi} + \hat{\Lambda}$ is a better estimator of $\text{Var}(\hat{\beta})$ than $\hat{\Phi}$ is. But $\hat{\Phi} + \hat{\Lambda}$ still tends to underestimate $\text{Var}(\hat{\beta})$. The bias is reduced by using an adjusted estimator

$$\hat{\Phi}_A = \hat{\Phi} + 2\hat{\Lambda}$$

(see Harville and Jeske 1992, Sect. 4.2).

Kenward and Roger (1997) use the test statistic T with $\hat{\text{Var}}(\hat{\beta}) = \hat{\Phi}_A$ and rescale it by dividing by ℓ :

$$F_{\text{KR}} = \frac{1}{\ell} \hat{\beta}' \mathbf{L} (\mathbf{L}' \hat{\Phi}_A \mathbf{L})^{-1} \mathbf{L}' \hat{\beta}. \tag{3.1}$$

4 Approximating the null distribution of the KR test statistic

Kenward and Roger (1997) approximate the null distribution of F_{KR} by supposing that there are positive numbers m and λ such that the null distribution of λF_{KR} is approximately an F distribution with ℓ numerator degrees of freedom and m denominator degrees of freedom. It is not required that m be an integer. The values of λ and m are determined by matching moments.

Generally there are no exact formulas for the moments $E(F_{\text{KR}})$ and $\text{Var}(F_{\text{KR}})$, but by using Taylor series expansions, Kenward and Roger (1997) obtain the following approximate formulas:

$$E(F_{\text{KR}}) \approx E^\# = 1 + \frac{1}{\ell} A_2 \quad \text{and} \quad \text{Var}(F_{\text{KR}}) \approx V^\# = \frac{2}{\ell} (1 + B) \tag{4.1}$$

where

$$\begin{aligned} A_2 &= \sum_{i=1}^r \sum_{j=1}^r \hat{w}_{ij} \text{tr}(\hat{\Psi} \hat{\mathbf{P}}_i \hat{\Psi} \hat{\mathbf{P}}_j) \\ B &= \frac{1}{2\ell} (A_1 + 6A_2) \\ A_1 &= \sum_{i=1}^r \sum_{j=1}^r \hat{w}_{ij} \text{tr}(\hat{\Psi} \hat{\mathbf{P}}_i) \text{tr}(\hat{\Psi} \hat{\mathbf{P}}_j) \\ \Psi &= \Psi(\theta) = \Phi \mathbf{L} (\mathbf{L}' \Phi \mathbf{L})^{-1} \mathbf{L}' \Phi, \quad \hat{\Psi} = \Psi(\hat{\theta}). \end{aligned}$$

We can approximate the null distribution of F_{KR} by

$$\lambda^\# F_{KR} \stackrel{d}{\approx} F(\ell, m^\#) \tag{4.2}$$

where $\lambda^\#$ and $m^\#$ are chosen so that the approximate moments of $\lambda^\# F_{KR}$ match the exact moments of $F(\ell, m^\#)$:

$$\begin{aligned} \lambda^\# E^\# &= E[F(\ell, m^\#)] = \frac{m^\#}{m^\# - 2} \\ (\lambda^\#)^2 V^\# &= \text{Var}[F(\ell, m^\#)] = 2 \left(\frac{m^\#}{m^\# - 2} \right)^2 \frac{(\ell + m^\# - 2)}{\ell(m^\# - 4)}. \end{aligned}$$

Solve for $\lambda^\#$ and $m^\#$:

$$m^\# = 4 + \frac{\ell + 2}{\ell \rho^\# - 1} \quad \text{and} \quad \lambda^\# = \frac{m^\#}{(m^\# - 2)E^\#} \tag{4.3}$$

where

$$\rho^\# = \frac{V^\#}{2(E^\#)^2}.$$

5 First special case: balanced ANOVA model

Consider a balanced one-way ANOVA fixed-effects model,

$$y_{ij} = \mu_i + e_{ij}$$

for $i = 1, \dots, t, j = 1, \dots, v$. The quantities μ_i are unknown fixed effects and the e_{ij} are unobservable i.i.d. random variables from a $N(0, \sigma^2)$ population. Let us test $H_0: \mu_1 = \dots = \mu_t$. This is a special case of the testing problem in Sect. 2 with $n = tv, p = t, r = 1$ and $\ell = t - 1$ (see A.1 in the ‘‘Appendix’’). One can calculate:

$$\begin{aligned} \hat{\Sigma} &= \hat{\sigma}^2 \mathbf{I}_n, \quad \hat{\sigma}^2 = \hat{\sigma}_{REML}^2 = \frac{\sum_{i=1}^t \sum_{j=1}^v (y_{ij} - \bar{y}_{i.})^2}{n - t} \\ \hat{\mu}_i &= \bar{y}_{i.}, \quad \hat{\Phi}_A = \hat{\Phi} = \left(\frac{\hat{\sigma}^2}{v} \right) \mathbf{I}_t \\ F_{KR} &= \frac{v \sum_{i=1}^t (\bar{y}_{i.} - \bar{y}_{..})^2 / (t - 1)}{\hat{\sigma}^2}. \end{aligned}$$

It is well known that in a normal balanced one-way ANOVA fixed-effects model, the null distribution of the statistic F_{KR} above is exactly the $F(t - 1, n - t)$ distribution (Kuehl 2000, p. 57). That is,

$$\lambda F_{KR} \stackrel{d}{=} F(\ell, m) \quad \text{with } m = n - t \quad \text{and } \lambda = 1. \tag{5.1}$$

To see whether the values of $m^\#$ and $\lambda^\#$ (defined in Sect. 4 above) are equal to the “ideal” values $n - t$ and 1, one can calculate:

$$E^\# = 1 + \frac{2}{n - t}, \quad V^\# = \frac{2}{t - 1} \left(1 + \frac{t + 5}{n - t} \right) \tag{5.2}$$

$$m^\# = 4 + (n - t) \frac{[1 + 2/(n - t)]^2}{1 - 4/[(t + 1)(n - t)]} > n - t \tag{5.3a}$$

$$\lambda^\# = \frac{m^\#(n - t)}{(m^\# - 2)(n - t + 2)} = \left(1 + \frac{2}{m^\# - 2} \right) \left(1 - \frac{2}{n - t + 2} \right) < 1. \tag{5.3b}$$

Neither $m^\#$ nor $\lambda^\#$ has the ideal value in this case.

6 Second special case: Hotelling T-squared test

Suppose $\mathbf{y}_1, \dots, \mathbf{y}_v$ are v independent and identically distributed random $p \times 1$ vectors with

$$\mathbf{y}_k \stackrel{d}{=} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_p)$$

for $k = 1, \dots, v$. Let us test $H_0: \boldsymbol{\mu} = \mathbf{0}$. This is a special case of the testing problem in Sect. 2 with $n = vp$, $p = p$, $r = p(p + 1)/2$ and $\ell = p$ (see A.5 in the “Appendix”). One can calculate:

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \mathbf{I}_v \otimes \mathbf{S}, \quad \mathbf{S} = \hat{\boldsymbol{\Sigma}}_{p\text{REML}} = \frac{\sum_{k=1}^v (\mathbf{y}_k - \bar{\mathbf{y}}.) (\mathbf{y}_k - \bar{\mathbf{y}}.)'}{v - 1} \\ \hat{\boldsymbol{\mu}} &= \bar{\mathbf{y}}., \quad \hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}} = \frac{1}{v} \mathbf{S} \\ F_{KR} &= \frac{v}{p} \bar{\mathbf{y}}.' \mathbf{S}^{-1} \bar{\mathbf{y}}. \end{aligned}$$

In this setting it is common to apply the one-sample Hotelling T-squared test, in which the test statistic is $T^2 = pF_{KR}$. It is known (Mardia et al. 1979, Section 5.2.1b) that the null distribution of $[(v - p)/(v - 1)p]T^2 = [(v - p)/(v - 1)]F_{KR}$ is exactly $F(p, v - p)$. That is,

$$\lambda F_{KR} \stackrel{d}{=} F(\ell, m) \quad \text{with } m = v - p \quad \text{and } \lambda = \frac{v - p}{v - 1}. \tag{6.1}$$

(We require $v > p$.) To see whether the values of $m^\#$ and $\lambda^\#$ are equal to the ideal values $v - p$ and $(v - p)/(v - 1)$, one can calculate:

$$E^\# = 1 + \frac{p + 1}{v - 1}, \quad V^\# = \frac{2}{p} \left(1 + \frac{3p + 4}{v - 1} \right) \tag{6.2}$$

$$m^\# = 4 + (v - p) \frac{[1 + 2p/(v - p)]^2}{1 - (p + 3)/[(p + 2)(v - p)]} > v - p \tag{6.3a}$$

$$\lambda^\# = \frac{m^\#(v - 1)}{(m^\# - 2)(v + p)}. \tag{6.3b}$$

Again in this second special case we see that $m^\#$ does not have the ideal value. And for most (if not all) choices of p and v , neither does $\lambda^\#$. For example, for $p = 2$ and $v = 10$ we get $\lambda^\# = 57/70 \neq 8/9 = (v - p)/(v - 1)$.

7 Kenward and Roger’s modification of the approximate null distribution

We see that in the two special cases the formulas (4.1) for $E^\#$ and $V^\#$, when plugged into formulas (4.3) for $m^\#$ and $\lambda^\#$, do not achieve the ideal values of m and λ . Kenward and Roger (1997) modified formulas (4.1) to obtain approximations E^* and V^* with the desirable property that the KR test in the two special cases coincides with the exact test.

In the ANOVA special case,

$$E(F_{KR}) = E[F(t - 1, n - t)] = \frac{n - t}{n - t - 2} = \frac{1}{1 - \frac{2}{n-t}} = \frac{1}{1 - \frac{A_2}{\ell}}.$$

In the Hotelling special case,

$$E(F_{KR}) = E \left[\left(\frac{v - 1}{v - p} \right) F(p, v - p) \right] = \frac{v - 1}{v - p - 2} = \frac{1}{1 - \frac{p+1}{v-1}} = \frac{1}{1 - \frac{A_2}{\ell}}.$$

Thus we are led to the formula

$$E^* = \frac{1}{1 - \frac{A_2}{\ell}}, \tag{7.1}$$

which Kenward and Roger apply to all models in the class described in Sect. 2. Note that formula (7.1) makes sense from an asymptotic viewpoint, because for large ℓ the quantity $\varepsilon = A_2/\ell$ becomes small, and for small ε we have $E^\# = 1 + \varepsilon \approx 1/(1 - \varepsilon) = E^*$.

Next consider $\text{Var}(F_{KR})$. The formula for $V^\#$ is a function of ℓ and B , so let us choose V^* to have this same feature. We will look at the exact values of $\text{Var}(F_{KR})$ under the null hypothesis in the two special cases and express them in terms of ℓ and B .

In the ANOVA special case, (5.1) states that, under the null hypothesis, $F_{KR} \stackrel{d}{=} F(\ell, m)$ with $m = n - t$, so:

$$\text{Var}(F_{KR}) = \text{Var}[F(\ell, m)] = 2 \left(\frac{m}{m-2} \right)^2 \frac{(\ell + m - 2)}{\ell(m-4)}.$$

This formula is in terms of ℓ and m , but we can express m in terms of ℓ and B . Calculate $B = (t + 5)/(n - t) = (\ell + 6)/m$ (see A.2 in the ‘‘Appendix’’) and write $1/m = B/(\ell + 6)$. Now

$$\begin{aligned} \text{Var}(F_{KR}) &= \frac{2}{\ell} \left(\frac{1}{1 - 2/m} \right)^2 \frac{[1 + (\ell - 2)/m]}{(1 - 4/m)} \\ &= \frac{2}{\ell} \frac{\left(1 + \frac{\ell-2}{\ell+6} B \right)}{\left(1 - \frac{2}{\ell+6} B \right)^2 \left(1 - \frac{4}{\ell+6} B \right)}. \end{aligned}$$

In the Hotelling special case, (6.1) implies that, under the null hypothesis, $F_{KR} \stackrel{d}{=} F(\ell, m)/\lambda$ with $m = v - p$ and $\lambda = (v - p)/(v - 1)$, so:

$$\text{Var}(F_{KR}) = \text{Var}[F(\ell, m)/\lambda] = 2 \left(\frac{m}{m-2} \right)^2 \frac{(\ell + m - 2)}{\ell(m-4)} \bigg/ \lambda^2.$$

Recall $p = \ell$, so that $\lambda = m/(m + \ell - 1)$ and

$$\text{Var}(F_{KR}) = \frac{2}{\ell} \left(\frac{m + \ell - 1}{m - 2} \right)^2 \frac{(m + \ell - 2)}{(m - 4)}.$$

To bring B into the formula, calculate $B = (3p + 4)/(v - 1) = (3\ell + 4)/(m + \ell - 1)$ (see A.6 in the ‘‘Appendix’’). Let $k = m + \ell - 1$ and write $1/k = B/(3\ell + 4)$. Now

$$\begin{aligned} \text{Var}(F_{KR}) &= \frac{2}{\ell} \left(\frac{k}{k - \ell - 1} \right)^2 \frac{(k - 1)}{(k - \ell - 3)} = \frac{2}{\ell} \left[\frac{1}{1 - (\ell + 1)/k} \right]^2 \frac{(1 - 1/k)}{[1 - (\ell + 3)/k]} \\ &= \frac{2}{\ell} \frac{\left(1 + \frac{-1}{3\ell+4} B \right)}{\left(1 - \frac{\ell+1}{3\ell+4} B \right)^2 \left(1 - \frac{\ell+3}{3\ell+4} B \right)}. \end{aligned}$$

In both cases, the formula for the exact value of the variance of F_{KR} has the form:

$$\text{Var}(F_{KR}) = \frac{2}{\ell} \frac{(1 + d_1 B)}{(1 - d_2 B)^2 (1 - d_3 B)}.$$

Let $\mathbf{d} = (d_1, d_2, d_3)$.

Case 1:

$$\mathbf{d} = \left(\frac{\ell-2}{\ell+6}, \frac{2}{\ell+6}, \frac{4}{\ell+6} \right)$$

Case 2:

$$\mathbf{d} = \left(\frac{-1}{3\ell+4}, \frac{\ell+1}{3\ell+4}, \frac{\ell+3}{3\ell+4} \right)$$

We need general formulas for these coefficients that reduce to the desired values in the two cases. Write $d_1 = g/h$.

Case 1:

$$g = \ell - 2, \quad h = \ell + 6$$

Case 2:

$$g = -1, \quad h = 3\ell + 4$$

Looking at the numerators of the ratios d_i , we see that in both cases

$$\mathbf{d} = \left(\frac{g}{h}, \frac{\ell - g}{h}, \frac{\ell - g + 2}{h} \right). \quad (7.2)$$

Formulas (4.1) for $E^\#$ and $V^\#$ derived by Kenward and Roger (1997) as initial approximations of $E(F_{KR})$ and $\text{Var}(F_{KR})$ follow naturally from Taylor expansions and from certain simplifying assumptions. The formulas that they constructed for g and h appear to be somewhat more improvised. Just like formulas (4.1), the formulas for g and h are functions of the quantities ℓ , A_1 , and A_2 . Kenward and Roger essentially chose to express g and h as linear functions of the ratio A_1/A_2 with coefficients that are functions of ℓ . That is,

$$g = a_0 + a_1 \frac{A_1}{A_2} \quad \text{and} \quad h = b_0 + b_1 \frac{A_1}{A_2}$$

for coefficients determined so that g and h have the desired values in the two special cases. It is equivalent if one expresses $h = c_0 + c_1 g$ as a linear function of g , and this leads to simpler coefficients.

Case 1:

$$\frac{A_1}{A_2} = \ell, \quad \ell - 2 = a_0 + a_1 \ell, \quad \ell + 6 = c_0 + c_1(\ell - 2)$$

Case 2:

$$\frac{A_1}{A_2} = \frac{2}{\ell + 1}, \quad -1 = a_0 + a_1 \frac{2}{\ell + 1}, \quad 3\ell + 4 = c_0 + c_1(-1)$$

These equations can be solved to obtain a_0, a_1, c_0, c_1 :

$$g = \frac{-(\ell + 4) + (\ell + 1)(A_1/A_2)}{\ell + 2} \quad \text{and} \quad h = 3\ell + 2 - 2g. \tag{7.3}$$

Thus we arrive at the formula for the Kenward–Roger approximation of the variance of their test statistic:

$$V^* = \frac{2}{\ell} \frac{(1 + d_1 B)}{(1 - d_2 B)^2(1 - d_3 B)} \tag{7.4}$$

where d_1, d_2, d_3 are given by formulas (7.2) and (7.3). The modified approximate null distribution of the Kenward–Roger test statistic is given by

$$\lambda^* F_{KR} \stackrel{d}{\approx} F(\ell, m^*) \tag{7.5}$$

where m^* and λ^* are calculated as in (4.3), replacing $E^\#$ and $V^\#$ by E^* and V^* . Approximation (7.5) is preferable to approximation (4.2) in so far as it reproduces the exact test in the two special cases. Moreover, in the simulation study reported in Sect. 11 below it is seen that, even in situations where no exact test is available, the modified approximation (7.5) does better than approximation (4.2).

8 An alternative modification

As will be seen in Sect. 11, the modification described in Sect. 7 is an important step in the development of the KR test. The essential idea of the modification is to find approximations E^* and V^* that lead to values m^* and λ^* such that (7.5) holds approximately under the null hypothesis and is exact in the two special cases. Some of the formulas in the modification, particularly (7.2), (7.3) and (7.4) that are used to calculate V^* , might appear to be somewhat arbitrary. Indeed there are alternative modifications that achieve the same goal. In this section we derive m^\dagger and λ^\dagger , different from m^* and λ^* , such that the null distribution of the KR test statistic is given approximately by

$$\lambda^\dagger F_{KR} \stackrel{d}{\approx} F(\ell, m^\dagger) \tag{8.1}$$

with exact equality in distribution for the two special cases.

We continue to use approximation E^* shown in (7.1). Rather than develop a formula for approximating $\text{Var}(F_{KR})$, we will obtain a formula for m^\dagger directly.

First, in the two special cases let us express m as a function of ℓ and B . Formulas for B are given in Sect. 7 and can be rearranged to obtain:

$$m = \frac{\ell + 6}{B} \text{ in case 1} \quad \text{and} \quad m = -(\ell - 1) + \frac{3\ell + 4}{B} \text{ in case 2.}$$

In both cases, m equals a linear function of $1/B$ with coefficients involving ℓ , so we choose m^\dagger to be such a function: $m^\dagger = e_0 + e_1/B$.

Case 1:

$$m = n - t, \quad e_0 = 0, \quad e_1 = \ell + 6$$

Case 2:

$$m = v - p, \quad e_0 = 1 - \ell, \quad e_1 = 3\ell + 4$$

The coefficients e_0 and e_1 depend on the model and to account for this we express them as functions of A_1/A_2 . For simplicity we choose linear functions:

$$e_0 = f_0 + f_1 \frac{A_1}{A_2} \quad \text{and} \quad e_1 = g_0 + g_1 \frac{A_1}{A_2}$$

in which the coefficients are functions of ℓ and are determined so that e_0 and e_1 have the desired values in the two special cases. It is equivalent if one expresses $e_1 = h_0 + h_1 e_0$ as a linear function of e_0 , and the coefficients are simpler.

Case 1:

$$\frac{A_1}{A_2} = \ell, \quad 0 = f_0 + f_1 \ell, \quad \ell + 6 = h_0 + h_1(0)$$

Case 2:

$$\frac{A_1}{A_2} = \frac{2}{\ell + 1}, \quad 1 - \ell = f_0 + f_1 \frac{2}{\ell + 1}, \quad 3\ell + 4 = h_0 + h_1(1 - \ell)$$

Solving for f_0, f_1, h_0, h_1 , we obtain

$$e_0 = \frac{\ell + 1}{\ell + 2} \left(\frac{A_1}{A_2} - \ell \right) \quad \text{and} \quad e_1 = \ell + 6 - 2e_0. \quad (8.2)$$

Therefore, our alternative modified approximation for the null distribution of the Kenward–Roger test statistic is given by (8.1) where

$$m^\dagger = e_0 + \frac{e_1}{B} \quad \text{and} \quad \lambda^\dagger = \frac{m^\dagger}{(m^\dagger - 2)E^\dagger}, \quad (8.3)$$

with e_0 and e_1 given in (8.2), and $E^\dagger = E^*$ in (7.1).

This alternative modification is still an improvisation but its formulas (see (7.1), (8.2), (8.3)) are simpler than those appearing in the modification derived by Kenward and Roger (1997); see formulas (7.1)–(7.4) and (4.3) above. Simpler formulas are more appealing, but a more important finding is that our alternative procedure performs very nearly the same as the original modification in the simulation study reported in Sect. 11 below. This reassures us that, even though some of the Kenward–Roger formulas may seem arbitrary and though equally justifiable alternative formulas exist, the particular choice of modification seems to have little effect on the performance of the KR test.

9 Another alternative modification

In this section we present another pair of values m^\ddagger and λ^\ddagger such that

$$\lambda^\ddagger F_{KR} \stackrel{d}{\approx} F(\ell, m^\ddagger) \tag{9.1}$$

under the null hypothesis, with exact equality in distribution for the two special cases. The initial, unmodified version of the KR test in Sect. 4 above uses $F(\ell, m^\#)/\lambda^\#$ as the approximate null distribution of F_{KR} where $m^\#$ and $\lambda^\#$ are calculated from $E^\#$ and $V^\#$, which are Taylor approximations to the mean and variance of F_{KR} . The improved, modified version of the KR test in Sect. 7 above uses $F(\ell, m^*)/\lambda^*$ where m^* and λ^* are calculated from the modified quantities E^* and V^* , which are extrapolated from the exact values of the mean and variance of F_{KR} in the two special cases. The alternative modification in Sect. 8 uses $F(\ell, m^\dagger)/\lambda^\dagger$ where m^\dagger is obtained directly by extrapolating from the exact values of m in the two special cases, and λ^\dagger is calculated from $E^\dagger = E^*$ and m^\dagger . Now in this section, a second alternative modification is described that uses $F(\ell, m^\ddagger)/\lambda^\ddagger$ where m^\ddagger is again obtained directly by extrapolating from the exact values of m in the two special cases but using a different extrapolation procedure.

In Sect. 8, m^\dagger is expressed as a linear function of $1/B$, in which the coefficients are linear functions of A_1/A_2 , in which the coefficients are functions of ℓ :

$$m^\dagger = \left(\frac{\ell + 1}{\ell + 2}\right) \left(\frac{A_1}{A_2} - \ell\right) + \left[(\ell + 6) - 2 \left(\frac{\ell + 1}{\ell + 2}\right) \left(\frac{A_1}{A_2} - \ell\right) \right] \frac{1}{B},$$

which can be rewritten as a ratio of two quadratic functions of A_1 and A_2 with coefficients that are functions of ℓ :

$$m^\dagger = \frac{2\ell(\ell + 2)(\ell + 6)A_2 + (\ell + 1)(A_1 - \ell A_2)(A_1 + 6A_2 - 4\ell)}{(\ell + 2)(A_1 + 6A_2)A_2}.$$

For the alternative modification presented in this section, m^\ddagger will be expressed as a ratio of two linear functions of A_1 and A_2 with coefficients that are functions of ℓ .

Formulas for the exact values of m in the two special cases are displayed in Sect. 8 in terms of ℓ and B . Note that each of the two formulas can be written as a ratio of two linear functions of A_1 and A_2 with coefficients that are functions of ℓ and with 1 as the constant term in the numerator function and 0 as the constant term in the denominator function:

Case 1:

$$m = \frac{\ell + 6}{B} = \frac{1}{\frac{1}{2\ell(\ell+6)}A_1 + \frac{3}{\ell(\ell+6)}A_2}$$

Case 2:

$$m = -(\ell - 1) + \frac{3\ell + 4}{B} = \frac{1 - \frac{\ell-1}{2\ell(3\ell+4)}A_1 - \frac{3(\ell-1)}{\ell(3\ell+4)}A_2}{\frac{1}{2\ell(3\ell+4)}A_1 + \frac{3}{\ell(3\ell+4)}A_2}$$

Both of these formulas have the form:

$$m = \frac{1 + c_1A_1 + c_2A_2}{d_1A_1 + d_2A_2}.$$

We can extrapolate from the two special cases by choosing c_1, c_2, d_1, d_2 such that:

$$c_1A_1 + c_2A_2 = 0 \tag{9.2a}$$

$$d_1A_1 + d_2A_2 = \frac{1}{2\ell(\ell + 6)}A_1 + \frac{3}{\ell(\ell + 6)}A_2 \tag{9.2b}$$

in case 1, that is, when $A_1 = 2\ell^2/m$ and $A_2 = 2\ell/m$ (see A.2 in the ‘‘Appendix’’), and such that:

$$c_1A_1 + c_2A_2 = -\frac{\ell - 1}{2\ell(3\ell + 4)}A_1 - \frac{3(\ell - 1)}{\ell(3\ell + 4)}A_2 \tag{9.2c}$$

$$d_1A_1 + d_2A_2 = \frac{1}{2\ell(3\ell + 4)}A_1 + \frac{3}{\ell(3\ell + 4)}A_2 \tag{9.2d}$$

in case 2, that is, when $A_1 = 2\ell/(m + \ell - 1)$ and $A_2 = \ell(\ell + 1)/(m + \ell - 1)$ (see A.6 in the ‘‘Appendix’’). It is convenient to divide each of the Eq. (9.2) by A_2 . Equations (9.2a) and (9.2c) become:

$$\ell c_1 + c_2 = 0, \quad \frac{2}{\ell + 1}c_1 + c_2 = -\frac{\ell - 1}{\ell(\ell + 1)}$$

which implies $c_1 = 1/[\ell(\ell + 2)]$ and $c_2 = -1/(\ell + 2)$. Equations (9.2b) and (9.2d) become:

$$\ell d_1 + d_2 = \frac{1}{2\ell}, \quad \frac{2}{\ell + 1}d_1 + d_2 = \frac{1}{\ell(\ell + 1)}$$

which implies $d_1 = 1/[2\ell(\ell + 2)]$ and $d_2 = 1/[\ell(\ell + 2)]$. Thus we take the denominator degrees of freedom in approximation (9.1) to be:

$$m^\ddagger = \frac{2\ell(\ell + 2) + 2(A_1 - \ell A_2)}{A_1 + 2A_2}. \tag{9.3}$$

Let

$$\lambda^\ddagger = \frac{m^\ddagger}{(m^\ddagger - 2)E^\ddagger}$$

where $E^\ddagger = E^*$ in (7.1).

10 When the three modifications are identical

Under certain conditions, the KR modification in Sect. 7 above and the two alternative modifications in Sects. 8 and 9 are identical. For proofs of the results in this section, see the “Appendix”.

Lemma 1 (a) *If $A_1/A_2 = \ell$, then the three modifications are identical:*

$$m^* = m^\dagger = m^\ddagger = 2\ell/A_2 \quad \text{and} \quad \lambda^* = \lambda^\dagger = \lambda^\ddagger = 1.$$

(b) *If $A_1/A_2 = 2/(\ell + 1)$, then the three modifications are identical:*

$$m^* = m^\dagger = m^\ddagger = \frac{\ell(\ell + 1)}{A_2} - (\ell - 1) \quad \text{and} \quad \lambda^* = \lambda^\dagger = \lambda^\ddagger = 1 - \frac{\ell - 1}{\ell(\ell + 1)} A_2.$$

Lemma 2 *When $\ell = 1$, then $A_1 = A_2$.*

Theorem 1 *When $\ell = 1$, the three modifications are identical, with denominator degrees of freedom $m^* = 2/A_2$ and scale factor $\lambda^* = 1$.*

The fact that $\ell = 1$ implies $\lambda^* = 1$ is stated in Kenward and Roger (1997, p. 988).

Suppose the design matrix of a model is partitioned as $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]$ so that $E(\mathbf{y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$. Note that

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

Lemma 3 *In a model partitioned as above, suppose:*

- (a) *the null hypothesis involves only the parameters in $\boldsymbol{\beta}_2$ so that $\mathbf{L}'\boldsymbol{\beta} = \mathbf{L}'_2\boldsymbol{\beta}_2$;*
- (b) $\mathbf{X}'_1\boldsymbol{\Sigma}^{-1}\mathbf{X}_2 = \mathbf{0}$;
- (c) $\mathbf{X}'_2\boldsymbol{\Sigma}^{-1}\mathbf{X}_2 = f(\boldsymbol{\theta})\mathbf{C}$

where f is a scalar-valued function and \mathbf{C} is a matrix not depending on $\boldsymbol{\theta}$. Then $A_1/A_2 = \ell$.

Consider a balanced incomplete block design (BIBD) with s blocks, each containing k plots, and with t treatments, each applied to r plots. For each pair of treatments, the number of blocks in which the two treatments appear together is the same number, say g , for all pairs. Suppose the treatment effects are fixed and the block effects are random. The model can be written as $y_{iju} = \mu + \tau_i + b_j + e_{iju}$ for $i = 1, \dots, t, j = 1, \dots, s, u = 1, \dots, n_{ij}$ (all n_{ij} are either 0 or 1) where the b_j 's and e_{iju} 's are independent of one another and are normally distributed with $E(b_j) = E(e_{iju}) = 0, \text{Var}(b_j) = \sigma_b^2, \text{Var}(e_{iju}) = \sigma_e^2$. In matrix notation, $\mathbf{y} = \mathbf{1}_n\mu + \mathbf{T}\boldsymbol{\tau} + \mathbf{B}\mathbf{b} + \mathbf{e}$ where $\mathbf{1}_n$ is an $n \times 1$ vector of 1's, $n = n_{..} = tr = sk, \boldsymbol{\tau} = (\tau_1, \dots, \tau_t)'$, $\mathbf{b} = (b_1, \dots, b_s)'$, $\mathbf{T}'\mathbf{T} = r\mathbf{I}_t, \mathbf{B}'\mathbf{B} = k\mathbf{I}_s, \mathbf{T}'\mathbf{B} = \mathbf{N} = [n_{ij}]_{t \times s}$. The design matrix for the fixed effects is $[\mathbf{1}_n \ \mathbf{T}]$, which does not have full column rank, and so to achieve the assumptions of model (2.1) we reparameterize by setting $\mu^* = \mu + \bar{\tau}$, and $\tau_i^* = \tau_i - \bar{\tau}$, for $i = 1, \dots, t - 1$.

Then $\mu + \tau_i = \mu^* + \tau_i^*$ for $i = 1, \dots, t - 1$ and $\mu + \tau_t = \mu^* - \tau_1^* - \dots - \tau_{t-1}^*$. This is an instance of model (2.1), with a full-column-rank design matrix that is partitioned as

$$\mathbf{X} = [\mathbf{1}_n \ \mathbf{T}^*], \quad \boldsymbol{\beta} = \begin{bmatrix} \mu^* \\ \boldsymbol{\tau}^* \end{bmatrix}$$

and with $\boldsymbol{\Sigma} = \sigma_b^2 \mathbf{B}\mathbf{B}' + \sigma_e^2 \mathbf{I}_n$.

A common null hypothesis for a block design is $H_0: \tau_1 = \dots = \tau_t$ or, in terms of the reparameterization, $H_0: \tau_1^* = \dots = \tau_{t-1}^* = 0$ or $H_0: \boldsymbol{\tau}^* = \mathbf{0}$. This testing problem satisfies the conditions of Lemma 3 with $\mathbf{X}_1 = \mathbf{1}_n$, $\mathbf{X}_2 = \mathbf{T}^*$, $\mathbf{L}_2 = \mathbf{I}_{t-1}$.

Theorem 2 *For testing the equality of the treatment effects in a BIBD model with random blocks, the three modifications are identical, with denominator degrees of freedom $m^* = 2(t - 1)/A_2$ and scale factor $\lambda^* = 1$.*

11 Simulation study

Simulations were run in order to compare the performance of the KR test with the two alternative KR-type tests as well as with the unmodified version of the KR test (see Sect. 4). The models we chose to simulate have incomplete block designs that are not balanced: complete block designs with missing observations, partially balanced incomplete block designs (PBIBDs), BIBDs with missing observations, and PBIBDs with missing observations.

The models can be written as $y_{iju} = \mu + \tau_i + b_j + e_{iju}$ for $i = 1, \dots, t, j = 1, \dots, s, u = 1, \dots, n_{ij}$ where all n_{ij} are either 0 or 1. The incidence matrix $\mathbf{N} = [n_{ij}]$ specifies the design. As in the preceding section, we can re-express the model as $y_{iju} = \mu^* + \tau_i^* + b_j + e_{iju}$ for $i = 1, \dots, t - 1$ and $y_{tju} = \mu^* - \tau_1^* - \dots - \tau_{t-1}^* + b_j + e_{tju}$ for $i = t$. The hypothesis of interest is $H_0: \tau_1^* = \dots = \tau_{t-1}^* = 0$, so the null model is $y_{iju} = \mu^* + b_j + e_{iju}$ all i, j, u . This model can be simulated by generating the b_j as an i.i.d. sample from the $N(0, \sigma_b^2)$ distribution and generating the e_{iju} as an i.i.d. sample from the $N(0, \sigma_e^2)$ distribution. For each design we considered five values of the ratio $\rho = \sigma_b/\sigma_e$, namely $\rho = 0.25, 0.5, 1, 2, 4$.

The following lemma implies there is no loss of generality in setting $\mu^* = 0$ and $\sigma_e = 1$ when simulating the null distribution of the KR test statistic. To make explicit the dependence of F_{KR} on the data let us write $F_{KR} = F_{KR}(\mathbf{y})$.

Lemma 4 *For the null model, $F_{KR}(\mathbf{y}) \stackrel{d}{=} F_{KR}(\mathbf{y}^{\S})$ where the components of \mathbf{y}^{\S} are $y_{iju}^{\S} = b_j^{\S} + e_{iju}^{\S}$ and can be simulated by generating the b_j^{\S} as an i.i.d. sample from the $N(0, \rho^2)$ distribution and generating the e_{iju}^{\S} as an i.i.d. sample from the $N(0, 1)$ distribution.*

Lemma 4 is an application of Lemma 5 below, which holds for the general testing problem described in Sect. 2.

Lemma 5 $F_{KR}(c\mathbf{y} + \mathbf{X}\mathbf{b}_0) = F_{KR}(\mathbf{y})$ for any $c \neq 0$ and any \mathbf{b}_0 satisfying $\mathbf{L}'\mathbf{b}_0 = \mathbf{0}$.

In other words, for any vector \mathbf{a} in the space $\{\mathbf{X}\mathbf{b}_0 : \mathbf{L}'\mathbf{b}_0 = \mathbf{0}\}$ of possible mean vectors in the null model, $F_{KR}(c\mathbf{y} + \mathbf{a}) = F_{KR}(\mathbf{y})$. For an incomplete block design as described above, let $c = \sigma_e^{-1}$, $\mathbf{a} = -\sigma_e^{-1}\mu^*\mathbf{1}$, and $\mathbf{y}^\S = c\mathbf{y} + \mathbf{a}$. Then $y_{iju}^\S = \sigma_e^{-1}y_{iju} - \sigma_e^{-1}\mu^* = \sigma_e^{-1}b_j + \sigma_e^{-1}e_{iju}$ and the $b_j^\S = \sigma_e^{-1}b_j$ are i.i.d. $N(0, \rho^2)$ and the $e_{iju}^\S = \sigma_e^{-1}e_{iju}$ are i.i.d. $N(0, 1)$.

The block designs we studied are listed below (n = number of observations, t = number of treatments, s = number of blocks, k = maximum block size):

D96 = a PBIBD with $n = 96, t = 16, s = 48, k = 2$ (Green 1974, p. 65).

D60 = a PBIBD with $n = 60, t = 15, s = 15, k = 4$ (Cochran and Cox 1957, p. 456).

D40 = a design with $n = 40, t = 6, s = 7, k = 6$, obtained from a complete block design with $t = 6$ and $s = 7$ by deleting two observations from different blocks and different treatments.

D21a = a design with $n = 21, t = 4, s = 9, k = 3$, obtained from a BIBD in Kuehl (2000, p. 317) by deleting run 10 and treatment 550.

D21b = a design with $n = 21, t = 9, s = 7, k = 3$, obtained from a PBIBD in Kuehl (2000, p. 329) by deleting blocks 8 and 9.

D18a = a design with $n = 18, t = 4, s = 5, k = 4$, obtained from a complete block design with $t = 4$ and $s = 5$ by deleting two observations from different blocks and different treatments.

D18b = a design with $n = 18, t = 7, s = 6, k = 3$, obtained from a BIBD in John (1971, p. 219) by deleting the last block.

D12 = a design with $n = 12, t = 6, s = 4, k = 3$, obtained from a cyclic design in Kuehl (2000, p. 346) by deleting blocks 5 and 6.

The test procedures we studied were:

KRU = the unmodified precursor to the KR test that uses the approximation developed in Sect. 4 and ignores the modification derived in Sect. 7.

KR = the test procedure presented in Kenward and Roger (1997).

KRA1 = a variation of the KR test using the alternative modification in Sect. 8.

KRA2 = a variation of the KR test using the alternative modification in Sect. 9.

For each of the 40 models (8 designs \times 5 values of ρ), we generated 10,000 independent data vectors \mathbf{y} under the null distribution. For each well-behaved (defined in Sect. 11.1 below) data vector we calculated $F_{KR}, m^\#, \lambda^\#, m^*, \lambda^*, m^\dagger, \lambda^\dagger, m^\ddagger, \lambda^\ddagger$. For each of the four tests, the p value of the test was approximated to be $p = \text{Prob}\{F(\ell, m) > \lambda F_{KR} \mid \mathbf{y}\}$ using the appropriate values of m and λ . We measure the adequacy of the approximation by the proportion of these p values that are smaller than 0.05. Ideally we want this proportion to be close to 0.05.

Table 1 The percentage of data vectors for which the REML algorithm converged, among 10,000 data vectors generated from each of 40 models (8 designs \times 5 values of ρ) under the null hypothesis. (100.0 denotes a percentage between 99.95 and 99.99% that has been rounded to one decimal place, whereas 100 denotes exactly 100%.)

	D96	D60	D40	D21a	D21b	D18a	D18b	D12
$\rho = 0.25$	100	99.9	100.0	96.1	89.0	99.5	99.6	92.1
$\rho = 0.5$	100	100.0	100.0	97.7	93.1	99.5	99.8	94.7
$\rho = 1$	100	100	100	99.5	98.0	99.9	100.0	97.5
$\rho = 2$	100	100	100	100.0	99.9	100.0	100.0	99.2
$\rho = 4$	100	100	100	100	100.0	100.0	100	99.9

11.1 Computational issues

The REML algorithm for computing estimates of the variance components sometimes failed to converge. The algorithm we used found a solution to the REML equations by iteratively applying equation (90) on p. 252 of Searle et al. (1992). For each of the 40 models, the percentage of data vectors for which the REML algorithm converged is shown in Table 1. For the three largest designs, convergence was achieved almost 100% of the time for all values of ρ . For the four smaller designs with $n = 21$ or 18, convergence was achieved almost 100% of the time when $\rho = 2$ or 4, that is when the variability of the block effects was high relative to the variability of the noise in the model. Convergence could possibly be improved by using suitable alternative numerical methods.

Another computational problem that can occur is numerical instability due to division by numbers close to zero. For example, one data set generated from the model with design D21a and $\rho = 0.25$ yielded $\lambda^* = -79700$ (to three significant digits). The value of A_2 for this data set is a very large positive number, 823000, so that $E^* = 1/(1 - A_2/\ell) = -0.00000365$ is a very small negative number, which leads to a very large negative value for $\lambda^* = [m^*/(m^* - 2)]/E^*$ due to division by E^* . Similarly, application of the two alternative KR-type procedures to this data set produced values of λ^\dagger and λ^\ddagger that were very large and negative.

The data sets generated from this model had a typical estimated scale factor λ^* of about 1, so the value $\lambda^* = -79700$ is definitely an outlier. The mere fact that it is negative conflicts with our objective of finding values of λ^* and m^* such that (7.5) is a good approximation, because an F distribution assumes only positive values. In the context of KR tests, let us say a data vector is *well-behaved with regard to the KR procedure* if (1) the REML algorithm converges, (2) the estimated expectation E^* is positive, and (3) the estimated denominator degrees of freedom m^* is > 4 . Condition (2) is appropriate because the test statistic F_{KR} is a quadratic form (see (3.1)) that is approximately positive definite, because $\hat{\Phi}_A = \hat{\Phi} + 2\hat{\Lambda}$ is approximately positive definite, because $\hat{\Phi}$ is positive definite and $\hat{\Lambda}$ is approximately equal to Λ , which is positive definite. Therefore it is reasonable to expect $E(F_{KR})$ to be positive. According to (7.1), condition (2) is equivalent to $A_2 < \ell$. Condition (3) is based on the fact that the derivation of the KR test uses the second moment of an F distribution, which requires

Table 2 The percentage of data vectors that were well-behaved with regard to the KR procedure, among 10,000 data vectors generated from each of the 25 smaller models (5 designs \times 5 values of ρ) under the null hypothesis. (100.0 denotes a percentage between 99.95 and 99.99% that has been rounded to one decimal place, whereas 100 denotes exactly 100%.)

	D21a	D21b	D18a	D18b	D12
$\rho = 0.25$	91.6	71.0	99.5	81.3	20.6
$\rho = 0.5$	95.3	81.5	99.5	88.5	19.8
$\rho = 1$	98.9	93.8	99.9	97.1	14.1
$\rho = 2$	100.0	99.5	100.0	99.8	4.9
$\rho = 4$	100	100.0	100.0	100	1.0

the ddf to be > 4 . From formula (4.3) we see that conditions (2) and (3) imply that λ^* is positive, which, as indicated in the discussion of the example above, is a sensible requirement. Replacing m^* in condition (3) by m^\dagger (or by m^\ddagger) gives the definition of *well-behaved with regard to the KRA1 (or KRA2) procedure*. The prevalence of good behavior with regard to the KR procedure is shown in Table 2; it was a problem only for the 25 smaller models (5 smallest designs \times 5 values of ρ).

The percentages of data vectors that were well-behaved with regard to the KRA1 procedure are exactly the same as the percentages for the KR procedure, for all 25 smaller models. For designs D21a and D18a, the percentages of data vectors that were well-behaved with regard to the KRA2 procedure are the same as the percentages for the KR procedure. For designs D21b and D18b, the KRA2 percentages are a little bit higher than the KR percentages but by no more than about 1 percentage point. For design D12, the KRA2 percentages are greater than the KR percentages by at most 3 percentage points.

We see that most data vectors generated from design D12 are ill-behaved. We also looked at a smaller design D10 with $n = 10, t = 5, s = 3, k = 4$, obtained from a PBIBD in Kuehl (2000, p. 323) by deleting treatment 6. For all five values of ρ , all 10,000 generated data vectors were ill-behaved. It could be worthwhile investigating other definitions of “well-behaved” in hope of including a larger proportion of “well-behaved” data vectors from smaller models.

11.2 Comparison of p values

For each of the 40 models (8 designs \times 5 values of ρ), we generated 10,000 independent data vectors \mathbf{y} under the null distribution. For each well-behaved data vector we calculated, for each of the four tests, the p value of the test. In Table 3 is displayed, for each model and each test, the observed percentage (restricting attention to well-behaved data sets) of p values that were less than 0.05. Ideally we want this percentage to be close to 5%. Of course we should keep in mind that for any given entry (that is, any given model and any given test procedure) in Table 3, if it were true that the long-run percentage of p values less than 0.05 was exactly 5%, we nevertheless could not expect the observed percentage to be exactly 5%, because of simulation error. The simulation standard error is $22\%/\sqrt{n}$ where n is the number of well-behaved data sets.

Table 3 The observed percentages of p values that were less than 0.05 for four test procedures calculated from the well-behaved data sets among 10,000 data sets generated from each of 40 models (8 designs \times 5 values of ρ)

	D96	D60	D40	D21a	D21b	D18a	D18b	D12
$\rho = 0.25$								
KRU	5.93	5.28	5.64	6.33	20.49	6.96	8.07	13.88
KR	5.68	4.66	5.21	5.47	6.66	5.57	5.78	0.83
KRA1	5.85	4.78	5.21	5.47	6.73	5.57	5.80	0.83
KRA2	5.73	4.68	5.21	5.47	6.60	5.57	5.76	0.85
$\rho = 0.5$								
KRU	5.85	5.36	5.10	7.04	16.75	6.91	8.14	13.84
KR	5.61	4.75	4.66	5.92	5.85	5.45	5.09	0.81
KRA1	5.79	4.83	4.66	5.93	5.98	5.46	5.13	0.81
KRA2	5.65	4.80	4.66	5.92	5.92	5.45	5.11	0.84
$\rho = 1$								
KRU	5.70	5.73	5.76	6.73	12.47	6.31	8.60	14.40
KR	5.40	5.21	5.24	5.45	5.73	4.90	5.05	0.92
KRA1	5.49	5.21	5.24	5.45	5.77	4.90	5.05	0.92
KRA2	5.43	5.21	5.24	5.45	5.75	4.90	5.05	0.92
$\rho = 2$								
KRU	5.57	5.48	5.85	6.58	9.74	6.22	9.10	13.16
KR	5.08	4.80	5.25	5.15	5.21	4.94	5.12	0.82
KRA1	5.11	4.80	5.25	5.15	5.22	4.94	5.12	0.82
KRA2	5.09	4.80	5.25	5.15	5.22	4.94	5.12	0.71
$\rho = 4$								
KRU	5.62	5.44	5.85	6.53	9.33	6.22	8.70	13.19
KR	4.93	4.84	5.27	5.04	4.89	4.90	4.92	0.00
KRA1	4.94	4.84	5.27	5.04	4.89	4.90	4.92	0.00
KRA2	4.93	4.84	5.27	5.04	4.89	4.90	4.92	0.00

In Table 3 we see that the three modified tests KR, KRA1 and KRA2 performed exactly the same (after rounding to two decimal places) in 22 out of the 40 models. In every one of the 40 models the percentages for the three modified KR tests were within 0.20% of one another. In many practical settings 0.20% could be regarded as being an unimportant difference and we could say that, in the models that were studied, when testing the equality of treatment effects in unbalanced incomplete block models with random blocks, the three modified KR tests gave very similar results. Moreover, recall from Sect. 10 that the three modified tests coincide when testing a one-dimensional hypothesis and when testing the equality of treatment effects in a BIBD model with random blocks.

It is clear from Tables 2 and 3 that none of the four test procedures perform adequately for models with design D12, so we will omit D12 from further discussion.

Table 3 shows that the modification step in Sect. 7 (or an alternative modification as in Sects. 8 or 9) is worth doing. The modifications in Sects. 7, 8 and 9 are all motivated by considering two special cases which can be regarded as “balanced” and admit well-known exact tests. The modifications change the formulas of KRU so as to produce these exact tests in the two special cases. In the “unbalanced” models that we simulated, the modifications do not produce exact tests but they do achieve an improvement in performance over the unmodified test. The unmodified KRU test is highly significantly liberal (i.e., its percentage in Table 3 is highly significantly larger than 5%) for 30 of the 35 models (omitting design D12). For all 35 models the modified tests are more conservative than for the unmodified test. For 23 of the 35 models the percentages for the modified tests are not significantly different from 5%, and for 11 models the modified tests are significantly liberal but less so than the unmodified test.

As would be expected, the similarity in performance of the KR, KRA1 and KRA2 tests is due in large part to the similarity of the values of m^* , m^\dagger and m^\ddagger and the values of λ^* , λ^\dagger and λ^\ddagger . See Tables 4 and 5 in A.14 in the “Appendix”. For each of the 35 models (omitting design D12), the average values of the scale factor shown in Table 5 for the three modified KR procedures are close to 1 (hence close to one another), or specifically, between 0.958 and 1.000 (when rounded to three decimal places). Recall from Sect. 10 that for all BIBD designs all three modified KR procedures have scale factor equal to 1. For each of the designs D96, D60, D40, D21a and D18a and each value of ρ , the average values of the scale factor for all three modified KR procedures are very close to 1, namely between 0.995 and 1.000. For each of the designs D21b and D18b and each value of ρ , the average scale factors for the three modified KR procedures are within 0.030 of one another. The average value of $\lambda^\#$ in the unmodified test KRU, which ranged between 0.813 and 0.995, was noticeably different from the average values of λ^* , λ^\dagger and λ^\ddagger in the three modified tests for most of the 35 models.

For all of the 35 models, the average values of m^* , m^\dagger and m^\ddagger shown in Table 4 for the three modified KR procedures were relatively close to one another. For 20 out of the 35 models, the average values of m^* , m^\dagger and m^\ddagger were close enough to one another to coincide when rounded to one decimal place. The average value of $m^\#$ in the unmodified test KRU, was noticeably different from the average values of m^* , m^\dagger and m^\ddagger in the three modified tests for most of the 35 models.

The simulations show that, for the models we studied, the modification step is an important part of the derivation of Kenward and Roger’s (1997) test and that the three modification methods presented in Sects. 7, 8 and 9 give very similar results. Our simulations are focused on block-design models—more simulation studies are needed to determine if similar statements can be made about Kenward–Roger-type tests of linear hypotheses on fixed effects in other normal mixed linear models.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

A Details

A.1 The ANOVA testing problem in Sect. 5 is a special case of Sect. 2

Testing the equality of group means in the balanced one-way ANOVA fixed-effects model in Sect. 5 is a special case of the testing problem in Sect. 2 with $\mathbf{y} = [y_{11} \ y_{12} \ \dots \ y_{1v} \ y_{21} \ \dots \ y_{tv}]'$, $\boldsymbol{\beta} = [\mu_1 \ \mu_2 \ \dots \ \mu_t]'$, $\mathbf{X} = \mathbf{I}_t \otimes \mathbf{1}_v$, $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$, $\boldsymbol{\theta} = \sigma^2$, $\mathbf{G}_1 = \mathbf{I}_n$, $\mathbf{L}' = [\mathbf{I}_{t-1} \ -\mathbf{1}_{t-1}]$.

A.2 Calculations for the ANOVA testing problem

One can calculate, $\hat{\boldsymbol{\beta}} = [\bar{y}_1. \ \bar{y}_2. \ \dots \ \bar{y}_t.]'$, $\hat{\boldsymbol{\Phi}} = (\hat{\sigma}^2/v)\mathbf{I}_t$, $\hat{w}_{11} = 2\hat{\sigma}^4/(n-t)$, $\hat{\mathbf{P}}_1 = -(v/\hat{\sigma}^4)\mathbf{I}_t$, $\hat{\boldsymbol{\Psi}} = (\hat{\sigma}^2/v)(\mathbf{I}_t - t^{-1}\mathbf{1}_t\mathbf{1}_t')$, $A_1 = 2(t-1)^2/(n-t)$, $A_2 = 2(t-1)/(n-t)$, $B = (t+5)/(n-t)$. Moreover, $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$ by the following lemma.

Lemma 6 *If the model satisfies Zyskind’s condition, then $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$.*

Note that every fixed-effects linear model satisfies Zyskind’s condition, because the column space of $\boldsymbol{\Sigma}\mathbf{X} = \sigma^2\mathbf{X}$ is contained in the column space of \mathbf{X} .

A.3 Proof of Lemma 6

We can show $\hat{\boldsymbol{\Lambda}} = \mathbf{0}$ by showing $\mathbf{Q}_{ij} - \mathbf{P}_i \boldsymbol{\Phi} \mathbf{P}_j = \mathbf{0}$. Zyskind’s condition is equivalent to the condition that $\boldsymbol{\Sigma}\mathbf{J} = \mathbf{J}\boldsymbol{\Sigma}$ for all allowable $\boldsymbol{\Sigma}$ where \mathbf{J} is the orthogonal projection operator on the column space of \mathbf{X} (Zyskind 1967, Theorem 2). The two technical assumptions mentioned in Sect. 2 imply that $\mathbf{G}_i\mathbf{J} = \mathbf{J}\mathbf{G}_i$ for all i . Moreover, $\boldsymbol{\Sigma}\mathbf{J} = \mathbf{J}\boldsymbol{\Sigma}$ implies $\boldsymbol{\Sigma}^{-1}\mathbf{J} = \mathbf{J}\boldsymbol{\Sigma}^{-1}$. Next, recall that the GLSE is $(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$ and the LSE is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. It follows from the remark in Sect. 3 that Zyskind’s condition is also equivalent to the condition that $(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ for all allowable $\boldsymbol{\Sigma}$, which implies $\mathbf{X}\boldsymbol{\Phi}\mathbf{X}'\boldsymbol{\Sigma}^{-1} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{J}$. Now

$$\begin{aligned} \mathbf{P}_i \boldsymbol{\Phi} \mathbf{P}_j &= \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{G}_j\boldsymbol{\Sigma}^{-1}\mathbf{X} \\ &= \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{J}\mathbf{G}_j\boldsymbol{\Sigma}^{-1}\mathbf{X} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{G}_j\boldsymbol{\Sigma}^{-1}\mathbf{J}\mathbf{X} \\ &= \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{G}_j\boldsymbol{\Sigma}^{-1}\mathbf{X} = \mathbf{Q}_{ij}. \end{aligned}$$

A.4 $m^\#$ and $\lambda^\#$ in the ANOVA case

Quantities from A.2 above can be substituted into (4.1) to obtain $E^\#$ and $V^\#$ shown in (5.2), which in turn can be used in (4.3) to obtain (5.3). The inequality in (5.3b) holds because $(m^\# - 2)(n - t + 2) = m^\#(n - t) + 2[m^\# - (n - t + 2)] > m^\#(n - t)$, because formula (5.3a) implies $m^\# > 4 + n - t$, which is $> n - t + 2$.

A.5 The Hotelling testing problem in Sect. 6 is a special case of Sect. 2

The one-sample Hotelling T-squared test in Sect. 6 is a special case of the testing problem in Sect. 2 with $\mathbf{y} = [y'_1 y'_2 \cdots y'_v]'$, $\boldsymbol{\beta} = \boldsymbol{\mu}$, $\mathbf{X} = \mathbf{1}_v \otimes \mathbf{I}_p$, $\boldsymbol{\Sigma} = \mathbf{I}_v \otimes \boldsymbol{\Sigma}_p$, $\boldsymbol{\Sigma}_p$ = an arbitrary positive definite $p \times p$ matrix, $\boldsymbol{\theta}$ = an $r \times 1$ vector whose entries are the entries on and above the main diagonal of $\boldsymbol{\Sigma}_p = [\sigma_{ij}]_{p \times p}$. To express the linear structure of the variance–covariance matrix it is convenient to use double subscripting: $\boldsymbol{\Sigma} = \sum\{\sigma_{ij}(\mathbf{I}_v \otimes \mathbf{G}_{pij}) : 1 \leq i \leq j \leq p\}$ where $\mathbf{G}_{pii} = \mathbf{u}_i \mathbf{u}'_i$, $\mathbf{G}_{pij} = \mathbf{u}_i \mathbf{u}'_j + \mathbf{u}_j \mathbf{u}'_i$ for $i \neq j$, \mathbf{u}_i = a $p \times 1$ vector with a 1 in the i th position and 0's elsewhere. Also, $\mathbf{L} = \mathbf{I}_p$.

A.6 Calculations for the Hotelling testing problem

One can calculate $\hat{\boldsymbol{\beta}} = \bar{\mathbf{y}}.$, $\hat{\boldsymbol{\Phi}} = (1/v)\mathbf{S}$, $\hat{w}_{ij,fg} = (\hat{\sigma}_{if}\hat{\sigma}_{jg} + \hat{\sigma}_{ig}\hat{\sigma}_{jf})/(v - 1)$ where $\hat{\sigma}_{ij}$ denotes an entry of the matrix $\mathbf{S} = \hat{\boldsymbol{\Sigma}}_{p\text{REML}}$, $\hat{\mathbf{P}}_{ii} = -v\mathbf{S}^{-1}\mathbf{u}_i \mathbf{u}'_i \mathbf{S}^{-1}$, $\hat{\mathbf{P}}_{ij} = -v\mathbf{S}^{-1}(\mathbf{u}_i \mathbf{u}'_j + \mathbf{u}_j \mathbf{u}'_i)\mathbf{S}^{-1}$ for $i < j$, $\hat{\boldsymbol{\Psi}} = (1/v)\mathbf{S}$, $A_1 = 2p/(v - 1)$, $A_2 = p(p + 1)/(v - 1)$, $B = (3p + 4)/(v - 1)$. In calculating A_1 and A_2 we use the identities $\text{tr}(\mathbf{u}_i \mathbf{u}'_j \mathbf{S}^{-1}) = \mathbf{u}'_j \mathbf{S}^{-1} \mathbf{u}_i = \hat{\sigma}^{ji}$ = the (j, i) entry of the matrix \mathbf{S}^{-1} and $\sum_{j=1}^p \hat{\sigma}_{ij} \hat{\sigma}^{js} = \delta_{is}$ = the Kronecker delta = the (i, g) entry of the matrix \mathbf{I} (because $\mathbf{S}\mathbf{S}^{-1} = \mathbf{I}$). Lemma 6 implies $\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}}$ because the model satisfies Zyskind's condition: the column space of $\boldsymbol{\Sigma}\mathbf{X} = (\mathbf{I}_v \otimes \boldsymbol{\Sigma}_p)(\mathbf{1}_v \otimes \mathbf{I}_p) = \mathbf{1}_v \otimes \boldsymbol{\Sigma}_p = (\mathbf{1}_v \otimes \mathbf{I}_p)\boldsymbol{\Sigma}_p = \mathbf{X}\boldsymbol{\Sigma}_p$ is contained in the column space of \mathbf{X} .

A.7 $m^\#$ and $\lambda^\#$ in the Hotelling case

Quantities from A.6 above can be substituted into (4.1) to obtain $E^\#$ and $V^\#$ shown in (6.2), which in turn can be used in (4.3) to obtain (6.3).

A.8 Proof of Lemma 1(a)

Suppose $A_1/A_2 = \ell$ as in special case 1, which implies $B = [(\ell + 6)A_2]/(2\ell)$. Formulas (7.2) and (7.3) were deliberately derived so that when $A_1/A_2 = \ell$, then $\mathbf{d} = (\ell - 2, 2, 4)/(\ell + 6)$, hence $\mathbf{d}\mathbf{B} = (\ell - 2, 2, 4)A_2/(2\ell)$. Thus (7.4) becomes

$$V^* = \frac{2}{\ell} \frac{(1 + \frac{\ell-2}{2\ell} A_2)}{(1 - \frac{1}{\ell} A_2)^2 (1 - \frac{2}{\ell} A_2)} = \frac{2(E^*)^2}{\ell} \frac{(1 + \frac{\ell-2}{2\ell} A_2)}{(1 - \frac{2}{\ell} A_2)}.$$

Via the formulas in (4.3) with superscripts # replaced by *, we obtain $\ell\rho^* = \{1 + [(\ell - 2)/(2\ell)]A_2\}/\{1 - (2/\ell)A_2\}$, $m^* = 2\ell/A_2$ and $\lambda^* = 1$.

Formulas (8.2) and (8.3) give us $e_0 = 0$, $e_1 = \ell + 6$, $m^\dagger = (\ell + 6)/(Q A_2) = 2\ell/A_2$. Formula (9.3) yields $m^\ddagger = 2\ell(\ell + 2)/[(\ell + 2)A_2] = 2\ell/A_2$. Thus $m^* = m^\dagger = m^\ddagger$. Because $E^* = E^\dagger = E^\ddagger$, this implies $\lambda^* = \lambda^\dagger = \lambda^\ddagger$.

A.9 Proof of Lemma 1(b)

Suppose $A_1/A_2 = 2/(\ell + 1)$ as in special case 2, which implies $B = [(3\ell + 4)A_2]/[\ell(\ell + 1)]$. Formulas (7.2) and (7.3) were deliberately derived so that when $A_1/A_2 = 2/(\ell + 1)$, then $\mathbf{d} = (-1, \ell + 1, \ell + 3)/(3\ell + 4)$, hence $\mathbf{d}B = (-1, \ell + 1, \ell + 3)A_2/[\ell(\ell + 1)]$. Thus (7.4) becomes

$$V^* = \frac{2(E^*)^2}{\ell} \frac{\left[1 - \frac{1}{\ell(\ell+1)}A_2\right]}{\left[1 - \frac{\ell+3}{\ell(\ell+1)}A_2\right]}.$$

Via the formulas in (4.3) with superscripts # replaced by *, we obtain $\ell\rho^* = \{1 - [1/[\ell(\ell + 1)]]A_2\}/\{1 - [(\ell + 3)/[\ell(\ell + 1)]]A_2\}$, and m^* and λ^* are as displayed in the lemma.

Formulas (8.2) and (8.3) give us $e_0 = 1 - \ell$, $e_1 = 3\ell + 4$, $m^\dagger = (1 - \ell) + (3\ell + 4)/(QA_2) = m^*$. We have $A_1 = [2/(\ell + 1)]A_2$ and so formula (9.3) yields $m^\ddagger = \{2\ell(\ell + 2) + 2[2/(\ell + 1) - \ell]A_2\}/\{[2/(\ell + 1) + 2]A_2\} = m^*$. So we see $m^* = m^\dagger = m^\ddagger$, which then implies $\lambda^* = \lambda^\dagger = \lambda^\ddagger$.

A.10 Proof of Lemma 2

It suffices to show $\text{tr}(\Psi\mathbf{P}_i)\text{tr}(\Psi\mathbf{P}_j) = \text{tr}(\Psi\mathbf{P}_i\Psi\mathbf{P}_j)$. For any $p \times p$ matrix \mathbf{M} , $\mathbf{L}'\mathbf{M}\mathbf{L}$ is a 1×1 matrix and hence is a scalar. Therefore

$$\begin{aligned} \text{tr}(\Psi\mathbf{P}_i\Psi\mathbf{P}_j) &= \text{tr}[\Phi\mathbf{L}(\mathbf{L}'\Phi\mathbf{L})^{-1}\mathbf{L}'\Phi\mathbf{P}_i\Phi\mathbf{L}(\mathbf{L}'\Phi\mathbf{L})^{-1}\mathbf{L}'\Phi\mathbf{P}_j] \\ &= (\mathbf{L}'\Phi\mathbf{L})^{-2} \text{tr}(\Phi\mathbf{L}\mathbf{L}'\Phi\mathbf{P}_i\Phi\mathbf{L}\mathbf{L}'\Phi\mathbf{P}_j) \\ &= (\mathbf{L}'\Phi\mathbf{L})^{-2} \text{tr}(\mathbf{L}'\Phi\mathbf{P}_i\Phi\mathbf{L}\mathbf{L}'\Phi\mathbf{P}_j\Phi\mathbf{L}) \\ &= (\mathbf{L}'\Phi\mathbf{L})^{-2}\mathbf{L}'\Phi\mathbf{P}_i\Phi\mathbf{L}\mathbf{L}'\Phi\mathbf{P}_j\Phi\mathbf{L} \\ &= [(\mathbf{L}'\Phi\mathbf{L})^{-1}\mathbf{L}'\Phi\mathbf{P}_i\Phi\mathbf{L}][(\mathbf{L}'\Phi\mathbf{L})^{-1}\mathbf{L}'\Phi\mathbf{P}_j\Phi\mathbf{L}] = \text{tr}(\Psi\mathbf{P}_i)\text{tr}(\Psi\mathbf{P}_j), \end{aligned}$$

because

$$\begin{aligned} \text{tr}(\Psi\mathbf{P}_i) &= \text{tr}[\Phi\mathbf{L}(\mathbf{L}'\Phi\mathbf{L})^{-1}\mathbf{L}'\Phi\mathbf{P}_i] = (\mathbf{L}'\Phi\mathbf{L})^{-1} \text{tr}(\Phi\mathbf{L}\mathbf{L}'\Phi\mathbf{P}_i) \\ &= (\mathbf{L}'\Phi\mathbf{L})^{-1} \text{tr}(\mathbf{L}'\Phi\mathbf{P}_i\Phi\mathbf{L}) = (\mathbf{L}'\Phi\mathbf{L})^{-1}\mathbf{L}'\Phi\mathbf{P}_i\Phi\mathbf{L}. \end{aligned}$$

A.11 Proof of Lemma 3

It suffices to show $\text{tr}(\Psi\mathbf{P}_i)\text{tr}(\Psi\mathbf{P}_j) = \ell \text{tr}(\Psi\mathbf{P}_i\Psi\mathbf{P}_j)$. The assumptions of the lemma imply

$$\mathbf{X}'\Sigma^{-1}\mathbf{X} = \begin{bmatrix} \mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & f(\theta)\mathbf{C} \end{bmatrix} \quad \text{and} \quad \Phi = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} = \begin{bmatrix} * & \mathbf{0} \\ \mathbf{0} & f(\theta)^{-1}\mathbf{C}^{-1} \end{bmatrix}.$$

We are assuming $\mathbf{L}' = [\mathbf{0} \ \mathbf{L}'_2]$, so

$$\Psi = \Phi \mathbf{L} (\mathbf{L}' \Phi \mathbf{L})^{-1} \mathbf{L}' \Phi = f(\theta)^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-1} \mathbf{L}_2 (\mathbf{L}'_2 \mathbf{C}^{-1} \mathbf{L}_2)^{-1} \mathbf{L}'_2 \mathbf{C}^{-1} \end{bmatrix}.$$

Note that \mathbf{P}_i in Sect. 3 can be expressed as $\mathbf{P}_i = (\partial/\partial\theta_i)(\mathbf{X}'\Sigma^{-1}\mathbf{X})$. Under the assumptions of the lemma,

$$\mathbf{P}_i = \begin{bmatrix} * & \mathbf{0} \\ \mathbf{0} & \frac{\partial f}{\partial \theta_i}(\theta) \mathbf{C} \end{bmatrix} \quad \text{and} \quad \Psi \mathbf{P}_i = g_i(\theta) \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix}$$

where $g_i(\theta) = f(\theta)^{-1}(\partial/\partial\theta_i)f(\theta)$ and $\mathbf{M} = \mathbf{C}^{-1} \mathbf{L}_2 (\mathbf{L}'_2 \mathbf{C}^{-1} \mathbf{L}_2)^{-1} \mathbf{L}'_2$. Now $\text{tr}(\Psi \mathbf{P}_i) = g_i(\theta) \text{tr}(\mathbf{M})$ and $\text{tr}(\mathbf{M}) = \text{tr}[(\mathbf{L}'_2 \mathbf{C}^{-1} \mathbf{L}_2)^{-1} \mathbf{L}'_2 \mathbf{C}^{-1} \mathbf{L}_2] = \text{tr}(\mathbf{I}_\ell) = \ell$. Therefore $\text{tr}(\Psi \mathbf{P}_i) \text{tr}(\Psi \mathbf{P}_j) = \ell^2 g_i(\theta) g_j(\theta)$. Also,

$$\Psi \mathbf{P}_i \Psi \mathbf{P}_j = g_i(\theta) g_j(\theta) \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^2 \end{bmatrix}$$

and $\mathbf{M}^2 = \mathbf{M}$, so $\text{tr}(\Psi \mathbf{P}_i \Psi \mathbf{P}_j) = \ell g_i(\theta) g_j(\theta)$.

A.12 Proof of Theorem 2

The theorem follows from Lemmas 1(a) and 3, if we show that a BIBD model satisfies the conditions of Lemma 3. First note that $\mathbf{B}\mathbf{B}' = k\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' = k\mathbf{P}_\mathbf{B}$ where $\mathbf{P}_\mathbf{B}$ denotes the orthogonal projection matrix on the column space of \mathbf{B} . Now we can write

$$\begin{aligned} \Sigma &= (k\sigma_b^2 + \sigma_e^2)\mathbf{P}_\mathbf{B} + \sigma_e^2(\mathbf{I}_n - \mathbf{P}_\mathbf{B}) \quad \text{and} \\ \Sigma^{-1} &= (k\sigma_b^2 + \sigma_e^2)^{-1}\mathbf{P}_\mathbf{B} + (\sigma_e^2)^{-1}(\mathbf{I}_n - \mathbf{P}_\mathbf{B}). \end{aligned}$$

Write $\mathbf{T}^* = \mathbf{T}\mathbf{U}$ where \mathbf{U} is the $t \times (t - 1)$ matrix obtained by subtracting the last column of \mathbf{I}_t from each of the first $t - 1$ columns of \mathbf{I}_t . Then

$$\mathbf{X}'_1 \Sigma^{-1} \mathbf{X}_2 = (k\sigma_b^2 + \sigma_e^2)^{-1} \mathbf{1}'_n \mathbf{P}_\mathbf{B} \mathbf{T} \mathbf{U} + (\sigma_e^2)^{-1} \mathbf{1}'_n (\mathbf{I}_n - \mathbf{P}_\mathbf{B}) \mathbf{T} \mathbf{U} = \mathbf{0},$$

because (1) $\mathbf{1}'_n \mathbf{P}_\mathbf{B} = \mathbf{1}'_n$ and (2) $\mathbf{1}'_n \mathbf{T} \mathbf{U} = \mathbf{0}$. (1) is true because $\mathbf{1}_n$ is in the column space of \mathbf{B} , and (2) is true because $\mathbf{1}'_n \mathbf{T} = r \mathbf{1}'_t$ and $\mathbf{1}'_t \mathbf{U} = \mathbf{0}$. Next write

$$\begin{aligned} \Sigma^{-1} &= [(k\sigma_b^2 + \sigma_e^2)^{-1} - (\sigma_e^2)^{-1}] \mathbf{P}_\mathbf{B} + (\sigma_e^2)^{-1} \mathbf{I}_n \\ &= \sigma_e^{-2} \{ \mathbf{I}_n - [\sigma_b^2 / (k\sigma_b^2 + \sigma_e^2)] \mathbf{B}\mathbf{B}' \}. \end{aligned}$$

It can be shown (Khuri et al. 1998, p. 176) that $\mathbf{N}\mathbf{N}' = (r - g)\mathbf{I}_t + g\mathbf{1}_t\mathbf{1}'_t$. Then

$$\begin{aligned} \mathbf{X}'_2\boldsymbol{\Sigma}^{-1}\mathbf{X}_2 &= \sigma_e^{-2}\{\mathbf{U}'\mathbf{T}'\mathbf{T}\mathbf{U} - [\sigma_b^2/(k\sigma_b^2 + \sigma_e^2)]\mathbf{U}'\mathbf{T}'\mathbf{B}\mathbf{B}'\mathbf{T}\mathbf{U}\}, \\ \mathbf{U}'\mathbf{T}'\mathbf{T}\mathbf{U} &= r\mathbf{U}'\mathbf{U}, \\ \mathbf{U}'\mathbf{T}'\mathbf{B}\mathbf{B}'\mathbf{T}\mathbf{U} &= \mathbf{U}'\mathbf{N}\mathbf{N}'\mathbf{U} = (r - g)\mathbf{U}'\mathbf{U}, \end{aligned}$$

and hence $\mathbf{X}'_2\boldsymbol{\Sigma}^{-1}\mathbf{X}_2 = f(\boldsymbol{\theta})\mathbf{C}$ where $f(\boldsymbol{\theta}) = \sigma_e^{-2}\{r - (r - g)\sigma_b^2/(k\sigma_b^2 + \sigma_e^2)\}$ and $\mathbf{C} = \mathbf{U}'\mathbf{U}$.

A.13 Proof of Lemma 5

We will assume that the REML estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ can be characterized as the unique solution of the REML equations (see Jiang 2007, p. 13):

$$\text{tr}\{\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{G}_i\} = \mathbf{y}'\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{G}_i\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{y} \quad \text{for } i = 1, \dots, r \tag{*}$$

where $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Phi}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$. The REML estimator is known to be location-invariant (Kackar and Harville 1984, p. 854); that is, $\hat{\boldsymbol{\theta}}(\mathbf{y} + \mathbf{X}\mathbf{b}) = \hat{\boldsymbol{\theta}}(\mathbf{y})$ for all $\mathbf{b} \in \mathbb{R}^p$. Moreover, it is scale-equivariant in the sense that $\hat{\boldsymbol{\theta}}(c\mathbf{y}) = c^2\hat{\boldsymbol{\theta}}(\mathbf{y})$ for $c \neq 0$, which can be verified as follows. The estimate $\hat{\boldsymbol{\theta}}(c\mathbf{y})$ is the unique solution to the REML equations (*) when the data vector is $c\mathbf{y}$: $\text{tr}\{\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(c\mathbf{y})]\mathbf{G}_i\} = (c\mathbf{y})'\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(c\mathbf{y})]\mathbf{G}_i\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(c\mathbf{y})](c\mathbf{y})$. If we can show that $c^2\hat{\boldsymbol{\theta}}(\mathbf{y})$ is also a solution, then we can conclude $\hat{\boldsymbol{\theta}}(c\mathbf{y}) = c^2\hat{\boldsymbol{\theta}}(\mathbf{y})$. So we want to show $\text{tr}\{\boldsymbol{\Gamma}[c^2\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{G}_i\} = (c\mathbf{y})'\boldsymbol{\Gamma}[c^2\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{G}_i\boldsymbol{\Gamma}[c^2\hat{\boldsymbol{\theta}}(\mathbf{y})](c\mathbf{y})$. First note that $\boldsymbol{\Gamma}(c\boldsymbol{\theta}) = c^{-1}\boldsymbol{\Gamma}(\boldsymbol{\theta})$ because $\boldsymbol{\Sigma}(c\boldsymbol{\theta}) = c\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Phi}(c\boldsymbol{\theta}) = c\boldsymbol{\Phi}(\boldsymbol{\theta})$. Now

$$\begin{aligned} (c\mathbf{y})'\boldsymbol{\Gamma}[c^2\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{G}_i\boldsymbol{\Gamma}[c^2\hat{\boldsymbol{\theta}}(\mathbf{y})](c\mathbf{y}) &= c^2\mathbf{y}'c^{-2}\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{G}_ic^{-2}\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{y} \\ &= c^{-2}\mathbf{y}'\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{G}_i\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{y} \stackrel{\text{by } (*)}{=} c^{-2} \text{tr}\{\boldsymbol{\Gamma}[\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{G}_i\} = \text{tr}\{\boldsymbol{\Gamma}[c^2\hat{\boldsymbol{\theta}}(\mathbf{y})]\mathbf{G}_i\}. \end{aligned}$$

For convenience one can combine the properties of location-invariance and scale-equivariance in a single equation: $\hat{\boldsymbol{\theta}}(c\mathbf{y} + \mathbf{X}\mathbf{b}) = c^2\hat{\boldsymbol{\theta}}(\mathbf{y})$.

By its definition $F_{\text{KR}} = \frac{1}{\ell}[\mathbf{L}'\hat{\boldsymbol{\beta}}(\mathbf{y})]'[\mathbf{L}'\hat{\boldsymbol{\Phi}}_A(\mathbf{y})\mathbf{L}]^{-1}[\mathbf{L}'\hat{\boldsymbol{\beta}}(\mathbf{y})]$. To compare this with $F_{\text{KR}}(c\mathbf{y} + \mathbf{X}\mathbf{b})$, first observe that $\hat{\boldsymbol{\Sigma}}(c\mathbf{y} + \mathbf{X}\mathbf{b}) = \boldsymbol{\Sigma}[\hat{\boldsymbol{\theta}}(c\mathbf{y} + \mathbf{X}\mathbf{b})] = \boldsymbol{\Sigma}[c^2\hat{\boldsymbol{\theta}}(\mathbf{y})] = c^2\boldsymbol{\Sigma}[\hat{\boldsymbol{\theta}}(\mathbf{y})] = c^2\hat{\boldsymbol{\Sigma}}(\mathbf{y})$. Therefore

$$\begin{aligned} \hat{\boldsymbol{\beta}}(c\mathbf{y} + \mathbf{X}\mathbf{b}) &= \{\mathbf{X}'[\hat{\boldsymbol{\Sigma}}(c\mathbf{y} + \mathbf{X}\mathbf{b})]^{-1}\mathbf{X}\}^{-1}\mathbf{X}'[\hat{\boldsymbol{\Sigma}}(c\mathbf{y} + \mathbf{X}\mathbf{b})]^{-1}(c\mathbf{y} + \mathbf{X}\mathbf{b}) \\ &= \{\mathbf{X}'[c^2\hat{\boldsymbol{\Sigma}}(\mathbf{y})]^{-1}\mathbf{X}\}^{-1}\mathbf{X}'[c^2\hat{\boldsymbol{\Sigma}}(\mathbf{y})]^{-1}(c\mathbf{y} + \mathbf{X}\mathbf{b}) \\ &= \{\mathbf{X}'[\hat{\boldsymbol{\Sigma}}(\mathbf{y})]^{-1}\mathbf{X}\}^{-1}\mathbf{X}'[\hat{\boldsymbol{\Sigma}}(\mathbf{y})]^{-1}(c\mathbf{y} + \mathbf{X}\mathbf{b}) \\ &= c\{\mathbf{X}'[\hat{\boldsymbol{\Sigma}}(\mathbf{y})]^{-1}\mathbf{X}\}^{-1}\mathbf{X}'[\hat{\boldsymbol{\Sigma}}(\mathbf{y})]^{-1}\mathbf{y} + \{\mathbf{X}'[\hat{\boldsymbol{\Sigma}}(\mathbf{y})]^{-1}\mathbf{X}\}^{-1}\mathbf{X}'[\hat{\boldsymbol{\Sigma}}(\mathbf{y})]^{-1}\mathbf{X}\mathbf{b} \\ &= c\hat{\boldsymbol{\beta}}(\mathbf{y}) + \mathbf{b} \end{aligned}$$

and $\mathbf{L}'\hat{\boldsymbol{\beta}}(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}_0) = \mathbf{L}'[c\hat{\boldsymbol{\beta}}(\mathbf{y}) + \mathbf{b}_0] = c\mathbf{L}'\hat{\boldsymbol{\beta}}(\mathbf{y})$ when $\mathbf{L}'\mathbf{b}_0 = \mathbf{0}$.

Next check that $\hat{\boldsymbol{\Phi}}(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}) = c^2\hat{\boldsymbol{\Phi}}(\mathbf{y})$, $\hat{\mathbf{P}}_i(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}) = c^{-4}\hat{\mathbf{P}}_i(\mathbf{y})$, $\hat{\mathbf{Q}}_{ij}(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}) = c^{-6}\hat{\mathbf{Q}}_{ij}(\mathbf{y})$, $\hat{\boldsymbol{\Gamma}}(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}) = c^{-2}\hat{\boldsymbol{\Gamma}}(\mathbf{y})$. Recall that $\hat{\mathbf{W}} = [\hat{w}_{ij}]_{r \times r} = \tilde{\mathbf{W}}[\hat{\boldsymbol{\theta}}(\mathbf{y})]$ where $\tilde{\mathbf{W}} = [\tilde{w}_{ij}]_{r \times r} = \mathfrak{I}^{-1}$ and \mathfrak{I} is the expected information matrix. We can write $\mathfrak{I} = [\tilde{w}^{ij}]_{r \times r}$ and $\tilde{w}^{ij} = \frac{1}{2} \text{tr}(\boldsymbol{\Gamma}\mathbf{G}_i\boldsymbol{\Gamma}\mathbf{G}_j)$ (see (1.21) in Jiang 2007). One can see that $\hat{w}^{ij}(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}) = c^{-4}\hat{w}^{ij}(\mathbf{y})$, $\hat{w}_{ij}(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}) = c^4\hat{w}_{ij}(\mathbf{y})$, $\hat{\boldsymbol{\Lambda}}(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}) = c^2\hat{\boldsymbol{\Lambda}}(\mathbf{y})$, and $\hat{\boldsymbol{\Phi}}_A(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}) = c^2\hat{\boldsymbol{\Phi}}_A(\mathbf{y})$. Now

$$\begin{aligned} F_{\text{KR}}(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}_0) &= \frac{1}{\ell}[\mathbf{L}'\hat{\boldsymbol{\beta}}(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}_0)]'[\mathbf{L}'\hat{\boldsymbol{\Phi}}_A(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}_0)\mathbf{L}]^{-1}[\mathbf{L}'\hat{\boldsymbol{\beta}}(\mathbf{c}\mathbf{y} + \mathbf{X}\mathbf{b}_0)] \\ &= \frac{1}{\ell}[c\mathbf{L}'\hat{\boldsymbol{\beta}}(\mathbf{y})]'[\mathbf{L}'c^2\hat{\boldsymbol{\Phi}}_A(\mathbf{y})\mathbf{L}]^{-1}[c\mathbf{L}'\hat{\boldsymbol{\beta}}(\mathbf{y})] \\ &= F_{\text{KR}}(\mathbf{y}). \end{aligned}$$

A.14 Tables of average values of simulated m and λ

The average values of the simulated m and λ are shown in Tables 4 and 5.

Table 4 The observed averages of the denominator degrees of freedom $m^\#$, m^* , m^\dagger and m^\ddagger for the four test procedures applied to the well-behaved data sets from among 10,000 data sets generated from each of 40 models (8 designs \times 5 values of ρ)

	D96	D60	D40	D21a	D21b	D18a	D18b	D12
$\rho = 0.25$								
KRU	83.7	48.9	37.3	23.9	19.7	20.4	21.6	42.1
KR	72.6	39.3	28.5	13.4	7.8	10.8	8.1	4.2
KRA1	82.3	41.7	28.5	13.5	8.6	10.8	8.4	4.3
KRA2	74.7	40.2	28.5	13.4	8.2	10.8	8.2	4.3
$\rho = 0.5$								
KRU	79.8	47.5	37.1	22.5	19.0	20.2	19.6	22.5
KR	69.0	38.5	28.3	12.7	7.7	10.6	7.8	4.2
KRA1	77.4	39.7	28.3	12.8	8.4	10.6	8.1	4.3
KRA2	70.9	38.9	28.3	12.8	8.2	10.6	8.0	4.3
$\rho = 1$								
KRU	65.6	43.2	36.9	21.0	17.5	19.9	17.3	17.7
KR	55.9	34.7	28.1	11.3	7.3	10.3	7.2	4.2
KRA1	59.9	34.9	28.1	11.3	7.6	10.3	7.3	4.2
KRA2	57.0	34.7	28.1	11.3	7.6	10.3	7.3	4.3
$\rho = 2$								
KRU	50.0	40.5	36.9	19.5	16.1	19.7	16.1	14.9
KR	41.3	32.1	28.0	9.8	6.6	10.1	6.5	4.2
KRA1	41.9	32.1	28.0	9.8	6.6	10.1	6.5	4.2
KRA2	41.5	32.1	28.0	9.8	6.6	10.1	6.5	4.3

Table 4 continued

	D96	D60	D40	D21a	D21b	D18a	D18b	D12
$\rho = 4$								
KRU	43.7	39.7	36.8	18.9	15.6	19.7	15.7	21.9
KR	35.3	31.3	28.0	9.2	6.2	10.0	6.1	4.2
KRA1	35.3	31.3	28.0	9.2	6.2	10.0	6.1	4.2
KRA2	35.3	31.3	28.0	9.2	6.2	10.0	6.1	4.3

Table 5 The observed averages of the scale factors $\lambda^\#$, λ^* , λ^\dagger and λ^\ddagger for the four test procedures applied to the well-behaved data sets from among 10,000 data sets generated from each of 40 models (8 designs \times 5 values of ρ)

	D96	D60	D40	D21a	D21b	D18a	D18b	D12
$\rho = 0.25$								
KRU	0.995	0.989	0.987	0.940	0.813	0.935	0.850	0.683
KR	0.998	0.997	1.000	0.998	0.988	1.000	0.994	0.956
KRA1	0.995	0.995	1.000	0.997	0.958	1.000	0.982	0.945
KRA2	0.997	0.998	1.000	0.997	0.958	1.000	0.985	0.925
$\rho = 0.5$								
KRU	0.995	0.992	0.987	0.942	0.842	0.933	0.886	0.694
KR	0.998	0.999	1.000	0.999	0.992	1.000	0.995	0.957
KRA1	0.995	0.998	1.000	0.998	0.968	1.000	0.987	0.946
KRA2	0.998	0.999	1.000	0.999	0.970	1.000	0.988	0.927
$\rho = 1$								
KRU	0.995	0.991	0.987	0.936	0.871	0.931	0.879	0.709
KR	0.999	1.000	1.000	1.000	0.998	1.000	0.998	0.960
KRA1	0.997	1.000	1.000	1.000	0.986	1.000	0.995	0.949
KRA2	0.998	1.000	1.000	1.000	0.987	1.000	0.996	0.932
$\rho = 2$								
KRU	0.993	0.990	0.987	0.925	0.874	0.929	0.871	0.713
KR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.962
KRA1	0.999	1.000	1.000	1.000	0.997	1.000	0.999	0.951
KRA2	1.000	1.000	1.000	1.000	0.998	1.000	0.999	0.935
$\rho = 4$								
KRU	0.992	0.990	0.987	0.919	0.886	0.928	0.864	0.705
KR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.964
KRA1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.955
KRA2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.941

References

- Arnau J, Bono R, Vallejo G (2009) Analyzing small samples of repeated measures data with the mixed-model adjusted F test. *Commun Stat Simul* 38:1083–1103
- Chen X (2006) The adjustment of random baseline measurements in treatment effect estimation. *J Stat Plan Infer* 136:4161–4175
- Cochran W, Cox G (1957) *Experimental designs*. Wiley, New York
- Green P (1974) On the design of choice experiments involving multifactor alternatives. *J Consum Res* 1:61–68
- Guiard V, Spilke J, Danicke S (2003) Evaluation and interpretation of results for three cross-over designs. *Arch Anim Nutr* 57:177–195
- Halekoh U, Højsgaard S (2014) A Kenward–Roger approximation and parametric bootstrap methods for tests in linear mixed models—the R package pbrtest. *J Stat Softw* 59(9):1–32
- Harville DA, Jeske DR (1992) Mean squared error of estimation or prediction under a general linear model. *J Am Stat Assoc* 87:724–731
- Jiang J (2007) *Linear and generalized linear mixed models and their applications*. Springer, New York
- John PWM (1971) *Statistical design and analysis of experiments*. Macmillan, New York
- Kackar RN, Harville DA (1984) Approximations for standard errors of estimators of fixed and random effects in mixed liner models. *J Am Stat Assoc* 79:853–862
- Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53:983–997
- Khuri AI, Mathew T, Sinha BK (1998) *Statistical tests for mixed linear models*. Wiley, New York
- Kowalchuk RK, Keselman HJ, Algina J, Wolfinger RD (2004) The analysis of repeated measurements with mixed-model adjusted F tests. *Educ Psychol Meas* 64:224–242
- Kuehl RO (2000) *Design of experiments: statistical principles of research design and analysis*, 2nd edn. Duxbury Press, Pacific Grove
- Livacic-Rojas P, Vallejo G, Fernandez P (2010) Analysis of type I error rates of univariate and multivariate procedures in repeated measures designs. *Commun Stat Simul* 39:624–640
- Mardia KV, Kent JT, Bibby JM (1979) *Multivariate analysis*. Academic Press, London
- SAS Institute Inc (2015) *SAS/STAT®14.1 User’s guide*. SAS Institute Inc., Cary
- Schaalje GB, McBride JB, Fellingham GW (2002) Adequacy of approximations to distributions of test statistics in complex mixed linear models. *J Agric Biol Environ Stat* 7:512–524
- Searle SR, Casella G, McCulloch CE (1992) *Variance components*. Wiley, New York
- Spilke J, Piepho H, Hu X (2005) 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* 10:374–389
- Stroup WW (1999) On using proc mixed for longitudinal data. In: Annual conference on applied statistics in agriculture, Paper 5. <http://newprairiepress.org/agstatconference/1999/proceedings/5>. Accessed 4 July 2016
- VanLeeuwen DM, Birkes DS, Seely JF (1999) Balance and orthogonality in designs for mixed classification models. *Ann Stat* 27:1927–1947
- Wimmer G, Witkovsky V (2007) Univariate linear calibration via replicated errors-in-variables model. *J Stat Comput Sim* 77:213–227
- Witkovsky V (2012) Estimation, testing, and prediction regions of the fixed and random effects by solving the Henderson’s mixed model equations. *Meas Sci Rev* 12:234–248
- Wulff SS, Robinson TJ (2009) Assessing the uncertainty of regression estimates in a response surface model for repeated measures. *Qual Technol Quant Manag* 6:309–324
- Zyskind G (1967) On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann Math Stat* 38:1092–1109