1-28-2015

On the Equivalence Between the LRT, RLRT and F-test for Testing Variance Components in the Generalized Split-plot Models

Fares Qeadan

Fares Qeadan
Candidate

Mathematics and Statistics
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Ronald Christensen, Chairperson

Yan Lu

Michael Sonksen

Huining Kang

On the Equivalence Between the LRT, RLRT and F-test for Testing Variance Components in the Generalized Split-plot Models

BY

FARES QEADAN

B.S., Mathematics and Computer Science, University of Nevada Reno, 2006

M.S., Mathematics, University of Nevada Reno, 2008

M.S., Statistics, Michigan State University, 2011

DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Statistics

The University of New Mexico

Albuquerque, New Mexico

December 2014

DEDICATION

This dissertation is dedicated to the memory of my beloved father, my first and most influential teacher, who taught me to never stop learning. I know he would have been proud of my achievement.

ACKNOWLEDGMENTS

Most of all, I would like to thank Professor Ronald Christensen, my dissertation advisor and chair, for his guidance and support over the past three years. Professor Christensen is a wonderful mentor and extremely knowledgeable in just about everything. I have learned a lot from him about statistics and life. His two Linear Models classes are unforgettable and made me a better statistician. It was a privilege to work with him and learn from him.

I would also like to extend my thanks to the members of my dissertation committee: Dr. Yan Lu, Dr. Michael Sonksen, and Dr. Huining Kang for all of their time, help, and guidance. Dr. Yan Lu was the one who inspired my dissertation topic, from her excellent Mixed Models class, and she is honoring me by being a member of my committee. Dr. Michael Sonksen taught me Bayesian statistics and he is an excellent Bayesian professor and friend to students. I am grateful to my external committee member, Dr. Huining Kang, from the Division of Epidemiology, Biostatistics, and Preventive Medicine for his interest in my dissertation.

I would also like to thank the chair, faculty and staff of the Mathematics and Statistics Department as well as my fellow graduate students. In particular, I would like to thank Dr. Guoyi Zhang for his tips for my talk rehearsal. My thanks go also to the Sandia National Labs for their generous support during my stay at UNM and in that regard, I would also like to thank Dr. Erick Erhardt who opened for me many doors of new opportunities when I started my graduate program at UNM.

I cannot end without thanking my family and wonderful wife Huan, on whose constant encouragement and love I have relied throughout my entire academic endeavour.

On the Equivalence Between the LRT, RLRT and F-test for Testing Variance Components in the Generalized Split-plot Models

by

Fares Qeadan

B.S., Mathematics and Computer Science, University of Nevada Reno, 2006

M.S., Mathematics, University of Nevada Reno, 2008

M.S., Statistics, Michigan State University, 2011

Ph.D, Statistics, University of New Mexico, 2014

Abstract

I study the equivalence between the Likelihood Ratio Test (LRT), Restricted Likelihood Ratio Test (RLRT) and the F-test when testing variance components within the class of generalized split-plot (GSP) models. In this work, I derive explicit expressions for both the maximum likelihood estimates (MLEs) and restricted maximum likelihood estimates (RMLEs) for the variance components of the GSP model and show the equivalence between the F-test, the LRT or the F-test and the RLRT when the level of the test, α, is less or equal to one minus the probability, p, that the LRT or the RLRT statistic is zero. However, when $\alpha > 1 - p$, I show that the F-test has a larger power than either the LRT or RLRT. Further, we derive the statistical distribution of these tests under both the null and alternative hypotheses H_0 and H_1 where H_0 is the hypothesis that the whole plot variance is zero.

To establish the power inequality for the case $\alpha > 1 - p$, I developed a new stochastic inequality involving a class of distributions that includes, for example, the F and Gamma distributions. I call random variables (r.v.s.) that inherit this inequality to be quantile-stochastic. The stochastic representation of the new inequality involves $\alpha, p \in (0, 1)$ such that if $p > \alpha$ and $k > 1$ with W being a random variable with an $F(\nu_1, \nu_2)$ or $Gamma(\tau, \theta)$ distribution then it's always true that

$$\frac{1}{p} P\left(W < \frac{W_p}{k}\right) > \frac{1}{\alpha} P\left(W < \frac{W_\alpha}{k}\right),$$

where $\gamma = P(W < W_\gamma)$. The inequality changes direction for $k \in [0, 1)$ and becomes equality for $k = 1$ and, trivially, for $k = \infty$.

KEY WORDS: MLEs, REMLs, LRT, RLRT, F-Test, generalized split-plot, mixed model, variance components, stochastic inequality, quantile-stochastic.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Recently, Lu and Zhang (2010) 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 both to a much broader family of models with one random effect, the generalized split-plot (GSP) models introduced in Christensen (1987) and to the generalized residual likelihood test. Generalized split-plot 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. The balanced one-way random effects model is the simplest generalized split plot model.

The GSP models are a special case of the general linear mixed models since they specify both fixed effects and random effects. A general linear mixed model can be written

$$Y = \tilde{X}\beta + X_1\gamma + \epsilon, \tag{1.1}$$

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

1.2 Notation

For the rest of this dissertation, 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 $\mathrm{diag}(V_1, \ldots, V_N)$ to denote a block diagonal matrix with square matrices $V_1, \ldots, V_N$ on its diagonal and Blk $\mathrm{diag}(V)$ 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. $L(.)$, $\ell(.)$, and $\ell_*(.)$ denote the likelihood, -2Log-likelihood and 2Log-likelihood functions respectively.

1.3 Generalized Split Plot (GSP) Models

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

$$Y = X\beta + \xi, \tag{1.2}$$

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 α is a vector of fixed cluster effects and γ is a vector of fixed non-cluster effects. Because it is a two-stage cluster sampling model, the error vector ξ has uncorrelated clusters and intraclass correlation structure and can be written with random effects as

$$\xi = X_1\eta + \epsilon, \tag{1.3}$$

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

$$Y = X_1\alpha + X_2\gamma + (X_1\eta + \epsilon), \tag{1.4}$$

with

$$V \equiv \text{Cov}(Y) = \sigma_w^2 X_1 X_1' + \sigma_s^2 I_n. \tag{1.5}$$

The GSP models are obtained *by imposing additional structure on the fixed cluster effects within the cluster sampling model* (1.4). To model the whole plot (cluster) effects we put a constraint on $C(X_1)$ by considering a reduced model

$$Y \quad = \quad X_*\delta + X_2\gamma + (X_1\eta + \epsilon), \qquad C(X_*) \subset C(X_1)$$

$$\equiv \quad \tilde{X}\beta_* + \xi, \tag{1.6}$$

where $\tilde{X} = [X_*, X_2]$ and $\beta_*' = [\delta', \gamma']$ and the covariance matrix remains as in (1.5). The δ_is will be the whole plots fixed effects and the γ_is will be the subplots fixed effects. In this model, η is the whole plot error and ϵ 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. δ 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 (1.6) are best linear unbiased estimates (BLUEs) and standard split-plot F and t statistics have null hypothesis F and t distributions, cf., Christensen (1987, 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 (1.6), with covariance structure (1.5) and assumption (a) we can rewrite (1.5) as

$$\begin{aligned} V &= \sigma_w^2 \text{Blk diag}(J_m J_m') + \sigma_s^2 I_n = \sigma_s^2 I_n + \sigma_w^2 m M_1 \\ &= \text{Blk diag}(\tilde{V}) = I_N \otimes \tilde{V}, \end{aligned} \tag{1.7}$$

where

$$\tilde{V} = \sigma_w^2 J_m J_m' + \sigma_s^2 I_m = \begin{bmatrix} \sigma_w^2 + \sigma_s^2 & \sigma_w^2 & \dots & \sigma_w^2 \\ \sigma_w^2 & \sigma_w^2 + \sigma_s^2 & \dots & \sigma_w^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_w^2 & \sigma_w^2 & \dots & \sigma_w^2 + \sigma_s^2 \end{bmatrix}. \tag{1.8}$$

In a general mixed model (1.1) exact statistical inference cannot typically be performed. 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 (1.6) reduces to the standard linear model

$$\begin{aligned} Y &= X_*\delta + X_2\gamma + \epsilon, \qquad \mathbb{E}(\epsilon) = 0, \mathrm{Cov}(\epsilon) = \sigma_e^2 I_n \\ &= \tilde{X}\beta_* + \epsilon. \end{aligned} \tag{1.9}$$

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 $k = 1, ..., mN_i$ instead of the usual $k = 1, ..., N_{ij}$. To test the appropriateness of the standard linear model (1.9), the hypothesis of interest is whether or not the covariance component for whole plots lies on the boundary of the parameter space,

$$H_0 : \sigma_w^2 = 0 \text{ vs. } H_1 : \sigma_w^2 > 0. \tag{1.10}$$

There are three competing test statistics for the null hypothesis in (1.10), 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), Molenberghsa and Verbeke (2007) and Scheipl et al. (2008)), however studies about their equivalnace in the context of split-plot designs are limited.

1.4 Objectives

In this work, I derive explicit expressions for both the maximum likelihood estimates (MLEs) and restricted maximum likelihood estimates (RMLEs) for the variance components of the GSP model. Further, I show that the LRT, RLRT and F tests are equivalent for testing (1.10) in GSP models when the size of the test, α, is reasonably small and I derive the exact distribution for the LRT and RLRT test statistics under both H_0 and H_1. In particular, I show that the three tests are equivalent when $\alpha \leq 1 - p$ where p is the probability that the LRT/RLRT statistic is zero and give a proof that the F test has a larger power when $\alpha > 1 - p$. As a byproduct of this research work, I develope a new stochastic inequality involving a class of distributions that includes the F and Gamma distributions and is called either the F-inequality or the G-inequality. I call random variables (r.v.s.) that inherit this property to be quantile-stochastic. This new theory of quantile-stochastic distributions will make it possible for the first time to compare distributions in terms of the ratio of their cumulative distribution function (CDF) and quantiles and it opens new windows to distribution theory. In fact, the work on these inequalities will be presented in separate series of publication due to its magnitude and special interest. These inequalities will be of interest to mathematicians and statisticians whose work involves stochastic inequalities, stochastic lower and upper bounds, power comparison, and distribution properties and classification.

1.5 Additional Notation, Means and Variance Components

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

$$M_1 = X_1(X_1^{'}X_1)^{-1}X_1^{'} = \text{Blk diag}\left(\frac{1}{m}J_m J_m^{\prime}\right) = \frac{1}{m}X_1 X_1^{\prime}, \tag{1.11}$$

and define

$$M_2 = (I - M_1)X_2\left(X_2^{\prime}(I - M_1)X_2\right)^{-} X_2^{\prime}(I - M_1) \tag{1.12}$$

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

$$\tilde{M} = M_* + M_2, \quad M_* M_1 = M_*, \quad M_1 M_2 = 0 \quad \text{and} \quad M = M_1 + M_2. \tag{1.13}$$

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

$$SSE(w) \equiv Y^{'}(M_1 - M_*)Y \quad \text{and} \quad SSE(s) \equiv Y^{'}(I - M)Y, \tag{1.14}$$

such that

$$SSE(s) = Y^{'}(I - M)Y = Y^{'}(I - \tilde{M})Y - Y^{'}(M_1 - M_*)Y. \tag{1.15}$$

The F statistic for testing (1.10) is defined as

$$F = \frac{MSE(w)}{MSE(s)} = \frac{SSE(w)/[r(X_1) - r(X_*)]}{SSE(s)/[n - r(X)]} = \frac{SSE(w)/[N - r(X_*)]}{SSE(s)/[n - r(X)]}. \tag{1.16}$$

1.6 Outline of the dissertation

In Chapter 2 we present the maximum likelihood estimators of the variance parameters both on and off the parameter boundary for σ_w^2 and σ_s^2. These are necessary for finding the LRT. It also contains the main result about the equivalence of the LRT and F-test and establishes the exact distribution of the LRT statistic, say Λ. Chapter 3 largely reiterates chapter 2 but for the REMLs and RLRT. Chapters, 2 and 3, stablishe respectively that the F-test has a larger power, when $\alpha > 1 - p$, than the LRT and RLRT. Monte Carlo simulations are presented in Chapter 4. In Chapter 5 I prove the F-inequality and the G-inequality respectively. All secondary proofs are deferred to an appendix.

Chapter 2

The Equivalence Between the LRT and F-test

2.1 Maximum Likelihood Estimators (MLEs)

The likelihood function for model (1.6) is:

$$L(\beta_*, \sigma_w^2, \sigma_s^2 | Y) = \frac{1}{(2\pi)^{n/2} |V|^{1/2}} e^{-1/2(Y-\tilde{X}\beta_*)'V^{-1}(Y-\tilde{X}\beta_*)}, \tag{2.1}$$

where

$$V = V(\sigma_w^2, \sigma_s^2) = m\sigma_w^2 M_1 + \sigma_s^2 I_n. \tag{2.2}$$

We use ℓ to denote minus 2 times the natural logarithm of the likelihood of Y (i.e. $\ell = -2 \log L$). We provide three lemmas

Lemma 2.1.1 *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*

$$(aI_n + bP)^{-1} = \frac{1}{a}\left(I_n - \frac{b}{a+b}P\right). \tag{2.3}$$

Proof of Lemma 2.1.1: See Proposition 12.11.1 in Christensen (2011, p.322). □

Lemma 2.1.2 *The determinant of $aI_n + bP$, where P is a PPO and a and b are real numbers, is*

$$|aI_n + bP| = a^{n-r(P)}(a+b)^{r(P)}. \tag{2.4}$$

Proof of Lemma 2.1.2: 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 A.2 contains an illustration.

Lemma 2.1.3 *For $q_1, q_2 > 0$, maximizing the function*

$$g(x_1, x_2) = -\left[constant + q_1 \log(x_1) + q_2 \log(x_2) + q_1 \left(\frac{Q_1}{x_1}\right) + q_2 \left(\frac{Q_2}{x_2}\right)\right] \tag{2.5}$$

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 $(x_1, x_2) = \left(\frac{q_1Q_1+q_2Q_2}{q_1+q_2}, \frac{q_1Q_1+q_2Q_2}{q_1+q_2}\right)$ when $x_2 = x_1$ (i.e. when (x_1, x_2) is on the boundary of the constraint).

The proof is in Appendix A.3.

Applying Lemmas 2.1.1 and 2.1.2 to V in (2.2) gives the following determinant and inverse covariance matrix:

$$|V| = \left|\sigma_s^2 + m\sigma_w^2\right|^N \left|\sigma_s^2\right|^{n-N}, \tag{2.6}$$

and

$$V^{-1} = \frac{1}{\sigma_s^2} I_n - \left(\frac{\sigma_w^2}{\sigma_s^2}\right) \frac{m}{\sigma_s^2 + m\sigma_w^2} M_1. \tag{2.7}$$

Substituting (2.6) in (2.1) and taking -2 times the natural logarithm leads to

$$\begin{aligned}\ell(\beta_*, \sigma_w^2, \sigma_s^2|Y) &= n\log(2\pi) + N\log\left(\sigma_s^2 + m\sigma_w^2\right) + N(m-1)\log(\sigma_s^2)\\ &\quad + (Y - \tilde{X}\beta_*)'V^{-1}(Y - \tilde{X}\beta_*). \qquad (2.8)\end{aligned}$$

Proposition 2.1.4 *The Maximum Likelihood estimators for β_*, σ_w^2 and σ_s^2 of model (1.6) are*

$$\hat{\beta}_* = \left(\tilde{X}'V^{-1}\tilde{X}\right)^{-}\tilde{X}'V^{-1}Y = \left(\tilde{X}'\tilde{X}\right)^{-}\tilde{X}'Y, \qquad (2.9)$$

$$\hat{\sigma}_w^2 = \frac{1}{m}max\left\{0, \frac{SSE(w)}{N} - \frac{SSE(s)}{n-N}\right\}, \ \ and \qquad (2.10)$$

$$\hat{\sigma}_s^2 = min\left\{\frac{SSE(s)}{n-N}, \frac{SSE(s)+SSE(w)}{n}\right\}, \qquad (2.11)$$

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 2.1.4. Differentiating (2.8) with respect to β_* and setting the partial derivative to zero, leads to

$$\hat{\beta}_* = \left(\tilde{X}'V^{-1}\tilde{X}\right)^{-1}\tilde{X}'V^{-1}Y. \qquad (2.12)$$

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 β_* so they are also the maximum likelihood estimates. Therefore, $\hat{\beta}_* = \left(\tilde{X}'V^{-1}\tilde{X}\right)^{-1}\tilde{X}'V^{-1}Y = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'Y$ and subsequently $\tilde{X}\hat{\beta}_*$ could be computed, using the OLSEs, as

$$\tilde{X}\hat{\beta}_* = \tilde{M}Y. \qquad (2.13)$$

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

$$\begin{aligned}\ell_*(\hat{\beta}_*, \sigma_w^2, \sigma_s^2|Y) \equiv -\ell(\hat{\beta}_*, \sigma_w^2, \sigma_s^2|Y) &= -[n\log(2\pi) + N(m-1)\log(\sigma_s^2) \\ &\quad + N\log\left(\sigma_s^2 + m\sigma_w^2\right) + \Psi(\sigma_w^2, \sigma_s^2)], \end{aligned} \tag{2.14}$$

where

$$\Psi(\sigma_w^2, \sigma_s^2) = (Y - \tilde{X}\hat{\beta}_*)' V^{-1} (Y - \tilde{X}\hat{\beta}_*). \tag{2.15}$$

To maximize this function we need to simplify Ψ. Thus, using (2.7) we can write $V^{-1} = aI + bM_1$ with $a = \frac{1}{\sigma_s^2}$ and $b = -\frac{\sigma_w^2}{\sigma_s^2}\frac{m}{\sigma_s^2 + m\sigma_w^2}$ which along with (1.13) and (1.14) gives

$$\begin{aligned}\Psi &= (Y - \tilde{M}Y)'V^{-1}(Y - \tilde{M}Y) = Y'(I - \tilde{M})(aI + bM_1)(I - \tilde{M})Y \\ &= Y'(I - \tilde{M})\left[a(I - \tilde{M})Y + bM_1(I - \tilde{M})Y\right] \\ &= aY'(I - \tilde{M})Y + bY'(M_1 - M_1\tilde{M})(I - \tilde{M})Y \\ &= aY'(I - \tilde{M})Y + bY'(M_1 - M_1M_* - M_1M_2)(I - \tilde{M})Y \\ &= aY'(I - \tilde{M})Y + bY'(M_1 - M_*)(I - \tilde{M})Y \\ &= aY'(I - \tilde{M})Y + bY'(M_1 - M_*)Y - bY'(M_1 - M_*)\tilde{M}Y \\ &= a\left[Y'(M_1 - M_*)Y + Y'(I - M)Y\right] + bY'(M_1 - M_*)Y \\ &= aY'(I - M)Y + (a + b)Y'(M_1 - M_*)Y \\ &= \frac{SSE(s)}{\sigma_s^2} + \frac{SSE(w)}{\sigma_s^2 + m\sigma_w^2}. \end{aligned} \tag{2.16}$$

Not simplifying and expressing Ψ in terms of the sum of square errors would make it impossible to, mathematically, maximize $\ell_*(\hat{\beta}_*, \sigma_w^2, \sigma_s^2|Y)$ and subsequently having closed form for the MLEs. Hence,

$$\ell_*(\hat{\beta}_*, \sigma_w^2, \sigma_s^2|Y) = -\left[n\log(2\pi) + N(m-1)\log(\sigma_s^2) + N\log\left(\sigma_s^2 + m\sigma_w^2\right)\right.$$

$$+\frac{SSE(s)}{\sigma_s^2}+\frac{SSE(w)}{\sigma_s^2+m\sigma_w^2}\Bigg]. \tag{2.17}$$

Now, 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. Applying Lemma 2.1.3 to ℓ_* in (2.17) gives the maximizers

$$\begin{aligned}
&\left(\hat{\sigma}_s^2, \hat{\sigma}_s^2 + m\hat{\sigma}_w^2\right) = \left(\frac{SSE(s)}{n-N}, \frac{SSE(w)}{N}\right)\\
\Longleftrightarrow\quad &\left(\hat{\sigma}_s^2, \hat{\sigma}_w^2\right) = \left(\frac{SSE(s)}{n-N}, \frac{1}{m}\left[\frac{SSE(w)}{N} - \frac{SSE(s)}{n-N}\right]\right)
\end{aligned} \tag{2.18}$$

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

$$\begin{aligned}
&\left(\hat{\sigma}_s^2, \hat{\sigma}_s^2 + m\hat{\sigma}_w^2\right) = \left(\frac{SSE(s)+SSE(w)}{n}, \frac{SSE(s)+SSE(w)}{n}\right)\\
\Longleftrightarrow\quad &\left(\hat{\sigma}_s^2, \hat{\sigma}_w^2\right) = \left(\frac{SSE(s)+SSE(w)}{n}, 0\right)
\end{aligned} \tag{2.19}$$

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

Now, suppose that the MLE of σ_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

$$\frac{SSE(s)}{n-N} \leq \left(\frac{n-N}{n}\right)\frac{SSE(s)}{n-N} + \left(\frac{N}{n}\right)\frac{SSE(w)}{N} = \frac{SSE(s)+SSE(w)}{n} \tag{2.20}$$

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 σ_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 (2.18) and (2.19) 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 (2.17) are

$$\frac{\partial \ell}{\partial \sigma_w^2} = -\left[\frac{n}{\sigma_s^2 + m\sigma_w^2} - \frac{mSSE(w)}{(\sigma_s^2 + m\sigma_w^2)^2}\right], \tag{2.21}$$

and

$$\frac{\partial \ell}{\partial \sigma_s^2} = -\left[\frac{N}{\sigma_s^2 + m\sigma_w^2} + \frac{N(m-1)}{\sigma_s^2} - \frac{SSE(s)}{(\sigma_s^2)^2} - \frac{SSE(w)}{(\sigma_s^2 + m\sigma_w^2)^2}\right]. \tag{2.22}$$

So, for varification purposes, pluging in the pair $\hat{\sigma}_w^2 = 0$ and $\hat{\sigma}_s^2 = \frac{SSE(s)+SSE(w)}{n}$ into (2.22) 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 (2.21) gives zero as desired.

2.2 Monotonic Relationship Between the LRT and F-test Statistics

We show that the LRT statistic Λ is a monotone function of the F-test statistic F for testing the null hypothesis in (1.10). When Λ is not 0, the monotone relationship is strict, so whenever the size of the test α 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 (1.9) versus the full model in (1.6). In particular, the sum of squares for errors under the reduced model is

$$SSE(e) \equiv SSE(w) + SSE(s). \tag{2.23}$$

Similarly, we use σ_e^2 to denote the variance parameter corresponding to the reduced model while σ_w^2 and σ_s^2 denote the variance parameters of the full model. Since the reduced and full models are nested we present the equivalance between the LRT and F-test under the full model. Since there are two cases for the MLEs of σ_w^2 and σ_s^2, the relationship between the statistics Λ and F will be decomposed into two cases as well.

Proposition 2.2.1 *The LRT statistic Λ, for (1.10), 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*

$$\Lambda = 0, \tag{2.24}$$

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

$$\begin{aligned} \Lambda &= n\log\left(\frac{m-1}{m}\right) + N\log\left(\frac{N(n-r(X))}{(n-N)(N-r(X_*))}\right) \\ &\quad + n\log\left(1 + \frac{N-r(X_*)}{n-r(X)}F\right) + N\log\left(\frac{1}{F}\right) \end{aligned} \tag{2.25}$$

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

$$\kappa = \frac{n-r(X)}{(m-1)(N-r(X_*))}. \tag{2.26}$$

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 Λ as a function of F is presented in Figure 2.1.

Proof of proposition (2.2.1). To examine the variation between plots, we test the whole plot error as in (1.10). Note that, under H_0 we get the reduced model (1.9) where $\epsilon \sim N(0, \sigma_e^2 I)$ with

$$L(\beta_*, \sigma_e^2|Y) = \frac{1}{(2\pi\sigma_e^2)^{n/2}} e^{-\frac{1}{2\sigma_e^2}(Y-\tilde{X}\beta_*)'(Y-\tilde{X}\beta_*)}, \tag{2.27}$$

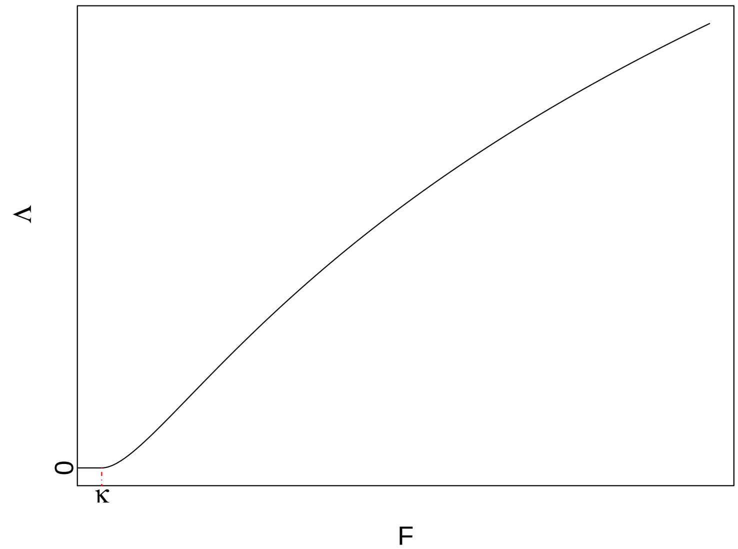

Figure 2.1: A plot of the LRT statistic Λ versus the F statistic according to (2.24) and (2.25). When $F \leq \kappa$, Λ is constant and equal to 0.

and

$$\ell(\hat{\beta}_*, \sigma_e^2|Y) = n\log(2\pi) + n\log(\sigma_e^2) + \frac{SSE(e)}{\sigma_e^2}, \tag{2.28}$$

so that the MLE of σ_e^2 is

$$\hat{\sigma}_e^2 = \frac{SSE(e)}{n}. \tag{2.29}$$

The likelihod ratio test statistic Λ is defined as the difference of -2Log-Likelihood for two nested models (reduced and full). In our case, (1.6) is the full model and (1.9) is the reduced one so that

$$\Lambda \;=\; \ell_R(\hat{\beta}_*, \hat{\sigma}_e^2|Y) - \ell_F(\hat{\beta}_*, \hat{\sigma}_w^2, \hat{\sigma}_s^2|Y), \tag{2.30}$$

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

(2.23), then plugging (2.29) into (2.28) gives

$$\ell_R(\hat{\beta}_*, \hat{\sigma}_e^2|Y) = n\log(2\pi) + n - n\log(n) + n\log(SSE(s) + SSE(w)). \qquad (2.31)$$

Further, if we let $\hat{\sigma}_s^2$ and $\hat{\sigma}_w^2$ be the MLEs for σ_s^2 and σ_w^2 respectively and plug them into (2.17) we get

$$\ell_F(\hat{\beta}_*, \hat{\sigma}_w^2, \hat{\sigma}_s^2|Y) = n\log(2\pi) + N(m-1)\log(\hat{\sigma}_s^2) + N\log(\hat{\sigma}_s^2 + m\hat{\sigma}_w^2) + \frac{SSE(s)}{\hat{\sigma}_s^2} + \frac{SSE(w)}{\hat{\sigma}_s^2 + m\hat{\sigma}_w^2}. \qquad (2.32)$$

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

$$\begin{aligned}
\Lambda &= n - n\log(n) + n\log(SSE(s) + SSE(w)) - Nm\log\left(\hat{\sigma}_s^2\right) \\
&\quad + N\log\left(\hat{\sigma}_s^2\right) - N\log\left(\hat{\sigma}_s^2 + m\cdot 0\right) - \frac{SSE(s)}{\left(\frac{SSE(s)+SSE(w)}{n}\right)} - \frac{SSE(w)}{\left(\frac{SSE(s)+SSE(w)}{n}\right)} \\
&= n - n\log(n) + n\log(SSE(s) + SSE(w)) \\
&\quad - n\log\left(\frac{SSE(s)+SSE(w)}{n}\right) - n \\
&= n - n\log(n) + n\log(n) - n = 0. \qquad (2.33)
\end{aligned}$$

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 (1.16), we get

$$\begin{aligned}
\Lambda &= n - n\log(n) + n\log(SSE(s) + SSE(w)) - Nm\log\left(\hat{\sigma}_s^2\right) \\
&\quad + N\log\left(\hat{\sigma}_s^2\right) - N\log\left(\hat{\sigma}_s^2 + m.\hat{\sigma}_s^2\right) - \frac{SSE(s)}{\hat{\sigma}_s^2} - \frac{SSE(w)}{\hat{\sigma}_s^2 + m\hat{\sigma}_w^2} \\
&= n - n\log(n) + n\log(SSE(s) + SSE(w)) - n\log\left(\frac{SSE(s)}{n-N}\right) \\
&\quad + N\log\left(\frac{SSE(s)}{n-N}\right) - N\log\left(\frac{SSE(s)}{n-N} + m\frac{1}{m}\left[\frac{SSE(w)}{N} - \frac{SSE(s)}{n-N}\right]\right) \\
&\quad - \frac{SSE(s)}{\left(\frac{SSE(s)}{n-N}\right)} - \frac{SSE(w)}{\left(\frac{SSE(s)}{n-N} + m\frac{1}{m}\left[\frac{SSE(w)}{N} - \frac{SSE(s)}{n-N}\right]\right)} \\
&= [n - n\log(n)] + n\log\left(1 + \frac{SSE(w)}{SSE(s)}\right) + n\log(n-N) \\
&\quad + N\log\left(\frac{SSE(s)}{n-N}\right) - N\log\left(\frac{SSE(w)}{N}\right) - (n-N) - N
\end{aligned}$$

$$
\begin{aligned}
&= [n - n\log(n) - n] + n\log\left(1 + \frac{SSE(w)}{SSE(s)}\right) \\
&\quad + N\log\left(\frac{N}{n-N}\frac{SSE(s)}{SSE(w)}\right) + n\log\left(n\frac{n-N}{n}\right) \\
&= [n - n\log(n) + n\log(n) - n] + n\log\left(\frac{n-N}{n}\right) \\
&\quad + n\log\left(1 + \frac{SSE(w)}{SSE(s)}\right) + N\log\left(\frac{N}{n-N}\frac{SSE(s)}{SSE(w)}\right) \\
&= n\log\left(\frac{m-1}{m}\right) + N\log\left(\frac{N(n-r(X))}{(n-N)(N-r(X_*))}\right) \\
&\quad + n\log\left(1 + \frac{N-r(X_*)}{n-r(X)}F\right) + N\log\left(\frac{1}{F}\right). \qquad (2.34)
\end{aligned}
$$

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

$$
\begin{aligned}
& \frac{1}{m}\left[\frac{SSE(w)}{N} - \frac{SSE(s)}{n-N}\right] > 0 \iff \frac{SSE(w)}{SSE(s)} > \frac{1}{m-1} \\
\iff \quad & \frac{N-r(X_*)}{n-r(X)}F > \frac{1}{m-1} \\
\iff \quad & F > \frac{n-r(X)}{(m-1)(N-r(X_*))} \\
\iff \quad & F > \kappa. \qquad (2.35)
\end{aligned}
$$

The function of Λ for Case II given in (2.34) is strictly increasing in F since

$$
\begin{aligned}
& \frac{\partial\Lambda}{\partial F} > 0 \iff \frac{n\frac{N-r(X_*)}{n-r(X)}}{1 + \frac{N-r(X_*)}{n-r(X)}F} - \frac{N}{F} > 0 \\
\iff \quad & \frac{n(N-r(X_*))F}{n-r(X) + (N-r(X_*))F} > N \\
\iff \quad & n(N-r(X_*))F > N\left[n-r(X) + (N-r(X_*))F\right] \\
\iff \quad & (n-N)(N-r(X_*))F > N(n-r(X)) \\
\iff \quad & F > \frac{N(n-r(X))}{(n-N)(N-r(X_*))} \\
\iff \quad & F > \frac{n-r(X)}{(m-1)(N-r(X_*))} \\
\iff \quad & F > \kappa. \qquad (2.36)
\end{aligned}
$$

$\square$

2.3 The Distribution of the F-test and LRT Statistics

The form of relation between the Λ and F helps us better understand the distribution of Λ. In particular, this relation implies an important lemma (see Lemma 2.3.2) on the distribution of Λ for (1.6).

Proposition 2.3.1 *The distribution of the $F-$ratio in (1.16) for the model in (1.6) is a constant multiple of an F distribution,*

$$F = \frac{MSE(w)}{MSE(s)} \sim \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} F(N - r(X_*), n - r(X)). \tag{2.37}$$

Proof of Proposition (2.3.1): It has been shown in Chapter 11.2 of Christensen (2011) that

$$W_1 := \frac{SSE(w)}{\sigma_s^2 + m\sigma_w^2} = \frac{Y'(M_1 - M_*)Y}{\sigma_s^2 + m\sigma_w^2} \sim \chi^2_{N-r(X_*)}, \tag{2.38}$$

$$W_2 := \frac{SSE(s)}{\sigma_s^2} = \frac{Y'(I - M)Y}{\sigma_s^2} \sim \chi^2_{n-r(X)}, \tag{2.39}$$

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

$$\begin{aligned} F &= \frac{SSE(w)/[N - r(X_*)]}{SSE(s)/[n - r(X)]} \\ &= \left(\frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2}\right) \frac{W_1/[N - r(X_*)]}{W_2/[n - r(X)]} \\ &\sim \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} F(N - r(X_*), n - r(X)). \end{aligned} \tag{2.40}$$

□

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 2.3.2 *The distribution of the LRT statistic Λ for the model in (1.6) is determined by the relationship in (2.41) where $W \sim F\left(N - r(X_*), n - r(X)\right)$ and a and τ are as described.*

$$\Lambda \sim \begin{cases} 0 & \text{if} \quad W \leq \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2} \\ \tau + N\log\left(\frac{(1+aW)^m}{W}\right) & \text{if} \quad W > \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2} \end{cases} \tag{2.41}$$

such that

$$p_m = Pr\left(W \leq \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right), \tag{2.42}$$

where

$$a = \frac{N - r(X_*)}{n - r(X)}\frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2}, \tag{2.43}$$

and

$$\tau = N\log\left(\frac{(m-1)^{m-1}}{m^m}\frac{1}{a}\right). \tag{2.44}$$

Proof of Lemma 2.3.2: From case-I of Proposition 2.2.1 we know that $\Lambda = 0$ iff $F \leq \kappa$ so

$$\begin{aligned} p_m &\equiv P(\Lambda = 0) = P\left(F \leq \kappa\right) \\ &= P\left(\frac{MSE(w)}{MSE(s)} \leq \kappa\right) \\ &= P\left(\frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2}W \leq \kappa\right) \\ &= P\left(W \leq \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right). \end{aligned} \tag{2.45}$$

The equality in the third line of (2.45) holds due to Proposition 2.3.1. Now, from case-II of Proposition 2.2.1 we also know that, if $F > \kappa$ (i.e. when $W > \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}$),

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

$$+n\log\left(1+\frac{N-r(X_*)}{n-r(X)}F\right)+N\log\left(\frac{1}{F}\right). \tag{2.46}$$

So, if we let $W \sim F\left(N-r(X_*), n-r(X)\right)$ then substituting F from (2.37) into (2.46) gives

$$\begin{aligned}
\Lambda &= n\log\left(\frac{m-1}{m}\right)+N\log\left(\frac{N(n-r(X))}{(n-N)(N-r(X_*))}\right) \\
&\quad +n\log\left(1+\frac{N-r(X_*)}{n-r(X)}F\right)+N\log\left(\frac{1}{F}\right) \\
&= Nm\log\left(\frac{m-1}{m}\right)+N\log\left(\frac{N(n-r(X))}{(Nm-N)(N-r(X_*))}\right) \\
&\quad +Nm\log\left(1+\frac{N-r(X_*)}{n-r(X)}F\right)-N\log(F) \\
&\sim N\log\left(\frac{(m-1)^m}{m^m}\right)+N\log\left(\frac{n-r(X)}{(m-1)(N-r(X_*))}\right) \\
&\quad +N\log\left[\left(1+\frac{N-r(X_*)}{n-r(X)}\frac{\sigma_s^2+m\sigma_w^2}{\sigma_s^2}W\right)^m\right]-N\log\left(\frac{\sigma_s^2+m\sigma_w^2}{\sigma_s^2}W\right) \\
&\sim N\log\left[\frac{(m-1)^{m-1}}{m^m}\frac{n-r(X)}{N-r(X_*)}\right]+N\log\left[(1+aW)^m\right] \\
&\quad -N\log\left(\frac{\sigma_s^2+m\sigma_w^2}{\sigma_s^2}\right)-N\log(W) \\
&\sim N\log\left[\frac{(m-1)^{m-1}}{m^m}\frac{n-r(X)}{N-r(X_*)}\frac{\sigma_s^2}{\sigma_s^2+m\sigma_w^2}\right]+N\log\left[\frac{(1+aW)^m}{W}\right] \\
&\sim N\log\left[\frac{(m-1)^{m-1}}{m^m}\frac{1}{a}\right]+N\log\left[\frac{(1+aW)^m}{W}\right] \\
&\sim \tau+N\log\left[\frac{(1+aW)^m}{W}\right],
\end{aligned} \tag{2.47}$$

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 $\lim_{W\to\frac{\kappa\sigma_s^2}{\sigma_s^2+m\sigma_w^2}}\Lambda=0$. 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 (2.47). In practice, the probability mass at zero for the likelihood ratio test in (2.42) should be numerically estimated by using the MLEs as follows.

$$p_m \equiv P(\Lambda=0)=P\left(F\le\kappa\right)$$

$$
\begin{aligned}
&= P\left(W \leq \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\
&\approx P\left(W \leq \frac{\kappa\hat{\sigma_s}^2}{\hat{\sigma_s}^2 + m\hat{\sigma}_w^2}\right).
\end{aligned} \tag{2.48}
$$

The distribution of Λ under the null hypothesis is obtained by substituting zero for σ_w^2 in Lemma 2.3.2. This implies that, under H_0, the probability that Λ is zero does not depend on the parameters σ_w^2 and σ_s^2 and equals

$$
p_m = P\left(W \leq \kappa\right). \tag{2.49}
$$

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

$$
p_c = P\left(\frac{\sum_{s=1}^{K}\mu_{s,n}w_s^2}{\sum_{s=1}^{n-\tilde{p}}w_s^2} \leq \frac{1}{n}\sum_{i=1}^{K}\xi_{s,n}\right), \tag{2.50}
$$

where $\mu_{s,n}$ and $\xi_{s,n}$ are the K eigenvalues of the $K \times K$ matrices $X_1^{'}P_0X_1$ and $X_1^{'}X_1$ respectively, where $w_i \sim N(0,1)$, $P_0 = I_n - \tilde{X}\left(\tilde{X}^{'}\tilde{X}\right)^{-1}\tilde{X}^{'}$ and $\tilde{p}$ is the dimensionality of the vector β_* in view of (1.6). We are assuming that they had a typo in defining $\tilde{p}$, and that they meant $\tilde{p}$ is the rank of the design matrix $\tilde{X}$ instead of the dimensionality of the vector β_*. We present the equivalance between the two formulas in (2.49) and (2.50) in the discussion that follows.

Considering the model in (1.6), 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 (1.13), it's immediate that $M_* \perp (M - M_1)$ and $M_1 \subset M$ so that

$$
r(\tilde{X}) \quad = \quad r(\tilde{M})
$$

$$
\begin{aligned}
&= r(M_* + M_2) \\
&= r(M_* + (M - M_1)) \\
&= r(M_*) + r(M - M_1) \\
&= r(M_*) + r(M) - r(M_1) \\
&= r(X_*) + r(X) - N. \qquad (2.51)
\end{aligned}
$$

Thus, if we let $w_s \sim N(0,1)$ then according to (2.50), the probability mass at zero for the likelihood ratio test in testing (1.10) is

$$
\begin{aligned}
& p_c = P\left(\frac{\sum_{s=1}^{K} \mu_{s,n} w_s^2}{\sum_{s=1}^{n-p} w_s^2} \le \frac{1}{n}\sum_{i=1}^{K} \xi_{s,n}\right) \\
\Longleftrightarrow \quad & p_c = P\left(\frac{\sum_{s=1}^{N-r(X_*)} m w_s^2}{\sum_{s=1}^{n-(r(X_*)+r(X)-N)} w_s^2} \le \frac{1}{n}\sum_{i=1}^{N} m\right) \\
\Longleftrightarrow \quad & p_c = P\left(\frac{\sum_{s=1}^{N-r(X_*)} m w_s^2}{\sum_{s=1}^{N-r(X_*)} w_s^2 + \sum_{s=N-r(X_*)+1}^{n-(r(X_*)+r(X)-N)} w_s^2} \le 1\right) \\
\Longleftrightarrow \quad & p_c = P\left(\sum_{s=1}^{N-r(X_*)} m w_s^2 \le \sum_{s=1}^{N-r(X_*)} w_s^2 + \sum_{s=N-r(X_*)+1}^{n-(r(X_*)+r(X)-N)} w_s^2\right) \\
\Longleftrightarrow \quad & p_c = P\left((m-1)\sum_{s=1}^{N-r(X_*)} w_s^2 \le \sum_{s=N-r(X_*)+1}^{n-(r(X_*)+r(X)-N)} w_s^2\right) \\
\Longleftrightarrow \quad & p_c = P\left(\frac{\sum_{s=1}^{N-r(X_*)} w_s^2}{\sum_{s=N-r(X_*)+1}^{n-(r(X_*)+r(X)-N)} w_s^2} \le \frac{1}{m-1}\right) \\
\Longleftrightarrow \quad & p_c = P\left(\frac{W_1}{W_2} \le \frac{1}{m-1}\right) \\
\Longleftrightarrow \quad & p_c = P\left(\frac{W_1/[N-r(X_*)]}{W_2/[n-r(X)]} \le \frac{1}{m-1}\frac{n-r(X)}{N-r(X_*}\right) \\
\Longleftrightarrow \quad & p_c = P\left(W \le \kappa\right), \qquad (2.52)
\end{aligned}
$$

which is the same as (2.49).

The formula in (2.49) is an easier way of getting the probability mass at zero for the likelihood ratio test under H_0 than (2.50). 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 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 2.3.2, 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.

2.4 Power Comparison

This section includes four subsections. In the first section, we illustrate the steps for computing the critical value and power of the F-test and give a concrete example. In the second section, 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 second section, don't utilize the relationship between the two test statistics Λ and F. Thus, the second section is concluded with a very short theorem on the equivalence between the two test for the case when $\alpha \leq 1 - p_m$. The third section gives a detailed proof, without numerical examples, that the F-test has a lerger power than the LRT when $\alpha > 1 - p_m$. The fourth section discusses the practicality of the second case when $\alpha > 1 - p_m$.

2.4.1 The Power of the F-test

The F-test statistic, by Proposition 2.3.1, 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 $W \sim F\left(N-r(X_*), n-r(X)\right)$. Then, at a given significance level α, the critical value C is computed under H_0 as

$$\begin{aligned} & \alpha = P\left(F \geq C|H_0 \text{ is true }\right) \iff \alpha = P\left(\frac{\sigma_s^2+m\sigma_w^2}{\sigma_s^2}W \geq C|\sigma_w^2=0\right) \\ \iff & \alpha = P\left(W \geq W_\alpha\right) \\ \iff & W_\alpha = G^{-1}(1-\alpha), \end{aligned} \tag{2.53}$$

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 α F-test is

$$\begin{aligned} & \Xi_F = P\left(F \geq C|H_a \text{ is true }\right) \iff \Xi_F = P\left(\frac{\sigma_s^2+m\sigma_w^2}{\sigma_s^2}W \geq W_\alpha\right) \\ \iff & \Xi_F = 1-G\left(\frac{\sigma_s^2}{\sigma_s^2+m\sigma_w^2}W_\alpha\right). \end{aligned} \tag{2.54}$$

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 $\Xi_F = 1-G\left(\frac{3}{31}\times 3.86255\right) = 0.7741$.

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

The LRT statistic, by Lemma 2.3.2, has a mixtute distribution as

$$\Lambda \sim \begin{cases} 0 & \text{w.p} \quad p_m \\ \tau + N\log\left(\frac{(1+aW)^m}{W}\right) & \text{w.p} \quad 1-p_m \end{cases} \tag{2.55}$$

such that p_m, τ and a are defined in Lemma 2.3.2. Thus, at a given significance level α, the critical value C' is computed under H_0 as

$$\begin{aligned}
&\alpha = P\left(\Lambda \geq C'|H_0 \text{ is true }, W > \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\
\iff &\alpha = P\left(\Lambda \geq C'|\sigma_w^2 = 0, W > \kappa\right) \\
\iff &(1-p_m)P\left(\tau + N\log\left(\frac{(1+aW)^m}{W}\right) \geq C'|\sigma_w^2 = 0, W > \kappa\right) = \alpha \\
\iff &P\left(\tau + N\log\left(\frac{(1+aW)^m}{W}\right) \geq C'|\sigma_w^2 = 0, W > \kappa\right) = \frac{\alpha}{1-p_m} \\
\iff &P\left(\tau + N\log\left(\frac{(1+aW)^m}{W}\right) < C'|\sigma_w^2 = 0, W > \kappa\right) = 1 - \frac{\alpha}{1-p_m} \\
\iff &C' = G'^{-1}(1 - \frac{\alpha}{1-p_m}),
\end{aligned} \tag{2.56}$$

where G' is the CDF of the transformed random variable $\tau + N\log\left(\frac{(1+aW)^m}{W}\right)$ for $W \sim F\left(N - r(X_*), n - r(X)\right)$ 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

$$\Lambda = \begin{cases} 0 & \text{w.p} \quad 0.56371 \\ 6\log\left[\frac{81}{256}\frac{\left(1+\frac{W}{3}\right)^4}{W}\right] & \text{w.p} \quad 0.43629 \end{cases} \tag{2.57}$$

so that C' is found, by numerical simulation or numerical integration after transformation, as $C' = G'^{-1}(1 - \frac{\alpha}{1-p_m}) = G^{-1}(0.8853973) = 4.848$. 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 α LRT is

$$\begin{aligned}
&\Xi_{LRT} = P\left(\Lambda \geq C'|H_a \text{ is true }, W > \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\
\iff &\Xi_{LRT} = P\left(\Lambda \geq C'|\sigma_w^2 > 0, W > \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\
\iff &\Xi_{LRT} = (1-p_m)P\left(\tau + N\log\left(\frac{(1+aW)^m}{W}\right) \geq C'|\sigma_w^2 > 0, W > \frac{\kappa\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\
\iff &\Xi_{LRT} = (1-p_m)\left[1 - G"\left(C'\right)\right]
\end{aligned} \tag{2.58}$$

where G" is the CDF of the transformed random variable $\tau + N\log\left(\frac{(1+aW)^m}{W}\right)$ for $W \sim F\left(N - r(X_*), n - r(X)\right)$ 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

$$\Lambda \sim \begin{cases} 0 & \text{w.p} \quad 0.04014 \\ 6\log\left[\frac{243}{7936}\frac{\left(1+\frac{31W}{9}\right)^4}{W}\right] & \text{w.p} \quad 0.95986 \end{cases} \tag{2.59}$$

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

Note that both test statistics Λ 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 consist of positive values where Λ is a strictly increasing function of the F, thus we have

Proposition 2.4.1 *Let α be the size of the test. If $\alpha \leq 1 - p_m$ where $p = P(\Lambda = 0|\sigma_w^2 = 0)$ then the F-test and LRT are equivalent and hence have the same power.*

Proof of Proposition (2.4.1): Since Λ can be written as

$$\Lambda \sim \begin{cases} 0 & \text{w.p} \quad p_m \\ g(F) & \text{w.p} \quad 1 - p_m \end{cases} \tag{2.60}$$

where $g(.)$ 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

$$\begin{aligned} & \Xi_{LRT} = P\left(\Lambda \geq C' | H_a \text{ is true }\right) \iff \Xi_{LRT} = P\left(g(F) \geq C' | \sigma_w^2 > 0\right) \\ \iff \quad & \Xi_{LRT} = P\left(F \geq C\text{"} | \sigma_w^2 > 0\right) = \Xi_F. \end{aligned} \tag{2.61}$$

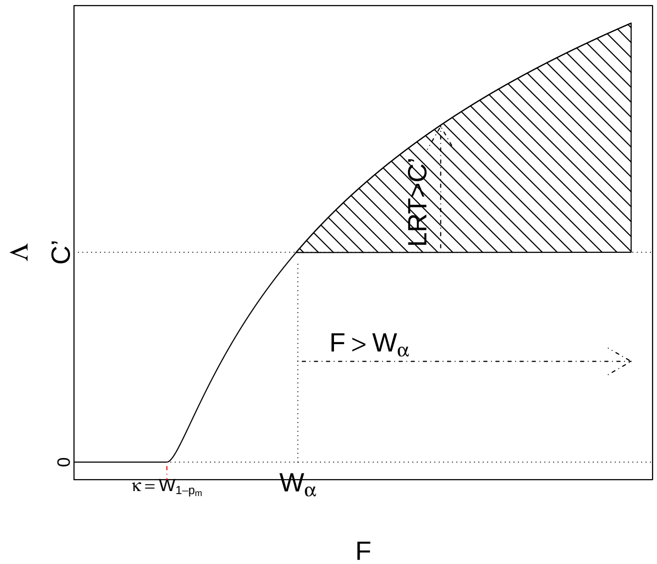

Figure 2.2: A plot of the LRT statistic versus the F-ratio showing their equivalence whenever $\alpha \leq 1 - p_m$. $W_{1-p_m} = \kappa$ is the minimal critical value at which the two tests are equivalent.

$\square$

Figure 2.2 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_α 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 $\lim_{W \to \kappa} \Lambda = 0$.

2.4.3 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 α F-test in terms of p_m the probabilty that $\Lambda = 0$ under H_0 as follows.

$$\begin{aligned}
\Xi_F &= P\left(F \geq W_\alpha\right) = P(F \geq W_{1-p_m}) + P(W_\alpha \leq F \leq W_{1-p_m}) \\
&= P\left(kW \geq W_{1-p_m}\right) + P\left(W_\alpha \leq kW \leq W_{1-p_m}\right) \\
&= P\left(W \geq \frac{1}{k}W_{1-p_m}\right) + P\left(W \leq \frac{1}{k}W_{1-p_m}\right) - P\left(W \leq \frac{1}{k}W_\alpha\right). \quad (2.62)
\end{aligned}$$

Note that the second equality in (2.62) 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.

$$\phi(\Lambda) = \begin{cases} 1 & \text{if} \quad \Lambda > 0 \\ \gamma & \text{if} \quad \Lambda = 0 \\ 0 & \text{if} \quad \Lambda < 0 \end{cases} \iff \phi(F) = \begin{cases} 1 & \text{if} \quad F > W_{1-p_m} \\ \gamma & \text{if} \quad F \leq W_{1-p_m} \end{cases} \quad (2.63)$$

where γ is determined according to the size of the test as

$$\begin{aligned}
\alpha \quad &= \quad E_{H_0}\phi(\Lambda) \iff \alpha = P(\Lambda > 0|\sigma_w^2 = 0) + \gamma P(\Lambda = 0|\sigma_w^2 = 0) \\
&\iff \alpha = (1 - p_m) + \gamma p_m \\
&\iff \gamma = \frac{\alpha - (1 - p_m)}{p_m}. \quad (2.64)
\end{aligned}$$

Hence, the power of the LRT is

$$\begin{aligned}
\Xi_{LRT} &= P\left(F \geq W_{1-p_m}\right) + \frac{\alpha - (1 - p_m)}{p_m} P\left(F \leq W_{1-p_m}\right) \\
&= P\left(W \geq \frac{1}{k}W_{1-p_m}\right) + \frac{\alpha - (1 - p_m)}{p_m} P\left(W \leq \frac{1}{k}W_{1-p_m}\right). \quad (2.65)
\end{aligned}$$

Proposition 2.4.2 *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* α *F-test is larger than that of the LRT in testing* $\sigma_w^2 = 0$*.*

Proof of Proposition 2.4.2: It's sufficient to show that

$$\begin{aligned}
& P\left(W \le \frac{1}{k}W_{1-p_m}\right) - P\left(W \le \frac{1}{k}W_\alpha\right) > \frac{\alpha - (1-p_m)}{p_m} P\left(W \le \frac{1}{k}W_{1-p_m}\right) \\
\Longleftrightarrow\ & P\left(W \le \frac{1}{k}W_{1-p_m}\right)\left[1 - \frac{\alpha - (1-p_m)}{p_m}\right] > P\left(W \le \frac{1}{k}W_\alpha\right) \\
\Longleftrightarrow\ & (1-\alpha)P\left(W \le \frac{1}{k}W_{1-p_m}\right) > p_m P\left(W \le \frac{1}{k}W_\alpha\right) \\
\Longleftrightarrow\ & \frac{1}{p_m}P\left(W \le \frac{1}{k}W_{1-p_m}\right) > \frac{1}{1-\alpha}P\left(W \le \frac{1}{k}W_\alpha\right), \qquad (2.66)
\end{aligned}$$

which is true, since $k = \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} > 1$, according to the F-Inequality in Chapter 5. □

2.4.4 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 α's satisfying the inequality $\alpha \le P\left(W > \frac{n-r(X)}{(m-1)(N-r(X_*))}\right)$. Table 2.1 presents the maximal values of α 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 α satisfying the inequality $\alpha \le P\left(W > \frac{n-r(X)}{(m-1)(N-r(X_*))}\right)$, it's sufficient to provide another Table (see Table 2.2) for the case when $m = 4$ to explain the pattern in which those maximal values of α behave as a function of m. The highlighted cells of Table 2.1 in red 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 2.2 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 2.4.3 *For GSP models, if $\alpha > 1-p_m$ then, for a size α test, $\Xi_{LRT} \uparrow \Xi_F$ in testing $\sigma_w^2 = 0$ as the whole plots size m approaches infinity.*

Proof of Proposition 2.4.3: Recall that

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

and

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

From Proposition 2.4.2, 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 $\Xi_{LRT} \uparrow \Xi_F$. In fact for $m = \infty$ we have $\Xi_{LRT} = \Xi_F = 1$ since $\lim_{k\to+\infty} P\left(W \geq \frac{1}{k}W_{1-p_m}\right) = 1$. $\square$

$df_2 \backslash df_1$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**15**	**20**	**30**	**40**	**60**	**120**
1	0.50	0.71	0.82	0.88	0.92	0.95	0.97	0.98	0.99	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	0.88	0.91	0.94	0.96	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	0.78	0.84	0.88	0.91	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	0.69	0.76	0.81	0.86	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	0.59	0.67	0.74	0.79	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	0.50	0.58	0.66	0.72	0.77	0.93	0.98	1.00	1.00	1.00	1.00
7	*0.03*	0.09	0.16	0.24	0.33	0.42	0.50	0.58	0.65	0.71	0.90	0.97	1.00	1.00	1.00	1.00
8	*0.02*	0.06	0.12	0.19	0.26	0.34	0.42	0.50	0.57	0.64	0.86	0.95	1.00	1.00	1.00	1.00
9	*0.01*	*0.04*	0.09	0.14	0.21	0.28	0.35	0.43	0.50	0.57	0.82	0.93	0.99	1.00	1.00	1.00
10	*0.01*	*0.03*	0.06	0.11	0.16	0.23	0.29	0.36	0.43	0.50	0.77	0.91	0.99	1.00	1.00	1.00
15	*0.00*	*0.01*	*0.01*	*0.03*	0.05	0.07	0.10	0.14	0.18	0.23	0.50	0.73	0.95	0.99	1.00	1.00
20	*0.00*	*0.00*	*0.00*	*0.01*	*0.01*	*0.02*	*0.03*	0.05	0.07	0.09	0.27	0.50	0.85	0.97	1.00	1.00
30	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.01*	0.05	0.15	0.50	0.80	0.99	1.00
40	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.03*	0.20	0.50	0.92	1.00
60	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	0.08	0.50	1.00
120	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	0.50

Table 2.1: The maximal values of α 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 $m = 2$.

$df_2 \backslash df_1$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**15**	**20**	**30**	**40**	**60**	**120**
1	0.67	0.87	0.94	0.97	0.99	0.99	1.00	1.00	1.00	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	0.98	0.99	1.00	1.00	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	0.97	0.98	0.99	1.00	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	0.95	0.97	0.98	0.99	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	0.92	0.96	0.97	0.99	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	0.90	0.94	0.96	0.98	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	0.86	0.91	0.95	0.97	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	0.83	0.89	0.93	0.96	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	0.79	0.86	0.91	0.94	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	0.76	0.83	0.89	0.92	0.95	1.00	1.00	1.00	1.00	1.00	1.00
15	*0.04*	0.12	0.22	0.33	0.45	0.56	0.66	0.74	0.81	0.86	0.98	1.00	1.00	1.00	1.00	1.00
20	*0.02*	0.06	0.12	0.20	0.29	0.39	0.49	0.58	0.67	0.74	0.94	0.99	1.00	1.00	1.00	1.00
30	*0.00*	*0.01*	*0.03*	0.06	0.11	0.16	0.23	0.31	0.39	0.47	0.80	0.95	1.00	1.00	1.00	1.00
40	*0.00*	*0.00*	*0.01*	*0.02*	*0.04*	0.06	0.09	0.14	0.19	0.25	0.58	0.83	0.99	1.00	1.00	1.00
60	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.01*	*0.02*	*0.03*	0.05	0.21	0.48	0.89	0.99	1.00	1.00
120	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	0.14	0.48	0.96	1.00

Table 2.2: The maximal values of α 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 $m = 4$.

Chapter 3

The Equivalence Between the RLRT and F-test

3.1 The Restricted Maximum Likelihood Estimators (REMLs)

Restricted (or residual, or reduced) maximum likelihood (REML) is a well known method for improving the bias of the MLEs in their estimation of variance components in LMMs. The REMLs of σ_w^2 and σ_s^2 for a GSP model are obtained by maximizing the likelihood function of $K'Y$ where K is any $n \times [n - r(\tilde{X})]$ full-rank matrix satisfying $K'\tilde{X} = 0$. Christensen (2011) has shown that the maximum of the likelihood does not depend on the choice of K. Since $K'Y \sim N(0, K'VK)$ where V is given in (1.7), then the likelihood function is free of β_* and can be written

$$L(\sigma_w^2, \sigma_s^2 | K'Y) = (2\pi)^{-0.5[n-r(\tilde{X})]} |K'VK|^{-0.5} e^{-0.5(K'Y)'(K'VK)^{-1}(K'Y)} \tag{3.1}$$

so that minus 2 times the log-likelihood is

$$\ell(\sigma_w^2, \sigma_s^2 | K'Y) = [n - r(\tilde{X})] \log(2\pi) + \log|K'VK| + (K'Y)'(K'VK)^{-1}(K'Y). \quad (3.2)$$

However, as a consequence of Lemmas 12.6.2 and 12.6.3 of Christensen (2011), the term $(K'Y)'(K'VK)^{-1}(K'Y)$ is the same as $\Psi = (Y - \tilde{X}\hat{\beta}_*)'V^{-1}(Y - \tilde{X}\hat{\beta}_*)$. Thus, ℓ in (3.2) can be simplified to

$$\ell(\sigma_w^2, \sigma_s^2 | K'Y) = [n - r(\tilde{X})] \log(2\pi) + \log|K'VK| + \frac{SSE(s)}{\sigma_s^2} + \frac{SSE(w)}{\sigma_s^2 + m\sigma_w^2}. \quad (3.3)$$

Further, from (1.13), one can verify that we have an orthogonal decomposition of R^n based on orthogonal ppos

$$I = M_* + (M_1 - M_*) + M_2 + (I - M), \quad (3.4)$$

where the right hand side are ppos on to, respectively, the whole plot effect space the whole plot error space the subplot effect space and the subplot error space. Thus, given condition (c) of the definition of a GSP model and (3.4), we can decompose the column spaces of $\tilde{X}$ and its orthogonal complement as

$$C(\tilde{X}) = C(M_* + M_2), \quad (3.5)$$

and

$$C(\tilde{X})^{\perp} = C((I - M) + (M_1 - M_*)). \quad (3.6)$$

So, if we define $K := [K_1, K_2]$ with $K'K = I$ then the column space of K could also be decomposed as

$$C(K) = C((I - M) + (M_1 - M_*)), \quad (3.7)$$

with $C(K_1) = C(I - M)$, $C(K_2) = C(M_1 - M_*) \subset C(M_1)$, $K_1'K_1 = I_{n-r(X)}$, $K_2'K_2 = I_{N-r(X_*)}$, $K_1'K_2 = 0_{(n-r(X))\times(N-r(X_*))}$ and $K_2'K_1 = 0_{(N-r(X_*))\times(n-r(X))}$. We use this decompostition to simplify the restricted likelihood. Using condition (a), $X_1X_1' = mM_1$ so that for $V = \sigma_s^2 I_n + \sigma_w^2 X_1X_1'$ we have

$$
\begin{aligned}
|K'VK| &= \left| \begin{bmatrix} K_1' \\ K_2' \end{bmatrix} (\sigma_s^2 I_n + m\sigma_w^2 M_1) \begin{bmatrix} K_1 & K_2 \end{bmatrix} \right| \\
&= \begin{vmatrix} (\sigma_s^2 K_1'K_1 + m\sigma_w^2 K_1'M_1K_1) & (\sigma_s^2 K_1'K_2 + m\sigma_w^2 K_1'M_1K_2) \\ (\sigma_s^2 K_2'K_1 + m\sigma_w^2 K_2'M_1K_1) & (\sigma_s^2 K_2'K_2 + m\sigma_w^2 K_2'M_1K_2) \end{vmatrix} \\
&= \begin{vmatrix} \sigma_s^2 I_{n-r(X)} & 0 \\ 0 & (\sigma_s^2 + m\sigma_w^2) I_{N-r(X_*)} \end{vmatrix} \\
&= (\sigma_s^2)^{n-r(X)} (\sigma_s^2 + m\sigma_w^2)^{N-r(X_*)}. \qquad (3.8)
\end{aligned}
$$

Hence, $-\ell$ in (3.3) becomes

$$
\begin{aligned}
\ell_*(\sigma_w^2, \sigma_s^2 | K'Y) &= -\Big[(n - r(\tilde{X})) \log(2\pi) + (n - r(X)) \log(\sigma_s^2) \\
&\quad + (N - r(X_*)) \log(\sigma_s^2 + m\sigma_w^2) + \frac{SSE(s)}{\sigma_s^2} + \frac{SSE(w)}{\sigma_s^2 + m\sigma_w^2}\Big]. \qquad (3.9)
\end{aligned}
$$

Proposition 3.1.1 *The Restricted Maximum Likelihood estimators for σ_w^2 and σ_s^2 of model (1.6) are*

$$\hat{\sigma_w^2} = \frac{1}{m} \max\left\{0, \frac{SSE(w)}{N - r(X_*)} - \frac{SSE(s)}{n - r(X)}\right\}, \text{ and} \qquad (3.10)$$

$$\hat{\sigma_s^2} = \min\left\{\frac{SSE(s)}{n - r(X)}, \frac{SSE(s) + SSE(w)}{n - r(\tilde{X})}\right\}, \qquad (3.11)$$

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

Proof of Proposition 3.1.1: let $q_1 = n-r(X)$, $q_2 = N-r(X_*)$, $x_1 = \sigma_s^2$, $x_2 = \sigma_s^2+m\sigma_w^2$, $Q_1 = \frac{SSE(s)}{q_1}$ 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 and therefore applying Lemma 2.1.3 to ℓ_* in (3.9) gives the maximizers

$$
\begin{aligned}
& \left(\hat{\sigma}_s^2, \hat{\sigma}_s^2 + m\hat{\sigma}_w^2\right) = \left(\frac{SSE(s)}{n-r(X)}, \frac{SSE(w)}{N-r(X_*)}\right) \\
\iff \quad & \left(\hat{\sigma}_s^2, \hat{\sigma}_w^2\right) = \left(\frac{SSE(s)}{n-r(X)}, \frac{1}{m}\left[\frac{SSE(w)}{N-r(X_*)} - \frac{SSE(s)}{n-r(X)}\right]\right) \qquad (3.12)
\end{aligned}
$$

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

$$
\begin{aligned}
& \left(\hat{\sigma}_s^2, \hat{\sigma}_s^2 + m\hat{\sigma}_w^2\right) = \left(\frac{SSE(s)+SSE(w)}{n-r(X)+N-r(X_*)}, \frac{SSE(s)+SSE(w)}{n-r(X)+N-r(X_*)}\right) \\
\iff \quad & \left(\hat{\sigma}_s^2, \hat{\sigma}_s^2 + m\hat{\sigma}_w^2\right) = \left(\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}, \frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}\right) \\
\iff \quad & \left(\hat{\sigma}_s^2, \hat{\sigma}_w^2\right) = \left(\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}, 0\right) \qquad (3.13)
\end{aligned}
$$

when $\frac{SSE(w)}{N-r(X_*)} \leq \frac{SSE(s)}{n-r(X)}$. Suppose the REML of σ_w^2 is $\hat{\sigma}_w^2 = \frac{1}{m}\left(\frac{SSE(w)}{N-r(X_*)} - \frac{SSE(s)}{n-r(X)}\right)$ then $\frac{SSE(s)}{n-r(X)} \leq \frac{SSE(w)}{N-r(X_*)}$ so that

$$
\begin{aligned}
\frac{SSE(s)}{n-r(X)} \quad & \leq \quad \left(\frac{n-r(X)}{n-r(\tilde{X})}\right)\frac{SSE(s)}{n-r(X)} + \left(\frac{N-r(X_*)}{n-r(\tilde{X})}\right)\frac{SSE(w)}{N-r(X_*)} \\
& = \frac{SSE(s)+SSE(w)}{n-r(\tilde{X})} \qquad (3.14)
\end{aligned}
$$

with the first inequality true because the term in the middle is a weighted average, since $r(\tilde{X}) = r(X_*) + r(X) - N$, so has to be larger than the smaller of the two things being averaged, therefore the REML of σ_s^2 is the smaller of the terms $\frac{SSE(s)}{n-r(X)}$ and $\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}$. That is, the larger term between 0 and $\frac{1}{m}\left(\frac{SSE(w)}{N-r(X_*)} - \frac{SSE(s)}{n-r(X)}\right)$ forces the answer to be the smaller term between $\frac{SSE(s)}{n-r(X)}$ and $\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}$ and vice versa. Hence (3.12) and (3.13) could be written via max and min as

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

and

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

$\square$

Note that the partial derivatives for ℓ_* in (3.9) are

$$\frac{\partial \ell}{\partial \sigma_w^2} = -\left[\frac{m(N - r(X_*))}{\sigma_s^2 + m\sigma_w^2} - \frac{mSSE(w)}{(\sigma_s^2 + m\sigma_w^2)^2}\right], \tag{3.15}$$

and

$$\frac{\partial \ell}{\partial \sigma_s^2} = -\left[\frac{N - r(X_*)}{\sigma_s^2 + m\sigma_w^2} + \frac{n - r(X)}{\sigma_s^2} - \frac{SSE(s)}{(\sigma_s^2)^2} - \frac{SSE(w)}{(\sigma_s^2 + m\sigma_w^2)^2}\right]. \tag{3.16}$$

So, for varification purposes, plugging in the pair $\hat{\sigma}_w^2 = 0$ and $\hat{\sigma}_s^2 = \frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}$ into (3.16) and the other pair $\hat{\sigma}_w^2 = \frac{1}{m}\left[\frac{SSE(w)}{N-r(X_*)} - \frac{SSE(s)}{n-r(X)}\right]$ and $\hat{\sigma}_s^2 = \frac{SSE(s)}{n-r(X)}$ into (3.15) gives zero as desired.

3.2 Monotonic Relationship Between the RLRT and F-test Statistics

We show that the restricted likelihood ratio test statistic Λ^r is a monotonic function of the F-test statistic F for testing the null hypothesis in (1.10). When the Λ^r is not 0, the monotone relationship is strict, so whenever the size of the test α is smaller than the probability $1 - p_r$ that $\Lambda^r \neq 0$, the tests are equivalent. We also examine the behavior of the tests when they are not equivalent (i.e. when $\alpha > 1 - p_r$). Firstly, under the reduced model in (1.9), the -2Log restricted likelihood function and REML estimate for σ_e^2 are, respectively,

$$\ell^r(\sigma_e^2|K'Y) = (n - r(\tilde{X}))\log(2\pi) + (n - r(\tilde{X}))\log(\sigma_e^2) + \frac{SSE(e)}{\sigma_e^2}, \tag{3.17}$$

and

$$\hat{\sigma}_e^2 = \frac{SSE(e)}{n - r(\tilde{X})}. \tag{3.18}$$

Secondly, the -2Log restricted likelihood function for the full model in (1.6) is

$$\begin{aligned}\ell^r(\sigma_w^2, \sigma_s^2|K'Y) &= (n - r(\tilde{X}))\log(2\pi) + (n - r(X))\log(\sigma_s^2) \\ &+(N - r(X_*))\log(\sigma_s^2 + m\sigma_w^2) + \frac{SSE(s)}{\sigma_s^2} + \frac{SSE(w)}{\sigma_s^2 + m\sigma_w^2} \quad (3.19)\end{aligned}$$

Proposition 3.2.1 *The RLRT statistic Λ^r, for testing (1.10), is a monotone function of the F-test statistic F. In particular, (case-I) when $\hat{\sigma}_w^2 = 0$ and $\hat{\sigma}_s^2 = \frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}$ we have*

$$\Lambda^r = 0, \tag{3.20}$$

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

$$\begin{aligned}\Lambda^r &= \zeta + (n - r(X))\log\left(1 + \frac{N - r(X_*)}{n - r(X)}F\right) \\ &+(N - r(X_*))\log\left(1 + \frac{n - r(X)}{N - r(X_*)}\frac{1}{F}\right), \quad (3.21)\end{aligned}$$

where

$$\zeta = (n - r(X))\log\left(\frac{n - r(X)}{n - r(\tilde{X})}\right) + (N - r(X_*))\log\left(\frac{N - r(X_*)}{n - r(\tilde{X})}\right) \tag{3.22}$$

and case-I occurs only when $F \leq 1$ while case-II occurs when $F > 1$.

Proof of proposition (3.2.1) The RLRT statistic Λ^r is defined as the difference of -2Log restricted likelihood for two nested models (reduced and full). In our case, (1.6) is the full model and (1.9) is the reduced one so that

$$\Lambda^r = \ell_R^r(\hat{\sigma}_e^2|K'Y) - \ell_F^r(\hat{\sigma}_w^2, \hat{\sigma}_s^2|K'Y), \tag{3.23}$$

where $\ell_R^r(\hat{\sigma}_e^2|K'Y) \equiv \sup \ell^r(\hat{\sigma}_e^2|K'Y)$ and $\ell_F^r(\hat{\sigma}_w^2, \hat{\sigma}_s^2|K'Y) \equiv \sup \ell^r(\hat{\sigma}_w^2, \hat{\sigma}_s^2|K'Y)$ such that the R and F subscripts refere to the reduced and full models respectively. Since $SSE(e) = SSE(s) + SSE(w)$, then plugging (3.18) into (3.17) gives

$$\ell_R^r = (n-r(\tilde{X}))\log(2\pi) + (n-r(\tilde{X})) + (n-r(\tilde{X}))\log\left(\frac{SSE(s) + SSE(w)}{n - r(\tilde{X})}\right). \tag{3.24}$$

Further, if we let $\hat{\sigma}_s^2$ and $\hat{\sigma}_w^2$ be the RMLEs for σ_s^2 and σ_w^2 respectively and plug them into (3.19) we get

$$\begin{aligned}
\ell_F^r &= (n-r(\tilde{X}))\log(2\pi) + (n-r(X))\log(\hat{\sigma}_s^2) + (N-r(X_*))\log(\hat{\sigma}_s^2+m\hat{\sigma}_w^2) \\
&\quad + \frac{SSE(s)}{\hat{\sigma}_s^2} + \frac{SSE(w)}{\hat{\sigma}_s^2+m\hat{\sigma}_w^2}. \qquad (3.25)
\end{aligned}$$

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

$$\begin{aligned}
\Lambda^r &= (n-r(\tilde{X})) + (n-r(\tilde{X}))\log\left(\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}\right) \\
&\quad -(n-r(X))\log\left(\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}\right) \\
&\quad -(N-r(X_*))\log\left(\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}\right) - \frac{SSE(s)}{\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}} \\
&= (n-r(\tilde{X}))\log\left(\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}\right) \\
&\quad -(n+N-r(X)-r(X_*))\log\left(\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}\right) \\
&= 0. \qquad (3.26)
\end{aligned}$$

Case-II: If $\hat{\sigma}_w^2 = \frac{1}{m}\left[\frac{SSE(w)}{N-r(X_*)} - \frac{SSE(s)}{n-r(X)}\right]$ and $\hat{\sigma}_s^2 = \frac{SSE(s)}{n-r(X)}$ then, by (1.16), we get

$$\begin{aligned}
\Lambda^r &= (n+r(\tilde{X})) + (n-r(X))\log\left(\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}\right) \\
&\quad -(n-r(X))\log\left(\frac{SSE(s)}{n-r(X)}\right) \\
&\quad -(N-r(X_*))\log\left(\frac{SSE(s)}{n-r(X)} + m\frac{1}{m}\left(\frac{SSE(w)}{N-r(X_*)} - \frac{SSE(s)}{n-r(X)}\right)\right) \\
&\quad -\frac{SSE(s)}{\frac{SSE(s)}{n-r(X)}} - \frac{SSE(w)}{\frac{SSE(s)}{n-r(X)} + m\frac{1}{m}\left(\frac{SSE(w)}{N-r(X_*)} - \frac{SSE(s)}{n-r(X)}\right)} \\
&= (n-r(\tilde{X})) - (n-r(X)) + (n-r(\tilde{X}))\log\left(\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}\right) \\
&\quad -(n-r(X))\log\left(\frac{SSE(s)}{n-r(X)}\right) \\
&\quad -(N-r(X_*))\log\left(\frac{SSE(w)}{N-r(X_*)}\right) - (N-r(X_*))
\end{aligned}$$

$$
\begin{aligned}
&= (n-r(X)+N-r(X_*))\log\left(\frac{SSE(s)+SSE(w)}{n-r(\tilde{X})}\right) \\
&\quad -(n-r(X))\log\left(\frac{SSE(s)}{n-r(X)}\right)-(N-r(X_*))\log\left(\frac{SSE(w)}{N-r(X_*)}\right) \\
&= (n-r(X))\log\left(\frac{n-r(X)}{n-r(\tilde{X})}\left(1+\frac{SSE(w)}{SSE(s)}\right)\right) \\
&\quad +(N-r(X_*))\log\left(\frac{N-r(X_*)}{n-r(\tilde{X})}\right)\left(1+\frac{SSE(s)}{SSE(w)}\right) \\
&= (n-r(X))\log\left(\frac{n-r(X)}{n-r(\tilde{X})}\left(1+\frac{N-r(X_*)}{n-r(X)}F\right)\right) \\
&\quad +(N-r(X_*))\log\left(\frac{N-r(X_*)}{n-r(\tilde{X})}\right)\left(1+\frac{n-r(X)}{N-r(X_*)}\frac{1}{F}\right) \\
&= (n-r(X))\log\left(\frac{n-r(X)}{n-r(\tilde{X})}\right)+(N-r(X_*))\log\left(\frac{N-r(X_*)}{n-r(\tilde{X})}\right) \\
&\quad +(n-r(X))\log\left(1+\frac{N-r(X_*)}{n-r(X)}F\right) \\
&\quad +(N-r(X_*))\log\left(1+\frac{n-r(X)}{N-r(X_*)}\frac{1}{F}\right) \\
&= \zeta+(n-r(X))\log\left(1+\frac{N-r(X_*)}{n-r(X)}F\right) \\
&\quad +(N-r(X_*))\log\left(1+\frac{n-r(X)}{N-r(X_*)}\frac{1}{F}\right). \qquad (3.27)
\end{aligned}
$$

This case holds only when $\hat{\sigma}_w^2 > 0$, which in turn means that F must be larger than 1 since

$$
\begin{aligned}
& \frac{1}{m}\left[\frac{SSE(w)}{N-r(X_*)}-\frac{SSE(s)}{n-r(X)}\right]>0 \iff \frac{SSE(w)}{SSE(s)}>\frac{N-r(X_*)}{n-r(X)} \\
\iff \quad & \frac{N-r(X_*)}{n-r(X)}F>\frac{N-r(X_*)}{n-r(X)} \\
\iff \quad & F>1. \qquad (3.28)
\end{aligned}
$$

The Λ^r (3.27) is strictly increasing in F for case II since

$$
\begin{aligned}
& \frac{\partial\Lambda^r}{\partial F}>0 \iff \frac{N-r(X_*)}{1+\frac{N-r(X_*)}{n-r(X)}F}-\frac{(n-r(X))\frac{1}{F^2}}{1+\frac{n-r(X)}{N-r(X_*)}\frac{1}{F}}>0 \\
\iff \quad & \frac{n-r(X)}{(n-r(X))+(N-r(X_*))F}>\frac{\frac{n-r(X)}{N-r(X_*)}\frac{1}{F^2}}{\frac{(N-r(X_*))F+(n-r(X))}{(N-r(X_*))F}}
\end{aligned}
$$

$$
\begin{aligned}
&\iff \quad \frac{n - r(X)}{(n - r(X)) + (N - r(X_*))F} > \frac{\frac{n-r(X)}{F}}{(n - r(X)) + (N - r(X_*))F} \\
&\iff \quad F > 1. \qquad (3.29)
\end{aligned}
$$

□

3.3 The Distribution of the RLRT Statistic

Lemma 3.3.1 *The distribution of the RLRT statistic Λ^r for the model in (1.6) is determined by the relationship in (3.30) where $W \sim F\left(N - r(X_*), n - r(X)\right)$, φ, ζ and Υ are as described.*

$$
\Lambda^r \sim \begin{cases} 0 & \text{if} \quad W \leq \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2} \\ \Upsilon + (n - r(\tilde{X}))\log\left(1 + \varphi W\right) - (N - r(X_*))\log(W) & \text{if} \quad W > \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2} \end{cases} \tag{3.30}
$$

such that

$$
p_r = Pr\left(W \leq \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right), \tag{3.31}
$$

where

$$
\varphi = \frac{N - r(X_*)}{n - r(X)} \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2}, \tag{3.32}
$$

$$
\zeta = (n - r(X))\log\left(\frac{n - r(X)}{n - r(\tilde{X})}\right) + (N - r(X_*))\log\left(\frac{N - r(X_*)}{n - r(\tilde{X})}\right), \tag{3.33}
$$

and

$$
\Upsilon = \zeta - (N - r(X_*))\log(\varphi). \tag{3.34}
$$

Proof of Lemma (3.3.1): From case-I of Proposition 3.2.1 we know that $\Lambda^r = 0$ iif $F \leq 1$ so

$$
p_r \equiv P(\Lambda^r = 0) = P\left(F \leq 1\right)
$$

$$
\begin{aligned}
&= P\left(\frac{MSE(w)}{MSE(s)} \leq 1\right) \\
&= P\left(\frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} W \leq 1\right) \\
&= P\left(W \leq \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right). \qquad (3.35)
\end{aligned}
$$

The equality in the third line of (3.35) holds due to Proposition 2.3.1. Now, from case-II of Proposition 3.2.1 we also know that, if $F > 1$ (i.e. when $W > \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}$),

$$
\begin{aligned}
\Lambda^r &= \zeta + (n - r(X)) \log\left(1 + \frac{N - r(X_*)}{n - r(X)} F\right) \\
&\quad + (N - r(X_*)) \log\left(1 + \frac{n - r(X)}{N - r(X_*)} \frac{1}{F}\right). \qquad (3.36)
\end{aligned}
$$

So, if we let $h = \frac{N - r(X_*)}{n - r(X)} F$ and $W \sim F\left(N - r(X_*), n - r(X)\right)$ then substituting F from (2.37) into (3.36) gives

$$
\begin{aligned}
\Lambda^r &= \zeta + (n - r(X)) \log(1 + h) + (N - r(X_*)) \log\left(1 + \frac{1}{h}\right) \\
&= \zeta + (n - r(X)) \log(1 + h) + (N - r(X_*)) \log(1 + h) - (N - r(X_*)) \log(h) \\
&= \zeta + (n - r(\tilde{X})) \log(1 + h) - (N - r(X_*)) \log(h) \\
&= \zeta + (n - r(\tilde{X})) \log\left(1 + \frac{N - r(X_*)}{n - r(X)} F\right) - (N - r(X_*)) \log\left(\frac{N - r(X_*)}{n - r(X)} F\right) \\
&= \zeta - (N - r(X_*)) \log\left(\frac{N - r(X_*)}{n - r(X)}\right) + (n - r(\tilde{X})) \log\left(1 + \frac{N - r(X_*)}{n - r(X)} F\right) \\
&\quad - (N - r(X_*)) \log(F) \\
&\sim \zeta - (N - r(X_*)) \log\left(\frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} \frac{N - r(X_*)}{n - r(X)}\right) - (N - r(X_*)) \log(W) \\
&\quad + (n - r(\tilde{X})) \log\left(1 + \frac{N - r(X_*)}{n - r(X)} \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} W\right) \\
&\sim \zeta - (N - r(X_*)) \log(\varphi) + (n - r(\tilde{X})) \log(1 + \varphi W) - (N - r(X_*)) \log(W) \\
&\sim \Upsilon + (n - r(\tilde{X})) \log(1 + \varphi W) - (N - r(X_*)) \log(W), \qquad (3.37)
\end{aligned}
$$

where $\varphi = \frac{N - r(X_*)}{n - r(X)} \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2}$, $\zeta = (n - r(X)) \log\left(\frac{n - r(X)}{n - r(\tilde{X})}\right) + (N - r(X_*)) \log\left(\frac{N - r(X_*)}{n - r(\tilde{X})}\right)$, and $\Upsilon = \zeta - (N - r(X_*)) \log(\varphi)$. $\square$

In practice, the probability mass at zero for the likelihood ratio test in (3.31) should be numerically estimated by using the REMLEs or MSEs as follows.

$$\begin{aligned} p_r &\equiv P(\Lambda^r = 0) = P\left(F \leq 1\right) \\ &= P\left(W \leq \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\ &\approx P\left(W \leq \frac{\hat{\sigma_s}^2}{\hat{\sigma_s}^2 + m\hat{\sigma}_w^2}\right), \end{aligned} \tag{3.38}$$

or

$$\begin{aligned} p_r &\equiv P(\Lambda^r = 0) = P\left(F \leq 1\right) \\ &= P\left(W \leq \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\ &\approx P\left(W \leq \frac{MSE(s)}{MSE(w)}\right). \end{aligned} \tag{3.39}$$

The approximation in the third line of (3.39) holds since $MSE(w)$ and $MSE(s)$ are an unbiased estimators for $\sigma_s^2 + m\sigma_w^2$ and σ_s^2 respectively as shown in the following Lemma.

Lemma 3.3.2 *For the model in (1.6), $MSE(w) = \frac{SSE(w)}{N-r(X_*)}$ and $MSE(s) = \frac{SSE(s)}{n-r(X)}$ are an unbiased estimators for $\sigma_s^2 + m\sigma_w^2$ and σ_s^2 respectively.*

Proof of Lemma (3.3.2):

$$\begin{aligned} EMSE_w &= E(MSE(w)) \\ &= E\left(\frac{Y'(M_1 - M_*)Y}{N - r(X_*)}\right) \\ &= \frac{1}{N - r(X_*)} E\left(Y'(M_1 - M_*)Y\right) \\ &= \frac{1}{N - r(X_*)} \left[\mathrm{tr}\left[(M_1 - M_*)\,\mathrm{Cov}(Y)\right] + \mu^{'}(M_1 - M_*)\mu\right], \quad \mu = E(Y) \\ &= \frac{1}{N - r(X_*)} \left[\mathrm{tr}\left[(M_1 - M_*)V\right]\right] \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N-r(X_*)}\left[\mathrm{tr}\left[(M_1-M_*)\left(\sigma_w^2 \mathrm{Blk\ diag}\,(J_mJ_m') + \sigma_s^2 I_n\right)\right]\right] \\
&= \frac{1}{N-r(X_*)}\left[\mathrm{tr}\left[(M_1-M_*)\left(\sigma_w^2 mM_1 + \sigma_s^2 I_n\right)\right]\right] \\
&= \frac{1}{N-r(X_*)}\left[\mathrm{tr}\left(M_1\sigma_w^2 mM_1 + M_1\sigma_s^2 I_n - M_*\sigma_w^2 mM_1 - M_*\sigma_s^2 I_n\right)\right] \\
&= \frac{1}{N-r(X_*)}\left[m\sigma_w^2 r(M_1) + \sigma_s^2 r(M_1) - m\sigma_w^2 r(M_*) - \sigma_s^2 r(M_*)\right] \\
&= \frac{1}{N-r(X_*)}(\sigma_s^2 + m\sigma_w^2)(r(M_1) - r(M_*)) \\
&= \frac{1}{N-r(X_*)}(\sigma_s^2 + m\sigma_w^2)(N - r(X_*)) \\
&= \sigma_s^2 + m\sigma_w^2, \qquad (3.40)
\end{aligned}
$$

and

$$
\begin{aligned}
EMSE_s &= E(MSE(s)) \\
&= E\left(\frac{Y'(I-M)Y}{n-r(X)}\right) \\
&= \frac{1}{n-r(X)}E\left(Y'(I-M)Y\right) \\
&= \frac{1}{n-r(X)}\left[\mathrm{tr}\left[(I-M)\,\mathrm{Cov}(Y)\right] + \mu^{'}(I-M)\mu\right], \quad \mu = E(Y) \\
&= \frac{1}{n-r(X)}\left[\mathrm{tr}\left[(I-M)V\right]\right] \\
&= \frac{1}{n-r(X)}\left[\mathrm{tr}\left[(I-M)(\sigma_w^2 \mathrm{Blk\ diag}\,(J_mJ_m') + \sigma_s^2 I_n)\right]\right] \\
&= \frac{1}{n-r(X)}\left[\mathrm{tr}\left[(I-M)(m\sigma_w^2 + \sigma_s^2 I_n)\right]\right] \\
&= \frac{1}{n-r(X)}\left[\mathrm{tr}\left(m\sigma_w^2 M_1 + \sigma_s^2 I_n - m\sigma_w^2 MM_1 - M\sigma_s^2 I_n\right)\right] \\
&= \frac{1}{n-r(X)}\left[m\sigma_w^2 r(M_1) + \sigma_s^2 r(I_n) - m\sigma_w^2 r(MM_1) - \sigma_s^2 r(M)\right] \\
&= \frac{1}{n-r(X)}\left[m\sigma_w^2 N + n\sigma_s^2 - m\sigma_w^2 r(MM_1) - \sigma_s^2 r(X)\right] \\
&= \frac{1}{n-r(X)}\left[mN\sigma_w^2 + n\sigma_s^2 - mN\sigma_w^2 - \sigma_s^2 r(X)\right] \\
&= \frac{1}{n-r(X)}\left[\sigma_s^2(n-r(X))\right] \\
&= \sigma_s^2. \qquad (3.41)
\end{aligned}
$$

The equality in line 9 of (3.40) is true because $M_* M_1 = M_*$ and the equality in line 11 of (3.41) is true because $M = M_1 + M_2$ and $M_1 M_2 = 0$ (see (1.13)) so that

$$\begin{aligned}
r(MM_1) &= r\left[(M_1 + M_2)M_1\right] \\
&= r(M_1 M_1 + M_2 M_1) \\
&= r(M_1 M_1) + r(M_2 M_1) \\
&= r(M_1) + r(M_1 M_2) = N. \qquad (3.42)
\end{aligned}$$

□

Under H_0, the probability that the $RLRT$ is zero does not depend on the parameters σ_w^2 and σ_s^2 and equals to

$$p_r = Pr\left(W \leq 1\right). \qquad (3.43)$$

This formula agrees with the findings of Crainiceanu and Ruppert (2004) who suggested, to compute the probability mass at zero for the Restricted likelihood ratio test in linear mixed models (LMM) with one variance component, the computation

$$p_{cr} = P\left(\frac{\sum_{s=1}^{K} \mu_{s,n} w_s^2}{\sum_{s=1}^{n-\tilde{p}} w_s^2} \leq \frac{1}{n-p} \sum_{i=1}^{K} \mu_{s,n}\right), \qquad (3.44)$$

where $\mu_{s,n}$, w_is and $\tilde{p}$ are defined in section 2.2. The simplification in (3.45) presents the equivalance between the two formulas in (3.43) and (3.44).

$$\begin{aligned}
& p_{cr} = P\left(\frac{\sum_{s=1}^{K} \mu_{s,n} w_s^2}{\sum_{s=1}^{n-p} w_s^2} \leq \frac{1}{n-p} \sum_{i=1}^{K} \mu_{s,n}\right) \\
\Longleftrightarrow \quad & p_{cr} = P\left(\frac{\sum_{s=1}^{N-r(X_*)} m w_s^2}{\sum_{s=1}^{n-(r(X_*)+r(X)-N)} w_s^2} \leq \frac{1}{n-r(\tilde{X})} \sum_{i=1}^{N-r(X_*)} m\right) \\
\Longleftrightarrow \quad & p_{cr} = P\left(\frac{\sum_{s=1}^{N-r(X_*)} m w_s^2}{\sum_{s=1}^{N-r(X_*)} w_s^2 + \sum_{s=N-r(X_*)+1}^{n-(r(X_*)+r(X)-N)} w_s^2} \leq \frac{m(N-r(X_*))}{n-r(\tilde{X})}\right)
\end{aligned}$$

$$
\begin{aligned}
\iff \quad & p_{cr} = P\left(\frac{n-r(\tilde{X})}{N-r(X_*)}\sum_{s=1}^{N-r(X_*)} w_s^2 \le \sum_{s=1}^{N-r(X_*)} w_s^2 + \sum_{s=N-r(X_*)+1}^{n-(r(X_*)+r(X)-N)} w_s^2\right) \\
\iff \quad & p_{cr} = P\left(\frac{n-r(\tilde{X})-(N-r(X_*))}{N-r(X_*)}\sum_{s=1}^{N-r(X_*)} w_s^2 \le \sum_{s=N-r(X_*)+1}^{n-(r(X_*)+r(X)-N)} w_s^2\right) \\
\iff \quad & p_{cr} = P\left(\frac{\sum_{s=1}^{N-r(X_*)} w_s^2}{\sum_{s=N-r(X_*)+1}^{n-(r(X_*)+r(X)-N)} w_s^2} \le \frac{N-r(X_*)}{n-[r(X_*)+r(X)-N]-(N-r(X_*))}\right) \\
\iff \quad & p_{cr} = P\left(\frac{W_1}{W_2} \le \frac{N-r(X_*)}{n-r(X)}\right) \\
\iff \quad & p_{cr} = P\left(W \le 1\right).
\end{aligned} \tag{3.45}
$$

3.4 Power Comparison

The power function of the F-test was shown in Section 2.4.1.

3.4.1 The power of the RLRT when $\alpha \le 1 - p_r$

The RLRT statistic Λ^r, by Lemma 3.3.1, has a mixtute distribution as

$$
\Lambda^r \sim \begin{cases} 0 & \text{w.p} \quad p_r \\ \Upsilon + (n-r(\tilde{X}))\log(1+\varphi W) - (N-r(X_*))\log(W) & \text{w.p } 1-p_r \end{cases} \tag{3.46}
$$

such that p_r, Υ, and φ are defined in Lemma 3.3.1 . Thus, at a given significance level α, the critical value K' is computed under H_0 as

$$
\begin{aligned}
& \alpha = P\left(\Lambda^r \ge K' | H_0 \text{ is true }, W > \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\
& \iff \alpha = P\left(\Lambda^r \ge K' | \sigma_w^2 = 0, W > 1\right) \\
\iff \quad & (1-p_r)P\Big(\Upsilon + (n-r(\tilde{X}))\log(1+\varphi W) \\
& -(N-r(X_*))\log(W) \ge K' | \sigma_w^2 = 0, W > 1\Big) = \alpha
\end{aligned}
$$

$$
\begin{aligned}
\iff\quad & P\Big(\Upsilon + (n - r(\tilde{X}))\log(1+\varphi W) \\
& -(N - r(X_*))\log(W) \geq K'|\sigma_w^2 = 0, W > 1\Big) = \frac{\alpha}{1-p_r} \\
\iff\quad & P\Big(\Upsilon + (n - r(\tilde{X}))\log(1+\varphi W) \\
& -(N - r(X_*))\log(W) \geq K'|\sigma_w^2 = 0, W > 1\Big) = \frac{\alpha}{1-p_r} \\
\iff\quad & K' = O'^{-1}(1 - \frac{\alpha}{1-p_r}),
\end{aligned} \tag{3.47}
$$

where O' is the CDF of the transformed random variable $\Upsilon+(n-r(\tilde{X}))\log(1+\varphi W)-(N-r(X_*))\log(W)$ for $W \sim F(N-r(X_*), n-r(X))$ when $W > 1$ and $\sigma_w^2 = 0$. For example, if we let $N-r(X_*) = 3$, $n-r(X) = 9$, and $n-r(\tilde{X}) = 12$ then, for $\alpha = 0.05$ under H_0, $p_r = 0.56371$ and

$$
\Lambda^r \sim \begin{cases} 0 & \text{w.p} \quad 0.56371 \\ 3\log\left[\frac{81}{256}\frac{\left(1+\frac{W}{3}\right)^4}{W}\right] & \text{w.p} \quad 0.43629 \end{cases} \tag{3.48}
$$

so that K' is found, by numerical simulation or numerical integration after transformation, as $K' = O'^{-1}(1 - \frac{\alpha}{1-p_r}) = O^{-1}(0.8853973) = 2.436$.

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

$$
\begin{aligned}
& \Xi_{RLRT} = P\left(\Lambda^r \geq K'|H_a \text{ is true }, W > \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\
\iff\quad & \Xi_{RLRT} = P\left(\Lambda^r \geq K'|\sigma_w^2 > 0, W > \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right) \\
\iff\quad & \Xi_{RLRT} = (1-p_r)P\Big(\Upsilon + (n - r(\tilde{X}))\log(1+\varphi W) \\
& -(N - r(X_*))\log(W) \geq K'|\sigma_w^2 > 0, W > \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\Big) \\
\iff\quad & \Xi_{RLRT} = (1-p_r)\Big[1 - P\Big(\Upsilon + (n - r(\tilde{X}))\log(1+\varphi W) \\
& -(N - r(X_*))\log(W) < K'|\sigma_w^2 > 0, W > \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\Big)\Big] \\
\iff\quad & \Xi_{RLRT} = (1-p_r)\left[1 - O"(K')\right]
\end{aligned} \tag{3.49}
$$

where $O"$ is the CDF of the transformed random variable $\Upsilon+(n-r(\tilde{X}))\log(1+\varphi W)-$

$(N - r(X_*)) \log(W)$ for $W \sim F\left(N - r(X_*), n - r(X)\right)$ when $W > \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}$ and $\sigma_w^2 > 0$. For example, if we let $N - r(X_*) = 3$, $n - r(X) = 9$, $n - r(\tilde{X}) = 12$, $m = 4$, $\sigma_s^2 = 3$ and $\sigma_w^2 = 7$ then, for $\alpha = 0.05$ under H_1, $p_1 = 0.04014$ and

$$\Lambda^r \sim \begin{cases} 0 & \text{w.p} \quad 0.04014 \\ 3 \log \left[\frac{243}{7936} \frac{\left(1 + \frac{31W}{9}\right)^4}{W} \right] & \text{w.p} \quad 0.95986 \end{cases} \tag{3.50}$$

so that Ξ_{RLRT} is found, by numerical simulation or numerical integration after transformation, as $\Xi_{RLRT} = (1 - p_r)\left[1 - O\text{''}(K')\right] = 0.95986\left[1 - O\text{''}(2.436)\right] = 0.7741$. Note that the power of the RLRT, for this example, is the same as that of the LRT and equals to 0.7741; the power of the F-test. However, it's not always true that the RLRT and LRT have the very same power. The exception in here is due to the fact that for this particular example $\alpha \leq 1 - p_r = 1 - p_m$ so the three tests are equivalent and hence have the same power.

Note that both test statistics Λ^r and F are nonnegative and whenever the $\Lambda^r \neq 0$ there is a strict monotonic relationship and thus when the RLRT critical region does not include 0, the tests are the same. In fact, in the case when $\alpha \leq 1 - p_r$, the critical region will consist of positive values where Λ^r is a strictly increasing function of the F, thus we have

Proposition 3.4.1 *Let α be the size of the test. If $\alpha \leq 1 - p_r$ where $p_r = P(\Lambda^r = 0|\sigma_w^2 = 0)$ then the F-test and RLRT are equivalent and hence have the same power.*

Proof of Proposition (3.4.1): Since the LRT can be written as

$$\Lambda^r \sim \begin{cases} 0 & \text{w.p} \quad p_r \\ g(F) & \text{w.p} \quad 1 - p_r \end{cases} \tag{3.51}$$

where $g(.)$ is a strictly increasing function, then the critical region of the RLRT when

$\alpha \leq 1 - p_r$ doesn't involve 0 and hence the power can be calculated as

$$\Xi_{RLRT} = P\left(\Lambda^r \geq K' | H_a \text{ is true }\right) \iff \Xi_{RLRT} = P\Big(g(F) \geq K' | \sigma_w^2 > 0\Big)$$
$$\iff \quad \Xi_{RLRT} = P\Big(F \geq C" | \sigma_w^2 > 0\Big) = \Xi_F. \tag{3.52}$$

$\square$

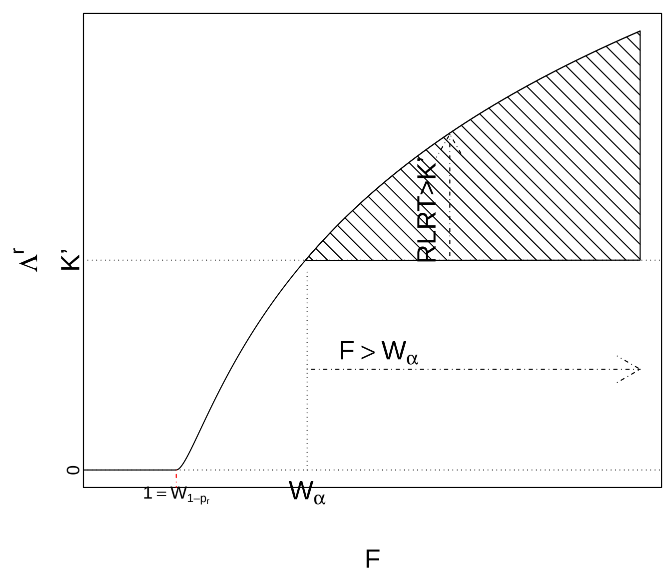

Figure 3.1: A plot of the *RLRT* statistic versus the F-ratio showing their equivalence whenever $\alpha \leq 1 - p_r$. $W_{1-p_r} = 1$ is the minimal critical value at which the two tests are equivalent.

Figure 3.1 illustrates the equivalence of the F-test and RLRT whenever $\alpha \leq 1 - p_r$ where $p_r = P(\Lambda^r = 0 | \sigma_w^2 = 0)$. Further, it clarifies why the two tests are equivalent as long as the critical value W_α of the F-test is larger than $W_{1-p_r} = 1$; the minimal critical value at which the two tests are equivalent.

3.4.2 Power Comparison when $\alpha > 1 - p_r$

In the case when $\alpha > 1 - p_r$, the critical region of the RLRT involves $\Lambda^r = 0$ and hence it involves randomization. We show mathematically , for this case, that the power of the F-test is larger than that of the RLRT. Firstly, we rewrite the power of a size α F-test in terms of p_r the probabilty that $\Lambda^r = 0$ under H_0 as follows.

$$\begin{aligned} \Xi_F = P\left(F \geq W_\alpha\right) &= P(F \geq W_{1-p_r}) + P(W_\alpha \leq F \leq W_{1-p_r}) \\ &= P\left(kW \geq W_{1-p_r}\right) + P\left(W_\alpha \leq kW \leq W_{1-p_r}\right) \\ &= P\left(W \geq \frac{1}{k}W_{1-p_r}\right) + P\left(W \leq \frac{1}{k}W_{1-p_r}\right) - P\left(W \leq \frac{1}{k}W_\alpha\right). \end{aligned} \tag{3.53}$$

Note that the second equality in (3.53) is due to the probabilistic identity $P(E) + P(E^c) = 1$. Secondly, we rewrite the randomized test for the RLRT in terms of the F-test according to the monotonic relationship between them and the smallest critical value, $W_{1-p_r} = 1$, where the F and RLRT tests are equivalent as follows.

$$\phi(\Lambda^r) = \begin{cases} 1 & \text{if} \quad \Lambda^r > 0 \\ \gamma & \text{if} \quad \Lambda^r = 0 \\ 0 & \text{if} \quad \Lambda^r < 0 \end{cases} \iff \phi(F) = \begin{cases} 1 & \text{if} \quad F > W_{1-p_r} \\ \gamma & \text{if} \quad F \leq W_{1-p_r} \end{cases} \tag{3.54}$$

where γ is determined according to the size of the test as

$$\begin{aligned} \alpha = E_{H_0}\phi(\Lambda^r) &\iff \alpha = P(\Lambda^r > 0|\sigma_w^2 = 0) + \gamma P(\Lambda^r = 0|\sigma_w^2 = 0) \\ \iff & \quad \alpha = (1 - p_r) + \gamma p_r \\ \iff & \quad \gamma = \frac{\alpha - (1 - p_r)}{p_r}. \end{aligned} \tag{3.55}$$

Hence, the power of the RLRT is

$$\begin{aligned} \Xi_{RLRT} &= P\left(F \geq W_{1-p_r}\right) + \frac{\alpha - (1 - p_r)}{p_r} P\left(F \leq W_{1-p_r}\right) \\ &= P\left(W \geq \frac{1}{k}W_{1-p_r}\right) + \frac{\alpha - (1 - p_r)}{p_r} P\left(W \leq \frac{1}{k}W_{1-p_r}\right). \end{aligned} \tag{3.56}$$

Proposition 3.4.2 *Let* $p_r = P(\Lambda^r = 0|\sigma_w^2 = 0)$*. For GSP models with a finite whole plots size* m*, if* $\alpha > 1 - p_r$ *then the power of the size* α *F-test is larger than that of the RLRT in testing* $\sigma_w^2 = 0$*.*

Proof of Proposition 3.4.2: It's sufficient to show that

$$\begin{aligned}
&P\left(W \leq \frac{1}{k}W_{1-p_r}\right) - P\left(W \leq \frac{1}{k}W_\alpha\right) > \frac{\alpha - (1-p_r)}{p_r} P\left(W \leq \frac{1}{k}W_{1-p_r}\right) \\
\Longleftrightarrow\ & P\left(W \leq \frac{1}{k}W_{1-p_r}\right)\left[1 - \frac{\alpha - (1-p_r)}{p_r}\right] > P\left(W \leq \frac{1}{k}W_\alpha\right) \\
\Longleftrightarrow\ & (1-\alpha)P\left(W \leq \frac{1}{k}W_{1-p_r}\right) > p_r P\left(W \leq \frac{1}{k}W_\alpha\right) \\
\Longleftrightarrow\ & \frac{1}{p_r}P\left(W \leq \frac{1}{k}W_{1-p_r}\right) > \frac{1}{1-\alpha}P\left(W \leq \frac{1}{k}W_\alpha\right), \qquad (3.57)
\end{aligned}$$

which is true, since $k = \frac{\sigma_s^2 + m\sigma_w^2}{\sigma_s^2} > 1$, according to the F-Inequality in Chapter 5. □

3.4.3 Is $\alpha > 1 - p_r$ Practical?

The RLRT and F-test are equivalent as long as the level of the test is smaller or equal to $P(W > 1)$ where $W \sim F_{(N-r(X_*), n-r(X))}$. That is, the two tests are equivalent for all α's satisfying the inequality $\alpha \leq P\,(W > 1)$. Table 3.2 presents the maximal values of α satisfying this inequality for different combinations of the degrees of freedom $df_1 = N - r(X_*)$ and $df_2 = n - r(X)$ for any m. Table 3.2 shows that for a commonly used α values the case $\alpha > 1 - p_r$ is not practical. This implies that the F-test would have a larger power than the RLRT only when α is larger the 0.30; a situation which never occurs in practice. So, for practical purposes, the F-test and RLRT are equivalent. We observed from simulation, and hence give a mathematical proof, that as m increases the power of the RLRT approaches that of the F-test.

Proposition 3.4.3 *For GSP models, if* $\alpha > 1 - p_r$ *then for a size* α *test* $\Xi_{RLRT} \uparrow \Xi_F$ *in testing* $\sigma_w^2 = 0$ *as the whole plots size* m *approaches infinity.*

Proof of Proposition 3.4.3: Recall that

$$\Xi_F = P\left(W \geq \frac{1}{k}W_{1-p_r}\right) + P\left(W \leq \frac{1}{k}W_{1-p_r}\right) - P\left(W \leq \frac{1}{k}W_{\alpha}\right),$$

and

$$\Xi_{RLRT} = P\left(W \geq \frac{1}{k}W_{1-p_r}\right) + \frac{\alpha - (1-p_r)}{p_r}P\left(W \leq \frac{1}{k}W_{1-p_r}\right).$$

From Proposition 3.4.2, we have established for a finite whole plot size m

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

If we let $m \uparrow \infty$ then $k \uparrow \infty$ so that $P\left(W \leq \frac{1}{k}W_{1-p_r}\right) = P\left(W \leq \frac{1}{k}W_{\alpha}\right) = P\left(W \leq \frac{1}{k}W_{1-p_r}\right) = 0$ and thus the inequality becomes equality and as a result $\Xi_{RLRT} \uparrow \Xi_F$. In fact for $m = \infty$ we have $\Xi_{RLRT} = \Xi_F = 1$ since $\lim_{k \to +\infty} P\left(W \geq \frac{1}{k}W_{1-p_r}\right) = 1$. □

3.5 Simultaneous Comparisons

Now, we consider the comparisons between the three tests simultaneously. Based on the previous section, we have 6 possible cases as shown in Table 3.1.

Power Comparisons	
$\alpha \leq 1 - p_m$ and $\alpha \leq 1 - p_r$	$\Xi_F = \Xi_{LRT} = \Xi_{RLRT}$
$\alpha \leq 1 - p_m$ and $\alpha > 1 - p_r$	$\Xi_F = \Xi_{LRT} > \Xi_{RLRT}$
$\alpha > 1 - p_m$ and $\alpha \leq 1 - p_r$	$\Xi_F = \Xi_{RLRT} > \Xi_{LRT}$
$\alpha > 1 - p_m = 1 - p_r$	$\Xi_F > \Xi_{LRT} = \Xi_{RLRT}$
$\alpha > 1 - p_m > 1 - p_r$	$\Xi_F > \Xi_{LRT} > \Xi_{RLRT}$
$\alpha > 1 - p_r > 1 - p_m$	$\Xi_F > \Xi_{RLRT} > \Xi_{LRT}$

Table 3.1: All possible power comparisons between the F-test, LRT and RLRT.

$df_2 \backslash df_1$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**15**	**20**	**30**	**40**	**60**	**120**
1	0.50	0.58	0.61	0.63	0.64	0.64	0.65	0.65	0.66	0.66	0.67	0.67	0.67	0.68	0.68	0.68
2	0.42	0.50	0.54	0.56	0.57	0.58	0.59	0.59	0.59	0.60	0.61	0.61	0.62	0.62	0.63	0.63
3	0.39	0.46	0.50	0.52	0.54	0.55	0.55	0.56	0.56	0.57	0.58	0.59	0.59	0.60	0.60	0.60
4	0.37	0.44	0.48	0.50	0.51	0.52	0.53	0.54	0.54	0.55	0.56	0.57	0.58	0.58	0.59	0.59
5	0.36	0.43	0.46	0.49	0.50	0.51	0.52	0.53	0.53	0.53	0.55	0.56	0.57	0.57	0.57	0.58
6	0.36	0.42	0.45	0.48	0.49	0.50	0.51	0.51	0.52	0.52	0.54	0.55	0.56	0.56	0.57	0.57
7	0.35	0.41	0.45	0.47	0.48	0.49	0.50	0.51	0.51	0.52	0.53	0.54	0.55	0.55	0.56	0.57
8	0.35	0.41	0.44	0.46	0.47	0.49	0.49	0.50	0.51	0.51	0.53	0.53	0.54	0.55	0.55	0.56
9	0.34	0.41	0.44	0.46	0.47	0.48	0.49	0.49	0.50	0.50	0.52	0.53	0.54	0.54	0.55	0.56
10	0.34	0.40	0.43	0.45	0.47	0.48	0.48	0.49	0.50	0.50	0.52	0.52	0.53	0.54	0.55	0.55
15	0.33	0.39	0.42	0.44	0.45	0.46	0.47	0.47	0.48	0.48	0.50	0.51	0.52	0.53	0.53	0.54
20	0.33	0.39	0.41	0.43	0.44	0.45	0.46	0.47	0.47	0.48	0.49	0.50	0.51	0.52	0.52	0.53
30	0.33	0.38	0.41	0.42	0.43	0.44	0.45	0.46	0.46	0.47	0.48	0.49	0.50	0.51	0.51	0.52
40	0.32	0.38	0.40	0.42	0.43	0.44	0.45	0.45	0.46	0.46	0.47	0.48	0.49	0.50	0.51	0.52
60	0.32	0.37	0.40	0.41	0.43	0.43	0.44	0.45	0.45	0.45	0.47	0.48	0.49	0.49	0.50	0.51
120	0.32	0.37	0.40	0.41	0.42	0.43	0.43	0.44	0.44	0.45	0.46	0.47	0.48	0.48	0.49	0.50

Table 3.2: The maximal values of α satisfying the inequality $\alpha \leq P(W > 1)$ for different combinations of the degrees of freedom $df_1 = N - r(X_*)$ and $df_2 = n - r(X)$ for any m.

Chapter 4

Illustration Via Simulation

4.1 Illustration

Consider the following split-plot model with a completely randomized design (CRD) for whole plot:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \phi_{k(i)} + \epsilon_{ijk}, \tag{4.1}$$

where $i = 1, \ldots, a$, $j = 1, \ldots, m$, $k = 1, \ldots, c$, $\phi_{k(i)} \sim N(0, \sigma_w^2)$ and $\epsilon_{ijk} \sim N(0, \sigma_s^2)$. In particular, y_{ijk} denotes observation k in level i of factor A and level j of factor B, μ denotes the overall mean, α_i denotes the effect of level i of factor A, β_j denotes the effect of level j of factor B, $(\alpha\beta)_{ij}$ the effect of the ij^{th} interaction of $A \times B$, $\phi_{k(i)}$ denotes the whole plot error and ϵ_{ijk} denotes the split-plot error. The ANOVA table for model (4.1) is presented in Table 4.1. Let $\alpha = [\alpha_1, \ldots, \alpha_a]$, $\beta = [\beta_1, \ldots, \beta_m]$, $\alpha\beta = [\alpha\beta_{(11)}, \ldots, \alpha\beta_{(am)}]$ and $\phi = [\phi_{1(1)}, \ldots, \phi_{c(1)}, \ldots, \phi_{c(a)}]$. Then, in view of (1.6), model (4.1) could be written as

$$Y = X_*[\mu : \alpha]' + X_2[\beta : \alpha\beta]' + (X_1[\phi]' + \epsilon)\,, \tag{4.2}$$

	Df	Sum Sq	F value
Whole plot factor: A	$a-1$	$SSE(A)$	$MS(A)/MS(w)$
Whole plot error: plot(A)	$a(c-1)$	$SSE(w)$	$MS(w)/MS(s)$
Subplot factor: B	$m-1$	$SSE(B)$	$MS(B)/MS(s)$
A× B	$(a-1)(m-1)$	$SSE(AB)$	$MS(AB)/MS(s)$
Subplot error: B× plot(A)	$a(m-1)(c-1)$	$SSE(s)$	

Table 4.1: ANOVA table for for model (4.1).

where $n = amc$, $N = ac$, $\delta = [\mu : \alpha]'$, $\gamma = [\beta : \alpha\beta]'$ and $\eta = [\phi]'$ such that $\delta_{(a+1)\times 1}$, $\gamma_{(m+am)\times 1}$, $\eta_{N\times 1}$, $X_{*n\times(a+1)}$, $X_{2n\times(m+am)}$, $X_{1n\times N}$ with

$$X_* = [J_a \otimes J_m \otimes J_c : I_a \otimes J_m \otimes J_c], \tag{4.3}$$

$$X_2 = [J_c \otimes I_m \otimes J_a : I_a \otimes J_c \otimes I_m], \text{ and} \tag{4.4}$$

$$X_1 = [I_a \otimes I_c \otimes J_m]. \tag{4.5}$$

Thus, $r(X) = r([X_1, X_2]) = ac + am - a$, $r(X_*) = a$ and $r(X_1) = ac$ so that

$$\frac{N - r(X_*)}{n - r(X)} = \frac{ac - a}{amc - (ac + am - a)} = \frac{1}{m-1}, \tag{4.6}$$

and subsequently $\kappa = 1$ and a in (2.43) is

$$a = \frac{\sigma_s^2 + m\sigma_w^2}{(m-1)\sigma_s^2}, \tag{4.7}$$

τ in (2.44) is

$$\tau = N \log\left[\left(\frac{m-1}{m}\right)^m \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right], \tag{4.8}$$

W in Lemma 2.3.2 is

$$W \sim F\left(a(c-1), a(m-1)(c-1)\right), \tag{4.9}$$

and p_m in (2.42) is

$$p_m = P\left(F\left(a(c-1), a(m-1)(c-1)\right) \leq \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\right). \tag{4.10}$$

Note that the degrees of freedom of the F distribution in (4.10) could be obtained directly from the ANOVA table (see Table 4.1) when testing for the whole plot error (i.e. $\sigma_w^2 = 0$). Thus, according to proposition 2.2.1, the LRT statistic Λ as a function of F is expressed in the following monotone relation

$$\Lambda \quad \sim \quad \begin{cases} 0 \\ n\log\left(\frac{m-1}{m}\right) + n\log\left(1 + \frac{F}{m-1}\right) + N\log\left(\frac{1}{F}\right) \end{cases} \tag{4.11}$$

such that the 0 case accurs when $\hat{\sigma}_w^2 = 0$ and $\hat{\sigma}_s^2 = \frac{SSE(s)+SSE(w)}{n}$ and the > 0 case occurs when $\hat{\sigma}_w^2 = \frac{1}{m}\left[\frac{SSE(w)}{N} - \frac{SSE(s)}{n-N}\right]$ and $\hat{\sigma}_s^2 = \frac{SSE(s)}{n-N}$. By lemma 2.3.2 Λ has the distribution

$$\Lambda \sim \begin{cases} 0 & \text{w.p.} \quad p_m \\ W_* & \text{w.p.} \quad 1 - p_m \end{cases} \tag{4.12}$$

where p_m is computed in (4.10) and W_* is the random variable

$$W_* = N\log\left(\left(\frac{m-1}{m}\right)^m \frac{\sigma_s^2}{\sigma_s^2 + m\sigma_w^2}\left[\frac{(1+aW)^m}{W}\right]\right). \tag{4.13}$$

To demonstrate the relation in (4.11), under H_1, we conduct a simulation. The results of 1,000 runs of a monte carlo simulation are presented in Figure 4.1. Further, this numerical simulation leads to the exact distribution of Λ, under H_1, as shown in Figure 4.2.

For the simulated examples in Figures 4.1 and 4.2 the empirical value of p_m, under H_1, was found to be $p_{1imp} = 0.247$ and $p_{2imp} = 0.045$ respectively for the cases ($\sigma_w^2 = 3, \sigma_s^2 = 7$) and ($\sigma_w^2 = 7, \sigma_s^2 = 3$). These numbers are very similar to the theoretical value of p_m as $p_1 = P(F_{3,9} \leq \frac{7}{7+4\times 3}) = 0.22$ and $p_2 = P(F_{3,9} \leq \frac{3}{3+4\times 7}) = 0.04$.

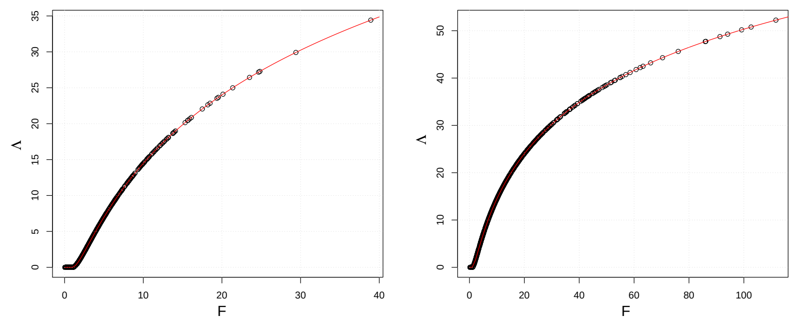

Figure 4.1: Two cases of relation of Λ and F with $1,000$ runs under H_1. Left panel: samples are from a split-plot design with $a = 3$, $m = 4$, $c = 2$, $\sigma_w^2 = 3$ and $\sigma_s^2 = 7$; Right panel: samples are from a split-plot design with $a = 3$, $m = 4$, $c = 2$, $\sigma_w^2 = 7$ and $\sigma_s^2 = 3$. Solid line on both graphs indicates the relation found in (4.11).

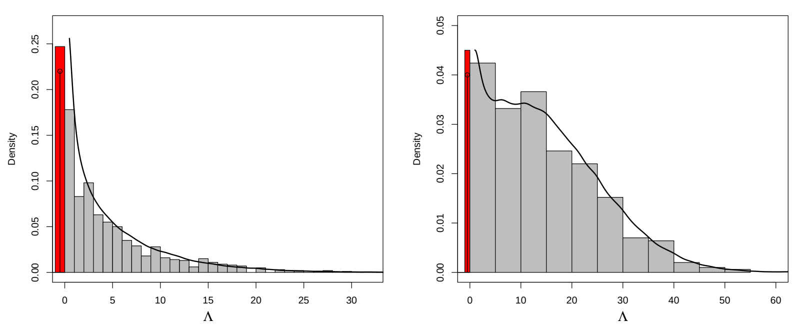

Figure 4.2: Two cases of the Empirical versus Theoretical mixed density function of Λ under H_1. Left panel: the empirical density is obtained from a split-plot design sample with $a = 3$, $m = 4$, $c = 2$, $\sigma_w^2 = 3$ and $\sigma_s^2 = 7$. Right panel: the empirical density is obtained from a split-plot design sample with $a = 3$, $m = 4$, $c = 2$, $\sigma_w^2 = 7$ and $\sigma_s^2 = 3$. The theoretical density in both panels, drawn in solid line and point mass, is obtained according to (4.12).

For instance, the model probability distribution of Λ for the two simulated examples under the full model are as follows.

For $\sigma_w^2 = 3$ and $\sigma_s^2 = 7$, the model probability distribution of Λ is

$$\Lambda \sim \begin{cases} 0 & \text{with} \quad p_m = 0.22 \\ 6\log\left[\left(\frac{567}{4864}\right)\frac{(1+\frac{19}{21}W)^4}{W}\right] & \text{with} \quad p_m = 0.78 \end{cases} \tag{4.14}$$

and for $(\sigma_w^2 = 7, \sigma_s^2 = 3)$, the model probability distribution of Λ is

$$\Lambda \sim \begin{cases} 0 & \text{with.} \quad p_m = 0.04 \\ 6\log\left[\left(\frac{243}{7936}\right)\frac{(1+\frac{31}{9}W)^4}{W}\right] & \text{with} \quad p_m = 0.96 \end{cases} \tag{4.15}$$

where $W \sim F(3,9)$.

The results in (4.14) and (4.15) should not be surprising since $\Lambda = 0$ is most likely to happen when σ_w^2 is much less than σ_s^2. This observation is due to the fact that $\frac{MSE(w)}{MSE(s)}$ estimates $\frac{\sigma_s^2+m\sigma_w^2}{\sigma_s^2}$ as was shown in Lemma (3.3.2).

This conducted simulation is important when one is interested in the power of the test. However, to conduct the test through the Neymen-Pearson approach, we only need to know the distribution of the LRT statistic under the null hypothesis. Knowing the sampling distribution of the test-statistic under H_0 helps us compute the p-value that tells us how much evidence we have against the null model. Thus, to demonstrate the relation in (4.11), under the null, we conduct a simulation. The results of 1,000 runs of a monte carlo simulation are presented in Figure 4.3. The model probability distribution of Λ under the reduced model for the example in Figure (4.3) is

$$\Lambda \sim \begin{cases} 0 & \text{with} \quad p_m = 0.56 \\ 6\log\left[\left(\frac{81}{256}\right)\frac{(1+\frac{1}{3}W)^4}{W}\right] & \text{with} \quad p_m = 0.44 \end{cases} \tag{4.16}$$

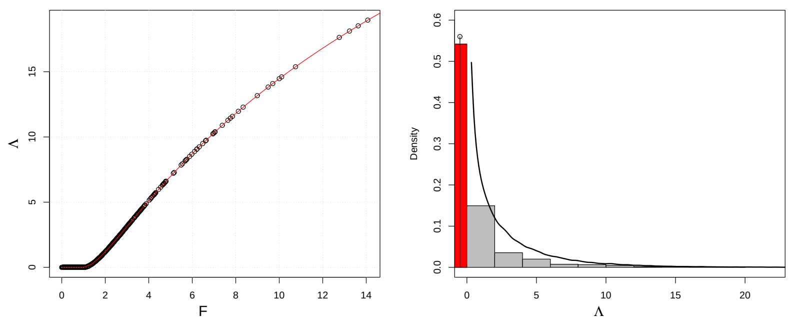

Figure 4.3: Left panel: relation of LRT and F with $1,000$ runs under the reduced model. Right panel: the Empirical versus Theoretical mixed density function of LRT under H_0. The samples are from a split-plot design with $a = 3$, $m = 4$, $c = 2$, $\sigma_w^2 = 0$ and $\sigma_s^2 = 3$. Solid line on left panel indicates the relation found in (4.11).The theoretical density in right panel is drawn in solid line and point mass according to (4.12).

For this example, the empirical value of p_m, under H_0, was found to be $p_{imp} = 0.542$. This number is very similar to the theoretical value of p_m as $p = P(F_{3,9} \leq 1) = 0.56$.

Chapter 5

New Stochastic Inequalities

5.1 Background and Motivation

We derive a new inequality involving either the F or Gamma distribution and their quantiles and call it either the F-Inequality or G-Inequality. The stochastic representation of the new inequality involves $\alpha, p \in (0,1)$ such that if $p > \alpha$ and $k > 1$ with W being a random variable with an $F(\nu_1, \nu_2)$ or $Gamma(\tau, \theta)$ distribution then it's always true that

$$\frac{1}{p} P\left(W < \frac{W_p}{k}\right) > \frac{1}{\alpha} P\left(W < \frac{W_\alpha}{k}\right),$$

where for any γ between 0 and 1, W_γ is defined by $\gamma = P(W < W_\gamma)$. The inequality changes direction for $k \in [0,1)$ and becomes equality for $k = 1$ and, trivially, for $k = \infty$. This inequality seems to hold for a larger class of distributions that include for example the F, Gamma, Cauchy, and special cases of the Beta distribution and perhaps others. However, we provide rigorous proofs only for the F and Gamma distribution.

One of theses inequalities, the F-inequality, came up while proving that for the generalized split-plot model the F-test for the whole plot variance being zero has larger power than either the likelihhod ratio test (LRT) or the restricted likelihood ratio test (RLRT) when the level of the test, α, is larger than one minus the probability, p, that the LRT or the RLRT statistic is zero(see Sections 2.4.3 and 3.4.2 for more details). Beside their use in power comparisons for testing variance parameters on the boundaries of the parameter space, these inequalities promise value in developing stochastic lower and upper bounds in many fields. Moreover, strikingly, they present a solid example demonstrating the case when the unjustified inequality $a/c > b/d$ for $a > b > 0$ and $c > d > 0$ holds, see Cloud and Drachman (1998) for more details about dividing inequalities. The next section presents two lemmas needed for the proofs. Section 5.3 gives the F-inequality and Section 5.4 establishes the G-inequality.

5.2 Hypergeometric Functions

We now present two lemma related to special hypergeometric functions that we need to prove the inequalities.

Definition 5.2.1 *For real numbers α, β and γ with $\gamma \neq 0, -1, -2, \ldots,$ the Gauss Hypergeometric Function, commonly written ${}_2F_1(\alpha, \beta, \gamma, x)$ but that for clarity we write here as $H(\alpha, \beta, \gamma, x)$, is defined as*

$$\begin{aligned} H(\alpha,\beta,\gamma,x) &= 1 + \frac{\alpha.\beta}{\gamma}\frac{x}{1!} + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)}\frac{x^2}{2!} \\ &\quad + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{\gamma(\gamma+1)(\gamma+2)}\frac{x^3}{3!} + \ldots, \end{aligned} \tag{5.1}$$

for $|x| < 1$ (Miller and Mocanu, 1990).

Lemma 5.2.2 *Let $k > 0$, F_p be the pth quantile of the $F(\nu_1, \nu_2)$ distribution such that $p = P(F < F_p)$ and define the function $s(p|\nu_1, \nu_2, k)$ in terms of the Gauss*

Hypergeometric Function as

$$s(p|\nu_1,\nu_2,k) = H\left(\frac{\nu_1+\nu_2}{2}, 1, \frac{\nu_1}{2}+1, \tilde{F}_{p,1}\right) - H\left(\frac{\nu_1+\nu_2}{2}, 1, \frac{\nu_1}{2}+1, \tilde{F}_{p,k}\right),$$

where

$$\tilde{F}_{p,k} \equiv \frac{F_p}{F_p + \frac{k\nu_2}{\nu_1}} \tag{5.2}$$

then

Case I: if $k \in (0,1)$ *then* $s(p|\nu_1,\nu_2,k) < 0$*.*

Case II: if $k = 1$ *then* $s(p|\nu_1,\nu_2,k) = 0$*.*

Case III: if $k > 1$ *then* $s(p|\nu_1,\nu_2,k) > 0$*.*

Proof of Lemma 5.2.2: By Definition 5.2.1, $H(\alpha,\beta,\gamma,x)$ is increasing in x for positive α,β and γ. In our case, $\alpha := \frac{\nu_1+\nu_2}{2} > 0$, $\beta := 1 > 0$ and $\gamma := \frac{\nu_1}{2}+1 > 0$. Thus

Case I: if $k \in (0,1)$ then $\tilde{F}_{p,1} < \tilde{F}_{p,k}$ and subsequently

$$s(p|\nu_1,\nu_2,k) = H\left(\alpha,\beta,\gamma,\tilde{F}_{p,1}\right) - H\left(\alpha,\beta,\gamma,\tilde{F}_{p,k}\right) < 0.$$

Case II: if $k = 1$ then $\tilde{F}_{p,1} = \tilde{F}_{p,k}$ and subsequently

$$s(p|\nu_1,\nu_2,k) = H\left(\alpha,\beta,\gamma,\tilde{F}_{p,1}\right) - H\left(\alpha,\beta,\gamma,\tilde{F}_{p,k}\right) = 0.$$

Case III: if $k > 1$ then $\tilde{F}_{p,1} > \tilde{F}_{p,k}$ and subsequently

$$s(p|\nu_1,\nu_2,k) = H\left(\alpha,\beta,\gamma,\tilde{F}_{p,1}\right) - H\left(\alpha,\beta,\gamma,\tilde{F}_{p,k}\right) > 0.$$

$\square$

Definition 5.2.3 *For real numbers α and β with $\beta \neq 0, -1, -2, \ldots,$ the Kummer Confluent Hypergeometric Function, commonly written ${}_1F_1(\alpha, \beta, x)$, but that for clarity we write here as $M(\alpha, \beta, x)$, is defined as*

$$\begin{aligned} M(\alpha, \beta, x) &= 1 + \frac{\alpha}{\beta}\frac{x}{1!} + \frac{\alpha(\alpha+1)}{\beta(\beta+1)}\frac{x^2}{2!} \\ &\quad + \frac{\alpha(\alpha+1)(\alpha+2)}{\beta(\beta+1)(\beta+2)}\frac{x^3}{3!} + \ldots, \end{aligned}$$

for $-\infty < x < \infty$ (Miller and Mocanu, 1990).

Lemma 5.2.4 *Let $k > 0$, G_p be the pth quantile of the Gamma distribution with positive shape parameter τ and scale parameter θ such that $p = P(G < G_p)$ and define the function $S(p|\tau, \theta, k)$ in terms of the Kummer's confluent hypergeometric function such that*

$$S(p|\tau, \theta, k) = M\left(1, \tau+1, \frac{G_p}{\theta}\right) - M\left(1, \tau+1, \frac{G_p}{k\theta}\right),$$

then

Case I: if $k \in [0, 1)$ then $S(p|\tau, \theta, k) < 0$.

Case II: if $k = 1$ then $S(p|\tau, \theta, k) = 0$.

Case III: if $k > 1$ then $S(p|\tau, \theta, k) > 0$.

Proof of Lemma 5.2.4: By Definition 5.2.3, $M(\alpha, \beta, x)$ is increasing in x for positive α and β. In our case, $\alpha := 1 > 0$ and $\beta := \tau + 1 > 0$. Thus

Case I: if $k \in (0, 1)$ then $G_p/\theta < G_p/k\theta$ so that $M(1, \tau+1, G_p) - M\left(1, \tau+1, \frac{G_p}{k}\right) < 0$ and hence $S(p|\tau, \theta, k) < 0$.

Case II: if $k = 1$ then $G_p/\theta = G_p/k\theta$ so that $M(1, \tau+1, G_p) - M\left(1, \tau+1, \frac{G_p}{k}\right) = 0$ and hence $S(p|\tau, \theta, k) = 0$.

Case III: if $k > 1$ then $G_p/\theta > G_p/k\theta$ so that $M(1, \tau+1, G_p) - M\left(1, \tau+1, \frac{G_p}{k}\right) > 0$ and hence $S(p|\tau, \theta, k) > 0$. $\square$

5.3 The F-Inequality

The F-inequality exploits the well-known relationship between the F distribution and the Beta distribution, namely that if $F \sim F(\nu_1, \nu_2)$ then $\frac{F}{F+\frac{\nu_2}{\nu_1}} \sim Beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)$.

Definition 5.3.1 *For $a, b > 0$, the regularized incomplete beta function $I_x(a,b)$ is defined in terms of the incomplete beta function $B(x;a,b)$ and the complete beta function $B(a,b)$ as*

$$I_x(a,b) = \frac{B(x;a,b)}{B(a,b)},$$

where

$$B(x;a,b) = \int_0^x t^{a-1}(1-t)^{b-1}dt$$

for $x \in [0,1]$ and

$$B(a,b) := B(1;a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt,$$

see O'Connor (2011).

Note that $I_x(a,b)$ is the cumulative distribution function (CDF) of a $Beta(a,b)$ distribution.

Proposition 5.3.2 *Let F be a random variable with an $F(\nu_1, \nu_2)$ distribution. Define the FR-function as $h(p|k) = \frac{1}{p}B\left(\tilde{F}_{p,k}, \frac{\nu_1}{2}, \frac{\nu_2}{2}\right)$ where $p \in (0,1)$. Then, $h(p|k)$ is a monotonic function for $k \geq 0$ such that:*

Case I: if $0 \leq k < 1$ then $h(p|k)$ is strictly decreasing in p.

Case II: if $k = 1$ then $h(p|k) = B(\nu_1/2, \nu_2/2)$ (constant).

Case III: if $k > 1$ then $h(p|k)$ is strictly increasing in p.

Proof of Proposition 5.3.2: The idea is to establish that $\frac{\partial}{\partial p}h(p|k) = 0$ if and only if $k = 1$. Since $\frac{\partial}{\partial p}h(p|k)$ is continuous in k, this means that for $k \in [0, 1)$, $h(p|k)$ is monotonic and for $k > 1$, $h(p|k)$ is also monotonic. Moreover, for $k \in [0, 1)$, we show that $h(p|k)$ is monotonically decreasing in p and for $k > 1$, that $h(p|k)$ is monotonically increasing in p.

Case II: The CDF of the F-distribution with ν_1 and ν_2 degrees of freedom is

$$F(x; \nu_1, \nu_2) := I_{\frac{x}{x+\frac{\nu_2}{\nu_1}}}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) = I_{\frac{\nu_1 x}{\nu_1 x + \nu_2}}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right),$$

where I is the Beta CDF. Thus, incorporating equation (5.2),

$$\begin{aligned} p = P\left(F \leq F_p\right) &= \frac{B\left(\frac{F_p}{F_p+\frac{\nu_2}{\nu_1}}, \frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}{B(\nu_1/2, \nu_2/2)} \\ \iff \quad \frac{1}{p}B\left(\tilde{F}_{p,1}, \frac{\nu_1}{2}, \frac{\nu_2}{2}\right) &= B(\nu_1/2, \nu_2/2). \end{aligned} \tag{5.3}$$

The left hand side of this equality is precisely $h(p|k = 1)$ which completes the proof for this case.

The result of Case II tells us that $\frac{\partial}{\partial p}h(p|k = 1) = 0$ for all p. If we let f denote the density function of the F random variable we have $\frac{\partial}{\partial p}F_p = \frac{1}{f(F_p)}$ so that

$$\frac{\partial}{\partial p}h(p|k) = \frac{-1}{p^2}B\left(\tilde{F}_{p,k}, \frac{\nu_1}{2}, \frac{\nu_2}{2}\right) + \frac{1}{p}\left(\tilde{F}_{p,k}\right)^{\frac{\nu_1}{2}-1}\left(1 - \tilde{F}_{p,k}\right)^{\frac{\nu_2}{2}-1}\frac{\nu_2 k}{\nu_1 f(F_p)\left(F_p + \frac{\nu_2 k}{\nu_1}\right)^2}, \tag{5.4}$$

and for $k = 1$,

$$\begin{aligned}
0 &= \frac{\partial}{\partial p} h(p|k=1) \\
&= -\frac{1}{p^2} B\left(\tilde{F}_{p,1}, \frac{\nu_1}{2}, \frac{\nu_2}{2}\right) \\
&\quad + \frac{1}{p}\left(\tilde{F}_{p,1}\right)^{\nu_1/2-1} \times \left(1-\tilde{F}_{p,1}\right)^{\nu_2/2-1} \frac{\nu_2}{\nu_1 f(F_p)\left(F_p + \frac{\nu_2}{\nu_1}\right)^2},
\end{aligned}$$

and thus

$$\frac{\nu_2}{\nu_1 f(F_p)} = \frac{B\left(\tilde{F}_{p,1}, \frac{\nu_1}{2}, \frac{\nu_2}{2}\right)\left(F_p + \frac{\nu_2}{\nu_1}\right)^2}{p\left(\tilde{F}_{p,1}\right)^{\nu_1/2-1}\left(1-\tilde{F}_{p,1}\right)^{\nu_2/2-1}}. \tag{5.5}$$

One can verify the equality in (5.5) as an exercise by using (5.3) and the substitution of F_p in the density function of the F distribution.

Before proceeding in the proof of cases I and III, we provide an expression for $\frac{\partial}{\partial p} h(p|k)$ as follows. Letting $a := \frac{\nu_2}{\nu_1}$ and substituting (5.5) in (5.4) gives

$$\begin{aligned}
\frac{\partial}{\partial p} h(p|k) &= \frac{-1}{p^2} B\left(\tilde{F}_{p,k}, \frac{\nu_1}{2}, \frac{\nu_2}{2}\right) \\
&\quad + \frac{k\left(F_p + a\right)^2 \left(\tilde{F}_{p,k}\right)^{\frac{\nu_1}{2}-1}\left(1-\tilde{F}_{p,k}\right)^{\frac{\nu_2}{2}-1} B\left(\tilde{F}_{p,1}, \frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}{p^2\left(F_p + ka\right)^2\left(\tilde{F}_{p,1}\right)^{\frac{\nu_1}{2}-1}\left(1-\tilde{F}_{p,1}\right)^{\frac{\nu_2}{2}-1}} \\
&= \frac{-1}{p^2} B_k + \frac{k}{p^2} \frac{\left(\tilde{F}_{p,k}\right)^{\frac{\nu_1}{2}-1}\left(1-\tilde{F}_{p,k}\right)^{\frac{\nu_2}{2}-1}}{\left(\tilde{F}_{p,1}\right)^{\frac{\nu_1}{2}-1}\left(1-\tilde{F}_{p,1}\right)^{\frac{\nu_2}{2}-1}} \frac{(F_p + a)^2}{(F_p + ka)^2} B_1,
\end{aligned}$$

where

$$B_k \equiv B\left(\tilde{F}_{p,k}, \frac{\nu_1}{2}, \frac{\nu_2}{2}\right).$$

Observe that, from (5.2)

$$\frac{\left(\tilde{F}_{p,k}\right)^{\frac{\nu_1}{2}-1}\left(1-\tilde{F}_{p,k}\right)^{\frac{\nu_2}{2}-1}}{\left(\tilde{F}_{p,1}\right)^{\frac{\nu_1}{2}-1}\left(1-\tilde{F}_{p,1}\right)^{\frac{\nu_2}{2}-1}} = \frac{(F_p + a)^{\frac{\nu_1}{2}+\frac{\nu_2}{2}-2} k^{\frac{\nu_2}{2}-1}}{(F_p + ka)^{\frac{\nu_1}{2}+\frac{\nu_2}{2}-2}}.$$

So,

$$\begin{aligned}
\frac{\partial}{\partial p}h(p|k) &= \frac{-1}{p^2}B_k + \frac{k^{\frac{\nu_2}{2}}}{p^2}\frac{(F_p+a)^{\frac{\nu_1+\nu_2}{2}-2}(F_p+a)^2}{(F_p+ka)^{\frac{\nu_1+\nu_2}{2}-2}(F_p+ka)^2}B_1 \\
&= \frac{-1}{p^2}B_k + \frac{k^{\frac{\nu_2}{2}}}{p^2}\frac{(F_p+a)^{\frac{\nu_1+\nu_2}{2}}}{(F_p+ka)^{\frac{\nu_1+\nu_2}{2}}}B_1 \\
&= \frac{-1}{p^2}B_k + \frac{1}{p^2}k^{\frac{\nu_2}{2}}\left(\frac{F_p+ka}{F_p+a}\right)^{-\frac{\nu_1+\nu_2}{2}}B_1. \qquad (5.6)
\end{aligned}$$

As shown in Dutka (1981), the Hypergeometric representation of the incomplete Beta function is $B(x;a,b) = \frac{x^a(1-x)^b}{a}H(a+b,1;a+1;x)$. Using this representation reduces (5.6) to

$$\begin{aligned}
\frac{\partial}{\partial p}h(p|k) &= \frac{-2}{\nu_1 p^2}\left(\tilde{F}_{p,k}\right)^{\frac{\nu_1}{2}}\left(1-\tilde{F}_{p,k}\right)^{\frac{\nu_2}{2}}H\left(\frac{\nu_1+\nu_2}{2},1,\frac{\nu_1}{2}+1,\tilde{F}_{p,k}\right) \\
&\quad + \frac{2}{\nu_1 p^2}k^{\frac{\nu_2}{2}}\left(\frac{F_p+ka}{F_p+a}\right)^{-\frac{\nu_1+\nu_2}{2}}\left(\tilde{F}_{p,1}\right)^{\frac{\nu_1}{2}}\left(1-\tilde{F}_{p,1}\right)^{\frac{\nu_2}{2}} \\
&\quad \times H\left(\frac{\nu_1+\nu_2}{2},1,\frac{\nu_1}{2}+1,\tilde{F}_{p,1}\right) \\
&= \frac{-2}{\nu_1 p^2}\left(\tilde{F}_{p,1}\right)^{\frac{\nu_1}{2}}\left(1-\tilde{F}_{p,1}\right)^{\frac{\nu_2}{2}}\left[-\left(\frac{\tilde{F}_{p,k}}{\tilde{F}_{p,1}}\right)^{\frac{\nu_1}{2}}\left(\frac{1-\tilde{F}_{p,k}}{1-\tilde{F}_{p,1}}\right)^{\frac{\nu_2}{2}}H_k\right. \\
&\quad \left. + k^{\frac{\nu_2}{2}}\left(\frac{F_p+ka}{F_p+a}\right)^{-\frac{\nu_1+\nu_2}{2}}H_1\right],
\end{aligned}$$

where

$$H_k \equiv H\left(\frac{\nu_1+\nu_2}{2},1,\frac{\nu_1}{2}+1,\tilde{F}_{p,k}\right).$$

Observe that, from (5.2)

$$\left(\frac{\tilde{F}_{p,k}}{\tilde{F}_{p,1}}\right)^{\frac{\nu_1}{2}}\left(\frac{1-\tilde{F}_{p,k}}{1-\tilde{F}_{p,1}}\right)^{\frac{\nu_2}{2}} = k^{\frac{\nu_2}{2}}\left(\frac{F_p+ka}{F_p+a}\right)^{-\frac{\nu_1+\nu_2}{2}}$$

and

$$\left(\tilde{F}_{p,1}\right)^{\frac{\nu_1}{2}}\left(1-\tilde{F}_{p,1}\right)^{\frac{\nu_2}{2}}k^{\frac{\nu_2}{2}}\left(\frac{F_p+ka}{F_p+a}\right)^{-\frac{\nu_1+\nu_2}{2}} = (ka)^{\frac{\nu_2}{2}}F_p^{\frac{\nu_1}{2}}(F_p+ka)^{-\frac{\nu_1+\nu_2}{2}}.$$

So,

$$\begin{aligned}\frac{\partial}{\partial p}h(p|k) &= \frac{2}{\nu_1 p^2}\left(\tilde{F}_{p,1}\right)^{\frac{\nu_1}{2}}\left(1-\tilde{F}_{p,1}\right)^{\frac{\nu_2}{2}} k^{\frac{\nu_2}{2}}\left(\frac{F_p+ka}{F_p+a}\right)^{-\frac{\nu_1+\nu_2}{2}}[H_1-H_k] \\ &= \frac{2}{\nu_1 p^2}F_p^{\frac{\nu_1}{2}}(F_p+ka)^{-\frac{\nu_1+\nu_2}{2}}(ka)^{\frac{\nu_2}{2}} \times s(p|\nu_1,\nu_2,k), \qquad (5.7)\end{aligned}$$

where $s(p|\nu_1,\nu_2,k)$ was defined in Lemma 5.2.2.

Case I: if $k \in [0,1)$ then $\frac{2}{\nu_1}F_p^{\frac{\nu_1}{2}}(F_p+ka)^{-\frac{\nu_1+\nu_2}{2}}(ka)^{\frac{\nu_2}{2}} > 0$ and by Case I of Lemma 5.2.2, $s(p|\nu_1,\nu_2,k) < 0$ and hence $\frac{\partial}{\partial p}h(p|k) < 0$ so that $h(p|k)$ is decreasing in p which completes the proof for this case.

Case III: if $k > 1$ then $\frac{2}{\nu_1}F_p^{\frac{\nu_1}{2}}(F_p+ka)^{-\frac{\nu_1+\nu_2}{2}}(ka)^{\frac{\nu_2}{2}} > 0$ and by Case III of Lemma 5.2.2, $s(p|\nu_1,\nu_2,k) > 0$ and hence $\frac{\partial}{\partial p}h(p|k) > 0$ so that $h(p|k)$ is increasing in p which completes the proof for this case. □

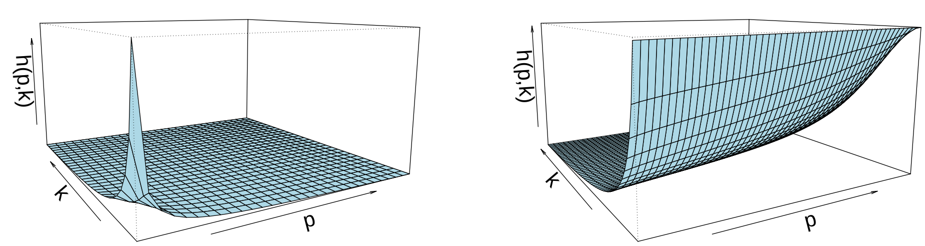

Figure 5.1: Left Panel is the bimonotonic surface of the function $h(p,k)$ for $k \in [0,1)$ and $p \in (0,1)$. Right Panel is the bimonotonic surface of the function $h(p,k)$ for $k > 1$ and $p \in (0,1)$.

Figure 5.1 illusrates the behaviour of $h(p|k)$ for different p and k values. It's very clear from Figure 5.1 that $h(p|k)$ is decreasing in p for $k \in [0, 1)$ and increasing in p for $k > 1$ while $h(p|k)$ is decreasing in k for any p.

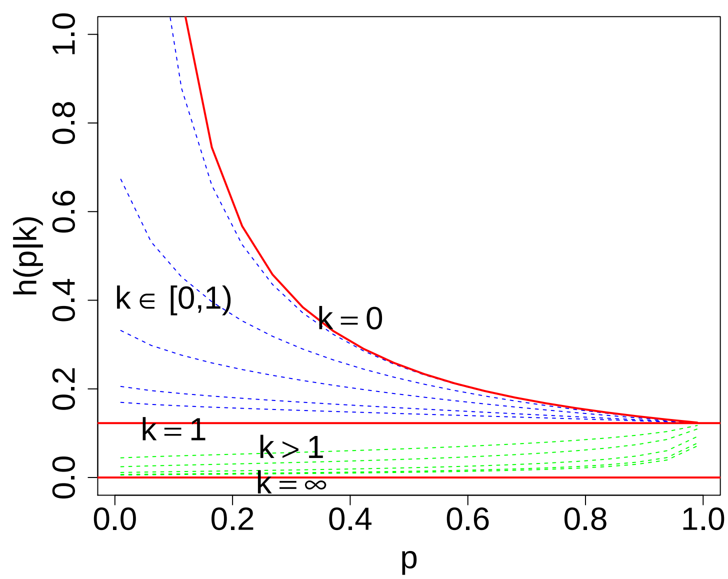

Figure 5.2: The plot of $h(p|k)$. Solid lines are the bounds at which $h(p|k)$ changes its monotonicity.

Figure 5.2 illustrates the behaviour of $h(p, k)$ in p for fixed k values in a one-dimensional plot. For example, for $k \in [0, 1)$ the dashed decreasing lines represent $h(p|k)$ as a function of changing p and they are bounded from below by the constant line $h(p|k = 1) = B(\nu_1/2, \nu_2/2)$ and from above by the decreasing curve $h(p|k = 0) = \frac{1}{p}B(\nu_1/2, \nu_2/2)$. On the other hand, for $k > 1$ the dashed increasing lines represent $h(p|k)$ as a function of changing p and they are bounded by the constant lines $h(p|k=1) = B(\nu_1/2, \nu_2/2)$ and $h(p|k=\infty) = 0$. If we look at $h(p|k)$ as a function of two variables p and k, it's easily seen that $h(p|k)$ is a continuous function with no non-degenerate stationary points. In fact

$$
\begin{aligned}
& \frac{\partial}{\partial k} h(p,k) = 0 \iff \left[\frac{\partial}{\partial x} B(x, \nu_1/2, \nu_2/2)/p\Big|_{x=\tilde{F}_{p,k}}\right] \times \left[\frac{\partial}{\partial k}\tilde{F}_{p,k}\right] = 0 \\
\iff \quad & \frac{1}{p}\left(\tilde{F}_{p,k}\right)^{\nu_1/2-1}\left(1-\tilde{F}_{p,k}\right)^{\nu_2/2-1} \frac{-F_p}{(F_p+ka)^2}\left(\frac{\nu_2}{\nu_1}\right) = 0 \\
\iff \quad & k = 0, \qquad (5.8)
\end{aligned}
$$

and we already have established that $\frac{\partial}{\partial p} h(p,k) = 0 \iff k = 1$. That is, there is no point (p,k) at which $\frac{\partial}{\partial p} h(p,k) = \frac{\partial}{\partial k} h(p,k) = 0$ which means that the surface of $h(p,k)$ is bimonotonic. Further, (5.8) implies that $h(k|p)$ is monotonic in k with a degenerate stationary point on the boundary of the domain when $k = 0$. Also, one should expect $h(p|k)$ to be decreasing in p for fixed $k \in [0,1)$ without examining the sign of the derivative for the following reasons. Note that $h(p|k=0) = \frac{1}{p}B(\nu_1/2, \nu_2/2)$, $h(p|k=1) = B(\nu_1/2, \nu_2/2)$ and $h(p|k=1) < h(p|k=0)$ so that $\lim_{p\to 1} h(p|k=0) = h(p|k=1)$ (i.e. $h(p,k)$ is bounded by a constant from below and a monotonically decreasing function from above). Thus since $h(p|k)$ is monotonic for $k \in [0,1)$ it must be be decreasing in p otherwise, by definition of infimum, it cannot be bounded by a constant from below and a monotonically decreasing function from above.

Theorem 5.3.3 *Let $\alpha, p \in (0,1)$ such that $p > \alpha$ (i.e. $F_p > F_\alpha$ where $p = P(F < F_p)$ and $\alpha = P(F < F_\alpha)$). For $k > 1$, if F is a random variable with an $F(\nu_1, \nu_2)$ distribution then the following inequality always holds:*

$$
\frac{1}{p} \times P\left(F < \frac{1}{k}F_p\right) \quad > \quad \frac{1}{\alpha} \times P\left(F < \frac{1}{k}F_\alpha\right), \qquad (5.9)
$$

or alternatively,

$$P\left(F < \frac{F_p}{k}\middle| F < F_p\right) \quad > \quad P\left(F < \frac{F_\alpha}{k}\middle| F < F_\alpha\right). \tag{5.10}$$

The inequality changes direction for $k \in [0,1)$ and becomes equality for $k = 1$ and, trivially, when $k = \infty$.

Proof of Theorem 5.3.3: Recall that the CDF of the F-distribution with ν_1 and ν_2 degrees of freedom is

$$F(x;\nu_1,\nu_2) = I_{\frac{\nu_1 x}{\nu_1 x + d_2}}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right),$$

where I is the reglarized incomplete beta function as defined in Definition 5.3.1. Thus, one can write the inequality in (5.9) in terms of the incomplete beta function as

$$\frac{1}{p} \times B\left(\tilde{F}_{p,k}; \frac{\nu_1}{2}, \frac{\nu_2}{2}\right) \quad > \quad \frac{1}{\alpha} \times B\left(\tilde{F}_{\alpha,k}; \frac{\nu_1}{2}, \frac{\nu_2}{2}\right). \tag{5.11}$$

Now, since $k > 1$ and $p > \alpha$, the inequality in (5.11) follows immediately from Case III of Proposition 5.3.2. If $k \in [0,1)$ then Case I of the proposition changes the direction of the inequality while, for $k = 1$, Case II changes the inequality to equality. For $k = \infty$, both sides of the inequality become zero making it equality. □

5.4 The G-Inequality

Similar arguments establish the G-inequality for Gamma distributions.

Definition 5.4.1 *For $a > 0$, the lower incomplete gamma function $\gamma(a,x)$ is defined as*

$$\gamma(a,x) = \int_0^x t^{a-1} e^{-t} dt$$

for $x \geq 0$ *and the gamma function* $\Gamma(a)$ *is defined as*

$$\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt,$$

see O'Connor (2011).

Recall that the CDF of the $Gamma(\tau, \theta)$ distribution is

$$F(x; \tau, \theta) = \frac{1}{\Gamma(\tau)} \gamma\left(\tau, \frac{x}{\theta}\right).$$

Proposition 5.4.2 *Let* G *be a random variable with the* $Gamma(\tau, \theta)$ *distribution. Define the GR-function as* $\tilde{h}(p|k) = \frac{1}{p}\gamma\left(\tau, \frac{G_p}{k\theta}\right)$ *where* $p \in (0,1)$ *and* G_p *is the pth quantile of the gamma distribution such that* $p = P(G < G_p)$*. Then,* $\tilde{h}(p|k)$ *is a monotonic function for* $k \geq 0$ *such that:*

Case I: if $0 \leq k < 1$ *then* $\tilde{h}(p|k)$ *is strictly decreasing in* p*.*

Case II: if $k = 1$ *then* $\tilde{h}(p|k) = \Gamma(\tau)$ *(constant).*

Case III: if $k > 1$ *then* $\tilde{h}(p|k)$ *is strictly increasing in* p*.*

Proof of Proposition 5.4.2: The outline of the proof is identical to that of Proposition 5.3.2 and therefore we start it with case II as follows.

Case II:

$$\begin{aligned} & p = P\left(G \leq G_p\right) = \frac{1}{\Gamma(\tau)} \gamma\left(\tau, \frac{G_p}{\theta}\right) \\ \Longleftrightarrow \quad & \frac{1}{p}\gamma\left(\tau, \frac{G_p}{\theta}\right) = \Gamma(\tau). \end{aligned} \tag{5.12}$$

But the left hand side of this equality is precisely $\tilde{h}(p|k = 1)$ which completes the proof for this case.

The result of Case II tells us that $\frac{\partial}{\partial p}\tilde{h}(p|k = 1) = 0$ for all p. So, if we let f denote the density function of the G random variable we have $\frac{\partial}{\partial p}G_p = \frac{1}{f(G_p)}$ so that the following identity is always true.

$$\begin{aligned}
& \frac{\partial}{\partial p}\tilde{h}(p|k = 1) = 0 \\
\iff \quad & -\frac{1}{p^2}\gamma\left(\tau, \frac{G_p}{\theta}\right) + \frac{1}{p}\left(\frac{G_p}{\theta}\right)^{\tau-1} e^{\frac{-G_p}{\theta}} \frac{1}{\theta f(G_p)} = 0 \\
\iff \quad & \left(\frac{G_p}{\theta}\right)^{\tau-1} \frac{1}{\theta f(G_p)} = \frac{1}{p}\gamma\left(\tau, \frac{G_p}{\theta}\right) e^{\frac{G_p}{\theta}}.
\end{aligned} \tag{5.13}$$

One can verify the equality in (5.13) as an exercise by using (5.12) and the substitution of G_p in the density function of the Gamma distribution.

Using this identity in deriving $\frac{\partial}{\partial p}\tilde{h}(p|k)$ gives

$$\begin{aligned}
\frac{\partial}{\partial p}\tilde{h}(p|k) &= -\frac{1}{p^2}\gamma\left(\tau, \frac{G_p}{k\theta}\right) + \frac{1}{p}\left(\frac{G_p}{k\theta}\right)^{\tau-1} \frac{1}{k\theta f(G_p)} e^{\frac{-G_p}{k\theta}} \\
&= -\frac{1}{p^2}\gamma\left(\tau, \frac{G_p}{k\theta}\right) + \left[\frac{1}{p}\left(\frac{G_p}{\theta}\right)^{\tau-1} \frac{1}{\theta f(G_p)}\right] \frac{1}{k^\tau} e^{\frac{-G_p}{k\theta}} \\
&= -\frac{1}{p^2}\gamma\left(\tau, \frac{G_p}{k\theta}\right) + \frac{1}{p^2}\gamma\left(\tau, \frac{G_p}{\theta}\right) e^{\frac{G_p}{\theta}} e^{\frac{-G_p}{k\theta}} k^{-\tau}.
\end{aligned} \tag{5.14}$$

As shown in Cuyt et al. (2008), the connection between the lower incomplete gamma function and the Kummer's confluent hypergeometric function is

$$\gamma(a, x) = a^{-1}x^a e^{-x} M(1, a + 1, x).$$

Using this representation and arguments similar to those for the F-inequality reduce (5.14) to

$$\frac{\partial}{\partial p}\tilde{h}(p|k) = \frac{1}{p^2}\tau^{-1}\left(\frac{G_p}{k\theta}\right)^{\tau} e^{\frac{-G_p}{k\theta}} \times S(p|\tau,\theta,k), \tag{5.15}$$

where $S(p|\tau,\theta,k)$ was defined in Lemma 5.2.4.

Case I: if $k \in [0,1)$ then $\frac{1}{p^2}\tau^{-1}\left(\frac{G_p}{k\theta}\right)^{\tau} e^{\frac{-G_p}{k\theta}} > 0$ and by Case I of Lemma 5.2.2, $S(p|\tau,\theta,k) < 0$ and hence $\frac{\partial}{\partial p}\tilde{h}(p|k) < 0$ so that $\tilde{h}(p|k)$ is decreasing in p which completes the proof for this case.

Case III: if $k > 1$ then $\frac{1}{p^2}\tau^{-1}\left(\frac{G_p}{k\theta}\right)^{\tau} e^{\frac{-G_p}{k\theta}} > 0$ and by Case III of Lemma 5.2.2, $S(p|\tau,\theta,k) > 0$ and hence $\frac{\partial}{\partial p}\tilde{h}(p|k) > 0$ so that $\tilde{h}(p|k)$ is increasing in p which completes the proof for this case. $\square$

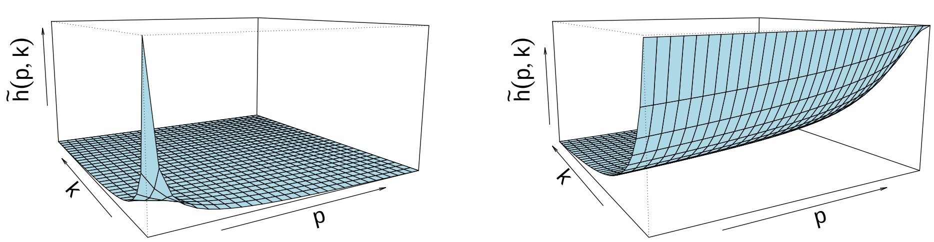

Figure 5.3: Left Panel is the bimonotonic surface of the function $\tilde{h}(p,k)$ for $k \in [0,1)$ and $p \in (0,1)$. Right Panel is the bimonotonic surface of $\tilde{h}(p,k)$ for $k > 1$ and $p \in (0,1)$.

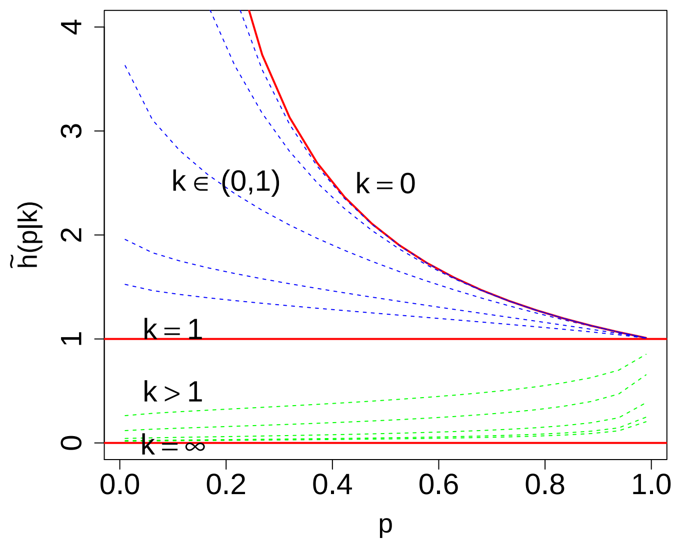

Figure 5.4: The plot of $\tilde{h}(p|k)$. Solid lines are the bounds at which $\tilde{h}(p|k)$ changes its monotonicity.

Figure 5.3 illusrates the behaviour of $\tilde{h}(p,k)$ for different p and k values. It's very clear from figure 5.3 that $\tilde{h}(p,k)$ is decreasing in p for $k \in [0,1)$ and increasing in p for $k > 1$ while $\tilde{h}(p,k)$ is decreasing in k for any p. Figure 5.4 illustrates the behaviour of $\tilde{h}(p,k)$ in p for fixed k values in a one-dimensional plot. For example, for $k \in [0,1)$ the dashed decreasing lines represent $\tilde{h}(p|k)$ as a function of changing p and they are bounded from below by the constant line $\tilde{h}(p|k=1) = \Gamma(\tau)$ and from above by the decreasing line $\tilde{h}(p|k=0) = \frac{1}{p}\Gamma(\tau)$. On the other hand, for $k > 1$ the dashed increasing lines represent $h(p|k)$ as a function of changing p and they are bounded by the constant lines $\tilde{h}(p|k=1) = \Gamma(\tau)$ and $\tilde{h}(p|k=\infty) = 0$. If we look at $\tilde{h}(p,k)$ as a function of two variables p and k, it's easily seen that $\tilde{h}(p,k)$ is a continuous function with no non-degenerate stationary points. In fact

$$\begin{aligned} \frac{\partial}{\partial k}\tilde{h}(p,k) = 0 \iff & \left[\frac{\partial}{\partial x}\gamma(\tau,x)/p\Big|_{x=\frac{G_p}{k\theta}}\right] \times \left[\frac{\partial}{\partial k}\frac{G_p}{k\theta}\right] = 0 \\ \iff \quad & \frac{1}{p}\left(\frac{G_p}{k\theta}\right)^{\tau-1} e^{\frac{-G_p}{k\theta}}\left(\frac{-G_p}{\theta k^2}\right) = 0 \\ \iff \quad & k = \infty, \end{aligned} \tag{5.16}$$

and we already have established that $\frac{\partial}{\partial p}\tilde{h}(p,k) = 0 \iff k = 1$. That is, there is no point (p,k) at which $\frac{\partial}{\partial p}\tilde{h}(p,k) = \frac{\partial}{\partial k}\tilde{h}(p,k) = 0$ which means that the surface of $\tilde{h}(p,k)$ is bimonotonic. Also, one should expect $\tilde{h}(p|k)$ to be decreasing in p for fixed $k \in [0,1)$ without examining the sign of the derivative for the following reasons. Note that from Cases I and II we know that $\tilde{h}(p|k=0) = \frac{1}{p}\Gamma(\tau)$, $\tilde{h}(p|k=1) = \Gamma(\tau)$ and $\tilde{h}(p|k=1) < \tilde{h}(p|k=0)$ so that $\lim_{p\to 1}\tilde{h}(p|k=0) = \tilde{h}(p|k=1)$ (i.e. $\tilde{h}(p,k)$ is bounded by a constant from below and a monotonically decreasing function from above). Thus since $\tilde{h}(p|k)$ is monotonic for $k \in [0,1)$ it must be be decreasing in p otherwise, by definition of infimum, it cannot be bounded by a constant from below and a monotonically decreasing function from above.

Theorem 5.4.3 *Let $\alpha, p \in (0,1)$ such that $p > \alpha$ (i.e. $G_p > G_\alpha$ where $p = P(G < G_p)$ and $\alpha = P(G < G_\alpha)$). For $k > 1$, if G is a gamma random variable with parameters τ and θ then the following inequality always hold:*

$$\frac{1}{p} \times P\left(G < \frac{G_p}{k}\right) \;>\; \frac{1}{\alpha} \times P\left(G < \frac{G_\alpha}{k}\right), \tag{5.17}$$

or alternatively,

$$P\left(G < \frac{G_p}{k}\Big| G < G_p\right) \;>\; P\left(G < \frac{G_\alpha}{k}\Big| G < G_\alpha\right). \tag{5.18}$$

The inequality changes direction for $k \in [0,1)$ and becomes equality for $k = 1$ and, trivially, when $k = \infty$.

Proof of Theorem 5.4.3: Recall that the CDF of the $Gamma(\tau, \theta)$ distribution is

$$F(x; \tau, \theta) = \frac{1}{\Gamma(\tau)} \gamma\left(\tau, \frac{x}{\theta}\right),$$

where γ is the lower incomplete gamma function and Γ is the gamma function as defined in Definition 5.4.1. Thus, one can write the inequality in (5.17) in terms of the lower incomplete gamma function as

$$\frac{1}{p} \gamma\left(\tau, \frac{G_p}{k\theta}\right) \;>\; \frac{1}{\alpha} \gamma\left(\tau, \frac{G_\alpha}{k\theta}\right). \tag{5.19}$$

Now, since $k > 1$ and $p > \alpha$ then the inequality in (5.19) follows immediately from Case III of Proposition 5.4.2. If $k \in [0, 1)$ then Case I of the very same Proposition changes the direction of the inequality while, for $k = 1$, Case II changes the inequality to equality. For $k = \infty$, both sides of the inequality become zero making it equality.

□

Appendix A

Proofs of secondary results

A.1 Proof for the PPOs Properties in (1.13)

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

$$\begin{aligned}
\tilde{M} &= [X_*, (I - M_1)X_2] \left(\begin{bmatrix} X_*' \\ [(I - M_1)X_2]' \end{bmatrix} [X_*, (I - M_1)X_2] \right)^{-1} \\
&\quad \times \begin{bmatrix} X_*' \\ [(I - M_1)X_2]' \end{bmatrix} \\
&= [X_*, (I - M_1)X_2] \begin{bmatrix} X_*'X_* & X_*'(I - M_1)X_2 \\ X_2'(I - M_1)X_* & X_2'(I - M_1)X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_*' \\ [(I - M_1)X_2]' \end{bmatrix} \\
&= [X_*, (I - M_1)X_2] \begin{bmatrix} \left(X_*'X_*\right)^{-1} & 0 \\ 0 & \left(X_2'(I - M_1)X_2\right)^{-1} \end{bmatrix} \begin{bmatrix} X_*' \\ [(I - M_1)X_2]' \end{bmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \left[X_* \left(X_*' X_*\right)^{-1}, (I - M_1)X_2 \left(X_2'(I - M_1)X_2\right)^{-1}\right] \begin{bmatrix} X_*' \\ X_2'(I - M_1) \end{bmatrix} \\
&= X_* \left(X_*' X_*\right)^{-1} X_*' + (I - M_1)X_2 \left(X_2'(I - M_1)X_2\right)^{-1} X_2'(I - M_1) \\
&= M_* + M_2. \qquad \text{(A.1)}
\end{aligned}
$$

The equality in the third line of (A.1) 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 $X_* = X_1 B$ for some matrix B and therefore, by defintion of PPO,

$$
\begin{aligned}
M_* &= X_* \left(X_*' X_*\right)^{-1} X_*' \\
&= X_1 B \left((X_1 B)' X_1 B\right)^{-1} (X_1 B)' \\
&= X_1 B \left(B' X_1' X_1 B\right)^{-1} B' X_1'. \qquad \text{(A.2)}
\end{aligned}
$$

Thus, using M_* from (A.2) and M_1 from (1.11) gives

$$
\begin{aligned}
M_* M_1 &= X_1 B \left(B' X_1' X_1 B\right)^{-1} B' X_1' X_1 \left(X_1' X_1\right)^{-1} X_1' \\
&= X_1 B \left(B' X_1' X_1 B\right)^{-1} B' X_1' \\
&= M_*. \qquad \text{(A.3)}
\end{aligned}
$$

Thirdly, $M_1 M_2 = 0$**:** This results is trivially obtained by simply multiplying M_1 from (1.11) and M_2 from (1.12).

Fourthly, $M = M_1 + M_2$**:** Let $X = [X_1, X_2]$ such that M is the PPO onto $C(X)$. Since $M_1 M_2 = 0$, 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.

$$\begin{aligned}
[X_1, (I - M_1)X_2] &= [X_1, X_2 - M_1X_2] \\
&= \left[X_1, X_2 - X_1(X_1'X_1)^{-1}X_1'X_2\right] \\
&= [X_1, X_2]\begin{bmatrix} I & -(X_1'X_1)^{-1}X_1'X_2 \\ 0 & I \end{bmatrix}
\end{aligned} \tag{A.4}$$

and

$$[X_1, X_2] = [X_1, (I - M_1)X_2]\begin{bmatrix} I & (X_1'X_1)^{-1}X_1'X_2 \\ 0 & I \end{bmatrix}. \tag{A.5}$$

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. □

A.2 Illustration for the proof of Lemma 2.1.2

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

$$\begin{aligned}
|aI_n + bP| &= |b\,(P - \lambda I_n)\,| \\
&= b^n|P - \lambda I_n|.
\end{aligned} \tag{A.6}$$

However, the determinant $|P - \lambda I_n|$ in (A.6) is the characteristic polynomial of P which equals to (A.7) since 1 and 0 are the eigenvalues for P with multiplicity $r(P)$ and $n - r(P)$ respectively.

$$|P - \lambda I_n| \equiv p_P(\lambda) = (-\lambda)^{n-r(P)}(1-\lambda)^{r(P)}. \tag{A.7}$$

Hence, substituting (A.7) in (A.6) gives the desired result

$$
\begin{aligned}
|aI_n + bP| &= b^n |P - \lambda I_n| \\
&= b^n (-\lambda)^{n-r(P)} (1-\lambda)^{r(P)} \\
&= b^n \left(\frac{a}{b}\right)^{n-r(P)} \left(1 + \frac{a}{b}\right)^{r(P)} \\
&= \frac{b^n a^{n-r(P)}}{b^{n-r(P)}} \frac{(a+b)^{r(P)}}{b^{r(P)}} \\
&= a^{n-r(P)} (a+b)^{r(P)}. \qquad \text{(A.8)}
\end{aligned}
$$

□

A.3 Proof of Lemma 2.1.3

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

$$\frac{\partial g}{\partial x_1} = 0 \iff \frac{q_1(Q_1 - x_1)}{x_1^2} = 0 \iff x_1 = Q_1, \qquad \text{(A.9)}$$

and

$$\frac{\partial g}{\partial x_2} = 0 \iff \frac{q_2(Q_2 - x_2)}{x_2^2} = 0 \iff x_2 = Q_2. \qquad \text{(A.10)}$$

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

$$
\begin{aligned}
D(x_1, x_2) &= g_{x_1x_1}(x_1, x_2) g_{x_2x_2}(x_1, x_2) - [g_{x_1x_2}(x_1, x_2)]^2 \\
&= \left(\frac{q_1}{x_1^2} - \frac{2q_1Q_1}{x_1^3}\right)\left(\frac{q_2}{x_2^2} - \frac{2q_2Q_2}{x_2^3}\right) \qquad \text{(A.11)}
\end{aligned}
$$

with $D(Q_1, Q_2) = \frac{q_1 q_2}{Q_1^2 Q_2^2} > 0$ and $g_{x_1x_1}(Q_1, Q_2) = \frac{-q_1}{Q_1^2} < 0$ 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

$$g(x) = -\left[\text{constant} + (q_1 + q_2)\log(x) + \frac{q_1Q_1 + q_2Q_2}{x}\right] \tag{A.12}$$

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

$$\frac{\partial g}{\partial x} = 0 \iff -\frac{q_1 + q_2}{x} + \frac{q_1Q_1 + q_2Q_2}{x^2} = 0 \iff x = \frac{q_1Q_1 + q_2Q_2}{q_1 + q_2}. \tag{A.13}$$

Now, using the second derivative test, we have

$$\frac{\partial g(x)}{\partial x^2} = \frac{q_1 + q_2}{x^2} - \frac{2(q_1Q_1 + q_2Q_2)}{x^3} \tag{A.14}$$

with

$$\frac{\partial g(x)}{\partial x^2}\,|_{x=\frac{q_1Q_1+q_2Q_2}{q_1+q_2}} = \frac{-(q_1 + q_2)^3}{(q_1Q_1 + q_2Q_2)^2} < 0 \tag{A.15}$$

so that $(x_1, x_2) = \left(\frac{q_1Q_1+q_2Q_2}{q_1+q_2}, \frac{q_1Q_1+q_2Q_2}{q_1+q_2}\right)$ 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

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

Further, when the maximum is on the boundary at $\left(\frac{q_1Q_1+q_2Q_2}{q_1+q_2}, \frac{q_1Q_1+q_2Q_2}{q_1+q_2}\right)$ it attains the value

$$g\left(\frac{q_1Q_1 + q_2Q_2}{q_1 + q_2}\right) = -\left[constant + (q_1 + q_2) + (q_1 + q_2)\log\left(\frac{q_1Q_1 + q_2Q_2}{q_1 + q_2}\right)\right].$$

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

$$\log\left[E(Q)\right] > E\left[\log(Q)\right]$$

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

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. $\square$

Bibliography

Castellacci, G. (2012). A Formula for the Quantiles of Mixtures of Distributions with Disjoint Supports. Available online: http://ssrn.com/abstract=2055022 (accessed on 15 April 2013).

Christensen, Ronald (1984). A note on ordinary least squares methods for two-stage sampling. *Journal of the American Statistical Association*. **79**, 720-72.

Christensen, Ronald (1987). The Analysis of Two-Stage Sampling Data by Ordinary Least Squares. *Journal of the American Statistical Association*. **82**, 492-498.

Christensen, Ronald (2011). *Plane Answers to Complex Questions: The Theory of Linear Models*. Fourth Edition, New York: Springer.

Cloud M. J., Drachman B. C. (1998). *Inequalities with Applications to Engineering*. First Edition. New York: Springer.

Crainiceanu, C. M. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *J. Royal Statist. Soc., Ser. B*. **66**, 16585.

Cuyt A., Petersen V.B., Verdonk B., Waadeland H., Jones W.B. (2008). *Handbook of Continued Fractions for Special Functions*, pp 319-341.

Dutka, J. (1981). The incomplete beta function - A historical profile. *Archive for History of the Exact Sciences*. **24**, 11–29.

Graybill Franklin A. (1961). *An introduction to linear statistical models*. Volume I. McGraw Hill, New York-Toronto-London.

Greven, S., Crainiceanu, C. M., Kchenhoff, H., Peters, A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics*. **17(4)**, 870891.

Harville David A. (1997). *Matrix Algebra From a Statistician's Perspective*. First Edition. New York: Springer.

Herbach Leon H. (1959). Properties of Model II–Type Analysis of Variance Tests, A: Optimum Nature of the F-Test for Model II in the Balanced Case. *The Annals of Mathematical Statistics*. **30(4)**, 939-959.

Miller S. S., Mocanu P. T. (1990). Univalence of Gaussian and confluent hypergeometric functions. *Proc. Amer. Math. Soc.* **110(2)**, 333-342.

Molenberghs, G., Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician*, **61(1)**, 22-27.

O'Connor A. N. (2011). *Probability Distributions Used in Reliability Engineering*, Maryland: RIAC.

Scheipl, Fabian, Greven S., Kchenhoff H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics and Data Analysis*. **52(7)**, 3283-3299.

Wiencierz, A., Greven, S., Kchenhoff, H. (2011). Restricted likelihood ratio testing in linear mixed models with general error covariance structure. *Electronic Journal of Statistics*. **5**, 1718-1734.

Yan Lu, and Guoyi Zhang (2010). The Equivalence between Likelihood Ratio Test and F-Test for Testing Variance Component in a Balanced One-Way Random Effects Model. *Journal of Statistical Computation and Simulation*. **80**, 443-450.