F approximations with Split Plot applications: A work in progress but mostly done.

Ronald Christensen

Department of Mathematics and Statistics

University of New Mexico

February 23, 2024

For linear models with general nonsingular *estimated* covariance matrices, we discuss the approximate null distributions proposed for the standard test statistics by Fai and Cornelius (1996), referred to as FC, and by Kenward and Roger (1997, 2009), referred to as KR1 and KR2, respectively, or as KR collectively. Much of the material relates to Christensen (2019) including the material on differentiation in his Appendix A.1. The proposed approximations are shown to give the exact distributions appropriate for use with Generalized Split Plot models. While there is a great deal of theory behind the approximations, there is also a great deal of seat-of-the-pants approximation involved.

1 Background

Consider a full linear model

$$Y = X\beta + e, \qquad e \sim N[0, V(\theta)] \tag{1.1}$$

and testing it against a reduced model

$$Y = X_0\beta_0 + e, \qquad C(X_0) \subset C(X).$$

Write

$$V \equiv V(\theta) = \sigma^2 V_*(\theta_*) \equiv \sigma^2 V_*$$

with $\theta' = (\sigma^2, \theta_*')$ a q vector of covariance parameters. Write oblique projection operators onto $C(X)$ and $C(X_0)$,

$$A \equiv X[X'V^{-1}X]^- X'V^{-1} \quad \text{and} \quad A_0 \equiv X_0[X_0'V^{-1}X_0]^- X_0'V^{-1}.$$

Note that these projection operators remain the same when V is replaced by V_*. Also, as in Christensen (2020, Exercise 2.5), the projection operators do not depend on the choice of the generalized inverse and

$$A'V^{-1}A = A'V^{-1} = V^{-1}A.$$

When using estimates $\tilde{\theta} = (\tilde{\sigma}^2, \tilde{\theta}_*')'$, write

$$\tilde{V} \equiv V(\tilde{\theta}) = \tilde{\sigma}^2 V_*(\tilde{\theta}_*) \equiv \tilde{\sigma}^2 \tilde{V}_*$$

and projection operators $\tilde{A}$ and $\tilde{A}_0$.

It is well known, e.g. Christensen (2020), that for known V the test statistic and null distribution can be written as

$$F = Y'(A-A_0)'V^{-1}(A-A_0)Y/r(A-A_0) \sim \chi^2[r(A-A_0)]/r(A-A_0) = F[r(A-A_0), \infty].$$

Also, for known V_* but unknown σ^2,

$$F = \frac{Y'(A-A_0)'V_*^{-1}(A-A_0)Y/r(A-A_0)}{Y'(I-A)'V_*^{-1}(I-A)Y/r(I-A)} \sim F[r(A-A_0), n-r(A)].$$

The intuitive basis for this change is that in the original F statistic,

$$Y'(A-A_0)'V^{-1}(A-A_0)Y = \frac{Y'(A-A_0)'V_*^{-1}(A-A_0)Y}{\sigma^2}$$

however we can replace the unknown parameter σ^2 using an unbiased estimate of it, $MSE \equiv Y'(I-A)'V_*^{-1}(I-A)Y/r(I-A)$. We can then show that for known V_*, this has the indicated F distribution. In fact, writing $\check{V} \equiv MSE \times V_*$,

$$F = \frac{Y'(A-A_0)'V_*^{-1}(A-A_0)Y/r(A-A_0)}{Y'(I-A)'V_*^{-1}(I-A)Y/r(I-A)}$$
$$= Y'(A-A_0)'\check{V}^{-1}(A-A_0)Y/r(A-A_0) \sim F[r(A-A_0), n-r(A)].$$

Again, with V_* known, we know both A and A_0.

The goal is to get better approximate distributions for F under the null model when V and V_* are unknown by incorporating information about the estimates of θ or θ_*. This involves finding an approximate number of degrees of freedom for the denominator of the F distribution. Later we will show that for generalized split plot models using REML estimates, these methods typically give the standard exact F statistics, i.e., that the approximation methods typically give the correct denominator degrees of freedom. (This fails when a REML estimate is 0.)

To work out the approximations we rewrite the problem as an ACOVA problem,

$$Y = X_0\beta_0 + X_1\beta_1 + e, \tag{1.2}$$

where $X = [X_0, X_1]$, $\beta' = [\beta_0', \beta_1']$. Our interest is in testing $\beta_1 = 0$, or more accurately, that we can drop β_1 from the model. (It is possible that there exists $\beta_1 \neq 0$ with $X_1\beta_1 \in C(X_0)$.)

As in Christensen (2020, Subsection 9.1.1),

$$A - A_0 = (I - A_0)X_1[X_1'(I - A_0)'V^{-1}(I - A_0)X_1]^{-}X_1'(I - A_0)'V^{-1} \equiv A_1,$$

so for known V,

$$\begin{aligned} F &= Y'(A - A_0)'V^{-1}(A - A_0)Y/r(A - A_0) \\ &= YA_1'V^{-1}A_1Y/r(A_1). \end{aligned}$$

For any generalized inverse,

$$\hat{\beta}_1(\theta) \equiv [X_1'(I - A_0)'V^{-1}(I - A_0)X_1]^{-}X_1'(I - A_0)'V^{-1}Y = \hat{\beta}_1(\theta_*)$$

is a generalized least squares estimate. The test statistic has the property that

$$YA_1'V^{-1}A_1Y = \hat{\beta}_1'[X_1'(I - A_0)'V^{-1}(I - A_0)X_1]\hat{\beta}_1.$$

Consider a (full rank version of the) singular value decomposition,

$$X_1'(I - A_0)'V^{-1}(I - A_0)X_1 = PD(\phi)P' = [P_1, \cdots, P_s]D(\phi)[P_1, \cdots, P_s]',$$

where

$$s = r[(I - A_0)X_1] = r(A_1),$$

and $\phi = (\phi_1, \ldots, \phi_s)'$ with $\phi_j > 0$, enabling us to write

$$YA_1'V^{-1}A_1Y = \sum_{j=1}^{s} \phi_j (P_j'\hat{\beta}_1)^2. \tag{1.3}$$

We want to show that the $P_j'\hat{\beta}_1$s are independent and that $P_j'\beta_1$ is estimable. For estimability, because it is a linear function of $X_0\beta_0 + X_1\beta_1$, clearly $(I - A_0)X_1\beta_1$ is estimable. Observe that $C(P) = C[X'(I - A_0)']$ so there exists Q such that $P' = Q'(I - A_0)X_1$ and $P'\beta_1$ is a linear function of an estimable parameter, hence estimable and each component of the vector is estimable. To check independence, by multivariate normality, it is enough to show that the covariance matrix is diagonal,

$$\begin{aligned}\mathrm{Cov}(P'\hat{\beta}_1) = \mathrm{Cov}[Q'(I - A_0)X_1\hat{\beta}_1] = \mathrm{Cov}(Q'A_1Y) = Q'A_1VA_1'Q \\ = P'[X_1'(I - A_0)'V^{-1}(I - A_0)X_1]^{-}P = P'[PD(\phi)^{-1}P']P = D(\phi)^{-1},\end{aligned}$$

where any generalized inverse when pre and post multiplied by P' and P will give the same result as using this Moore-Penrose generalized inverse $PD(\phi)^{-1}P'$.

Although $\hat{\beta}_1$ is unknown, and in practice must be replaced by

$$\tilde{\beta}_1 \equiv \hat{\beta}(\tilde{\theta}) = [X_1'(I - \tilde{A}_0)'\tilde{V}^{-1}(I - \tilde{A}_0)X_1]^{-}X_1'(I - \tilde{A}_0)'\tilde{V}^{-1}Y,$$

the FC Satterthwaite approximation works with an estimate of $\mathrm{Cov}(\hat{\beta}_1) = [X_1'(I - A_0)'V^{-1}(I - A_0)X_1]^{-}$ rather than an estimate of $\mathrm{Cov}(\tilde{\beta}_1)$. KR approximate $\mathrm{Cov}(\tilde{\beta}_1)$. (For the covariance matrices to behave uniquely, we need to restrict attention to looking at estimable functions of β_1.) We will also need $\tilde{P}$ and $\tilde{\phi}$ that come from a singular value decomposition,

$$[X_1'(I - \tilde{A}_0)'\tilde{V}^{-1}(I - \tilde{A}_0)X_1]^{-} = \tilde{P}D(\tilde{\phi})\tilde{P}'.$$

As discussed in the appendix, this term arises naturally when looking at $(X'\tilde{V}^{-1}X)^{-}$.

2 Satterthwaite Approximation

This Satterthwaite approximation was developed by Fai and Cornelius (1996) based on previous work by Giesbrecht and Burns (1985) [GB].

2.1 Fai and Cornelius

FC's approach to applying Satterthwaite approximations for F statistics is to find a denominator degrees of freedom ν for

$$F = Y\tilde{A}_1'\tilde{V}^{-1}\tilde{A}_1 Y/s \sim F(s,\nu)$$

by writing

$$F = \sum_{j=1}^{s} \tilde{\phi}_j(\tilde{P}_j'\tilde{\beta}_1)^2/s,$$

treating the terms $\sqrt{\tilde{\phi}_j}(\tilde{P}_j'\tilde{\beta}_1)$ as having $t(v_j)$ distributions, and then using the v_js to find an appropriate value for ν. The v_js are found by using GB's method which will be discussed later.

Letting $T(v_i)$ denote a random variable with a $t(v_i)$ distribution and recalling that $t^2(v_i) \sim F(1,v_i)$,

$$\mathrm{E}\left[\sum_{i=1}^{s} T^2(v_i)/s\right] = \sum_{i=1}^{s} v_i/(v_i-2)s$$

and the mean of an $F(s,\nu)$ distribution is $\nu/(\nu-2)$. The approximating F denominator degrees of freedom are taken as the value ν that satisfies

$$\nu/(\nu-2) = \sum_{i=1}^{s} v_i/(v_i-2)s$$

i.e.,

$$\nu = \frac{2\sum_{i=1}^{s} v_i/(v_i-2)s}{\sum_{i=1}^{s} v_i/(v_i-2)-s}.$$

2.2 Giesbrecht and Burns

We now examine GB's approach to finding the degrees of freedom v_j for an approximate t test based on

$$\sqrt{\tilde{\phi}_j}\tilde{P}_j'\tilde{\beta}_1 = \frac{\tilde{P}_j'\tilde{\beta}_1}{\sqrt{\tilde{P}_j'[X_1'(I-\tilde{A}_0)'\tilde{V}^{-1}(I-\tilde{A}_0)X_1]^{-}\tilde{P}_j}}.$$

Then, as indicated above, we use these v_js to determine an overall denominator degrees of freedom v for the F test.

To justify the approach, we will look at

$$\frac{P_j'\hat{\beta}_1}{\sqrt{P_j'[X_1'(I-A_0)'\tilde{V}^{-1}(I-A_0)X_1]^{-}P_j}}, \tag{2.1}$$

and then ultimately estimate all the pieces in (2.1) that are unknown.

We presume that $\tilde{\theta}$ is a function of $B'Y$ where B is a full rank matrix such that $C(B) = C(X)^{\perp}$. This is true for REML and MINQUE estimates. It follows easily from the formula for the projection operator A that $\mathrm{Cov}(AY, B'Y) = 0$, so from multivariate normality

$$AY \perp\!\!\!\perp B'Y$$

and

$$P_j'\hat{\beta}_1 \perp\!\!\!\perp P_j'[X_1'(I-A_0)'\tilde{V}^{-1}(I-A_0)X_1]^{-}P_j.$$

We know that $P_j'\hat{\beta}_1$ is normal so, for the statistic in (2.1) to have a $t(v_j)$ distribution, we need only show that

$$\frac{v_j\, P_j'[X_1'(I-A_0)'\tilde{V}^{-1}(I-A_0)X_1]^{-}P_j}{P_j'[X_1'(I-A_0)'V^{-1}(I-A_0)X_1]^{-}P_j} \sim \chi^2(v_j). \tag{2.2}$$

(None of these statements would be true if we were using $\tilde{\beta}_1$, $\tilde{P}_j$, or $\tilde{A}_0$.)

The $t(v_j)$ approximation is based on *assuming* that (2.2) is true and estimating the degrees of freedom v_j. If the $\chi^2(v_j)$ distribution in (2.2) is true, the variance has to satisfy

$$\frac{v_j^2\mathrm{Var}\{P_j'[X_1'(I-A_0)'\tilde{V}^{-1}(I-A_0)X_1]^{-}P_j\}}{\{P_j'[X_1'(I-A_0)'V^{-1}(I-A_0)X_1]^{-}P_j\}^2} = 2v_j.$$

or

$$v_j = 2\frac{\{P_j'[X_1'(I-A_0)'V^{-1}(I-A_0)X_1]^{-}P_j\}^2}{\mathrm{Var}\{P_j'[X_1'(I-A_0)'\tilde{V}^{-1}(I-A_0)X_1]^{-}P_j\}}. \tag{2.3}$$

The only "unknown" in equation (2.3) is $\text{Var}\{P_j'[X_1'(I-A_0)'\tilde{V}^{-1}(I-A_0)X_1]^- P_j\}$ for which we will use an asymptotic variance based on the asymptotic distribution of $\tilde{\theta}$ and the Delta Method. (In reality most of the terms in (2.3) are unknown and will eventually have to be estimated.)

For REML estimates

$$\sqrt{n}(\tilde{\theta}-\theta) \stackrel{\mathscr{L}}{\to} N[0, \mathbf{I}_r^{-1}(\theta)],$$

where, defining $\mathscr{I}_r(\theta)$ to be the information matrix from the restricted likelihood computed on $B'Y$, further define $\mathbf{I}_r(\theta)$ by the limit

$$\mathscr{I}_r(\theta)/n \to \mathbf{I}_r(\theta),$$

cf. Cressie and Lahiri (1993). Note that in practice this limit need not be appropriate. For example in a one-way ANOVA with random group effects, the estimated variance of the group effects will only be good asymptotically if the the number of groups goes to infinity. More generally, the method only requires a known asymptotic distribution for $\tilde{\theta}$.

Define

$$g(\theta) \equiv P_j'[X_1'(I-A_0)'V^{-1}(\theta)(I-A_0)X_1]^- P_j = P_j'[PD(1/\phi)P']^- P_j,$$

where we ignore the fact that P_j and A_0 are functions of θ. GB treated fixed estimable functions $\lambda'\beta$ but FC's application has estimable functions that depend on θ. GB need to compute the simple looking function $[X'V^{-1}(\theta)X]^-$ but, as discussed in the appendix, this involves $[X_1'(I-A_0)'V^{-1}(\theta)(I-A_0)X_1]^-$. Indeed, for the sake of taking derivatives, it may be better to actually apply GB's method to $\Lambda'\beta$ as discussed in the appendix.

Let $\mathbf{d}_\theta g(\theta)$ be the $1\times q$ row vector of partial derivatives of $g(\cdot)$. In particular, similar to Christensen (2019, Chapter 4), the kth component of the row vector is

$$\mathbf{d}_{\theta_k} g(\theta) \equiv P_j'[X_1'(I-A_0)'V^{-1}(\theta)\{\mathbf{d}_{\theta_k}V(\theta)\}V^{-1}(\theta)(I-A_0)X_1]^- P_j.$$

By the Delta Method

$$\sqrt{n}[g(\tilde{\theta}) - g(\theta)] \xrightarrow{\mathscr{L}} N[0, \mathbf{d}_\theta g(\theta)\mathbf{I}_r^{-1}(\theta)\mathbf{d}_\theta g(\theta)'].$$

Thus we take

$$v_j = 2n\frac{\{\tilde{P}_j'[X_1'(I-\tilde{A}_0)'\tilde{V}^{-1}(I-\tilde{A}_0)X_1]^-\tilde{P}_j\}^2}{\mathbf{d}_\theta g(\tilde{\theta})\mathbf{I}_r^{-1}(\tilde{\theta})\mathbf{d}_\theta g(\tilde{\theta})'}.$$

Replacing $\mathbf{I}_r^{-1}(\tilde{\theta})$ with $n\mathscr{I}_r^{-1}(\tilde{\theta})$, the formula becomes

$$v_j = 2\frac{\{\tilde{P}_j'[X_1'(I-\tilde{A}_0)'\tilde{V}^{-1}(I-\tilde{A}_0)X_1]^-\tilde{P}_j\}^2}{\mathbf{d}_\theta g(\tilde{\theta})\mathscr{I}_r^{-1}(\tilde{\theta})\mathbf{d}_\theta g(\tilde{\theta})'}.$$

Having found the derivative (which depends on j), it only remains to find $\mathscr{I}_r(\theta)$.

2.3 Derivation of $\mathscr{I}_r(\theta)$

Much of the material in this subsection is similar to Christensen (2019, Sections 4.3, 4.2). REML estimates are based on

$$U \equiv B'Y \sim N(0, V_r); \qquad V_r \equiv B'VB.$$

We know that, because $C(B) = C(X)^\perp$,

$$V^{-1}(\theta)[I - A(\theta)] = B(B'V(\theta)B)^{-1}B' = [I - A(\theta)]'V^{-1}(\theta).$$

By definition, $\mathscr{I}_r(\theta)$ is the covariance matrix of the score function $S(U,\theta)$ which is the transpose of the derivative of the log-likelihood. Our score vector has ith component of the form $S_i(U,\theta) = U'R_iU - h_i(\theta)$ with $\mathrm{E}(U'R_iU) = h_i(\theta)$. The i,j element of the information matrix is

$$\mathscr{I}_r(\theta)_{ij} \equiv \mathrm{E}\left\{[U'R_iU - h_i(\theta)][U'R_jU - h_j(\theta)]\right\} = \mathrm{Cov}[U'R_iU, U'R_jU]$$

From Christensen (2019, Section 4.3)

$$2h_j(\theta) = \mathrm{tr}\left\{[B'V(\theta)B]^{-1}[\mathbf{d}_{\theta_j}B'V(\theta)B]\right\}$$

and

$$2U'R_jU = Y'B[B'V(\theta)B]^{-1}[\mathbf{d}_{\theta_j}B'V(\theta)B][B'V(\theta)B]^{-1}B'Y\,,$$

or

$$2h_j(\theta) = \mathrm{tr}\left\{[B'V(\theta)B]^{-1}B'[\mathbf{d}_{\theta_j}V(\theta)]B\right\};$$

$$2U'R_jU = U'(B'V(\theta)B)^{-1}B'[\mathbf{d}_{\theta_j}V(\theta)]B[B'V(\theta)B]^{-1}U\,.$$

The proof of the following proposition is based on the Wishart matrix proof for Theorem 4.6.1 in Christensen (2019). See, e.g., http://stat.unm.edu/~fletcher/Wishart.pdf for the covariance matrix of the Vec of a Wishart.

Proposition. If $U \sim N(0, V_r)$, then

$$\mathrm{Cov}(U'R_iU, U'R_jU) = 2\mathrm{tr}(R_iV_rR_jV_r).$$

Using this proposition with the earlier implicit definition of R_i and the fact that $\mathbf{d}_{\theta_i}V^{-1}(\theta) = -V^{-1}\mathbf{d}_{\theta_i}V(\theta)V^{-1}$,

$$\begin{aligned}
&2\mathscr{I}_r(\theta)_{ij}\\
&= \mathrm{tr}\Big\{(B'V(\theta)B)^{-1}B'[\mathbf{d}_{\theta_i}V(\theta)]B[B'V(\theta)B]^{-1}B'VB(B'V(\theta)B)^{-1}B'[\mathbf{d}_{\theta_j}V(\theta)]\\
&\qquad \times B[B'V(\theta)B]^{-1}B'VB\Big\}\\
&= \mathrm{tr}\left\{(I-A)'V^{-1}[\mathbf{d}_{\theta_i}V(\theta)]V^{-1}(I-A)V(I-A)'V^{-1}[\mathbf{d}_{\theta_j}V(\theta)]V^{-1}(I-A)V\right\}\\
&= \mathrm{tr}\left\{(I-A)'[\mathbf{d}_{\theta_i}V^{-1}(\theta)](I-A)V(I-A)'[\mathbf{d}_{\theta_j}V^{-1}(\theta)](I-A)V\right\}\\
&= \mathrm{tr}\left\{[\mathbf{d}_{\theta_i}V^{-1}(\theta)](I-A)V(I-A)'[\mathbf{d}_{\theta_j}V^{-1}(\theta)](I-A)V(I-A)'\right\}\\
&= \mathrm{tr}\left\{[\mathbf{d}_{\theta_i}V^{-1}(\theta)](I-A)(I-A)V[\mathbf{d}_{\theta_j}V^{-1}(\theta)](I-A)(I-A)V\right\}\\
&= \mathrm{tr}\left\{[\mathbf{d}_{\theta_i}V^{-1}(\theta)](I-A)V[\mathbf{d}_{\theta_j}V^{-1}(\theta)](I-A)V\right\}
\end{aligned}$$

This agrees with KR1's formula (KR2 has a similar formula but contains a typo)

$$2\mathscr{I}_r(\theta)_{ij} = \text{tr}\left(\{\mathbf{d}_{\theta_i} V^{-1}(\theta)\} V \{\mathbf{d}_{\theta_j} V^{-1}(\theta)\} V\right) - \text{tr}\left(2\Phi Q_{ij} - \Phi P_i \Phi P_j\right),$$

where

$$\Phi = [X'V^{-1}(\theta)X]^{-1}, \qquad P_i = X'\{\mathbf{d}_{\theta_i} V^{-1}(\theta)\}X, \tag{2.4}$$

$$Q_{ij} = X'\{\mathbf{d}_{\theta_i} V^{-1}(\theta)\} V \{\mathbf{d}_{\theta_j} V^{-1}(\theta)\}X. \tag{2.5}$$

3 Kenward-Roger Approximation

As discussed in Christensen (2019, Subsection 4.7.1), under reasonable conditions

$$\text{E}\left[\widehat{\text{Var}}(\lambda'\hat{\beta}_1)\right] = \text{E}\left\{\lambda'[X_1'(I-\tilde{A}_0)'\tilde{V}^{-1}(I-\tilde{A}_0)X_1]^{-}\lambda\right\} \leq \text{Var}(\lambda'\hat{\beta}_1) \leq \text{Var}(\lambda'\tilde{\beta}_1).$$

FC and GB focus their methods on $\widehat{\text{Var}}(\lambda'\hat{\beta}_1)$ rather than trying to estimate the larger value $\text{Var}(\lambda'\tilde{\beta}_1)$. KR try to correct that.

KR begin with

$$\widehat{\text{Var}}(\lambda'\hat{\beta}_1) \equiv \lambda'[X_1'(I-\tilde{A}_0)'\tilde{V}^{-1}(I-\tilde{A}_0)X_1]^{-}\lambda = \lambda'\Phi(\tilde{\theta})\lambda,$$

where $\Phi \equiv \Phi(\theta)$ is defined implicitly. They seek an estimate of the bias term

$$\text{Var}(\lambda'\hat{\beta}_1) - \text{E}\left\{\lambda'[X_1'(I-\tilde{A}_0)'\tilde{V}^{-1}(I-\tilde{A}_0)X_1]^{-}\lambda\right\}$$

and they also seek an estimate of the bias term $\text{Var}(\lambda'\tilde{\beta}_1) - \text{Var}(\lambda'\hat{\beta}_1) \equiv \lambda'\Psi\lambda$. As far as I can tell, KR use the same second bias term estimate but for the first term the R programs `pbkrtest` and `lmerTest` only use results from KR1 and not from KR2.

In summery, if

$$\mathrm{E}(\tilde{\Phi}) = \Phi - \mathrm{E}(A)$$

then

$$\mathrm{E}(\tilde{\Phi} + A) = \Phi.$$

Since

$$\mathrm{Cov}(\tilde{\beta}) = \mathrm{Cov}(\hat{\beta}) + \mathrm{Cov}(\tilde{\beta} - \hat{\beta}) = \Phi + \Psi,$$

We estimate $\mathrm{Cov}(\tilde{\beta})$ with

$$\breve{\Phi} \equiv \tilde{\Phi} + \tilde{A} + \tilde{\Psi}. \tag{3.1}$$

Using the notation of (2.4) and (2.5), KR give this result as

$$\breve{\Phi} = \tilde{\Phi} + 2\tilde{\Phi}\left[\sum_{i=1}^{q}\sum_{j=1}^{q} w_{ij}\left(Q_{ij} - P_i \Phi P_j - \frac{1}{4}R_{ij}\right)\right]\tilde{\Phi},$$

where

$$R_{ij} = X'V^{-1}\mathbf{d}^2_{\theta_i\theta_j}V^{-1}X.$$

Note that if $V(\theta)$ is linear in θ, as in traditional variance component models, then $R_{ij} = 0$.

We examine the second bias term first.

3.1 Finding $\mathrm{Var}(\lambda'\tilde{\beta}_1) - \mathrm{Var}(\lambda'\hat{\beta}_1)$

Focus on a regression version of model (1.1) and all of β. Kacker and Harville (1984) argue that when using translation invariant estimates in $\tilde{\beta}$, we have $(\tilde{\beta} - \hat{\beta}) \perp\!\!\!\perp \hat{\beta}$, so $\mathrm{Cov}(\tilde{\beta}) = \mathrm{Cov}(\tilde{\beta} - \hat{\beta}) + \mathrm{Cov}(\hat{\beta})$. They proceed to estimate the bias term $\mathrm{Cov}(\tilde{\beta} - \hat{\beta})$. In particular, recalling from Christensen (2019, Subsection 4.7.1) that $\mathrm{E}(\tilde{\beta} - \hat{\beta}) = 0$,

$$\mathrm{Cov}(\tilde{\beta} - \hat{\beta}) = \mathrm{E}\left[(\tilde{\beta} - \hat{\beta})(\tilde{\beta} - \hat{\beta})'\right].$$

Recalling that $\tilde{\beta} \equiv \hat{\beta}(\tilde{\theta})$ and letting θ_0 be the true value of θ, use a first order Taylor's approximation to get

$$[\hat{\beta}(\tilde{\theta}) - \hat{\beta}(\theta_0)] \doteq \left[\mathbf{d}_\theta \hat{\beta}(\theta_0)\right](\tilde{\theta} - \theta_0).$$

Then,

$$\begin{aligned}
\mathrm{Cov}(\tilde{\beta} - \hat{\beta}) &= \mathrm{E}\left[(\tilde{\beta} - \hat{\beta})(\tilde{\beta} - \hat{\beta})'\right] \\
&\doteq \mathrm{E}\left\{\left[\mathbf{d}_\theta \hat{\beta}(\theta_0)\right](\tilde{\theta} - \theta_0)(\tilde{\theta} - \theta_0)'\left[\mathbf{d}_\theta \hat{\beta}(\theta_0)\right]'\right\} \\
&= \mathrm{E}\left\{\left[\mathbf{d}_\theta \hat{\beta}(\theta_0)\right]\mathrm{E}\left\{(\tilde{\theta} - \theta_0)(\tilde{\theta} - \theta_0)'\right\}\left[\mathbf{d}_\theta \hat{\beta}(\theta_0)\right]'\right\} \\
&\doteq \mathrm{E}\left\{\left[\mathbf{d}_\theta \hat{\beta}(\theta_0)\right]\mathrm{Cov}(\tilde{\theta})\left[\mathbf{d}_\theta \hat{\beta}(\theta_0)\right]'\right\} \\
&\equiv \Psi.
\end{aligned}$$

The third line follows from writing the previous expectation as first conditional on $\hat{\beta}$ and recognizing that $\hat{\beta}(\theta) \perp\!\!\!\perp \tilde{\theta}$. The penultimate equality requires $\tilde{\theta}$ to be unbiased, or at least asymptotically unbiased so we can use an asymptotic covariance matrix for $\tilde{\theta}$, typically $\mathscr{I}_r^{-1}(\tilde{\theta})$.

With $\beta = \mathrm{E}[\hat{\beta}(\theta)]$, typically $0 = \mathrm{E}\left[\mathbf{d}_\theta \hat{\beta}(\theta)\right]$, and the random vectors $\mathbf{d}_{\theta_j} \hat{\beta}(\theta)$ and $\mathbf{d}_\theta \hat{\beta}_i(\theta)'$ have mean 0. In particular, using the extension of a standard result on the expected value of quadratic forms, e.g, Christensen's (2020) Theorem 1.3.2 extended in his Exercise 10.1a, we compute

$$\begin{aligned}
\Psi_{rs} &= \mathrm{E}\left\{\left[\mathbf{d}_\theta \hat{\beta}_r(\theta_0)\right]\mathrm{Cov}(\tilde{\theta})\left[\mathbf{d}_\theta \hat{\beta}_s(\theta_0)\right]'\right\} \\
&= \mathrm{tr}\left\{\mathrm{Cov}(\tilde{\theta})\mathrm{Cov}\left[\mathbf{d}_\theta \hat{\beta}_s(\theta_0)', \mathbf{d}_\theta \hat{\beta}_r(\theta_0)'\right]\right\}.
\end{aligned}$$

To make further progress we need to examine properties of $\mathbf{d}_\theta \hat{\beta}(\theta)$. Using the notation of (2.4) and (2.5),

$$\begin{aligned}
\mathbf{d}_{\theta_j} \hat{\beta}(\theta) &= \mathbf{d}_{\theta_j}\left\{\left[X'V^{-1}(\theta)X\right]^{-1} X'V^{-1}(\theta)Y\right\} \\
&= \mathbf{d}_{\theta_j}\left\{\left[X'V^{-1}(\theta)X\right]^{-1}\right\} X'V^{-1}(\theta)Y + \left[X'V^{-1}(\theta)\right]^{-1}\, \mathbf{d}_{\theta_j} X'V^{-1}(\theta)Y
\end{aligned}$$

$$
\begin{aligned}
&= -\left[\Phi P_j \Phi\right] X' V^{-1}(\theta) Y + \Phi X' \left[\mathbf{d}_{\theta_j} V^{-1}(\theta)\right] Y \\
&= \left\{-\left[\Phi P_j \Phi\right] X' V^{-1}(\theta) + \Phi X' \left[\mathbf{d}_{\theta_j} V^{-1}(\theta)\right]\right\} Y.
\end{aligned}
$$

This gives

$$
\mathbf{d}_\theta \hat{\beta}(\theta) = \left\{-\left[\Phi P_1 \Phi\right] X' V^{-1} + \Phi X' \left[\mathbf{d}_{\theta_1} V^{-1}\right], \ldots, -\left[\Phi P_q \Phi\right] X' V^{-1} + \Phi X' \left[\mathbf{d}_{\theta_q} V^{-1}\right]\right\} Y.
$$

Write

$$
B_i \equiv -\left[\Phi P_i \Phi\right] X' V^{-1} + \Phi X' \left[\mathbf{d}_{\theta_i} V^{-1}\right]
$$

so

$$
\mathbf{d}_\theta \hat{\beta}(\theta) = \left\{B_1, \ldots, B_q\right\} Y.
$$

Let e_r be a vector with all 0s except a 1 in the rth place

$$
\operatorname{Cov}\left[\mathbf{d}_\theta \hat{\beta}_r(\theta)', \mathbf{d}_\theta \hat{\beta}_s(\theta)'\right] = \operatorname{Cov}\left\{[e_r' \mathbf{d}_\theta \hat{\beta}(\theta)]', [e_s' \mathbf{d}_\theta \hat{\beta}(\theta)]'\right\} = \operatorname{Cov}\left(\begin{bmatrix} e_r' B_1 Y \\ \vdots \\ e_r' B_q Y \end{bmatrix}, \begin{bmatrix} e_s' B_1 Y \\ \vdots \\ e_s' B_q Y \end{bmatrix}\right).
$$

The ij element of this covariance matrix is

$$
\begin{aligned}
&\operatorname{Cov}\left(e_r' B_i Y, e_s' B_j Y\right) \\
&= e_r B_i V B_j' e_s \\
&= e_r' \left(\left\{-\left[\Phi P_i \Phi\right] X' V^{-1} + \Phi X' \left[\mathbf{d}_{\theta_i} V^{-1}\right]\right\} V \left\{-\left[\Phi P_j \Phi\right] X' V^{-1} + \Phi X' \left[\mathbf{d}_{\theta_j} V^{-1}\right]\right\}'\right) e_s \\
&= e_r' \left\{\left[\Phi P_i \Phi\right] \Phi^{-1} \left[\Phi P_j \Phi\right] - \Phi P_i \Phi P_j \Phi - \Phi P_i \Phi P_j \Phi + \Phi Q_{ij} \Phi\right\} e_s \\
&= e_r' \left(\Phi Q_{ij} \Phi - \Phi P_i \Phi P_j \Phi\right) e_s \\
&= e_r' \Phi \left(Q_{ij} - P_i \Phi P_j\right) \Phi e_s.
\end{aligned}
$$

As shown earlier, with $W \equiv \operatorname{Cov}(\tilde{\theta})$, $\Psi_{rs} = \operatorname{tr}\left(W \operatorname{Cov}\left[\mathbf{d}_\theta \hat{\beta}_r(\theta)', \mathbf{d}_\theta \hat{\beta}_s(\theta)'\right]\right)$, so, recalling that for symmetric A and B, $\mathrm{tr}(AB) = \sum_i \sum_j a_{ij} b_{ij}$, we can now write

$$\begin{aligned}
\Psi_{rs} &= \mathrm{tr}\left(W\mathrm{Cov}\left[\mathbf{d}_\theta \hat{\beta}_r(\theta)', \mathbf{d}_\theta \hat{\beta}_s(\theta)'\right]\right) \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} w_{ij} e_r' \Phi \left(Q_{ij} - P_i \Phi P_j\right) \Phi e_s \\
&= e_r' \Phi \left[\sum_{i=1}^{q}\sum_{j=1}^{q} w_{ij} \left(Q_{ij} - P_i \Phi P_j\right)\right] \Phi e_s
\end{aligned}$$

so that, as stated in KR,

$$\Psi = \Phi \left[\sum_{i=1}^{q}\sum_{j=1}^{q} w_{ij} \left(Q_{ij} - P_i \Phi P_j\right)\right] \Phi.$$

This will be estimated by

$$\tilde{\Psi} = \tilde{\Phi} \left[\sum_{i=1}^{q}\sum_{j=1}^{q} \tilde{w}_{ij} \left(\tilde{Q}_{ij} - \tilde{P}_i \tilde{\Phi} \tilde{P}_j\right)\right] \tilde{\Phi}.$$

Kacker and Harville (1984) near *their* equation (2.3) indicate that this bias estimator is exactly correct under conditions that seem to hold for generalized split plot models. In particular, it holds when the covariance matrix of $\tilde{\theta}$ is exact, the generalized least squares estimate does not depend on θ, and their (2.3) is exact.

I think this idea may be useful

$$\mathbf{d}_\theta X'V(\theta)Y = \left[\mathbf{d}_{\theta_1} X'V(\theta)Y, \ldots, \mathbf{d}_{\theta_q} X'V(\theta)Y\right]$$

3.2 Finding $\mathrm{Var}(\lambda'\hat{\beta}_1) - \mathrm{E}\left\{\lambda'[X_1'(I-\tilde{A}_0)'\tilde{V}^{-1}(I-\tilde{A}_0)X_1]^{-}\lambda\right\}$

KR also estimate the bias $\mathrm{Var}(\lambda'\hat{\beta}_1) - \mathrm{E}\left\{\lambda'[X_1'(I-\tilde{A}_0)'\tilde{V}^{-1}(I-\tilde{A}_0)X_1]^{-}\lambda\right\}$

Using a second-order Taylor approximation,

$$\Phi_{ij}(\tilde{\theta}) \doteq \Phi_{ij}(\theta_0) + [\mathbf{d}_\theta \Phi_{ij}(\theta_0)](\tilde{\theta} - \theta_0) + \frac{1}{2}(\tilde{\theta} - \theta_0)'[\mathbf{d}^2_{\theta\theta} \Phi_{ij}(\theta_0)](\tilde{\theta} - \theta_0).$$

Assuming that $\tilde{\theta}$ is unbiased,

$$
\begin{aligned}
\mathrm{E}\left[\Phi_{ij}(\tilde{\theta})\right] &\doteq \Phi_{ij}(\theta_0) + [\mathbf{d}_{\theta}\Phi_{ij}(\theta_0)]\mathrm{E}\left[(\tilde{\theta}-\theta_0)\right] + \frac{1}{2}\mathrm{E}\left[(\tilde{\theta}-\theta_0)'[\mathbf{d}^2_{\theta\theta}\Phi_{ij}(\theta_0)](\tilde{\theta}-\theta_0)\right] \\
&= \Phi_{ij}(\theta_0) + \frac{1}{2}\mathrm{tr}\left\{W[\mathbf{d}^2_{\theta\theta}\Phi_{ij}(\theta_0)]\right] \qquad (3.2)
\end{aligned}
$$

As illustrated earlier,

$$
\mathbf{d}_{\theta_r}\Phi_{ij}(\theta) = e_i'\mathbf{d}_{\theta_r}\left\{\left[X'V^{-1}(\theta)X\right]^{-1}\right\}e_j = -e_i'\left[\Phi P_r \Phi\right]e_j,
$$

so

$$
\begin{aligned}
&\mathbf{d}^2_{\theta_r\theta_s}\Phi_{ij}(\theta) \\
&\quad= -e_i'\left[\mathbf{d}_{\theta_s}\Phi P_r\Phi\right]e_j \\
&\quad= -e_i'\left[\mathbf{d}_{\theta_s}\Phi\right]P_r\Phi e_j - e_i'\Phi\left[\mathbf{d}_{\theta_s}P_r\right]\Phi e_j - e_i'\Phi P_r\left[\mathbf{d}_{\theta_s}\Phi\right]e_j \\
&\quad= e_i'\left[\Phi P_s\Phi\right]P_r\Phi e_j - e_i'\Phi\left[\mathbf{d}_{\theta_s}P_r\right]\Phi e_j + e_i'\Phi P_r\left[\Phi P_s\Phi\right]e_j \\
&\quad= e_i'\Phi P_s\Phi P_r\Phi e_j - e_i'\Phi\left[\mathbf{d}_{\theta_s}P_r\right]\Phi e_j + e_i'\Phi P_r\Phi P_s\Phi e_j
\end{aligned}
$$

While it is possible that the covariance matrix might be specified in terms of $V^{-1}(\theta)$, more commonly it is specified as $V(\theta)$. Following KR, we have specified most derivatives in terms of $\mathbf{d}_{\theta_r}V^{-1}(\theta)$ but as in ALM-III,

$$
\mathbf{d}_{\theta_r}V^{-1}(\theta) = -V^{-1}(\theta)[\mathbf{d}_{\theta_r}V(\theta)]V^{-1}(\theta),
$$

so $\mathbf{d}_{\theta_r}V^{-1}(\theta)$ is easily found in terms of $\mathbf{d}_{\theta_r}V(\theta)$. Finding $e_i'\Phi\left[\mathbf{d}_{\theta_s}P_r\right]\Phi e_j$ in terms of derivatives of $V^{-1}(\theta)$ is easy using

$$
\mathbf{d}^2_{sr}V^{-1}(\theta) = \mathbf{d}_{\theta_s}[\mathbf{d}_{\theta_r}V^{-1}(\theta)],
$$

but finding it in terms of derivatives of $V(\theta)$ is considerably more work. We begin by looking at $\mathbf{d}^2_{sr}V^{-1}(\theta)$ in terms of $\mathbf{d}_{\theta_r}V^{-1}(\theta)$ and second derivatives of $V(\theta)$.

$$
\begin{aligned}
&\mathbf{d}_{\theta_s}[\mathbf{d}_{\theta_r}V^{-1}(\theta)] \\
&\quad = \mathbf{d}_{\theta_s}\{-V^{-1}(\theta)[\mathbf{d}_{\theta_r}V(\theta)]V^{-1}(\theta)\} \\
&\quad = -\left\{[\mathbf{d}_{\theta_s}V^{-1}(\theta)][\mathbf{d}_{\theta_r}V(\theta)]V^{-1} + V^{-1}[\mathbf{d}_{\theta_s}[\mathbf{d}_{\theta_r}V(\theta)]]V^{-1} + V^{-1}[\mathbf{d}_{\theta_r}V(\theta)][\mathbf{d}_{\theta_s}V^{-1}(\theta)]\right\} \\
&\quad = -\left\{-V^{-1}[\mathbf{d}_{\theta_s}V]V^{-1}[\mathbf{d}_{\theta_r}V]V^{-1} + V^{-1}[\mathbf{d}^2_{\theta_s\theta_r}V]V^{-1} - V^{-1}[\mathbf{d}_{\theta_r}V]V^{-1}[\mathbf{d}_{\theta_s}V]V^{-1}\right\} \\
&\quad = [\mathbf{d}_{\theta_s}V^{-1}]V[\mathbf{d}_{\theta_r}V^{-1}] - V^{-1}[\mathbf{d}^2_{\theta_s\theta_r}V(\theta)]V^{-1} - [\mathbf{d}_{\theta_r}V^{-1}]V[\mathbf{d}_{\theta_s}V^{-1}]
\end{aligned}
$$

so relating back to KR's notation,

$$X'\mathbf{d}_{\theta_s}[\mathbf{d}_{\theta_r}V^{-1}(\theta)]X = Q_{sr} - R_{sr} + Q_{rs}$$

and

$$e_i'\Phi\,[\mathbf{d}_{\theta_s}P_r]\,\Phi e_j = e_i'\Phi(Q_{sr} - R_{sr} + Q_{rs})\Phi e_j$$

and

$$\mathbf{d}^2_{\theta_r\theta_s}\Phi_{ij}(\theta) = e_i'\Phi P_s\Phi P_r\Phi e_j + e_i'\Phi P_r\Phi P_s\Phi e_j - e_i'\Phi(Q_{sr} - R_{sr} + Q_{rs})\Phi e_j.$$

which agrees with KR1.

From (3.2) our desired bias term is

$$
\begin{aligned}
\Phi_{ij}(\theta) - \mathrm{E}\left[\Phi_{ij}(\tilde{\theta})\right] &= -\frac{1}{2}\mathrm{tr}\left\{W[\mathbf{d}^2_{\theta\theta}\Phi_{ij}(\theta)]\right\} \\
&= -e_i'\Phi\left[\sum_r\sum_s w_{rs}\left(P_s\Phi P_r - Q_{sr} + \frac{1}{2}R_{sr}\right)\right]\Phi e_j \\
&= e_i'\Phi\left[\sum_r\sum_s w_{rs}\left(Q_{sr} - P_s\Phi P_r - \frac{1}{2}R_{sr}\right)\right]\Phi e_j
\end{aligned}
$$

and

$$\Phi(\theta) - \mathrm{E}\left[\Phi(\tilde{\theta})\right] \quad = \quad \Phi\left[\sum_r\sum_s w_{rs}\left(Q_{sr} - P_s\Phi P_r - \frac{1}{2}R_{sr}\right)\right]\Phi$$

Similar to (3.1), the estimate of $\mathrm{Cov}(\hat{\beta})$ is

$$\begin{aligned}
\breve{\Phi} &= \tilde{\Phi} + \tilde{\Phi}\left[\sum_r\sum_s \tilde{w}_{rs}\left(\tilde{Q}_{sr} - \tilde{P}_s\tilde{\Phi}\tilde{P}_r - \frac{1}{2}\tilde{R}_{sr}\right)\right]\tilde{\Phi} \\
&\quad + \tilde{\Phi}\left[\sum_{i=1}^{q}\sum_{j=1}^{q} \tilde{w}_{ij}\left(\tilde{Q}_{ij} - \tilde{P}_i\tilde{\Phi}\tilde{P}_j\right)\right]\tilde{\Phi} \\
&= \tilde{\Phi} + 2\tilde{\Phi}\left[\sum_r\sum_s \tilde{w}_{rs}\left(\tilde{Q}_{sr} - \tilde{P}_s\tilde{\Phi}\tilde{P}_r - \frac{1}{4}\tilde{R}_{sr}\right)\right]\tilde{\Phi}.
\end{aligned} \tag{3.3}$$

This is the result in KR1 that seems to be what is incorporated into the R programs `pbkrtest` and `lmerTest`. KR2 incorporate a first order bias approximation for $\mathrm{E}\left[(\tilde{\theta} - \theta_0)\right]$ based on the REML estimates (using the score function of the restricted data $B'Y$).

3.3 KR Test Statistic and Degrees of Freedom

Using (3.3) to define

$$\widehat{\mathrm{Cov}}(\tilde{\beta}) \equiv \breve{\Phi}.$$

the KR1 test statistic for testing $\Lambda'\beta = 0$ with $r(\Lambda) = s$, is

$$F_{KR} = \frac{1}{s}(\Lambda'\tilde{\beta})[\Lambda'\breve{\Phi}\Lambda]^{-}\beta(\Lambda'\tilde{\beta}).$$

To find the degrees of freedom, they find approximate values for $\mathrm{E}(F_{KR})$ and $\mathrm{Var}(F_{KR})$, use these to determine a coefficient of variation and take the denominator degrees of freedom ν to be the value for which an $F(s,\nu)$ distribution has the same coefficient of variation. In addition, they introduce a scale factor δ (that does not affect the coefficient of variation) so that

$$\mathrm{E}(\delta F_{KR}) = \mathrm{E}[F(s,\nu)] = \nu/(\nu-2),$$

and approximate

$$F_* \equiv \delta F_{KR} \sim F(s,\nu).$$

In more detail, $\mathrm{Var}[F(s,\nu)] = \{\mathrm{E}[F((s,\nu)]\}^2 2(s+\nu-2)/s(\nu-4)$, so ν is obtained by setting

$$\frac{\mathrm{Var}(F_{KR})}{[\mathrm{E}(F_{KR})]^2} = \frac{2(s+\nu-2)}{s(\nu-4)}$$

or

$$\nu = 4 + \frac{s+2}{\frac{s}{2}\frac{\mathrm{Var}(F_{KR})}{\{\mathrm{E}(F_{KR})\}^2} - 1}$$

It remains to find $\mathrm{E}(F_{KR})$ and $\mathrm{Var}(F_{KR})$ which is done through a (second order?) Taylor's approximation and

$$\mathrm{E}(F) = \mathrm{E}_{\breve{\Phi}}[E(F|\breve{\Phi})]$$

$$\mathrm{Var}(F) = \mathrm{E}_{\breve{\Phi}}[\mathrm{Var}(F|\breve{\Phi})] + \mathrm{Var}_{\breve{\Phi}}[\mathrm{E}(F|\breve{\Phi})]$$

I currently have no justification for the KR1 claims that follow.

$$\mathrm{E}(F) = 1/[1+\frac{A_2}{s}] \doteq [1+\frac{A_2}{s}]$$

$$\mathrm{Var}(F) = \frac{2}{s}\frac{1+c_1 B}{(1-c_2 B)^2(1-c_3 B)} \doteq \frac{2}{s}(1+B)$$

$$B = \frac{1}{2s}(A_1 + 6A_2)$$

$$A_1 = \sum_i^q \sum_j w_{ij} \mathrm{tr}(\Theta\Phi P_i \Phi)\mathrm{tr}(\Theta\Phi P_j \Phi)$$

$$A_2 = \sum_i \sum_j w_{ij} \text{tr}(\Theta\Phi P_i \Phi\Theta\Phi P_j \Phi)$$

$$\Theta = \Lambda[\Lambda'\Phi\Lambda]^- \Lambda'$$

$$c_1 = \frac{g}{3s+2(1-g)}, \quad c_2 = \frac{s-g}{3s+2(1-g)}, \quad c_3 = \frac{s+2-g}{3s+2(1-g)}.$$

$$g = \frac{(s+1)A_1 - (s-4)A_2}{(s+2)A_2}$$

If $\text{E}(F) = 1/[1+\frac{A_2}{s}] = \nu/(\nu-2)$ then $A_2 = -2s/\nu$. Also $\Theta\Phi P_i \Phi = A_\Lambda P_i \Phi$.

4 Application to Generalized Split Plot Models

Because of the nice structure of the split plot covariance matrix you can find V^{-1} easily. It looks like computing $Y'(A-A_0)V^{-1}(A-A_0)Y$ should just give the split plot tests. REML estimates should be the obvious ones. So program `lmer` works.

$$F = \frac{Y'(A-A_0)'V_*^{-1}(A-A_0)Y/r(A-A_0)}{Y'(I-A)'V_*^{-1}(I-A)Y/r(I-A)}$$

$$V = \sigma^2 V_*, \qquad \sigma^2 = \sigma_w^2 + \sigma_s^2, \qquad \rho = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_s^2}$$

$$V_* = (1-\rho)I + m\rho M_1$$

For any projection operator P,

$$[aI + bP]^{-1} = \frac{1}{a}\left[I + \frac{b}{a+b}P\right],$$

so

$$V_*^{-1} = \frac{1}{1-\rho}\left[I + \frac{m\rho}{(1-\rho)+m\rho}M_1\right]$$

Testing models is in PA 11.2.2

For model (8), $A - A_0 = M_* - M_{*0}$

$$\begin{aligned}
V^{-1}(A-A_0) &= \frac{1}{\sigma^2(1-\rho)}\left[I + \frac{m\rho}{(1-\rho)+m\rho}M_1\right](M_* - M_{*0}) \\
&= \frac{1}{\sigma^2(1-\rho)}\left[M_1 + \frac{m\rho}{(1-\rho)+m\rho}M_1\right](M_* - M_{*0}) \\
&= \frac{1}{\sigma^2(1-\rho)}\left[1 + \frac{m\rho}{(1-\rho)+m\rho}\right](M_* - M_{*0}) \\
&= \frac{1}{\sigma^2(1-\rho)}\left[\frac{1-\rho}{(1-\rho)+m\rho}\right](M_* - M_{*0}) \\
&= \frac{1}{\sigma^2}\left[\frac{1}{(1-\rho)+m\rho}\right](M_* - M_{*0})
\end{aligned}$$

$$\mathrm{E}[MSE(1)] = \sigma^2[(1-\rho)+m\rho]$$

For model (9), $A - A_0 = M_2 - M_3$

$$\begin{aligned}
V^{-1}(A-A_0) &= \frac{1}{\sigma^2(1-\rho)}\left[I + \frac{m\rho}{(1-\rho)+m\rho}M_1\right](M_2 - M_3) \\
&= \frac{1}{\sigma^2(1-\rho)}(M_2 - M_3)
\end{aligned}$$

$$\mathrm{E}[MSE(2)] = \sigma^2(1-\rho)$$

Qeadan (2014, Proposition 3.1.1) and Christensen (2019, Exercise 4.6) establish that if

$$MSE(1) > MSE(2)$$

then the split plot error estimates are REML estimates. In particular, we can reparameterize as $\theta_1 = \sigma_s^2$, $\theta_2 = \sigma_s^2 + m\sigma_w^2$, so

$$V = \theta_1(I - M_1) + \theta_2 M_1$$

For normal data, exact covariance matrix is variance of quadratic forms and independent.

Large sample covariance matrix must agree.

5 References

Christensen, Ronald (2019). *Advanced Linear Modeling: Statistical Learning and Dependent Data*, Third Edition. Springer-Verlag, New York.

Christensen, Ronald (2020). *Plane Answers to Complex Questions: The Theory of Linear Models*, Fifth Edition. Springer, New York.

Cressie, Noel and Lahiri, S.N. (1993). The asymptotic distribution of REML estimates. *Journal of Multivariate Analysis*, **45**, 217-233.

Fai, A.H. and Cornelius, P.L. (1996). Approximate F-Tests of Multiple Degree of Freedom Hypotheses in Generalised Least Squares Analyses of Unbalanced Split-Plot Experiments." *Journal of Statistical Computation and Simulation*, 54(4), 363-378.

Giesbrecht, F. G. and Burns, J. C. (1985). Two-Stage Analysis Based on a Mixed Model: Large-Sample Asymptotic Theory and Small-Sample Simulation Results. *Biometrics*, Vol. 41, No. 2 (Jun., 1985), pp. 477-486

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

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

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

Kenward, M.G., Roger, J.H., (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics and Data Analysis*, **53**, 2583-2595.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13). *Journal of Statistical Software*

Qeadan, Fares (2014). *On the Equivalence Between the LRT, RLRT and F-test for Testing Variance Components in the Generalized Split-plot Models.* Ph.D. Dissertation, University of New Mexico

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

6 Appendix: GLS — Testing models versus hypotheses

This section gives a direct demonstration that the sums of squares for testing a reduced model is the same as the sum of squares for testing a corresponding estimable hypothesis. In Christensen (2020, Section 3.8) this was done indirectly (and much more simply). To explore methods of approximating F denominatior degrees of freedom for tests with estimated covariance matrices, it is useful the know the direct derivation.

Consider a model

$$Y = X\beta + e, \qquad e \sim N[0, V(\theta)] \tag{6.1}$$

in partitioned form

$$Y = X_0\beta_0 + X_1\beta_1 + e, \qquad e \sim N[0, V(\theta)]. \tag{6.2}$$

With $M_0 \equiv X_0(X_0'X_0)^{-}X_0'$ being the Euclidean-distance perpendicular projection operator onto $C(X_0)$, testing the reduced model that does not include β_1 (for which it is sufficient but not necessary that $\beta_1 = 0$) is equivalent to testing $(I - M_0)X_1\beta_1 = 0$ or $(I - A_0)X_1\beta_1 = 0$. Note that each of these implies the other because

$$(I-M_0)(I-A_0)=(I-M_0); \qquad (I-A_0)(I-M_0)=(I-A_0).$$

Moreover, $(I-M_0)X_1\beta_1=0$ iff $(I-M_0)X_1\beta_1 \perp C[(I-M_0)X_1]$ iff $X_1'(I-M_0)X_1\beta_1=0$. This last form of the hypothesis is convenient because it does not involve θ and (typically) has fewer than n rows.

We want to show that the sum of squares for testing the reduced model is the same as the sum of squares for testing $X_1'(I-M_0)X_1\beta_1=0$. The sum of squares for testing the reduced model is $Y'A_1'V^{-1}A_1Y=\hat{\beta}_1'\{X_1'(I-A_0)'V^{-1}(I-A_0)X_1\}\hat{\beta}_1$ (as shown earlier) and with $\Lambda'=[0,X_1'(I-M_0)X_1]$ the sum of squares for testing the hypothesis is

$$(\Lambda'\hat{\beta})'[\Lambda'(X'V^{-1}X)^{-}\Lambda]^{-}(\Lambda'\hat{\beta}).$$

We now examine the various pieces of this last sum of squares.

It is easy to see that

$$\Lambda'\hat{\beta}=X_1'(I-M_0)X_1\hat{\beta}_1.$$

We now work on computing $[\Lambda'(X'V^{-1}X)^{-}\Lambda]^{-}$. Clearly,

$$(X'V^{-1}X)=\begin{bmatrix} X_0'V^{-1}X_0 & X_0'V^{-1}X_1 \\ X_1'V^{-1}X_0 & X_1'V^{-1}X_1 \end{bmatrix}.$$

From the formula for finding (generalized) inverses of partitioned matrices, e.g. Christensen (2020, Exercise B.21),

$$(X'V^{-1}X)^{-}=\begin{bmatrix} \cdots & \cdots \\ \cdots & \{X_1'V^{-1}X_1-X_1'V^{-1}X_0[X_0'V^{-1}X_0]^{-}X_0'V^{-1}X_1\}^{-} \end{bmatrix},$$

where, because of the structure of Λ, we do not need explicit forms except for the bottom right entry. Exploiting the structure of Λ,

$$\Lambda'(X'V^{-1}X)^{-}\Lambda = [X_1'(I-M_0)X_1]\{X_1'V^{-1}X_1 - X_1'V^{-1}X_0[X_0'V^{-1}X_0]^{-}X_0'V^{-1}X_1\}^{-}[X_1'(I-M_0)X_1].$$

The center term can be simplified using

$$\{X_1'V^{-1}X_1 - X_1'V^{-1}X_0[X_0'V^{-1}X_0]^{-}X_0'V^{-1}X_1\} = \{X_1'(I-A_0)'V^{-1}(I-A_0)X_1\},$$

so

$$\Lambda'(X'V^{-1}X)^{-}\Lambda = [X_1'(I-M_0)X_1]\{X_1'(I-A_0)'V^{-1}(I-A_0)X_1\}^{-}[X_1'(I-M_0)X_1]$$

and taking generalized inverses,

$$[\Lambda'(X'V^{-1}X)^{-}\Lambda]^{-} = [X_1'(I-M_0)X_1]^{-}\{X_1'(I-A_0)'V^{-1}(I-A_0)X_1\}[X_1'(I-M_0)X_1]^{-}.$$

This allows us to compute the sums of squares,

$$\hat{\beta}'\Lambda[\Lambda'(X'V^{-1}X)^{-}\Lambda]^{-}\Lambda'\hat{\beta} =$$
$$\hat{\beta}_1'[X_1'(I-M_0)X_1][X_1'(I-M_0)X_1]^{-}\{X_1'(I-A_0)'V^{-1}(I-A_0)X_1\}[X_1'(I-M_0)X_1]^{-}[X_1'(I-M_0)X_1]\hat{\beta}_1.$$

To further simplify the sums of squares, recall that for any matrix Q and B with $C(B) \subset C(Q)$, $QQ^{-}B = B$. Therefore, since $\{X_1'(I-A_0)'V^{-1}(I-A_0)X_1\} = \{X_1'(I-M_0)(I-A_0)'V^{-1}(I-A_0)(I-M_0)X_1\}$, we get

$$[X_1'(I-M_0)X_1][X_1'(I-M_0)X_1]^{-}\{X_1'(I-A_0)'V^{-1}(I-A_0)X_1\}[X_1'(I-M_0)X_1]^{-}[X_1'(I-M_0)X_1]$$
$$= \{X_1'(I-A_0)'V^{-1}(I-A_0)X_1\}$$

and

$$\hat{\beta}'\Lambda[\Lambda'(X'V^{-1}X)^{-}\Lambda]^{-}\Lambda'\hat{\beta} = \hat{\beta}_1'\{X_1'(I-A_0)'V^{-1}(I-A_0)X_1\}\hat{\beta}_1 = Y'A_1V^{-1}A_1Y.$$