



# On the equivalence between the LRT and F-test for testing variance components in a class of linear mixed models

Fares Qeadan<sup>1</sup> · Ronald Christensen<sup>2</sup>

Received: 3 May 2019 / Published online: 28 May 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

For the special case of balanced one-way random effects ANOVA, it has been established that the generalized likelihood ratio test (LRT) and Wald's test are largely equivalent in testing the variance component. We extend these results to explore the relationships between Wald's  $F$  test, and the LRT for a much broader class of linear mixed models; the generalized split-plot models. In particular, we explore when the two tests are equivalent and prove that when they are not equivalent, Wald's  $F$  test is more powerful, thus making the LRT test inadmissible. We show that inadmissibility arises in realistic situations with common number of degrees of freedom. Further, we derive the statistical distribution of the LRT under both the null and alternative hypotheses  $H_0$  and  $H_1$  where  $H_0$  is the hypothesis that the between variance component is zero. Providing an exact distribution of the test statistic for the LRT in these models will help in calculating a more accurate p-value than the traditionally used p-value derived from the large sample chi-square mixture approximations.

**Keywords** F-test · LRT · Generalized split-plot · Variance component · Random effect · Mixed model

**Mathematics Subject Classification** 62C15 · 62E15 · 62F03 · 62F10 · 62K99

## 1 Introduction

For linear mixed models with one variance component beyond the usual error, Wald proposed in 1947, for the first time, an exact  $F$  test for the variance component being

---

✉ Fares Qeadan  
fares.qeadan@utah.edu  
Ronald Christensen  
fletcher@stat.unm.edu

<sup>1</sup> University of Utah, Salt Lake City, UT, USA

<sup>2</sup> University of New Mexico, Albuquerque, NM, USA

zero. In 2010, Lu and Zhang have established the equivalence between the generalized likelihood ratio test and the traditional  $F$  test for no group effects in the balanced one-way random effects model, see also Herbach (1959). We extend this result to a much broader family of linear mixed effects models with one random effect, compound symmetry (CS) covariance structure, and other constraints—namely the generalized split-plot (GSP) models introduced in Christensen (1987). GSPs assume CS covariance structure indicating heterogeneous variances and constant correlations among repeated measures. GSP models include standard split-plot models with nearly any experimental design for the whole plot treatments and include some ability to incorporate covariates into split-plot models. Further, the standard multivariate growth curve models (cf. some standard multivariate text like Johnson and Wichern) with compound symmetry covariance matrices are special cases of generalized split plot models as shown by Christensen (2019, Section 11.7.1). The balanced one-way random effects model is the simplest GSP model.

A general linear mixed model that specifies both fixed effects and random effects can be written as

$$Y = \tilde{X}\beta + X_1\gamma + \varepsilon, \quad (1)$$

where  $Y$  is a vector of observations,  $\tilde{X}$  is a known model matrix for fixed effects,  $\beta$  is an unobservable parameter vector of fixed effects,  $X_1$  is a known model matrix for random effects,  $\gamma$  is an unobservable vector of random effects, and  $\varepsilon$  is a vector of residual errors with  $\mathbb{E}(\varepsilon) = 0$ ,  $\text{Cov}(\varepsilon) = R$ ,  $\mathbb{E}(\gamma) = 0$ ,  $\text{Cov}(\gamma) = D$ ,  $\text{Cov}(\varepsilon, \gamma) = 0$ , and therefore  $\text{Cov}(Y) = X_1DX_1' + R$ .

## 1.1 Notation

For the rest of this paper, we use the notation  $C(A)$  and  $r(A)$  to denote the column space and rank of the matrix  $A$  respectively. In addition, the perpendicular projection operator (PPO) onto  $C(A)$  is denoted by  $M_A$  unless otherwise specified. In matrix notation,  $J_r^c$  and  $0_{r \times c}$  denote a matrix of ones and a matrix of zeros respectively each of size  $r \times c$ . When  $c = 1$ , for simplicity we suppress  $c$  so that  $J_r$  and  $0_r$  denote a column vector of ones and a column vector of zeros respectively of length  $r$ .  $I_n$  is the identity matrix of size  $n$  while  $I$ , with no subscripts, has size that can be inferred from context. We use  $\text{diag}(V_1, \dots, V_N)$  to denote a block diagonal matrix with square matrices  $V_1, \dots, V_N$  on its diagonal and  $\text{Bdiag}(V)$  to denote a block diagonal matrix whose diagonal entries are all  $V$ . Vertical lines denote the determinant when enclosing a matrix and absolute value when enclosing a number.  $L(\cdot)$ , and  $\ell(\cdot)$  denote the likelihood, and  $-2\text{Log-likelihood}$  functions respectively while  $\ell_*(\cdot) \equiv -\ell(\cdot)$ .

## 1.2 A class of linear mixed models of interest

Our class of linear mixed models of interest is of the form

$$Y = X_*\delta + X_2\gamma + (X_1\eta + \varepsilon), \quad (2)$$

with  $\eta \sim N(0, \sigma_w^2 I_N)$ ,  $\epsilon \sim N(0, \sigma_s^2 I_n)$ ,  $\text{Cov}(\epsilon, \eta) = 0_{n \times N}$ ,  $n = Nm$ ,  $\delta$  and  $\gamma$  are fixed effects,  $C(X_*) \subset C(X_1)$ , and  $C([X_*, X_2]) = C(X_*, (I - M_1)X_2)$  where  $M_1$  is the PPO onto  $C(X_1)$ . If  $V = \text{Cov}(Y)$ , then for this class of models it's immediate that  $C(VX) \subset C(X)$  given that  $X_1$  is a matrix of indicator variables for the clusters with equal group sizes. GSPs are the class of such linear models for which  $V = (\sigma_w^2 + \sigma_s^2) [(1 - \rho)I_n + \rho J_{n \times n}]$  where  $\rho = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_s^2}$ ,  $\sigma_w^2$  is the variance between clusters or blocks, and  $\sigma_s^2$  is the variance within clusters or blocks.

### 1.2.1 Generalized split plot (GSP) models

We summarize the definition and analysis of GSP models from Christensen (1984, 1987 and 2011). Consider a two-stage cluster sampling model

$$Y = X\beta + \xi, \tag{3}$$

with  $n$  observations,  $m_i$  subjects from each cluster, and includes fixed effects for each cluster. Let  $X = [X_1, X_2]$  where the columns of  $X_1$  are indicators for the clusters and  $X_2$  contains the other effects. Write  $\beta' = [\alpha', \gamma']$  so that  $\alpha$  is a vector of fixed cluster effects and  $\gamma$  is a vector of fixed non-cluster effects. Because it is a two-stage cluster sampling model, the error vector  $\xi$  has uncorrelated clusters and intraclass correlation structure and can be written with random effects as

$$\xi = X_1\eta + \epsilon, \tag{4}$$

where  $\eta$  contains random cluster effects and  $\epsilon$  is a random error. Assume that  $\eta$  and  $\epsilon$  are independent such that  $\eta \sim N(0, \sigma_w^2 I_N)$ ,  $\epsilon \sim N(0, \sigma_s^2 I_n)$  with  $\text{Cov}(\epsilon, \eta) = 0_{n \times N}$  such that  $n = \sum_{i=1}^N m_i$  then we get the mixed model

$$Y = X_1\alpha + X_2\gamma + (X_1\eta + \epsilon), \tag{5}$$

with

$$V \equiv \text{Cov}(Y) = \sigma_w^2 X_1 X_1' + \sigma_s^2 I_n. \tag{6}$$

The GSP models are obtained by imposing additional structure on the fixed cluster effects within the cluster sampling model (5). To model the whole plot (cluster) effects we put a constraint on  $C(X_1)$  by considering a reduced model

$$\begin{aligned} Y &= X_*\delta + X_2\gamma + (X_1\eta + \epsilon), & C(X_*) \subset C(X_1) \\ &\equiv \tilde{X}\beta_* + \xi, \end{aligned} \tag{7}$$

where  $\tilde{X} = [X_*, X_2]$  and  $\beta_*' = [\delta', \gamma']$  and the covariance matrix remains as in (6). The  $\delta_i$ s will be the whole plots fixed effects and the  $\gamma_i$ s will be the subplots fixed effects. In this model,  $\eta$  is the whole plot error and  $\epsilon$  is the subplot error.

To develop a traditional split-plot analysis for GSP models, we need three assumptions:

- (a)  $m_i = m$ , i.e. all whole plots are of the same size,
- (b)  $C(X_*) \subset C(X_1)$ , i.e.  $\delta$  contains whole plot effects,
- (c)  $C(\tilde{X}) = C(X_*, (I - M_1)X_2)$  where  $M_1$  is the PPO onto  $C(X_1)$ , i.e. subplot effects are orthogonal to whole plots (not just whole plot effects).

In particular, with these three conditions, the least-squares estimates (LSEs) for model (7) are best linear unbiased estimates (BLUEs) and standard split-plot  $F$  and  $t$  statistics have null hypothesis  $F$  and  $t$  distributions, cf., Christensen (2011, Chapter 11).

We consider a GSP model with  $n$  total observations,  $r(M_1) = N$  whole plots and  $m$  subplots in each whole plot. In model (7), with covariance structure (6) and assumption (a) we can rewrite (6) as

$$V = \sigma_w^2 \text{Bdiag}(J_m J_m') + \sigma_s^2 I_n = \sigma_s^2 I_n + \sigma_w^2 m M_1. \tag{8}$$

Also,

$$V = \text{Bdiag}(\tilde{V}) = I_N \otimes \tilde{V}, \tag{9}$$

where

$$\tilde{V} = \sigma_w^2 J_m J_m' + \sigma_s^2 I_m = \begin{bmatrix} \sigma_w^2 + \sigma_s^2 & \sigma_w^2 & \dots & \sigma_w^2 \\ \sigma_w^2 & \sigma_w^2 + \sigma_s^2 & \dots & \sigma_w^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_w^2 & \sigma_w^2 & \dots & \sigma_w^2 + \sigma_s^2 \end{bmatrix}. \tag{10}$$

In a general mixed model (1) exact statistical inference cannot typically be performed noting that approximation for exact calculations could be made (Lavine et al. 2015). However, there are special cases such as Wald’s tests for variance components and split-plot models where exact inferences are available. The analysis of split-plot designs is complicated by having two different errors in the model. If  $\sigma_w^2 = 0$ , the model (7) reduces to the standard linear model

$$\begin{aligned} Y &= X_* \delta + X_2 \gamma + \epsilon, & \mathbb{E}(\epsilon) &= 0, \text{Cov}(\epsilon) = \sigma_e^2 I_n \\ &= \tilde{X} \beta_* + \epsilon. \end{aligned} \tag{11}$$

In particular, if the whole plot model is a one-way ANOVA, i.e., the whole plot design is a completely randomized design (CRD), and the subplot effects involve only subplot main effects and whole plot by subplot interaction then the model reduces to two-way ANOVA with interaction. Further, if the whole plot one-way model is unbalanced this might be an interesting two-way ANOVA model wherein the number of observations on each pair of factors  $ij$  is  $k = 1, \dots, mN_i$  instead of the usual  $k = 1, \dots, N_{ij}$  which makes whole plot treatments and subplot treatments orthogonal (Christensen 2011, Section 7.4). To test the appropriateness of the standard linear model (11), the

hypothesis of interest is whether or not the covariance component for whole plots lies on the boundary of the parameter space, i.e.,

$$H_0 : \sigma_w^2 = 0 \text{ vs. } H_1 : \sigma_w^2 > 0. \tag{12}$$

There are three competing test statistics for the null hypothesis in (12), the likelihood ratio test (LRT) and the traditional, exact, F-test (Wald’s test) and the likelihood ratio test based on the residual likelihood (RLRT). These tests have been studied extensively over the past few decades (see Crainiceanu and Ruppert 2004; Greven et al. 2008; Wiencierz et al. 2011; Molenberghs and Verbeke 2007; Scheipl et al. 2008), however studies about their equivalence in the context of split-plot designs are limited.

### 1.3 Objectives

In this work, we aim to show the inadmissibility of the LRT in testing the hypothesis that the between variance component in a linear mixed model with CS covariance structure and other constraints is zero (i.e. among GSP models). Firstly, we derive explicit expressions for the maximum likelihood estimates (MLEs) for the variance components of the GSP model. Further, we show that the LRT and F tests are equivalent for testing (12) in GSP models when the size of the test,  $\alpha$ , is reasonably small and we derive the exact distribution for the LRT test statistics under both  $H_0$  and  $H_1$ . In particular, we show that the two tests are equivalent when  $\alpha \leq 1 - p$  where  $p$  is the probability that the LRT statistic is zero and give a proof that the F test has a larger power when  $\alpha > 1 - p$ .

### 1.4 Additional notations

Let  $M, M_1, \tilde{M}$  and  $M_*$  be the PPOs onto  $C(X), C(X_1), C(\tilde{X})$  and  $C(X_*)$ , respectively, so that

$$M_1 = X_1(X_1'X_1)^{-1}X_1' = \text{Bdiag} \left( \frac{1}{m} J_m J_m' \right) = \frac{1}{m} X_1 X_1', \tag{13}$$

and define

$$M_2 = (I - M_1)X_2 \left( X_2'(I - M_1)X_2 \right)^{-1} X_2'(I - M_1) \tag{14}$$

to be the PPO onto  $C(X_1)^\perp_{C(X)}$  which under our assumptions is also the PPO onto  $C(X_*)^\perp_{C(\tilde{X})}$ . From Christensen (2011, Chapter 11) the projection operators satisfy

$$\tilde{M} = M_* + M_2, \quad M_*M_1 = M_*, \quad M_1M_2 = 0 \quad \text{and} \quad M = M_1 + M_2. \tag{15}$$

For completeness, these properties are reproven in Appendix 1. The sum of squares for whole plot error and subplot error are, respectively,

$$SSE(w) \equiv Y'(M_1 - M_*)Y \quad \text{and} \quad SSE(s) \equiv Y'(I - M)Y, \tag{16}$$

such that

$$SSE(s) = Y'(I - M)Y = Y'(I - \tilde{M})Y - Y'(M_1 - M_*)Y. \tag{17}$$

The F statistic for testing (12) is defined as

$$F = \frac{MSE(w)}{MSE(s)} = \frac{SSE(w)/[r(X_1) - r(X_*)]}{SSE(s)/[n - r(X)]} = \frac{SSE(w)/[N - r(X_*)]}{SSE(s)/[n - r(X)]}. \tag{18}$$

## 2 The equivalence between the LRT and F-test

### 2.1 Maximum likelihood estimators (MLEs)

The likelihood function for model (7) is:

$$L(\beta_*, \sigma_w^2, \sigma_s^2 | Y) = \frac{1}{(2\pi)^{n/2} |V|^{1/2}} e^{-1/2(Y - \tilde{X}\beta_*)'V^{-1}(Y - \tilde{X}\beta_*)}, \tag{19}$$

where

$$V = V(\sigma_w^2, \sigma_s^2) = m\sigma_w^2 M_1 + \sigma_s^2 I_n. \tag{20}$$

We provide three lemmas:

**Lemma 1** *The inverse of  $aI_n + bP$ , where  $P$  is a PPO and  $a$  and  $b$  are real numbers such that  $a \neq 0$  and  $a \neq -b$ , is*

$$(aI_n + bP)^{-1} = \frac{1}{a} \left( I_n - \frac{b}{a+b} P \right). \tag{21}$$

**Proof of Lemma 1** See Proposition 12.11.1 in Christensen (2011, p.322). □

**Lemma 2** *The determinant of  $aI_n + bP$ , where  $P$  is a PPO and  $a$  and  $b$  are real numbers, is*

$$|aI_n + bP| = a^{n-r(P)}(a + b)^{r(P)}. \tag{22}$$

**Proof of Lemma 2** Any nonzero vector in  $C(P)^\perp$  is an eigenvector for the eigenvalue  $a$ , so  $a$  has multiplicity  $n - r(P)$ . Similarly, any nonzero vector in  $C(P)$  is an eigenvector for the eigenvalue  $a + b$ , so  $a + b$  has multiplicity  $r(P)$ . The determinant is the product of the eigenvalues and hence  $|aI_n + bP| = a^{n-r(P)}(a + b)^{r(P)}$ . Appendix 2 contains an illustration. □

**Lemma 3** For  $q_1, q_2 > 0$ , maximizing the function

$$g(x_1, x_2) = - \left[ \text{constant} + q_1 \log(x_1) + q_2 \log(x_2) + q_1 \left( \frac{Q_1}{x_1} \right) + q_2 \left( \frac{Q_2}{x_2} \right) \right] \tag{23}$$

subject to the constraint  $x_2 \geq x_1 > 0$  gives a maximum at  $(x_1, x_2) = (Q_1, Q_2)$  when  $x_2 > x_1 > 0$  (i.e. when  $(x_1, x_2)$  is in the interior of the constraint) or a maximum at  $(x_1, x_2) = \left( \frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}, \frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2} \right)$  when  $x_2 = x_1$  (i.e. when  $(x_1, x_2)$  is on the boundary of the constraint).

The proof is in Appendix 3.

Applying Lemmas 1 and 2 to  $V$  in (20) gives the following determinant and inverse covariance matrix:

$$|V| = \left| \sigma_s^2 + m\sigma_w^2 \right|^N \left| \sigma_s^2 \right|^{n-N}, \tag{24}$$

and

$$V^{-1} = \frac{1}{\sigma_s^2} I_n - \left( \frac{\sigma_w^2}{\sigma_s^2} \right) \frac{m}{\sigma_s^2 + m\sigma_w^2} M_1. \tag{25}$$

Substituting (24) in (19) and taking  $-2$  times the natural logarithm leads to

$$\begin{aligned} \ell(\beta_*, \sigma_w^2, \sigma_s^2 | Y) &= n \log(2\pi) + N \log \left( \sigma_s^2 + m\sigma_w^2 \right) + N(m - 1) \log(\sigma_s^2) \\ &\quad + (Y - \tilde{X}\beta_*)' V^{-1} (Y - \tilde{X}\beta_*). \end{aligned} \tag{26}$$

**Proposition 1** The Maximum Likelihood estimators for  $\beta_*$ ,  $\sigma_w^2$  and  $\sigma_s^2$  of model (7) are

$$\hat{\beta}_* = \left( \tilde{X}' V^{-1} \tilde{X} \right)^{-1} \tilde{X}' V^{-1} Y = \left( \tilde{X}' \tilde{X} \right)^{-1} \tilde{X}' Y, \tag{27}$$

$$\hat{\sigma}_w^2 = \frac{1}{m} \max \left\{ 0, \frac{SSE(w)}{N} - \frac{SSE(s)}{n - N} \right\}, \text{ and} \tag{28}$$

$$\hat{\sigma}_s^2 = \min \left\{ \frac{SSE(s)}{n - N}, \frac{SSE(s) + SSE(w)}{n} \right\}, \tag{29}$$

such that the pair  $\hat{\sigma}_w^2 = 0$  and  $\hat{\sigma}_s^2 = \frac{SSE(s) + SSE(w)}{n}$  occurs when  $\frac{SSE(w)}{N} \leq \frac{SSE(s)}{n - N}$  and the other pair  $\hat{\sigma}_w^2 = \frac{1}{m} \left[ \frac{SSE(w)}{N} - \frac{SSE(s)}{n - N} \right]$  and  $\hat{\sigma}_s^2 = \frac{SSE(s)}{n - N}$  occurs otherwise.

**Proof of Proposition 1** Differentiating (26) with respect to  $\beta_*$  and setting the partial derivative to zero, leads to

$$\hat{\beta}_* = \left( \tilde{X}' V^{-1} \tilde{X} \right)^{-1} \tilde{X}' V^{-1} Y. \tag{30}$$

It is well known, for the fixed effects in mixed models, that the maximum likelihood estimates (MLEs) are also the best linear unbiased estimates (BLUEs) and Christensen (2011, Chapter 11) has shown, since  $C(V\tilde{X}) \subset C(\tilde{X})$ , that the ordinary least squares estimates (OLSEs) are the BLUEs for  $\beta_*$  so they are also the maximum likelihood estimates. Therefore,  $\hat{\beta}_* = (\tilde{X}'V^{-1}\tilde{X})^{-1}\tilde{X}'V^{-1}Y = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y$  and subsequently  $\tilde{X}\hat{\beta}_*$  could be computed, using the OLSEs, as

$$\tilde{X}\hat{\beta}_* = \tilde{M}Y. \tag{31}$$

Note that the least squares estimates do not depend on the variance parameters so to find  $\hat{\sigma}_w^2$  and  $\hat{\sigma}_s^2$ , we need to maximize

$$\begin{aligned} \ell_*(\hat{\beta}_*, \sigma_w^2, \sigma_s^2|Y) &\equiv -\ell(\hat{\beta}_*, \sigma_w^2, \sigma_s^2|Y) = -[n \log(2\pi) + N(m-1) \log(\sigma_s^2) \\ &\quad + N \log(\sigma_s^2 + m\sigma_w^2) + \Psi(\sigma_w^2, \sigma_s^2)], \end{aligned} \tag{32}$$

Not re-expressing  $\Psi$  in terms of the sum of square errors would make it impossible to find closed form solutions that maximize  $\ell_*(\hat{\beta}_*, \sigma_w^2, \sigma_s^2|Y)$ . Hence, since  $\Psi = (Y - \tilde{M}Y)'V^{-1}(Y - \tilde{M}Y)$ , we get

$$\begin{aligned} \ell_*(\hat{\beta}_*, \sigma_w^2, \sigma_s^2|Y) &= -\left[ n \log(2\pi) + N(m-1) \log(\sigma_s^2) + N \log(\sigma_s^2 + m\sigma_w^2) \right. \\ &\quad \left. + \frac{SSE(s)}{\sigma_s^2} + \frac{SSE(w)}{\sigma_s^2 + m\sigma_w^2} \right]. \end{aligned} \tag{33}$$

To apply Lemma 3 to  $\ell_*$ , let  $q_1 = N(m-1) = n - N$ ,  $q_2 = N$ ,  $x_1 = \sigma_s^2$ ,  $x_2 = \sigma_s^2 + m\sigma_w^2$ ,  $Q_1 = \frac{SSE(s)}{q_1}$  and  $Q_2 = \frac{SSE(w)}{q_2}$ . A key point is  $x_2 \geq x_1$  so our maximization has to be done subject to that constraint. Lemma 3 gives the maximizers

$$\begin{aligned} (\hat{\sigma}_s^2, \hat{\sigma}_s^2 + m\hat{\sigma}_w^2) &= \left( \frac{SSE(s)}{n - N}, \frac{SSE(w)}{N} \right) \\ \iff (\hat{\sigma}_s^2, \hat{\sigma}_w^2) &= \left( \frac{SSE(s)}{n - N}, \frac{1}{m} \left[ \frac{SSE(w)}{N} - \frac{SSE(s)}{n - N} \right] \right) \end{aligned} \tag{34}$$

when  $\frac{SSE(w)}{N} > \frac{SSE(s)}{n - N}$  and

$$\begin{aligned} (\hat{\sigma}_s^2, \hat{\sigma}_s^2 + m\hat{\sigma}_w^2) &= \left( \frac{SSE(s) + SSE(w)}{n}, \frac{SSE(s) + SSE(w)}{n} \right) \\ \iff (\hat{\sigma}_s^2, \hat{\sigma}_w^2) &= \left( \frac{SSE(s) + SSE(w)}{n}, 0 \right) \end{aligned} \tag{35}$$

when  $\frac{SSE(w)}{N} \leq \frac{SSE(s)}{n - N}$ .



Now, suppose that the MLE of  $\sigma_w^2$  is  $\hat{\sigma}_w^2 = \frac{1}{m} \left( \frac{SSE(w)}{N} - \frac{SSE(s)}{n-N} \right)$  then  $\frac{SSE(s)}{n-N} \leq \frac{SSE(w)}{N}$  so that

$$\frac{SSE(s)}{n-N} \leq \left( \frac{n-N}{n} \right) \frac{SSE(s)}{n-N} + \left( \frac{N}{n} \right) \frac{SSE(w)}{N} = \frac{SSE(s) + SSE(w)}{n} \tag{36}$$

with the first inequality true because the term in the middle is a weighted average so has to be larger than the smaller of the two things being averaged, therefore the MLE of  $\sigma_s^2$  is the smaller of the terms  $\frac{SSE(s)}{n-N}$  and  $\frac{SSE(s)+SSE(w)}{n}$ . That is, the larger term between 0 and  $\frac{1}{m} \left( \frac{SSE(w)}{N} - \frac{SSE(s)}{n-N} \right)$  forces the answer to be the smaller term between  $\frac{SSE(s)}{n-N}$  and  $\frac{SSE(s)+SSE(w)}{n}$  and vice versa. Hence (34) and (35) could be written via max and min as

$$\hat{\sigma}_w^2 = \frac{1}{m} \max \left\{ 0, \frac{SSE(w)}{N} - \frac{SSE(s)}{n-N} \right\}$$

and

$$\hat{\sigma}_s^2 = \min \left\{ \frac{SSE(s)}{n-N}, \frac{SSE(s) + SSE(w)}{n} \right\}$$

□

Note that the partial derivatives for (33) are

$$\frac{\partial \ell}{\partial \sigma_w^2} = - \left[ \frac{n}{\sigma_s^2 + m\sigma_w^2} - \frac{mSSE(w)}{(\sigma_s^2 + m\sigma_w^2)^2} \right], \tag{37}$$

and

$$\frac{\partial \ell}{\partial \sigma_s^2} = - \left[ \frac{N}{\sigma_s^2 + m\sigma_w^2} + \frac{N(m-1)}{\sigma_s^2} - \frac{SSE(s)}{(\sigma_s^2)^2} - \frac{SSE(w)}{(\sigma_s^2 + m\sigma_w^2)^2} \right]. \tag{38}$$

So, for varification purposes, plugging in the pair  $\hat{\sigma}_w^2 = 0$  and  $\hat{\sigma}_s^2 = \frac{SSE(s)+SSE(w)}{n}$  into (38) and the other pair  $\hat{\sigma}_w^2 = \frac{1}{m} \left[ \frac{SSE(w)}{N} - \frac{SSE(s)}{n-N} \right]$  and  $\hat{\sigma}_s^2 = \frac{SSE(s)}{n-N}$  into (37) gives zero as desired.

### 3 Monotonic relationship between the LRT and F-test statistics

We show that the LRT statistic  $\Lambda$  is a monotone function of the F-test statistic  $F$  for testing the null hypothesis in (12). When  $\Lambda$  is not 0, the monotone relationship is strict, so whenever the size of the test  $\alpha$  is smaller than the probability  $1 - p_m$  that  $\Lambda \neq 0$ , the tests are equivalent. We also examine the behavior of the tests when they are not equivalent (i.e. when  $\alpha > 1 - p_m$ ). To establish this monotone relationship we

need to distinguish between the sum of squared errors and model parameters under the reduced model in (11) versus the full model in (7). In particular, the sum of squares for errors under the reduced model is

$$SSE(e) \equiv SSE(w) + SSE(s). \tag{39}$$

Similarly, we use  $\sigma_e^2$  to denote the variance parameter corresponding to the reduced model while  $\sigma_w^2$  and  $\sigma_s^2$  denote the variance parameters of the full model. Since the reduced and full models are nested we present the equivalence between the LRT and F-test under the full model. Since there are two cases for the MLEs of  $\sigma_w^2$  and  $\sigma_s^2$ , the relationship between the statistics  $\Lambda$  and  $F$  will be decomposed into two cases as well.

**Proposition 2** *The LRT statistic  $\Lambda$ , for (12), is a monotone function of the  $F$  statistic. In particular, (case-I) when  $\hat{\sigma}_w^2 = 0$  and  $\hat{\sigma}_s^2 = \frac{SSE(s)+SSE(w)}{n}$  we have*

$$\Lambda = 0, \tag{40}$$

and (case-II) when  $\hat{\sigma}_w^2 = \frac{1}{m} \left[ \frac{SSE(w)}{N} - \frac{SSE(s)}{n-N} \right]$  and  $\hat{\sigma}_s^2 = \frac{SSE(s)}{n-N}$  we have

$$\begin{aligned} \Lambda &= n \log \left( \frac{m-1}{m} \right) + N \log \left( \frac{N(n-r(X))}{(n-N)(N-r(X_*))} \right) \\ &+ n \log \left( 1 + \frac{N-r(X_*)}{n-r(X)} F \right) + N \log \left( \frac{1}{F} \right) \end{aligned} \tag{41}$$

such that case-I occurs when  $F \leq \kappa$  and case-II occurs when  $F > \kappa$  where

$$\kappa = \frac{n-r(X)}{(m-1)(N-r(X_*))}. \tag{42}$$

In Case-II, the relationship is strictly monotone so the tests are equivalent as long as the alpha level is smaller than  $P(F > \kappa)$ . The plot of  $\Lambda$  as a function of  $F$  is presented in Fig. 1.

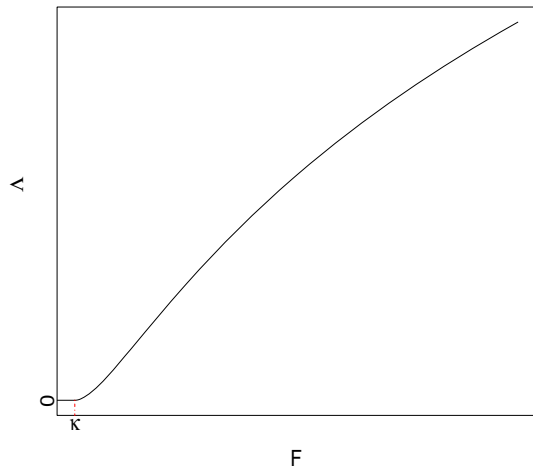
**Proof of proposition (2)** To examine the variation between plots, we test the whole plot error as in (12). Note that, under  $H_0$  we get the reduced model (11) where  $\epsilon \sim N(0, \sigma_e^2 I)$  with

$$L(\beta_*, \sigma_e^2 | Y) = \frac{1}{(2\pi\sigma_e^2)^{n/2}} e^{-\frac{1}{2\sigma_e^2}(Y-\tilde{X}\beta_*)'(Y-\tilde{X}\beta_*)}, \tag{43}$$

and

$$\ell(\hat{\beta}_*, \sigma_e^2 | Y) = n \log(2\pi) + n \log(\sigma_e^2) + \frac{SSE(e)}{\sigma_e^2}, \tag{44}$$

**Fig. 1** A plot of the LRT statistic  $\Lambda$  versus the  $F$  statistic according to (40) and (41). When  $F \leq \kappa$ ,  $\Lambda$  is constant and equal to 0



so that the MLE of  $\sigma_e^2$  is

$$\hat{\sigma}_e^2 = \frac{SSE(e)}{n}. \tag{45}$$

In our case, (7) is the full model and (11) is the reduced one so that

$$\Lambda = \ell_R(\hat{\beta}_*, \hat{\sigma}_e^2|Y) - \ell_F(\hat{\beta}_*, \hat{\sigma}_w^2, \hat{\sigma}_s^2|Y), \tag{46}$$

where  $\ell_R(\hat{\beta}_*, \hat{\sigma}_e^2|Y) \equiv \sup \ell(\hat{\beta}_*, \sigma_e^2|Y)$  and  $\ell_F(\hat{\beta}_*, \hat{\sigma}_w^2, \hat{\sigma}_s^2|Y) \equiv \sup \ell(\hat{\beta}_*, \sigma_w^2, \sigma_s^2|Y)$  such that the  $R$  and  $F$  subscripts refer to the reduced and full models respectively. Using (39), then plugging (45) into (44) gives

$$\ell_R(\hat{\beta}_*, \hat{\sigma}_e^2|Y) = n \log(2\pi) + n - n \log(n) + n \log(SSE(s) + SSE(w)). \tag{47}$$

Further, if we let  $\hat{\sigma}_s^2$  and  $\hat{\sigma}_w^2$  be the MLEs for  $\sigma_s^2$  and  $\sigma_w^2$  and plug them into (33) we get

$$\begin{aligned} \ell_F(\hat{\beta}_*, \hat{\sigma}_w^2, \hat{\sigma}_s^2|Y) &= n \log(2\pi) + N(m-1) \log(\hat{\sigma}_s^2) + N \log(\hat{\sigma}_s^2 + m\hat{\sigma}_w^2) \\ &+ \frac{SSE(s)}{\hat{\sigma}_s^2} + \frac{SSE(w)}{\hat{\sigma}_s^2 + m\hat{\sigma}_w^2}. \end{aligned} \tag{48}$$

**Case-I:** If  $\hat{\sigma}_w^2 = 0$  and  $\hat{\sigma}_s^2 = \frac{SSE(s)+SSE(w)}{n}$  then

$$\begin{aligned} \Lambda &= n - n \log(n) + n \log(SSE(s) + SSE(w)) \\ &- n \log\left(\frac{SSE(s) + SSE(w)}{n}\right) - n = 0. \end{aligned} \tag{49}$$

**Case-II:** If  $\hat{\sigma}_w^2 = \frac{1}{m} \left[ \frac{SSE(w)}{N} - \frac{SSE(s)}{n-N} \right]$  and  $\hat{\sigma}_s^2 = \frac{SSE(s)}{n-N}$  then, by (18), we get

$$\Lambda = n \log \left( \frac{m-1}{m} \right) + N \log \left( \frac{N(n-r(X))}{(n-N)(N-r(X_*))} \right) + n \log \left( 1 + \frac{N-r(X_*)}{n-r(X)} F \right) + N \log \left( \frac{1}{F} \right). \tag{50}$$

We note that this case holds only when  $\hat{\sigma}_w^2 > 0$  so  $F$  must be larger than  $\kappa$  since

$$\frac{1}{m} \left[ \frac{SSE(w)}{N} - \frac{SSE(s)}{n-N} \right] > 0 \iff F > \frac{n-r(X)}{(m-1)(N-r(X_*))} = \kappa. \tag{51}$$

The function of  $\Lambda$  for Case II given in (50) is strictly increasing in  $F$  since

$$\frac{\partial \Lambda}{\partial F} > 0 \iff \frac{n \frac{N-r(X_*)}{n-r(X)}}{1 + \frac{N-r(X_*)}{n-r(X)} F} - \frac{N}{F} > 0 \iff F > \kappa. \tag{52}$$

□

### 4 The distribution of the F-test and LRT statistics

The form of relation between the  $\Lambda$  and  $F$  helps us better understand the distribution of  $\Lambda$ . In particular, this relation implies an important lemma (Lemma 4) on the distribution of  $\Lambda$  for (7).

**Proposition 3** *The distribution of the F-ratio in (18) for the model in (7) is a constant multiple of a central F distribution,*

$$F = \frac{MSE(w)}{MSE(s)} \sim \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} F(N-r(X_*), n-r(X)). \tag{53}$$

**Proof of Proposition (3)** It has been shown in Chapter 11.2 of Christensen (2011) that

$$W_1 := \frac{SSE(w)}{\sigma_s^2 + m\sigma_w^2} = \frac{Y'(M_1 - M_*)Y}{\sigma_s^2 + m\sigma_w^2} \sim \chi_{N-r(X_*)}^2, \tag{54}$$

$$W_2 := \frac{SSE(s)}{\sigma_s^2} = \frac{Y'(I - M)Y}{\sigma_s^2} \sim \chi_{n-r(X)}^2, \tag{55}$$

and  $SSE(w)$  is independent of  $SSE(s)$ . Thus, since the F-distribution arises from the ratio of two independent chi-squared random variables, each divided by its respective degrees of freedom, we have

$$F = \frac{SSE(w)/[N-r(X_*)]}{SSE(s)/[n-r(X)]}$$

$$= \left( \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} \right) \frac{W_1/[N - r(X_*)]}{W_2/[n - r(X)]} \sim \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} F(N - r(X_*), n - r(X)). \tag{56}$$

□

Now, if we let  $W \sim F(N - r(X_*), n - r(X))$  then  $\hat{\sigma}_w^2 = 0 \iff F \leq \kappa \iff W \leq \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}$  where the first *if and only if* is an exact relationship and the second one is only a distributional relationship.

**Lemma 4** *The distribution of the LRT statistic  $\Lambda$  for the model in (7) is determined by the relationship in (57) where  $W \sim F(N - r(X_*), n - r(X))$ ,  $p_m$ ,  $a$ , and  $\tau$  are as described.*

$$\Lambda \sim \begin{cases} 0 & \text{if } W \leq \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2} \\ \tau + N \log \left( \frac{(1+aW)^m}{W} \right) & \text{if } W > \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2} \end{cases} \tag{57}$$

such that

$$p_m = Pr \left( W \leq \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2} \right), \tag{58}$$

where

$$a = \frac{N - r(X_*)}{n - r(X)} \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2}, \tag{59}$$

and

$$\tau = N \log \left( \frac{(m - 1)^{m-1}}{m^m} \frac{1}{a} \right). \tag{60}$$

**Proof of Lemma 4** From case-I of Proposition 2 we know that  $\Lambda = 0$  iff  $F \leq \kappa$  so

$$\begin{aligned} p_m &\equiv P(\Lambda = 0) = P(F \leq \kappa) \\ &= P \left( \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} W \leq \kappa \right) = P \left( W \leq \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2} \right). \end{aligned} \tag{61}$$

The equality in the second line of (61) holds due to Proposition 3.

Now, from case-II of Proposition 2 we also know that, if  $F > \kappa$  (i.e. when  $W > \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}$ ),

$$\begin{aligned} \Lambda &= n \log \left( \frac{m - 1}{m} \right) + N \log \left( \frac{N(n - r(X))}{(n - N)(N - r(X_*))} \right) \\ &\quad + n \log \left( 1 + \frac{N - r(X_*)}{n - r(X)} F \right) + N \log \left( \frac{1}{F} \right). \end{aligned} \tag{62}$$

So, if we let  $W \sim F(N - r(X_*), n - r(X))$  then substituting  $F$  from (53) into (62) gives

$$\begin{aligned} \Lambda &= n \log \left( \frac{m - 1}{m} \right) + N \log \left( \frac{N(n - r(X))}{(n - N)(N - r(X_*))} \right) \\ &\quad + n \log \left( 1 + \frac{N - r(X_*)}{n - r(X)} F \right) + N \log \left( \frac{1}{F} \right) \\ &\sim \tau + N \log \left[ \frac{(1 + aW)^m}{W} \right], \end{aligned} \tag{63}$$

where  $a = \frac{N - r(X_*)}{n - r(X)} \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2}$  and  $\tau = N \log \left( \frac{(m - 1)^{m - 1}}{m^m} \frac{1}{a} \right)$ . □

We note that  $\lim_{W \rightarrow \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}} \Lambda = 0$ . This result can be verified by plugging in  $W = \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}$  in the third equality of (63). In practice, the probability mass at zero for the likelihood ratio test in (58) should be numerically estimated by using the MLEs as follows.

$$\begin{aligned} p_m &\equiv P(\Lambda = 0) = P(F \leq \kappa) \\ &= P \left( W \leq \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2} \right) \approx P \left( W \leq \frac{\kappa\hat{\sigma}_s^2}{\hat{\sigma}_s^2 + m\hat{\sigma}_w^2} \right). \end{aligned} \tag{64}$$

Providing an exact distribution of the test statistic for the LRT in these models will help in calculating a more accurate p-value than the traditionally used p-value from the large sample chi-square mixture approximations introduced by Self and Liang (1987). The distribution of  $\Lambda$  under the null hypothesis is obtained by substituting zero for  $\sigma_w^2$  in Lemma 4. This implies that, under  $H_0$ , the probability that  $\Lambda$  is zero does not depend on the parameter  $\sigma_s^2$  and equals

$$p_m = P(W \leq \kappa). \tag{65}$$

GSP models offer remarkable advantages in the ability to perform exact calculations. For example, Crainiceanu and Ruppert (2004) derive the probability mass at zero for the likelihood ratio test in linear mixed models (LMM) with one variance component as

$$p_c = P \left( \frac{\sum_{s=1}^K \mu_{s,n} w_s^2}{\sum_{s=1}^{n - \tilde{p}} w_s^2} \leq \frac{1}{n} \sum_{i=1}^K \xi_{s,n} \right), \tag{66}$$

where  $\mu_{s,n}$  and  $\xi_{s,n}$  are the  $K$  eigenvalues of the  $K \times K$  matrices  $X_1' P_0 X_1$  and  $X_1' X_1$  respectively, where  $w_i \sim N(0, 1)$ ,  $P_0 = I_n - \tilde{X} (\tilde{X}' \tilde{X})^{-1} \tilde{X}'$  and  $\tilde{p}$  is the dimensionality of the vector  $\beta_*$  in view of (7). We note that  $\tilde{p}$  is the rank of the design

matrix  $\tilde{X}$  instead of the dimensionality of the vector  $\beta_*$  when the generalized inverse is used. We present the equivalence between the two formulas in (65) and (66) in the discussion that follows.

Considering the model in (7), one can show that the eigenvalues  $\xi_{s,n}$ , of  $X_1'X_1$ , are  $m$  of multiplicity  $N$  and the eigenvalues  $\mu_{s,n}$ , of  $X_1'P_0X_1$ , are  $m$  of multiplicity  $N - r(X_*)$  and zero of multiplicity  $r(X_*)$ . Further, from (15), it's immediate that  $M_* \perp (M - M_1)$  and  $M_1 \subset M$  so

$$\begin{aligned} r(\tilde{X}) &= r(\tilde{M}) = r(M_* + M_2) = r(M_* + (M - M_1)) \\ &= r(M_*) + r(M - M_1) = r(M_*) + r(M) - r(M_1) = r(X_*) + r(X) - N. \end{aligned} \tag{67}$$

Thus, if we let  $w_s \sim N(0, 1)$  then according to (66), the probability mass at zero for the likelihood ratio test in testing (12) is

$$\begin{aligned} p_c &= P\left(\frac{\sum_{s=1}^K \mu_{s,n} w_s^2}{\sum_{s=1}^{n-p} w_s^2} \leq \frac{1}{n} \sum_{i=1}^K \xi_{s,n}\right) = P\left(\frac{\sum_{s=1}^{N-r(X_*)} m w_s^2}{\sum_{s=1}^{n-(r(X_*)+r(X)-N)} w_s^2} \leq \frac{1}{n} \sum_{i=1}^N m\right) \\ \iff p_c &= P\left(\frac{\sum_{s=1}^{N-r(X_*)} w_s^2}{\sum_{s=N-r(X_*)+1}^{n-(r(X_*)+r(X)-N)} w_s^2} \leq \frac{1}{m-1}\right) = P\left(\frac{W_1}{W_2} \leq \frac{1}{m-1}\right) \\ \iff p_c &= P(W \leq \kappa), \end{aligned} \tag{68}$$

which is the same as (65).

The formula in (65) is an easier way of getting the probability mass at zero for the likelihood ratio test under  $H_0$  than (66). In fact, our results stand out relative to the derivation of Crainiceanu and Ruppert (2004) for being nice and compact and because it's not possible in general to give such a short and straightforward expression for computing the probability mass at zero for the LRT statistic. Also, to the best of our knowledge, it is the first time that an explicit mathematical form, Lemma 4, has been presented for the LRT for any variance component in a linear mixed model under the full model. This allows us compute the power of the test through a formula instead of a Monte Carlo simulation.

### 5 Power comparison

This section includes four subsections. In the Sect. 1, we illustrate the steps for computing the critical value and power of the F-test and give a concrete example. In the Sect. 2, we illustrate the steps for computing the critical value and power of the LRT when  $\alpha \leq 1 - p_m$  (i.e. when there is no randomized test) and give a concrete example; we use the very same example that we use for the F-test to show the equivalence in power through a numerical example. These illustrative steps and example, of the Sect. 2, don't use the relationship between the two test statistics  $\Lambda$  and  $F$ . Thus, the Sect. 2 is concluded with a very short theorem on the equivalence between the two test

for the case when  $\alpha \leq 1 - p_m$ . The Sect. 4 gives a detailed proof, without numerical examples, that the F-test has a larger power than the LRT when  $\alpha > 1 - p_m$ . The Sect. 5 discusses the practicality of the second case when  $\alpha > 1 - p_m$ .

### 5.1 The power of the F-test

The F-test statistic, by Proposition 3, has a distribution that is a constant multiple of an F distribution as  $F \sim \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} F(N - r(X_*), n - r(X))$ . Since the F-test statistic is denoted by  $F$  and the F-distribution is traditionally known by the symbol  $F$ , to eliminate ambiguities, we let  $W \sim F(N - r(X_*), n - r(X))$ . Then, at a given significance level  $\alpha$ , the critical value  $C$  is computed under  $H_0$  as

$$\begin{aligned} \alpha = P(F \geq C | H_0 \text{ is true}) &\iff \alpha = P\left(\frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} W \geq C | \sigma_w^2 = 0\right) \\ &\iff \alpha = P(W \geq W_\alpha) \\ &\iff W_\alpha = G^{-1}(1 - \alpha), \end{aligned} \tag{69}$$

$G$  is the CDF for  $F_{(N-r(X_*), n-r(X))}$ . For example, if we let  $N - r(X_*) = 3$  and  $n - r(X) = 9$  then, for  $\alpha = 0.05$ ,  $C$  is found as  $C = G^{-1}(1 - \alpha) = G^{-1}(0.95) = 3.86255$ .

If  $m = 4$ ,  $\sigma_s^2 = 3$  and  $\sigma_w^2 = 7$ , then the power of a size  $\alpha$  F-test is

$$\begin{aligned} \mathcal{E}_F = P(F \geq C | H_a \text{ is true}) &\iff \mathcal{E}_F = P\left(\frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} W \geq W_\alpha\right) \\ &\iff \mathcal{E}_F = 1 - G\left(\frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2} W_\alpha\right). \end{aligned} \tag{70}$$

For example, if we let  $N - r(X_*) = 3$ ,  $n - r(X) = 9$ ,  $m = 4$ ,  $\sigma_s^2 = 3$  and  $\sigma_w^2 = 7$  then, for  $\alpha = 0.05$ , the critical value is 3.86255 and the power is  $\mathcal{E}_F = 1 - G\left(\frac{3}{31} \times 3.86255\right) = 0.7741$ .

### 5.2 The power of the LRT when $\alpha \leq 1 - p_m$

The LRT statistic, by Lemma 4, has a mixture distribution as

$$\Lambda \sim \begin{cases} 0 & \text{w.p. } p_m \\ \tau + N \log\left(\frac{(1+aW)^m}{W}\right) & \text{w.p. } 1 - p_m \end{cases} \tag{71}$$

such that  $p_m$ ,  $\tau$  and  $a$  are defined in Lemma 4. Thus, at a given significance level  $\alpha$ , the critical value  $C'$  is computed under  $H_0$  as

$$\alpha = P\left(\Lambda \geq C' | H_0 \text{ is true}, W > \frac{\kappa \sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right)$$



$$\begin{aligned} &\iff (1 - p_m)P\left(\tau + N \log\left(\frac{(1 + aW)^m}{W}\right) \geq C' \mid \sigma_w^2 = 0, W > \kappa\right) = \alpha \\ &\iff C' = G'^{-1}\left(1 - \frac{\alpha}{1 - p_m}\right), \end{aligned} \tag{72}$$

where  $G'$  is the CDF of the transformed random variable  $\tau + N \log\left(\frac{(1+aW)^m}{W}\right)$  for  $W \sim F(N - r(X_*), n - r(X))$  when  $W > \kappa$  and  $\sigma_w^2 = 0$ . For example, if we let  $N - r(X_*) = 3$  and  $n - r(X) = 9$  then, for  $\alpha = 0.05$  under  $H_0, \kappa = 1, p_m = 0.56371$  and

$$\Lambda = \begin{cases} 0 & \text{w.p. } 0.56371 \\ 6 \log\left[\frac{81}{256} \frac{\left(1 + \frac{W}{3}\right)^4}{W}\right] & \text{w.p. } 0.43629 \end{cases} \tag{73}$$

so that  $C'$  is found, by numerical simulation or numerical integration after transformation, as  $C' = G'^{-1}\left(1 - \frac{\alpha}{1 - p_m}\right) = G^{-1}(0.8853973) = 4.848$ . See Castellacci (2012) for more details on computing the quantiles of mixture distributions.

If  $m = 4, \sigma_s^2 = 3$  and  $\sigma_w^2 = 7$  then the power of a size  $\alpha$  LRT is

$$\begin{aligned} \mathcal{E}_{LRT} &= P\left(\Lambda \geq C' \mid H_a \text{ is true}, W > \frac{\kappa \sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\ &\iff \mathcal{E}_{LRT} \\ &= (1 - p_m)P\left(\tau + N \log\left(\frac{(1 + aW)^m}{W}\right) \geq C' \mid \sigma_w^2 > 0, W > \frac{\kappa \sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\ &\iff \mathcal{E}_{LRT} = (1 - p_m) \left[1 - G''(C')\right] \end{aligned} \tag{74}$$

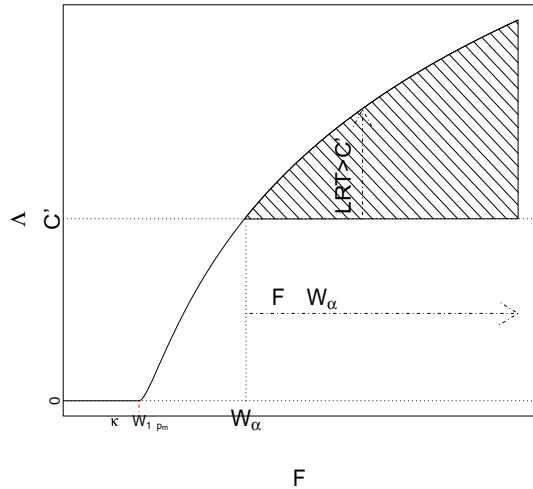
where  $G''$  is the CDF of the transformed random variable  $\tau + N \log\left(\frac{(1+aW)^m}{W}\right)$  for  $W \sim F(N - r(X_*), n - r(X))$  when  $W > \frac{\kappa \sigma_s^2}{\sigma_s^2 + m\sigma_w^2}$  and  $\sigma_w^2 > 0$ . For example, if we let  $N - r(X_*) = 3, n - r(X) = 9, m = 4, \sigma_s^2 = 3$  and  $\sigma_w^2 = 7$  then, for  $\alpha = 0.05$  under  $H_1, \kappa = \frac{3}{4}, p_m = 0.04014$  and

$$\Lambda \sim \begin{cases} 0 & \text{w.p. } 0.04014 \\ 6 \log\left[\frac{243}{7936} \frac{\left(1 + \frac{31W}{9}\right)^4}{W}\right] & \text{w.p. } 0.95986 \end{cases} \tag{75}$$

so that  $\mathcal{E}_{LRT}$  is found, by numerical simulation or numerical integration after transformation, as  $\mathcal{E}_{LRT} = (1 - p_m) \left[1 - G''(C')\right] = 0.95986 \left[1 - G''(4.848)\right] = 0.7741$ .

Note that both test statistics  $\Lambda$  and  $F$  are nonnegative and whenever  $\Lambda \neq 0$  there is a strict monotonic relationship and thus when the LRT critical region does not include 0, the tests are the same. In fact, in the case when  $\alpha \leq 1 - p_m$ , the critical region will

**Fig. 2** A plot of the *LRT* statistic versus the *F*-ratio showing their equivalence whenever  $\alpha \leq 1 - p_m$ .  $W_{1-p_m} = \kappa$  is the minimal critical value at which the two tests are equivalent



consist of positive values where  $\Lambda$  is a strictly increasing function of the  $F$ , thus we have

**Proposition 4** *Let  $\alpha$  be the size of the test. If  $\alpha \leq 1 - p_m$  where  $p_m = P(\Lambda = 0 | \sigma_w^2 = 0)$  then the *F*-test and *LRT* are equivalent and hence have the same power.*

**Proof of Proposition (4)** Since  $\Lambda$  can be written as

$$\Lambda \sim \begin{cases} 0 & \text{w.p. } p_m \\ g(F) & \text{w.p. } 1 - p_m \end{cases} \tag{76}$$

where  $g(\cdot)$  is a strictly increasing function, then the critical region of the *LRT* when  $\alpha \leq 1 - p_m$  doesn't involve 0 and hence the power can be calculated as

$$\begin{aligned} \mathcal{E}_{LRT} &= P(\Lambda \geq C' | H_a \text{ is true}) \iff \mathcal{E}_{LRT} = P(g(F) \geq C' | \sigma_w^2 > 0) \\ &\iff \mathcal{E}_{LRT} = P(F \geq C'' | \sigma_w^2 > 0) = \mathcal{E}_F. \end{aligned} \tag{77}$$

□

Figure 2 illustrates the equivalence of the *F*-test and *LRT* whenever  $\alpha \leq 1 - p_m$  where  $p_m = P(\Lambda = 0 | \sigma_w^2 = 0)$ . Further, it clarifies why the two tests are equivalent as long as the critical value  $W_\alpha$  of the *F*-test is larger than  $W_{1-p_m}$ ; the minimal critical value at which the two tests are equivalent. In fact, under the null hypothesis of  $\sigma_w^2 = 0$  we have  $\lim_{W \rightarrow \kappa} \Lambda = 0$ .

### 5.3 Power comparison when $\alpha > 1 - p_m$

In the case when  $\alpha > 1 - p_m$ , the critical region of the *LRT* involves  $\Lambda = 0$  hence it involves randomization. We show mathematically, for this case, that the power of the

F-test is larger than that of the LRT. Let  $k := \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2}$ . Firstly, we rewrite the power of a size  $\alpha$  F-test in terms of  $p_m$  the probability that  $\Lambda = 0$  under  $H_1$  as follows.

$$\begin{aligned} \mathcal{E}_F &= P(F \geq W_\alpha) = P(F \geq W_{1-p_m}) + P(W_\alpha \leq F \leq W_{1-p_m}) \\ &= P(kW \geq W_{1-p_m}) + P(W_\alpha \leq kW \leq W_{1-p_m}) \\ &= P\left(W \geq \frac{1}{k}W_{1-p_m}\right) + P\left(W \leq \frac{1}{k}W_{1-p_m}\right) - P\left(W \leq \frac{1}{k}W_\alpha\right). \end{aligned} \tag{78}$$

Note that the second equality in (78) is due to the probabilistic identity  $P(E) + P(E^c) = 1$ . Secondly, we rewrite the randomized test for the LRT in terms of the F-test according to the monotonic relationship between their test statistics and the smallest critical value,  $W_{1-p_m} = \kappa$ , where the F and LRT tests are equivalent as follows.

$$\phi(\Lambda) = \begin{cases} 1 & \text{if } \Lambda > 0 \\ \gamma & \text{if } \Lambda = 0 \\ 0 & \text{if } \Lambda < 0 \end{cases} \iff \phi(F) = \begin{cases} 1 & \text{if } F > W_{1-p_m} \\ \gamma & \text{if } F \leq W_{1-p_m} \end{cases} \tag{79}$$

where  $\gamma$  is determined according to the size of the test as

$$\begin{aligned} \alpha &= E_{H_0}\phi(\Lambda) \iff \alpha = P(\Lambda > 0 | \sigma_w^2 = 0) + \gamma P(\Lambda = 0 | \sigma_w^2 = 0) \\ &\iff \alpha = (1 - p_m) + \gamma p_m \\ &\iff \gamma = \frac{\alpha - (1 - p_m)}{p_m}. \end{aligned} \tag{80}$$

Hence, the power of the LRT is

$$\begin{aligned} \mathcal{E}_{LRT} &= P(F \geq W_{1-p_m}) + \frac{\alpha - (1 - p_m)}{p_m} P(F \leq W_{1-p_m}) \\ &= P\left(W \geq \frac{1}{k}W_{1-p_m}\right) + \frac{\alpha - (1 - p_m)}{p_m} P\left(W \leq \frac{1}{k}W_{1-p_m}\right). \end{aligned} \tag{81}$$

**Proposition 5** Let  $p_m = P(\Lambda = 0 | \sigma_w^2 = 0)$ . For GSP models with a finite whole plots size  $m$ , if  $\alpha > 1 - p_m$  then the power of the size  $\alpha$  F-test is larger than that of the LRT in testing  $\sigma_w^2 = 0$ .

**Proof of Proposition 5** It's sufficient to show that

$$\begin{aligned} P\left(W \leq \frac{1}{k}W_{1-p_m}\right) - P\left(W \leq \frac{1}{k}W_\alpha\right) &> \frac{\alpha - (1 - p_m)}{p_m} P\left(W \leq \frac{1}{k}W_{1-p_m}\right) \\ \iff \frac{1}{p_m} P\left(W \leq \frac{1}{k}W_{1-p_m}\right) &> \frac{1}{1 - \alpha} P\left(W \leq \frac{1}{k}W_\alpha\right). \end{aligned} \tag{82}$$

Since  $k = \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} > 1$ , this follows from Qeadan and Christensen (2014). □

### 5.4 Is $\alpha > 1 - p_m$ practical?

The LRT and F-test are equivalent as long as the level of the test is smaller or equal to  $P(W > \kappa)$  where  $\kappa = \frac{n-r(X)}{(m-1)(N-r(X_*)})$  and  $W \sim F_{N-r(X_*), n-r(X)}$ . That is, the two tests are equivalent for all  $\alpha$ 's satisfying the inequality  $\alpha \leq P\left(W > \frac{n-r(X)}{(m-1)(N-r(X_*)})\right)$ . Table 1 presents the maximal values of  $\alpha$  satisfying this inequality for different combinations of the degrees of freedom  $df_1 = N - r(X_*)$  and  $df_2 = n - r(X)$  when  $m = 2$ . Since the increase in  $m$ , for being in the denominator of  $\frac{n-r(X)}{(m-1)(N-r(X_*)})$ , increases the maximal values of  $\alpha$  satisfying the inequality  $\alpha \leq P\left(W > \frac{n-r(X)}{(m-1)(N-r(X_*)})\right)$ , it's sufficient to provide another Table (see Table 2) for the case when  $m = 4$  to explain the pattern in which those maximal values of  $\alpha$  behave as a function of  $m$ . The highlighted cells of Table 1 in italic represent the combination of degrees of freedom for which the F-test has a larger power than the LRT at the 5% significance level when  $m = 2$ . The very same thing is true for Table 2 when  $m = 4$ . We observed from simulation, and below give a mathematical proof, that as  $m$  increases the power of the LRT approaches that of the F-test. Typically, the degrees of freedom for subplot error  $df_2$  are much larger than the degrees of freedom for whole plots error  $df_1$ ; for GSP models. So, the  $\alpha > 1 - p_m$  case is very practical.

**Proposition 6** For GSP models, if  $\alpha > 1 - p_m$  then, for a size  $\alpha$  test,  $\Xi_{LRT} \uparrow \Xi_F$  in testing  $\sigma_w^2 = 0$  as the whole plots size  $m$  approaches infinity.

**Table 1** The maximal values of  $\alpha$  satisfying the inequality  $\alpha \leq P\left(W > \frac{n-r(X)}{(m-1)(N-r(X_*)})\right)$  for different combinations of the degrees of freedom  $df_1 = N - r(X_*)$  and  $df_2 = n - r(X)$  when  $m = 2$

$df_2 \setminus df_1$	1	2	3	4	5	10	15	20	30	40	60	120
1	0.50	0.71	0.82	0.88	0.92	0.99	1.00	1.00	1.00	1.00	1.00	1.00
2	0.29	0.50	0.65	0.75	0.82	0.97	0.99	1.00	1.00	1.00	1.00	1.00
3	0.18	0.35	0.50	0.62	0.71	0.94	0.99	1.00	1.00	1.00	1.00	1.00
4	0.12	0.25	0.38	0.50	0.60	0.89	0.97	0.99	1.00	1.00	1.00	1.00
5	0.08	0.18	0.29	0.40	0.50	0.84	0.95	0.99	1.00	1.00	1.00	1.00
6	0.05	0.12	0.22	0.31	0.41	0.77	0.93	0.98	1.00	1.00	1.00	1.00
7	<i>0.03</i>	0.09	0.16	0.24	0.33	0.71	0.90	0.97	1.00	1.00	1.00	1.00
8	<i>0.02</i>	0.06	0.12	0.19	0.26	0.64	0.86	0.95	1.00	1.00	1.00	1.00
9	<i>0.01</i>	<i>0.04</i>	0.09	0.14	0.21	0.57	0.82	0.93	0.99	1.00	1.00	1.00
10	<i>0.01</i>	<i>0.03</i>	0.06	0.11	0.16	0.50	0.77	0.91	0.99	1.00	1.00	1.00
15	<i>0.00</i>	<i>0.01</i>	<i>0.01</i>	<i>0.03</i>	0.05	0.23	0.50	0.73	0.95	0.99	1.00	1.00
20	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.01</i>	<i>0.01</i>	0.09	0.27	0.50	0.85	0.97	1.00	1.00
30	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.01</i>	0.05	0.15	0.50	0.80	0.99	1.00
40	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.01</i>	<i>0.03</i>	0.20	0.50	0.92	1.00
60	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.01</i>	0.08	0.50	1.00
120	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	0.50

Italic values represent the combination of degrees of freedom for which the F-test has a larger power than the LRT at the 5% significance

**Table 2** The maximal values of  $\alpha$  satisfying the inequality  $\alpha \leq P\left(W > \frac{n-r(X)}{(m-1)(N-r(X_*))}\right)$  for different combinations of the degrees of freedom  $df_1 = N - r(X_*)$  and  $df_2 = n - r(X)$  when  $m = 4$

$df_2 \backslash df_1$	1	2	3	4	5	10	15	20	30	40	60	120
1	0.67	0.87	0.94	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.50	0.75	0.88	0.94	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	0.39	0.65	0.80	0.89	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	0.31	0.56	0.73	0.84	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	0.25	0.49	0.67	0.79	0.87	0.99	1.00	1.00	1.00	1.00	1.00	1.00
6	0.21	0.42	0.60	0.74	0.83	0.99	1.00	1.00	1.00	1.00	1.00	1.00
7	0.17	0.37	0.54	0.69	0.79	0.98	1.00	1.00	1.00	1.00	1.00	1.00
8	0.14	0.32	0.49	0.63	0.75	0.97	1.00	1.00	1.00	1.00	1.00	1.00
9	0.12	0.27	0.44	0.58	0.70	0.96	1.00	1.00	1.00	1.00	1.00	1.00
10	0.10	0.24	0.39	0.53	0.66	0.95	1.00	1.00	1.00	1.00	1.00	1.00
15	<i>0.04</i>	0.12	0.22	0.33	0.45	0.86	0.98	1.00	1.00	1.00	1.00	1.00
20	<i>0.02</i>	0.06	0.12	0.20	0.29	0.74	0.94	0.99	1.00	1.00	1.00	1.00
30	<i>0.00</i>	<i>0.01</i>	<i>0.03</i>	0.06	0.11	0.47	0.80	0.95	1.00	1.00	1.00	1.00
40	<i>0.00</i>	<i>0.00</i>	<i>0.01</i>	<i>0.02</i>	<i>0.04</i>	0.25	0.58	0.83	0.99	1.00	1.00	1.00
60	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	0.05	0.21	0.48	0.89	0.99	1.00	1.00
120	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.01</i>	0.14	0.48	0.96	1.00

Italic values represent the combination of degrees of freedom for which the F-test has a larger power than the LRT at the 5% significance

**Proof of Proposition 6** Recall that

$$\mathcal{E}_F = P\left(W \geq \frac{1}{k}W_{1-p_m}\right) + P\left(W \leq \frac{1}{k}W_{1-p_m}\right) - P\left(W \leq \frac{1}{k}W_\alpha\right),$$

and

$$\mathcal{E}_{LRT} = P\left(W \geq \frac{1}{k}W_{1-p_m}\right) + \frac{\alpha - (1 - p_m)}{p_m} P\left(W \leq \frac{1}{k}W_{1-p_m}\right).$$

From Proposition 5, we have established for a finite whole plot size  $m$

$$P\left(W \leq \frac{1}{k}W_{1-p_m}\right) - P\left(W \leq \frac{1}{k}W_\alpha\right) > \frac{\alpha - (1 - p_m)}{p_m} P\left(W \leq \frac{1}{k}W_{1-p_m}\right).$$

If we let  $m \uparrow \infty$  then  $k \uparrow \infty$  so that

$$P\left(W \leq \frac{1}{k}W_{1-p_m}\right) = P\left(W \leq \frac{1}{k}W_\alpha\right) = P\left(W \leq \frac{1}{k}W_{1-p_m}\right) = 0,$$

and thus the inequality becomes equality and as a result  $\mathcal{E}_{LRT} \uparrow \mathcal{E}_F$ . In fact for  $m = \infty$  we have  $\mathcal{E}_{LRT} = \mathcal{E}_F = 1$  since  $\lim_{k \rightarrow +\infty} P\left(W \geq \frac{1}{k}W_{1-p_m}\right) = 1$ .  $\square$

## 6 Conclusion

For a class of linear mixed models with CS covariance structure, we show the equivalence between the F-test and LRT when the level of the test,  $\alpha$ , is less or equal to one minus the probability,  $p$ , that the LRT statistic is zero. However, when  $\alpha > 1 - p$ , we show that the F-test has a larger power than the LRT and thus proving that the LRT test is inadmissible. Further, we derive the statistical distribution of the LRT under both the null and alternative hypotheses  $H_0$  and  $H_1$  where  $H_0$  is the hypothesis that the between component variance is zero. This work proves important property on the relationship between the F-test and LRT for a large class of Mixed effect models. Further, providing an exact distribution of the test statistic for the LRT in these models will help in calculating a more accurate p-value than the traditionally used p-value from the large sample chi-square mixture approximations introduced by Self and Liang (1987).

**Acknowledgements** We acknowledge the reviewers greatly for their valuable time in reviewing the paper.

## Appendix 1

Supplementary material includes (i) Proof for the PPOs Properties in (15), (ii) Illustration for the proof of Lemma 2, and (iii) Proof of Lemma 3.

### Proof for the PPOs Properties in (15)

Firstly,  $\tilde{M} = M_* + M_2$ : This result is an immediate consequence of conditions (b) and (c) of Sect. 1.3. In particular, since  $C(\tilde{X}) = C(X_*, (I - M_1)X_2)$  then by definition of PPO

$$\begin{aligned} \tilde{M} &= [X_*, (I - M_1)X_2] \left( \begin{bmatrix} X_*' \\ [(I - M_1)X_2]' \end{bmatrix} [X_*, (I - M_1)X_2] \right)^{-1} \\ &\quad \times \begin{bmatrix} X_*' \\ [(I - M_1)X_2]' \end{bmatrix} \\ &= X_* \left( X_*' X_* \right)^{-1} X_*' + (I - M_1)X_2 \left( X_2'(I - M_1)X_2 \right)^{-1} X_2'(I - M_1) \\ &= M_* + M_2. \end{aligned} \tag{83}$$

The equality in the third line of (83) is due to condition (b) which implies that  $(I - M_1)X_* = X_*'(I - M_1) = 0$ .

Secondly,  $M_*M_1 = M_*$ : This result is an immediate consequence of condition (b). In particular, since  $C(X_*) \subset C(X_1)$  then  $X_* = X_1B$  for some matrix  $B$  and therefore, by definition of PPO,

$$M_* = X_* \left( X_*' X_* \right)^{-1} X_*' = X_1B \left( B' X_1' X_1 B \right)^{-1} B' X_1'. \tag{84}$$

Thus, using  $M_*$  from (84) and  $M_1$  from (13) gives

$$\begin{aligned} M_*M_1 &= X_1B \left( B'X_1'X_1B \right)^{-1} B'X_1'X_1 \left( X_1'X_1 \right)^{-1} X_1' \\ &= X_1B \left( B'X_1'X_1B \right)^{-1} B'X_1' = M_*. \end{aligned} \tag{85}$$

Thirdly,  $M_1M_2 = 0$ : This results is trivially obtained by simply multiplying  $M_1$  from (13) and  $M_2$  from (14).

Fourthly,  $M = M_1 + M_2$ : Let  $X = [X_1, X_2]$  such that  $M$  is the PPO onto  $C(X)$ . Since  $M_1M_2 = 0$ , then  $C(M_1) \perp C(M_2)$  and hence  $M = M_1 + M_2$  is a PPO onto  $C(M_1, M_2)$  by Theorem B.45 of Christensen (2011). But  $C(M_1, M_2) = C(X_1, (I - M_1)X_2)$  since  $C(M_1) = C(X_1)$  and  $C(M_2) = C((I - M_1)X_2)$ . So, it remains to prove that  $C(X_1, X_2) = C(X_1, (I - M_1)X_2)$  to complete the proof. To do so, we use the fact that  $C(A_1) = C(A_2)$  iff there exist  $B_1$  and  $B_2$  such that  $A_1 = A_2B_2$  and  $A_2 = A_1B_1$  as follows.

$$\begin{aligned} [X_1, (I - M_1)X_2] &= [X_1, X_2 - M_1X_2] \\ &= \left[ X_1, X_2 - X_1(X_1'X_1)^{-1}X_1'X_2 \right] \\ &= [X_1, X_2] \begin{bmatrix} I - (X_1'X_1)^{-1}X_1'X_2 \\ 0 \quad I \end{bmatrix} \end{aligned} \tag{86}$$

and

$$[X_1, X_2] = [X_1, (I - M_1)X_2] \begin{bmatrix} I (X_1'X_1)^{-1}X_1'X_2 \\ 0 \quad I \end{bmatrix}. \tag{87}$$

That is,  $C(X_1, (I - M_1)X_2) \subset C(X_1, X_2)$  and  $C(X_1, X_2) \subset C(X_1, (I - M_1)X_2)$  so that  $C(X_1, X_2) = C(X_1, (I - M_1)X_2)$  as desired.  $\square$

## Appendix 2

### Illustration for the proof of Lemma 2

Let  $\lambda = -\frac{a}{b}$ . Then

$$|aI_n + bP| = |b(P - \lambda I_n)| = b^n |P - \lambda I_n|. \tag{88}$$

However, the determinant  $|P - \lambda I_n|$  in (88) is the characteristic polynomial of  $P$  which equals to (89) since 1 and 0 are the eigenvalues for  $P$  with multiplicity  $r(P)$  and  $n - r(P)$  respectively.

$$|P - \lambda I_n| \equiv p_P(\lambda) = (-\lambda)^{n-r(P)}(1 - \lambda)^{r(P)}. \tag{89}$$

Hence, substituting (89) in (88) gives the desired result

$$\begin{aligned}
 |aI_n + bP| &= b^n |P - \lambda I_n| = b^n (-\lambda)^{n-r(P)} (1 - \lambda)^{r(P)} \\
 &= \frac{b^n a^{n-r(P)} (a + b)^{r(P)}}{b^{n-r(P)}} = a^{n-r(P)} (a + b)^{r(P)}. \tag{90}
 \end{aligned}$$

□

### Appendix 3

#### Proof of Lemma 3

When  $x_2 > x_1 > 0$ , we have a standard maximization problem for a function of two variables. Setting the partial derivatives to zero gives

$$\frac{\partial g}{\partial x_1} = 0 \iff \frac{q_1(Q_1 - x_1)}{x_1^2} = 0 \iff x_1 = Q_1, \tag{91}$$

and

$$\frac{\partial g}{\partial x_2} = 0 \iff \frac{q_2(Q_2 - x_2)}{x_2^2} = 0 \iff x_2 = Q_2. \tag{92}$$

Let  $g_{x_i x_j} = \frac{\partial}{\partial x_j} \left( \frac{\partial}{\partial x_i} g(x_i, x_j) \right)$  for  $i, j \in \{1, 2\}$ . Then, according to the second derivative test, we have

$$\begin{aligned}
 D(x_1, x_2) &= g_{x_1 x_1}(x_1, x_2)g_{x_2 x_2}(x_1, x_2) - [g_{x_1 x_2}(x_1, x_2)]^2 \\
 &= \left( \frac{q_1}{x_1^2} - \frac{2q_1 Q_1}{x_1^3} \right) \left( \frac{q_2}{x_2^2} - \frac{2q_2 Q_2}{x_2^3} \right) \tag{93}
 \end{aligned}$$

with  $D(Q_1, Q_2) = \frac{q_1 q_2}{Q_1^2 Q_2^2} > 0$  and  $g_{x_1 x_1}(Q_1, Q_2) = \frac{-q_1}{Q_1^2} < 0$  so that  $(x_1, x_2) = (Q_1, Q_2)$  is a maximum point. Thus, if  $Q_2 > Q_1 > 0$  then the point  $(Q_1, Q_2)$  is in the interior and maximizes the function within the interior; i.e. a local maximum.

When  $x_1 = x_2 := x$ , using direct substitution, the problem reduces to maximizing the function of one variable

$$g(x) = - \left[ \text{constant} + (q_1 + q_2) \log(x) + \frac{q_1 Q_1 + q_2 Q_2}{x} \right] \tag{94}$$

over  $R^+$ . So, setting the partial derivative of  $g(x)$  to zero gives

$$\frac{\partial g}{\partial x} = 0 \iff -\frac{q_1 + q_2}{x} + \frac{q_1 Q_1 + q_2 Q_2}{x^2} = 0 \iff x = \frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}. \tag{95}$$



Now, using the second derivative test, we have

$$\frac{\partial g(x)}{\partial x^2} = \frac{q_1 + q_2}{x^2} - \frac{2(q_1 Q_1 + q_2 Q_2)}{x^3} \tag{96}$$

with

$$\frac{\partial g(x)}{\partial x^2} \Big|_{x=\frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}} = \frac{-(q_1 + q_2)^3}{(q_1 Q_1 + q_2 Q_2)^2} < 0 \tag{97}$$

so that  $(x_1, x_2) = \left(\frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}, \frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}\right)$  is a maximum point on the boundary of the domain.

Now, we show that if  $Q_2 > Q_1 > 0$  then the maximum in the interior is a global maximum. Note that when the maximum is in the interior at  $(Q_1, Q_2)$  it attains the value

$$g(Q_1, Q_2) = - [constant + (q_1 + q_2) + \log(Q_1^{q_1} Q_2^{q_2})].$$

Further, when the maximum is on the boundary at  $\left(\frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}, \frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}\right)$  it attains the value

$$g\left(\frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}\right) = - \left[ constant + (q_1 + q_2) + (q_1 + q_2) \log\left(\frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}\right) \right].$$

Showing that  $g(Q_1, Q_2) > g\left(\frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}\right)$  is the same as showing

$$\log\left(\frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}\right) > \frac{q_1 \log(Q_1)}{q_1 + q_2} + \frac{q_2 \log(Q_2)}{q_1 + q_2},$$

which is true due to Jensen's Inequality:

Let  $Q$  be a r.v. such that  $P(Q = Q_1) = \frac{q_1}{q_1 + q_2}$  and  $P(Q = Q_2) = \frac{q_2}{q_1 + q_2}$  then by Jensen's Inequality we have

$$\begin{aligned} \log[E(Q)] &> E[\log(Q)] \\ \iff \log\left(\frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}\right) &> \frac{q_1 \log(Q_1)}{q_1 + q_2} + \frac{q_2 \log(Q_2)}{q_1 + q_2}. \end{aligned}$$

Now, we show that if  $Q_1 > Q_2 > 0$  then the maximum in the boundary is a global maximum. Note that if  $Q_1 > Q_2 > 0$ , there are no critical points of the function within the interior. Further, we know that  $g(x_1, x_2)$  goes to  $-\infty$  in both  $x_1$  and  $x_2$  which forces the maximum on the boundary at  $\left(\frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}, \frac{q_1 Q_1 + q_2 Q_2}{q_1 + q_2}\right)$  to be a global maximum. □

## References

- Castellacci G (2012) A formula for the quantiles of mixtures of distributions with disjoint supports. <http://ssrn.com/abstract=2055022>. Accessed 15 April 2013
- Christensen R (1984) A note on ordinary least squares methods for two-stage sampling. *J Am Stat Assoc* 79:720–72
- Christensen R (1987) The analysis of two-stage sampling data by ordinary least squares. *J Am Stat Assoc* 82:492–498
- Christensen R (2011) *Plane answers to complex questions: the theory of linear models*, 4th edn. Springer, New York
- Christensen R (2019) *Advanced linear modeling: statistical learning and dependent data*, 3rd edn. Springer, New York
- Cloud MJ, Drachman BC (1998) *Inequal Appl Eng*, 1st edn. Springer, New York
- Crainiceanu CM, Ruppert D (2004) Likelihood ratio tests in linear mixed models with one variance component. *J R Stat Soc Ser B* 66:165–85
- Cuyt A, Petersen VB, Verdonk B, Waadeland H, Jones WB (2008) *Handbook of continued fractions for special functions*, pp 319–341
- Dutka J (1981) The incomplete beta function—a historical profile. *Arch History Exact Sci* 24:11–29
- Graybill FA (1961) *An introduction to linear statistical models*, vol I. McGraw Hill, New York
- Greven S, Crainiceanu CM, Küchenhoff H, Peters A (2008) Restricted likelihood ratio testing for zero variance components in linear mixed models. *J Comput Gr Stat* 17(4):870–891
- Harville DA (1997) *Matrix algebra from a statistician's perspective*, 1st edn. Springer, New York
- Herbach LH (1959) Properties of model II-type analysis of variance tests, a: optimum nature of the  $F$ -test for model II in the balanced case. *Ann Math Stat* 30(4):939–959
- Johnson RA, Wichern DW (2002) *Applied multivariate statistical analysis*. Prentice hall, Upper Saddle River, NJ
- Miller SS, Mocanu PT (1990) Univalence of Gaussian and confluent hypergeometric functions. *Proc Am Math Soc* 110(2):333–342
- Lavine M, Bray A, Hodges J (2015) Approximately exact calculations for linear mixed models. *Electron J Stat* 9(2):2293–2323
- Molenberghs G, Verbeke G (2007) Likelihood ratio, score, and wald tests in a constrained parameter space. *Am Stat* 61(1):22–27
- O'Connor AN (2011) *Probability distributions used in reliability engineering*. RIAC, Maryland
- Qeadan F, Christensen R (2014) New stochastic inequalities involving the  $F$  and Gamma Distribut. *J Inequal Spec Funct* 5(4):22–33
- Scheipl F, Greven S, Küchenhoff H (2008) Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Comput Stat Data Anal* 52(7):3283–3299
- Self SG, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 82(398):605–610
- Wiencierz A, Greven S, Küchenhoff H (2011) Restricted likelihood ratio testing in linear mixed models with general error covariance structure. *Electron J Stat* 5:1718–1734
- Yan L, Zhang G (2010) The equivalence between likelihood ratio test and  $f$ -test for testing variance component in a balanced one-way random effects model. *J Stat Comput Simul* 80:443–450

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.