EVALUATING ADJUSTMENTS TO THE MEAN SQUARED ERROR DUE TO ESTIMATING VARIANCE PARAMETERS IN LINEAR MIXED MODELS

By

JAMIE MCCLAVE BALDWIN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2007
© 2007 Jamie McClave Baldwin
To my husband, Ian, and to my children, Caryss, Zachary, and “Baby” Baldwin
ACKNOWLEDGMENTS

First and most importantly, I thank my husband, Ian Baldwin, for providing endless patience over the last 10 years while I have pursued this degree. I would not have been able to endure this process without his unwavering support and constant belief in me, even when I couldn’t believe in myself. I thank him for maintaining both the house and my sanity when times were hard and for making me smile when days were too long to manage. I express my eternal love and gratitude to my children, Caryss, Zach, and Baby Number 3, for their infinite supply of laughter and hugs. They have inspired me to push beyond any limitation and live life to the fullest. They are the lights of my life and the most beautiful souls I have ever encountered and I give thanks for them everyday. My family is a daily demonstration of God’s Grace; there is nothing I could ever do to deserve their unconditional love and support.

I thank my parents, Jim and Mary Jay McClave, for their hard work, love, patience, and constant support. To my father, I am so grateful to have such a wonderful role model, both professionally and personally. I am forever indebted to him for continual editing and mathematical advice. I am so fortunate to have the opportunity to work with the smartest and bravest man that I know, not to mention the best consultant in the industry. To my mom, I am so thankful for her wise parenting advice, too many hours of childcare to even count, and girl time when it was most needed. I literally could not have finished this monstrous endeavor without her emergency child care services. I also thank them for all of the generosity they have shown me and my family in so many ways.

I would also like to thank my brother, Will McClave, his wife, Amber, and my awesome nephews, Jake and Luke. They have been a constant support for my family and I look forward to the next chapter of our lives together. I am so blessed to have such a wonderful, supportive family that lives so close by. It is a rarity in this world which I truly appreciate.
To the best friends anyone could ask for, I am thankful for their friendship, love, and encouragement through this entire process. I thank Holly Holly for always helping me see the bright side, laughing at my stories, listening to my endless complaints, and allowing me to learn so much about motherhood and balance from her wonderful example. I thank Heather Bristol for being our pioneer through the statistics program. She has shown me what true bravery is and taught me how to deal with the most unexpected of situations with grace and strength. I thank Dr. Patches Johnson for encouraging my continuation when I thought I would go no further, and for being my mentor in so many ways. I thank her for being a fantastic and fun traveling companion and for her contagious laugh. I thank Meghan Medlock Abraham, my college roommate, for a friendship that picks up right where it was left off no matter how much time has passed and for having a mouth big enough to eat her own fist. I thank Brooke Bloomberg, my “oldest” and dearest friend, for her sisterly bond and for beating me to the finish line; it was all the motivation I needed! It is amazing that we have both accomplished our childhood ambitions. We rock!!

I am grateful for other friends that have helped me along the way with their support and laughter. My sincerest appreciation goes to Andy and Tracy Bachmann, Eric and Ann Bloomberg Beshore, Nicole Provost and Kathy Carroll, Debbie Hagan, Lisa Jamba, Ron and Cindy Marks, Larry and Sandy Reimer, and Yvette and Scott Silvey. I consider each one of them family and I am so lucky to have them in my life.

Finally, I thank Dr. Ramon Littell for his patience and guidance during this seemingly endless project. Without his loyal support and enduring belief, this dissertation could not have been completed. This degree would not have been possible without his wealth of knowledge, generous spirit, and encouragement.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS .........................................................................................................................4

LIST OF TABLES ......................................................................................................................................8

LIST OF FIGURES ....................................................................................................................................9

ABSTRACT ...........................................................................................................................................12

CHAPTER

1 INTRODUCTION AND BACKGROUND ..................................................................................14

1.1 Introduction to Mixed Models ...............................................................................................14

1.2 The Linear Mixed Model ........................................................................................................15

1.3 Best Linear Unbiased Predictors (BLUPs) When $\theta$ is Known: Mixed Model Equations .......................................................................................................................15

1.4 Empirical Best Linear Unbiased Estimation (EBLUP) ........................................................16

1.5 Two-Stage Estimation and Unbiasedness ............................................................................17

2 LITERATURE REVIEW ...........................................................................................................19

2.1 Precision of EBLUPs ...............................................................................................................19

2.1.1 Mean Squared Error of Prediction (MSEP) of EBLUPs ..................................................19

2.1.2 Kackar-Harville Approximation to MSEP .......................................................................22

2.1.3 Estimating $C(\theta)$ ...........................................................................................................24

2.1.4 Estimating $M_1(\theta)$ .......................................................................................................25

2.1.5 Estimators of $M(\theta)$ Based on the Kackar-Harville Approximation .........................26

2.2 Hypothesis Testing for Single Dimension Linear Combinations ........................................27

2.3 Hypothesis Testing Methods for Multiple Dimension Linear Combinations ....................29

2.3.1 Kenward-Roger Multiple Dimension Hypothesis Testing Method ...............................29

2.3.2 Fai-Cornelius Multiple Dimension Hypothesis Testing Methods ...............................35

2.4 Negative Variance Component Estimates ............................................................................38

3 ESTIMATES OF MSEP FOR MULTIPLE DIMENSION LINEAR COMBINATIONS ....42

3.1 Reconciling Kenward-Roger (1997) and Fai-Cornelius (1996) Definitions of the Correction Term Approximation ..........................................................................................42

3.1.1 Linear Combinations of Fixed Effects ...........................................................................42

3.1.2 Linear Combinations of Fixed and Random Effects .......................................................44

3.2 Multiple Dimension Bias Correction ...................................................................................49

3.2.1 Bias Correction Estimator for Multiple Dimension Linear Combinations of Fixed Effects .........................................................................................................................49
3.2.2 Bias Correction Estimator for Multiple Dimension Linear Combinations of Fixed and Random Effects

3.2.3 Definiteness Properties of the Estimator for MSEP for Multiple-Dimension Linear Combinations of Fixed and Random Effects

3.3 Transform Invariance of MSEP Estimator

4 BALANCED ONE-WAY RANDOM EFFECTS MODEL

4.1 Kackar-Harville Approximation Based on REML Estimation Method

4.2 Kackar-Harville Approximation Based on ANOVA Estimation Methods

4.3 Bias Correction Term for the Balanced One-Way Random Effects Model

4.3.1 Bias Correction Term Approximation under $\theta$ Parameterization

4.3.2 Expected Value of $\text{BIAS}^{KR}(\theta)$ with REML Covariance Parameter Estimation

4.3.3 Expected Value of $\text{BIAS}^{KR}(\theta)$ with ANOVA Covariance Parameter Estimation

4.3.4 Impact of Negative Values of Bias Correction Term Approximation

4.4 Performance of Overall Estimators for the MSEP

4.5 Comparison of Prediction Interval Methods

5 THE EFFECT OF NEGATIVE VARIANCE COMPONENTS ON EBLUP

5.1 Negative Variance Component Estimates and the Balanced One Way Random Effects Model

5.2 Considering the Variance Parameter as a Covariance

5.3 BLUP derivation for Random Effects Model with Correlated Errors

6 CONCLUSION AND FUTURE WORK

APPENDIX DETAILS FOR MONTE CARLO SIMULATION STUDY ON PREDICTION INTERVALS

LIST OF REFERENCES

BIOGRAPHICAL SKETCH
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Summary of developments for MSEP estimation and hypothesis testing</td>
<td>40</td>
</tr>
<tr>
<td>4-1</td>
<td>Accuracy of Kackar-Harville estimator (4-17) for $C_r(\theta)$</td>
<td>71</td>
</tr>
<tr>
<td>4-2</td>
<td>Relative bias (%) of estimators of correction term</td>
<td>84</td>
</tr>
<tr>
<td>4-3</td>
<td>Relative bias (%) of MSEP estimators under REML and ANOVA estimation</td>
<td>100</td>
</tr>
<tr>
<td>4-4</td>
<td>Summary of prediction interval procedures</td>
<td>102</td>
</tr>
<tr>
<td>4-5</td>
<td>Effect of prediction interval procedures for $k = 6, n = 6, \sigma^2 = 1$ and $\sigma_a^2 = .5$</td>
<td>108</td>
</tr>
<tr>
<td>5-1</td>
<td>Generated data that produces negative variance component estimate</td>
<td>138</td>
</tr>
<tr>
<td>5-2</td>
<td>REML variance parameter estimates</td>
<td>138</td>
</tr>
<tr>
<td>5-3</td>
<td>Solution for random effects with REML variance parameter estimates</td>
<td>138</td>
</tr>
<tr>
<td>5-4</td>
<td>REML NOBOUND variance parameter estimates</td>
<td>139</td>
</tr>
<tr>
<td>5-5</td>
<td>Solution for random effects with REML NOBOUND variance parameter estimates</td>
<td>139</td>
</tr>
<tr>
<td>5-6</td>
<td>LSMEANS from PROC GLM</td>
<td>139</td>
</tr>
<tr>
<td>5-7</td>
<td>Data generated from distribution in Model (5-1)</td>
<td>145</td>
</tr>
<tr>
<td>5-8</td>
<td>Covariance parameter estimates under Model (5-1)</td>
<td>146</td>
</tr>
<tr>
<td>5-9</td>
<td>Solution for random effects under Model (5-1)</td>
<td>146</td>
</tr>
<tr>
<td>5-10</td>
<td>LSMEANS from PROC GLM</td>
<td>146</td>
</tr>
<tr>
<td>5-11</td>
<td>Variance parameter estimates under random effects model</td>
<td>147</td>
</tr>
<tr>
<td>5-12</td>
<td>Solution for random effects under random effects model</td>
<td>147</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-1</td>
<td>Equation (4-13) as a function of $\nu$ for several values of $k$, holding $n = 6$. Equation (4-14) holds when functions cross zero line.</td>
</tr>
<tr>
<td>4-2</td>
<td>Accuracy of approximations $tr[A(\theta)B_R(\theta)]$ and $tr[A(\theta)I^{-1}(\theta)]$ for the correction term, $C_R(\theta)$, under REML variance component estimation.</td>
</tr>
<tr>
<td>4-3</td>
<td>Accuracy of approximation $tr[A(\theta)B_R(\theta)]$ and estimator $E\left[tr[A(\hat{\theta})B_R(\hat{\theta})]\right]$ for the correction term, $C_R(\theta)$, under REML variance component estimation.</td>
</tr>
<tr>
<td>4-4</td>
<td>Accuracy of approximation $tr[A(\theta)I^{-1}(\theta)]$ and estimator $E\left[tr[A(\hat{\theta})I^{-1}(\hat{\theta})]\right]$ for the correction term, $C_R(\theta)$, under REML variance component estimation.</td>
</tr>
<tr>
<td>4-5</td>
<td>Comparison of accuracy of estimators for the correction term, $C_R(\theta)$, under REML variance component estimation.</td>
</tr>
<tr>
<td>4-6</td>
<td>Comparison of accuracy of approximations $tr[A(\theta)I^{-1}(\theta)]$, $tr[A(\theta)B_A(\theta)]$, and $tr[A(\theta)B_A(\theta)]$ for the correction term, $C_A(\theta)$, under ANOVA variance component estimation.</td>
</tr>
<tr>
<td>4-7</td>
<td>Accuracy of approximation $tr[A(\theta)B_A(\theta)]$ and estimator $E\left[tr[A(\hat{\theta})B_A(\hat{\theta})]\right]$ for the correction term, $C_A(\theta)$, under ANOVA variance component estimation.</td>
</tr>
<tr>
<td>4-8</td>
<td>Accuracy of approximation $tr[A(\theta)B_A^<em>(\theta)]$ and estimator $E\left[tr[A(\hat{\theta})B_A^</em>(\hat{\theta})]\right]$ for the correction term, $C_A(\theta)$, under ANOVA variance component estimation.</td>
</tr>
<tr>
<td>4-9</td>
<td>Accuracy of approximation $tr[A(\theta)I^{-1}(\theta)]$ and estimator $E\left[tr[A(\hat{\theta})I^{-1}(\hat{\theta})]\right]$ for the correction term, $C_A(\theta)$, under ANOVA variance component estimation.</td>
</tr>
<tr>
<td>4-10</td>
<td>Comparison of accuracy of estimators for the correction term, $C_A(\theta)$, under ANOVA variance component estimation.</td>
</tr>
<tr>
<td>4-11</td>
<td>Accuracy of estimators for the correction term under REML, $\hat{\theta}$, and ANOVA, $\tilde{\theta}$, variance component estimation.</td>
</tr>
</tbody>
</table>
4-12 Relative bias of estimators for the correction term under REML, \( \hat{\theta} \), or ANOVA, \( \tilde{\theta} \), variance component estimation........................................................................................................................................................................120

4-13 Bias produced by estimating \( M_i(\theta) \) with \( M_i(\hat{\theta}) \). ........................................................................................................................................................................121

4-14 Accuracy of approximation \( -\frac{1}{2} tr[\Lambda(\theta)I^{-1}(\theta)] \) and estimator
\( E\left\{ -\frac{1}{2} tr[\Lambda(\hat{\theta})I^{-1}(\hat{\theta})] \right\} \) for the true bias correction term, \( M_i(\theta) - E\left[ M_i(\hat{\theta}) \right] \), under REML variance component estimation. ........................................................................122

4-15 Accuracy of approximation \( tr[\nabla(\theta)I^{-1}(\theta)] \) and estimator \( E\left\{ tr[\nabla(\hat{\theta})I^{-1}(\hat{\theta})] \right\} \) for the true bias correction term, \( M_i(\theta) - E\left[ M_i(\hat{\theta}) \right] \), under REML variance component estimation........................................................................................................123

4-16 Comparing the accuracy of \( E\left\{ -\frac{1}{2} tr[\Lambda(\hat{\theta})I^{-1}(\hat{\theta})] \right\} \) and \( E\left\{ tr[\nabla(\hat{\theta})I^{-1}(\hat{\theta})] \right\} \) for the true bias correction term, \( M_i(\theta) - E\left[ M_i(\hat{\theta}) \right] \). ........................................................................................................................................................................124

4-17 Bias produced by estimating \( M_i(\theta) \) with \( M_i(\hat{\theta}) \). ........................................................................................................................................................................125

4-18 Accuracy of approximation \( -\frac{1}{2} tr[\Lambda(\theta)I^{-1}(\theta)] \) and estimator
\( E\left\{ -\frac{1}{2} tr[\Lambda(\hat{\theta})I^{-1}(\hat{\theta})] \right\} \) for true bias correction term, \( M_i(\theta) - E\left[ M_i(\hat{\theta}) \right] \), under ANOVA variance component estimation. ........................................................................................................126

4-19 Accuracy of approximation, \( tr[\nabla(\theta)I^{-1}(\theta)] \) and estimator \( E\left\{ tr[\nabla(\hat{\theta})I^{-1}(\hat{\theta})] \right\} \) for true bias correction term, \( M_i(\theta) - E\left[ M_i(\hat{\theta}) \right] \), under ANOVA variance component estimation........................................................................................................127

4-20 Comparing the accuracy of \( E\left\{ -\frac{1}{2} tr[\Lambda(\hat{\theta})I^{-1}(\hat{\theta})] \right\} \) and \( E\left\{ tr[\nabla(\hat{\theta})I^{-1}(\hat{\theta})] \right\} \) for the true bias correction term, \( M_i(\theta) - E\left[ M_i(\hat{\theta}) \right] \)........................................................................................................................................................................128

4-21 Comparison of estimators of the MSEP for the EBLUP of \( a_i \) utilizing REML variance component estimation........................................................................................................129
4-22 Comparison of estimators of the MSEP for the EBLUP of $a_i$ utilizing ANOVA variance component estimation.................................................................130
4-23 Relative bias of possible estimators for MSEP of EBLUP for $a_i$ .........................131
4-24 Simulated coverage rates of prediction interval procedures for $t = a_i$ with $k = 3$.........132
4-25 Simulated coverage rates of prediction interval procedures for $t = a_i$ with $k = 6$.........133
4-26 Simulated coverage rates of prediction interval procedures for $t = a_i$ with $k = 15$........134
4-27 Simulated coverage rates of prediction interval procedures for $t = a_i$ with $k = 30$........135
4-28 Simulated coverage rates for the Kenward-Roger prediction interval procedure.........136
EVALUATING ADJUSTMENTS TO THE MEAN SQUARED ERROR DUE TO
ESTIMATING VARIANCE PARAMETERS IN LINEAR MIXED MODELS

By

Jamie McClave Baldwin

December 2007

Chair: Ramon Littell
Major: Statistics

Several methods are commonly used to handle the propagation of variance stemming from estimating unknown variance components in Linear Mixed Models (LMM). The accuracy and consequences of using these methods, however, have not been thoroughly investigated. Empirical Best Linear Unbiased Predictors (EBLUP) for analyzing LMMs are widely used, yet the best way to evaluate the precision of the EBLUP is not generally understood. Many developments in the estimation of the Mean Squared Error of Prediction (MSEP) of the EBLUP, and the use of these estimates for hypothesis testing, have occurred during the last two decades. This dissertation begins with a thorough review of these developments.

Existing methodologies for evaluating the precision of the EBLUP are generalized to apply to multiple dimension linear combinations of fixed and random effects, and the definiteness properties of these methodologies are examined. The methods for evaluating the MSEP of the EBLUP are examined thoroughly for the balanced one-way random effects model to assess the accuracy of the methods, the effect of the parameterization of the model, and the impact of variance parameter estimation techniques on the components of the MSEP estimators. The impact of negative variance component estimates on EBLUP methodologies is examined, and an
alternative solution to account for negative variance estimates is developed for the one-way model.
1.1 Introduction to Mixed Models

In 1984, Kackar and Harville introduced a new method to handle the propagation of variance stemming from estimating unknown variance parameters in linear mixed models. In the nearly quarter century since this development, there have been many adjustments, improvements and new techniques resulting from their idea. Linear mixed models and the procedures used to analyze them are full of complexities and nuances that can confuse even the most experienced analyst. For example, complex covariance structures coupled with unbalanced data and variance parameter estimates on or near the boundary of the parameter space can produce confounding results with many mixed models procedures. This paper begins to unwind and isolate these complexities using simple cases to explore the best options available and determine when procedures for estimating the realized values of the random effects and evaluating the variability of these estimates succeed and fail.

Mixed models methodologies have developed over time to accommodate situations where prediction of the realized values of the random effects is required. These situations arise, for example, in agriculture, education, and clinical trials. A typical agriculture example is the need to assess the genetic contribution of a particular animal for breeding, which is considered to be drawn from a population of animals. An example from education is the desire to assess the effect of a particular teacher on the test scores of his or her pupils. A clinical trial with multiple clinics and varying patient dropout rates gives another arena where mixed models are frequently used to analyze data. Many of these analyses involve complex covariance structures and almost all analyses involve unknown covariance parameters. It is important not only to develop accurate methods for evaluating the realized value of the random effect, but also to evaluate the
associated mean squared error to determine the precision of the estimate. We begin by
describing the framework of the model.

### 1.2 The Linear Mixed Model

The general form of the linear mixed model is

$$y = X\beta + Zu + e$$  \hspace{1cm} (1-1)

where $X$ is a known $(n \times p)$ design matrix for the fixed effects, $\beta$ is a $(p \times 1)$ vector of fixed
effect parameters, $Z$ is a known $(n \times q)$ design matrix for the random effects, $u$ is a $(q \times 1)$
vector of random effects, and $e$ is an $(n \times 1)$ vector containing the random error terms. We will
assume that $E(u) = 0$, $V(u) = G$, and $E(e) = 0$, $V(e) = R$, $\text{cov}(u, e) = 0$. From the model in
Equation (1-1), it follows that $E(y) = X\beta$ and $V(y) = V = ZGZ' + R$. The matrices $G$ and $R$
are taken to have known structures but depend on an unknown dispersion parameter, $\theta \in \Omega$, for
some known parameter space $\Omega$. Multivariate normal distributions on $u$ and $e$ are often
assumed and many of the results are derived under normality assumptions; the validity of these
results under other distributional assumptions is questionable. For simplicity, all matrices are
assumed to be of full rank.

### 1.3 Best Linear Unbiased Predictors (BLUPs) When $\theta$ is Known: Mixed Model

Equations

Best Linear Unbiased Predictors (BLUPs) originated from the desire to estimate the
realized values of the random effects along with the fixed effects, or predictable linear
combinations of both types of effects. This method was originally developed by Henderson
(1950) for estimating genetic merits in animal breeding. Henderson approached the problem
using maximum likelihood methodology, although he later acknowledged that his method cannot
be called “maximum likelihood” since the “function being maximized is not a likelihood”
(Robinson 1991, p.18). Assuming that $u$ and $e$ are normally distributed, Henderson maximized
the joint density of $y$ and $u$ with respect to $\beta$ and $u$. This method results in Henderson’s “mixed models equations” (MMEs):

$$
X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{u} = X'R^{-1}y
$$
$$
Z'R^{-1}X\hat{\beta} + (Z'R^{-1}Z + G^{-1})\hat{u} = Z'R^{-1}y.
$$

(1-2)

Note here that $G$ and $R$ are assumed fully known. The solutions to these equations,

$$
\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y
$$
$$
\hat{u} = GZV^{-1}(y - X\hat{\beta}),
$$

(1-3)

are the Best Linear Unbiased Estimate (BLUE) and Best Linear Unbiased Predictor (BLUP), respectively: they are linear in the data, unbiased, and achieve minimum variance or mean squared error (MSE), respectively, among all linear, unbiased estimators or predictors (see, for example, Harville 1990 and Robinson 1991). Note that the solution for $\hat{\beta}$ in Equation (1-3) gives the GLS estimate. Robinson summarized several other justifications for the BLUP estimates, including a Bayesian method with a flat prior on $\beta$ and a $N(0, G)$ prior on $u$ yielding the same result as the MMEs, as well as justifications that do not require normality. A linear combination of fixed and random effects is predictable when the fixed effect portion is estimable. That is, for vectors $\lambda$ and $\delta$, $\lambda'\hat{\beta} + \delta'\hat{u}$ is predictable if $\lambda'\hat{\beta} = a'X\hat{\beta}$ for some $a$ (Littell, Milliken, Stroup, Wolfinger, and Schabenberger 2006, p. 211). When estimation of a predictable linear combination of fixed and random effects is desired, $\lambda'\hat{\beta} + \delta'\hat{u}$ gives the BLUP for the predictable function.

1.4 Empirical Best Linear Unbiased Estimation (EBLUP)

The estimates in Equation (1-3) assume $G$ and $R$ (and thus $V$) are fully known; this assumption, however, is rarely satisfied. A common alternative is to assume $G$ and $R$ are known
up to a dispersion parameter $\theta = (\theta_1, \ldots, \theta_r)'$, thus $G = G(\theta)$ and $R = R(\theta)$. The BLUP solutions depend on $G$ and $R$, and therefore $\theta$, which is typically unknown. The standard way of resolving this issue is to employ “two-stage estimation;” that is, first estimate the dispersion parameter $\theta$ by, say, $\hat{\theta}$, then use these estimates in the equations for the BLUP of the predictable function, yielding $\lambda'\hat{\beta} + \delta'u$, where $\hat{\beta}$ and $\hat{u}$ are the solutions to Equations (1-2) with $\hat{\theta}$ estimating $\theta$:

$$
\hat{\theta} \Rightarrow \hat{G} = G(\hat{\theta}) \Rightarrow \hat{V} = V(\hat{\theta}) \Rightarrow \hat{\beta} = \left(X'\hat{V}^{-1}X\right)^{-1}X'\hat{V}^{-1}y, \\
\hat{u} = \hat{G}Z\hat{V}^{-1}(y - X\hat{\beta}).
$$

These predictors are often referred to as Empirical BLUPs (EBLUP), although they are no longer linear functions of the data, they are not necessarily unbiased, and they no longer necessarily have minimum MSE.

There are many ways to estimate $\theta$, including ANOVA, MINQUE, MIVQUE, maximum likelihood (ML), and restricted maximum likelihood (REML). The REML methods tend to be favored from a theoretical perspective (Searle, Casella, and McCulloch, 1992). However, from a data analytic perspective it is unclear which method is most advantageous. The technicalities of these methods will not be reviewed here. For an overview of the technical details of these estimation methods see, for example, Khuri and Sahai (1985) and Robinson (1987). We will review the advantages and disadvantages of the methods as necessary in the context of evaluating the precision of the EBLUP methodology.

### 1.5 Two-Stage Estimation and Unbiasedness

The consequences of two-stage estimation on the estimates of the linear combination, $\lambda'\hat{\beta} + \delta'u$, are discussed in Kackar and Harville (1981). They show that when the dispersion...
parameter estimates are translation invariant, even functions of the data, and the data are symmetrically distributed, the two-stage estimation procedure resulting in the EBLUPs provides unbiased predictors of linear combinations of effects; i.e., $E[\lambda \hat{\beta} + \delta \hat{u}] = \lambda \hat{\beta}$. Khatri and Shah (1981) provide similar results. Kackar and Harville (1981) show that these conditions are satisfied by assuming normality of the random effects and error and using REML, ML, ANOVA, MINQUE or MIVQUE variance component estimation methods. Note that this result can be extended to the broader class of location equivariant estimators, as shown in Kackar and Harville (1984), citing a theorem by Wolfe (1973).

In evaluating the precision of EBLUPs, we must account for sampling variability contributed by the estimation of $\theta$. When this added variability is ignored, the true precision is often grossly underestimated, as demonstrated by Kackar and Harville (1984), Harville and Jeske (1992), and Tuchscherer, Herrendorfer, and Tuchscherer (1998). Goals of this dissertation include clarifying and comparing existing methods of estimating the precision of the EBLUP, assessing the accuracy of these methods and evaluating the methods of hypothesis testing which utilize the various methods of estimating the precision. After reviewing the related literature in Chapter 2, we generalize results on evaluating the precision of the EBLUP to include multiple dimensional cases and random effects in Chapter 3. A case study of the balanced one-way random effects model in Chapter 4 sheds light on the problems with existing methods and suggests possibilities for improvement. Chapter 5 tackles the issue of negative variance component estimates and the impact on EBLUP methodology. Conclusions and future work are summarized in Chapter 6.
CHAPTER 2
LITERATURE REVIEW

2.1 Precision of EBLUPs

2.1.1 Mean Squared Error of Prediction (MSEP) of EBLUPs

Let \( t = \lambda \hat{\beta} + \delta'\hat{u} \) and let \( \hat{t} = \hat{t}(\theta) = \lambda \hat{\beta} + \delta'\hat{u} \) denote the estimate of \( t \) when \( \theta \) (and hence \( V \)) is known. When \( \theta \) is unknown, denote the estimate of \( t \), the EBLUP, by \( \hat{t} = \hat{t}(\hat{\theta}) = \lambda \hat{\beta} + \delta'\hat{u} \). The prediction error for the EBLUP is \( \hat{t} - t \) and the mean squared error of prediction (MSEP) is \( M(\theta) = E[(\hat{t} - t)^2] \). By decomposing the prediction error into four components, each assuming a different, successive state of knowledge, Harville (1985) shows how each component adds to the variability.

The first state of knowledge supposes full knowledge of the joint distribution of \((y, t)\). The unbiased minimum MSE predictor of \( t \) in this case is clearly

\[
E[t | y] = \lambda \hat{\beta} + \delta'E[u | y] \quad (\text{Harville 1985, sec. 2.1})
\]

This is known as the *Best Predictor*.

The second state of knowledge supposes the first and second moments are known, i.e., \( \beta \) and \( \theta \) are known but the functional form of the joint distribution of \((y, t)\) is unknown. In this case the predictor with minimum MSE among all linear predictors is

\[
\gamma = \lambda'\hat{\beta} + \delta'GZ'(ZGZ' + R)^{-1} (y - X\hat{\beta}) \\
= \eta + \delta'GZ'(ZGZ' + R)^{-1} y
\]

where

\[
\eta = \lambda'\hat{\beta} - \delta'GZ'(ZGZ' + R)^{-1} X\hat{\beta}
\]

(\text{Harville 1985, sec. 2.2}). This is known as the *Best Linear Predictor*. 
The third state of knowledge assumes only the variances and covariances are known, i.e., \( \theta \) is known but \( \beta \) and the functional form of the joint distribution of \((y, t)\) are unknown. Since \( \eta \) depends on \( \beta \) we must replace \( \beta \) with an estimate. Choosing \( \hat{\beta} \), the BLUE for \( \beta \), results in

\[
\tilde{\tau} = \tilde{\eta}_\beta(y) + \delta'GZ'(ZGZ' + R)^{-1}y
\]

where

\[
\tilde{\eta}_\beta(y) = \lambda'\hat{\beta} - \delta'GZ'(ZGZ' + R)^{-1}X\hat{\beta}
\]

(Harville 1985, sec. 2.3). This is the Best Linear Unbiased Predictor (BLUP).

The fourth state of knowledge represents the typical situation in data analysis: all moments and the form of the joint distribution of \((y, t)\) are unknown. That is, now \( \theta \) is also unknown. Because the BLUP in Equation (2-2) depends on \( \theta \) we cannot directly use it without further adjustment. Let \( z \) be maximal invariant, that is, let \( z = L'y \) where \( L \) is an \( n \times (n - p) \) matrix with \( \text{rank}(L) = n - p \) and \( L'X = 0 \). A function \( g(y) \) is translation invariant if and only if \( g(y) = h(L'y) \). Thus let \( \hat{\theta}(z) = \hat{\theta} \) be an arbitrary translation invariant estimate of \( \theta \). Then

\[
\hat{\tau}_\beta(y) = \int_{\Omega} \hat{\tau}_\beta(y; \omega) dP(\omega; z), \quad \text{where} \quad P(\omega; z) \quad \text{is an arbitrary probability distribution on} \quad \Omega, \quad \text{gives a feasible estimate of} \quad t. \quad \text{The Empirical Best Linear Unbiased Predictor, EBLUP,} \quad \hat{\tau}, \quad \text{corresponds to} \quad \hat{\tau}_\beta(y; \hat{\theta}), \quad \text{the value of} \quad \hat{\tau}_\beta(y) \quad \text{when} \quad P(\cdot; z) \quad \text{is the degenerate probability distribution assigning all mass to} \quad \hat{\theta} \quad \text{(Harville 1985, sec. 2.4).} \quad \text{Note that} \quad \hat{\tau}_\beta(y) \quad \text{is translation invariant, thus preserving the unbiasedness of the predictor.}

With this notation, we now have the four components of the prediction error:

\[
\hat{\tau} - t = E(t|y) - t + \tau - E(t|y) + \tilde{\eta}_\beta - \eta + \hat{\tau}_\beta(y; \hat{\theta}) - \hat{\tau}_\beta.
\]
Each successive component is associated with moving from one state of knowledge to the next. This decomposition allows analysis of the role of each component in the MSEP.

Component (1) of Equation (2-3) is uncorrelated with all other components by the following result (Harville 1985, p. 135):

\[
\text{cov}(h(y), E(t \mid y) - t) = E[h(y)(E(t \mid y) - t)] \\
= E\{E[h(y)(E(t \mid y) - t)] \mid y\} \\
= 0. \tag{2-4}
\]

Components (2) and (3) of Equation (2-3) are also uncorrelated since

\[
\text{cov}(y, \tau - E(t \mid y)) = 0 \quad \text{and} \quad \tilde{\eta}_B \text{ is linear in } y \quad \text{(Harville 1985, p. 135).}
\]

Thus the Mean Squared Error of Prediction (MSEP), \(M(\theta)\), can be written as the following six components:

\[
M(\theta) = \left[\hat{\tau} - t\right]^2 \\
= E\left[E(t \mid y) - t\right]^2 + E\left[\tau - E(t \mid y)\right]^2 + E\left[\tilde{\eta}_B - \eta\right]^2 + E\left[\hat{\tau}_B \left(\hat{\theta}\right) - \tilde{\tau}_B (\theta)\right]^2 \tag{2-5}
\]

\[
+ E\left[\left(\tau - E(t \mid y)\right)(\hat{\tau}_B \left(\hat{\theta}\right) - \tilde{\tau}_B (\theta))\right] + E\left[\left(\tilde{\eta}_B - \eta\right)(\hat{\tau}_B \left(\hat{\theta}\right) - \tilde{\tau}_B (\theta))\right].
\]

In the development of estimators for \(M(\theta)\), it is often assumed that the data follow a normal distribution. When the joint distributions of \((z, \tau - t)\) and \((z, \tilde{\eta}_B(y))\) are multivariate normal, as is the case when \(u\) and \(e\) are normally distributed, then components (5) and (6) of Equation (2-5) are 0. However, without normality, there is no guarantee that these cross-product terms disappear.

The remainder of this chapter will review the developments over the last two decades for evaluating the MSEP of the EBLUP and conducting tests of hypotheses. Table 2-1 contains a
summary of the developments reviewed herein and stands as a helpful reference for these developments and the notation used throughout.

2.1.2 Kackar-Harville Approximation to MSEP

Kackar and Harville (1984) and Harville and Jeske (1992) assume normality in developing an approximation and estimator for $M(\theta)$. Thus, only components (1) – (4) of Equation (2-5) need to be assessed for the approximation. Components (1), (2), and (3) of Equation (2-5) comprise the MSEP when $\theta$ is known. This is referred to as the naïve MSEP and will be denoted by $M_1(\theta) = E[\hat{\tau}_\theta - \tau]^2$. In the mixed model case with $V(u) = G$ and $V(e) = R$, this is

$$M_1(\theta) = \delta'G\delta - \delta'GZV^{-1}ZG\delta + (\lambda' - \delta'GZV^{-1}X)(X'V^{-1}X)^{-1}(\lambda' - X'V^{-1}ZG\delta).$$

When normality is assumed, the cross-product terms are zero, and the only contribution to the MSEP of estimating $\theta$ is component (4) of Equation (2-5). We will refer to this component as the correction term and denote it by $C(\theta)$. $M_1(\theta)$ is often used as an approximation for the MSEP. When the data are normally distributed, this can seriously underestimate the MSEP since it ignores the contribution of estimating $\theta$ captured by $C(\theta)$, as shown by Kackar and Harville (1984), Harville and Jeske (1992), and Tuchscherer et al. (1998). Peixoto and Harville (1986) expand on results in Kackar and Harville (1984) to derive exact, closed-form expressions of the MSEP for certain predictors for the special case of the balanced one-way random effects model. We use similar methods in Chapter 4 to enable comparisons of approximations and estimators to the exact MSEP in the balanced one-way random effects model.

Assuming normality, Kackar and Harville (1984) provide an approximation to $C(\theta)$ using Taylor series methods. Using a Taylor series expansion of $\hat{\tau} = \bar{\tau}(\hat{\theta})$ about $\theta$
yields \( \hat{\tau}(\theta) = \hat{\tau}(\theta) + d(y;\theta)'(\hat{\theta} - \theta) \), where \( d(y;\theta) = \partial \hat{\tau}(y;\theta)/\partial \theta \). This leads to the approximation,

\[
[\hat{\tau} - \tau]^2 = [\hat{\tau}(\theta) - \tau(\theta)]^2 = [d(y;\theta)'(\hat{\theta} - \theta)]^2.
\] (2-6)

Note that when \( \tau \) is linear in \( \theta \) this Taylor-series approximation is exact. The validity of the Taylor-series approximation is contingent on two conditions: First, the function being expanded must have derivatives of all orders defined in a neighborhood of the point about which the function is being expanded; and second, the remainder term must tend quickly to zero (Khuri 1993, p. 111). Thus if the parameter space, \( \Omega \), is restricted (i.e., bounded below by zero) and the true value of \( \theta \) falls on the boundary of the parameter space, the Taylor series expansion is not valid. Assessing the effect of the parameterization and parameter estimates on the estimates of the MSEP is one goal of this dissertation.

When \( \hat{\theta} \) is unbiased for \( \theta \) and \( \text{cov}[d(y;\theta)d(y;\theta)',(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] = 0 \) we have

\[
E[\hat{\tau} - \tau]^2 = E[d(y;\theta)'(\hat{\theta} - \theta)]^2 = E[d(y;\theta)d(y;\theta)']E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'].
\]

Noting that \( E[d(y;\theta)] = 0 \), Kackar and Harville (1984) propose the following approximation to \( C(\theta) \):

\[
C(\theta) \cong \text{tr}[A(\theta)B(\theta)]
\] (2-7)

where, \( A(\theta) = \text{var}[d(y;\theta)] = E[d(y;\theta)]^2 \), and \( B(\theta) \) is either an approximation to or the exact value of the MSE matrix, \( E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] \). The inverse of the observed information matrix, \( I^{-1}(\theta) \), is commonly used for \( B(\theta) \). Equation (2-7) will be referred to as the Kackar-Harville
approximation for the correction term. Thus the approximation of the MSEP of $\hat{\tau}$ proposed by Kackar and Harville (1984) is

$$M_1(\theta) + tr[A(\theta)B(\theta)].$$

(2-8)

The accuracy of the Kackar-Harville approximation of the correction term in Equation (2-7) depends on the following conditions:

1. $\hat{\tau}$ linear in $\theta$;
2. the validity of Taylor series expansion of $\hat{\tau}$ about $\theta$;
3. $\hat{\theta}$ unbiased for $\theta$;
4. $\text{cov}
\begin{bmatrix}
  d(y;\theta) \cdot d(y;\theta)',
\end{bmatrix}
\text{tr}
\begin{bmatrix}
  (\hat{\theta} - \theta)(\hat{\theta} - \theta)',
\end{bmatrix}
= 0$;
5. the tractability of the exact MSE of $\hat{\theta}$.

If all of these criteria are met, the Kackar-Harville approximation exactly equals the correction term it is approximating.

2.1.3 Estimating $C(\theta)$

Because the calculation of the Kackar-Harville approximation to the MSEP in Equation (2-8) requires knowing $\theta$, it is not directly functional. An estimator based on the Kackar-Harville approximation was developed by Harville and Jeske (1990) and Prasad and Rao (1990) and is now implemented in the MIXED procedure of SAS®. Following developments from Prasad and Rao (1990), an estimator for $\text{tr}\left[A(\theta)B(\theta)\right]$ is given by $\text{tr}\left[A(\hat{\theta})B(\hat{\theta})\right]$; where $\hat{\theta}$ is substituted for $\theta$ in the Kackar-Harville approximation for the correction term. The justification for this substitution is not directly evident. Prasad and Rao (1990) show that in the special case of the Fay-Herriot small area model, in which $V(e) = R$ is known, the order of approximation is $o(q^{-1})$ where $q$ is the number of small areas, or in general the number of levels of the random effect. This result has not been generalized for the typical case when $R$ is unknown. The
validity of this substitution and the consequences on the accuracy of the MSEP estimate will be
examined for the balanced one-way random effects model in Chapter 4.

2.1.4 Estimating $M_1(\theta)$

Historically, $M_1(\theta)$ has often been estimated by $M_1(\hat{\theta})$, that is, by substituting $\hat{\theta}$ for $\theta$
in $M_1(\theta)$; however, $M_1(\hat{\theta})$ may have bias of the same or higher order as the correction term
(Prasad and Rao 1990; Booth and Hobert 1998). Consequently, it is just as important to account
for the bias introduced by estimating $M_1(\theta)$ with $M_1(\hat{\theta})$ as it is to account for $C(\theta)$. Prasad
and Rao (1990), Harville and Jeske (1992), and Kenward and Roger (1997) all approximate the
bias of $M_1(\hat{\theta})$ by taking the expectation of a Taylor series of $M_1(\hat{\theta})$ about $\theta$ (assuming $\hat{\theta}$ is
unbiased for $\theta$). Harville and Jeske (1992) developed the following expression for the bias of
$M_1(\hat{\theta})$:

$$E[M_1(\hat{\theta})] \approx M_1(\theta) + \frac{1}{2} tr[A(\theta)B(\theta)]$$ (2-9)

where $A(\theta) = \frac{\partial^2 M_1(\theta)}{\partial \theta^2}$. They show that when $\text{cov}(y,t)$, $\text{var}(t)$ and $\text{var}(y)$ are all linear in
$\theta$, $A(\theta) = -2A(\theta)$ leading to

$$E[M_1(\hat{\theta})] \approx M_1(\theta) - tr[A(\theta)B(\theta)].$$ (2-10)

Prasad and Rao (1990) and Kenward and Roger (1997) developed expressions similar to
Equation (2-10).

Again by substituting $\hat{\theta}$ for $\theta$, possible estimates for $M_1(\theta)$ are
\[ M_i(\hat{\theta}) - \frac{1}{2} \text{tr} \left[ A(\hat{\theta})B(\hat{\theta}) \right] \]  
(2-11)

or

\[ M_i(\hat{\theta}) + \text{tr} \left[ A(\hat{\theta})B(\hat{\theta}) \right]. \]  
(2-12)

Notice again that there is bias introduced by substituting \( \hat{\theta} \) for \( \theta \) in Equations (2-11) and (2-12). This bias is often assumed to be negligible, although the only noted justification, previously mentioned in Section 2.1.3, comes from Prasad and Rao (1990) and only applies to Equation (2-12) in the Fay-Herriot model. We will study this issue further through a special case in Chapter 4.

### 2.1.5 Estimators of \( M(\theta) \) Based on the Kackar-Harville Approximation

Based on Equations (2-11) and (2-12), respectively, the following estimators of the MSEP follow from the Kackar-Harville approximation and the bias correction approximations:

\[ M_i(\hat{\theta}) - \frac{1}{2} \text{tr} \left[ A(\hat{\theta})B(\hat{\theta}) \right] + \text{tr} \left[ A(\hat{\theta})B(\hat{\theta}) \right] \]  
(2-13)

and

\[ M_i(\hat{\theta}) + 2\text{tr} \left[ A(\hat{\theta})B(\hat{\theta}) \right]. \]  
(2-14)

The DDFM = KR option of the MIXED procedure in SAS® utilizes Equation (2-13) which reduces to Equation (2-14) when the model covariance structure is linear in \( \theta \).

Chapter 3 explores the Kenward and Roger (1997) and Fai and Cornelius (1996) expansion of Equation (2-13) into multiple dimensions. Therein we establish the equality of the Kenward-Roger and Fai-Cornelius approximations to the correction term and expand the term to include linear combinations containing random effects. We also provide a proof that the most
The general form of the correction term approximation is non-negative definite and discuss the definiteness properties of the overall MSEP estimator.

### 2.2 Hypothesis Testing for Single Dimension Linear Combinations

Several methods of approximating the distributions for test statistics have been developed for single and multiple dimension linear combinations. The accuracy of these methods has been largely untested. Following is a summary of the available methods for approximating the distributions for single dimension linear combinations.

Giesbrecht and Burns (1985) and Jeske and Harville (1988) provide approximations for \(t\)-type test statistics based on the Satterthwaite (1941) method. Giesbrecht and Burns (1985) present a test statistic for fixed effects with the “naïve” variance estimate:

\[
t_{GB} = \frac{\hat{\lambda}' \hat{\beta}}{\sqrt{\hat{\lambda}' (X'\hat{V}^{-1}X)^{-1} \hat{\lambda}}} \sim t_{df_{GB}},
\]

(2-15)

When \(V\) is known and used in place of \(\hat{V}\) in Equation (2-15), the test statistic has a normal distribution. Thus, it is assumed that Equation (2-15) has a \(t\)-distribution with the degrees of freedom derived from the approximate distribution of the denominator. To estimate the degrees of freedom, a “Satterthwaite” method is used. The variance of the denominator is matched with that of the approximating \(\chi^2\)-distribution as follows:

\[
\text{var} \left[ \frac{\lambda'(X'\hat{V}^{-1}X)^{-1} \lambda}{\lambda'(X'V^{-1}X)^{-1} \lambda} \right] = 2df_{GB} \Rightarrow \frac{(df_{GB})^2}{\lambda'(X'V^{-1}X)^{-1} \lambda^2} \text{var} \left[ \frac{\lambda'(X'\hat{V}^{-1}X)^{-1} \lambda}{\lambda'(X'V^{-1}X)^{-1} \lambda} \right] = 2df_{GB}
\]

\[
\Rightarrow df_{GB} \approx \frac{2 \left[ \frac{\lambda'(X'\hat{V}^{-1}X)^{-1} \lambda}{\lambda'(X'V^{-1}X)^{-1} \lambda} \right]^2}{\text{var} \left[ \frac{\lambda'(X'\hat{V}^{-1}X)^{-1} \lambda}{\lambda'(X'V^{-1}X)^{-1} \lambda} \right]}
\]

(2-16)
The term \( \text{var}\left[ \beta' \left( X' \hat{V}^{-1} X \right)^{\dagger} \lambda \right] \) is evaluated using a Taylor series approximation and the asymptotic covariance matrix of \( \hat{\theta} \). In the Giesbrecht-Burns case, this yields

\[
\text{var}\left[ \beta' \left( X' \hat{V}^{-1} X \right)^{\dagger} \lambda \right] \approx \sum_{j=1}^{r} \sum_{i=1}^{r} \text{cov}\left( \hat{\beta}_i, \hat{\beta}_j \right) \left[ \beta' X' \hat{V}^{-1} Z' \hat{V}^{-1} X \left( X' \hat{V}^{-1} X \right)^{\dagger} \lambda \right] * \left[ \beta' X' \hat{V}^{-1} Z' \hat{V}^{-1} X \left( X' \hat{V}^{-1} X \right)^{\dagger} \lambda \right].
\] (2-17)

The underlying assumption in developing this procedure is that the distribution of \( \beta' \left( X' \hat{V}^{-1} X \right)^{\dagger} \lambda \) is a multiple of a \( \chi^2 \)-distribution.

Jeske and Harville (1988) also use a Satterthwaite method to approximate the degrees of freedom of a \( t \)-statistic; however, they include fixed and random effects in their test statistic with a corrected version of the MSEP estimate. The approximate \( t \)-test is given by

\[
t_{JH} = \frac{\hat{\tau}}{\sqrt{M_1(\hat{\theta}) + tr\left[ A(\hat{\theta}) B(\hat{\theta}) \right]}} \approx t_{df_{JH}},
\] (2-18)

where

\[
df_{JH} \approx \frac{2 \left[ M_1(\hat{\theta}) + tr\left[ A(\hat{\theta}) B(\hat{\theta}) \right] \right]^2}{\text{var}\left[ M_1(\hat{\theta}) + tr\left[ A(\hat{\theta}) B(\hat{\theta}) \right] \right]}. \] (2-19)

As in the Giesbrecht-Burns \( t \)-test, the denominator of Equation (2-19) is approximated by a Taylor series expansion yielding

\[
\text{var}\left[ M_1(\hat{\theta}) + tr\left[ A(\hat{\theta}) B(\hat{\theta}) \right] \right] \approx h(\theta)' B(\theta) h(\theta),
\] (2-20)

where

\[
h(\theta) = \frac{\partial \left[ M_1(\hat{\theta}) + tr\left[ A(\hat{\theta}) B(\hat{\theta}) \right] \right]}{\partial \theta} \] and \( B(\theta) \) is an approximation of \( \text{var}(\hat{\theta}) \).
Again, an underlying assumption is that the distribution of $M_t(\hat{\theta}) + tr[A(\hat{\theta})B(\hat{\theta})]$ follows a multiple of a $\chi^2$-distribution. An extension of this method would be to incorporate the bias correction into the MSEP as seen in Equations (2-13) and (2-14). This extension is considered in Chapter 4.

2.3 Hypothesis Testing Methods for Multiple Dimension Linear Combinations

Kenward and Roger (1997) and Fai and Cornelius (1996) offer $F$-tests to handle multiple dimension as well as single dimension cases; however, the methods for determining the distribution of the test statistics differ. One of the Fai and Cornelius methods (DDFM = SATTERTH option) and the Kenward-Roger method (DDFM = KR option) are available in the SAS® MIXED procedure. A goal of this dissertation is to compare these methods with other modified methods from Fai and Cornelius (1996) for the balanced one-way random effects model. We will also assess the effect of sample size, parameterization, and parameter value on these hypothesis testing methods.

2.3.1 Kenward-Roger Multiple Dimension Hypothesis Testing Method

The test statistic developed by Kenward and Roger (1997) utilizes an MSEP estimate for multiple dimension linear combinations that is adjusted for the correction term and the bias of $M_t(\hat{\theta})$. This MSEP estimate is thoroughly discussed in Chapter 3, but for now will be denoted by $M_{kr}(\hat{\theta})$. Note that these methods are only developed for tests of fixed effects; thus the MSEP can be thought of as a variance. Theoretically these methods could easily extend to include random effects. We include a simulation study for the one-way random effects illustration to determine the validity of including random effects in these procedures.
Consider the hypothesis

\[ H_0 : L\beta = 0 \text{ vs. } H_a : L\beta \neq 0, \]

where \( L \) is an \( l \)-dimension matrix of constants representing the linear combination of interest.

Kenward and Roger (1997) propose the following Wald-type test statistic:

\[
F_{KR} = \frac{1}{l}(\hat{\beta} - \beta)' L' (L M^{KR}(\hat{\theta}) L)'^{-1} L (\hat{\beta} - \beta).
\]

An \( F \)-distribution is approximated for the test statistic to take into account the random structure of \( M^{KR}(\hat{\theta}) \) and to match known cases where the test statistic has an exact \( F \)-distribution. We will briefly review the procedure to allow comparisons to alternative methods and to examine any shortcomings.

Kenward and Roger (1997) use a Taylor series expansion to approximate

\[ (L M^{KR}(\hat{\theta}) L)'^{-1} \]; however, it is unclear about which point the series is expanded and which higher order terms are dropped. It appears that the series is expanded about \( M(\theta) \) rather than the true value, \( M(\hat{\theta}) \). This may be justified by dropping higher order terms; however, this step is the only place where the corrected MSEP is taken into account in the distribution of the test statistic. Using this Taylor series approximation, Kenward and Roger (1997) use the identities

\[
E[F_{KR}] = E\left[ E\left[ F_{KR} | M^{KR}(\hat{\theta}) \right] \right]
\]

\[
V[F_{KR}] = E\left[ V\left[ F_{KR} | M^{KR}(\hat{\theta}) \right] \right] + V\left[ E\left[ F_{KR} | M^{KR}(\hat{\theta}) \right] \right]
\]

to approximate the moments of the test statistic. A two-moment-matching method is then used to obtain estimates for a coefficient, \( \kappa \), and degrees of freedom, \( m \), so that

\[
\kappa F_{KR} \sim F_{l,df_{KR}}.
\]
To make the approximating distribution match the exact cases where the distribution is known, selected higher degree terms are added to the Taylor series expansion. The following summarizes the results of the Kenward-Roger procedure:

\[ \text{df}_{KR} = 4 + \frac{l + 2}{l \rho - 1} \quad \text{and} \quad \kappa = \frac{\text{df}_{KR}}{E[F_{KR}](\text{df}_{KR} - 2)} \]

where

\[ \rho = \frac{V[F_{KR}]}{2E[F_{KR}]^2}, \]

and

\[ E[F_{KR}] = \left(1 - \frac{A_2}{l}\right)^{-1} \quad \text{and} \quad V[F_{KR}] = \frac{2}{l} \left[ \frac{1 + c_1 B}{(1 - c_2 B)^2 (1 - c_3 B)} \right] \]

where

\[ A_1 = \sum_{i=1}^{r} \sum_{j=1}^{r} \left( B_{ij} tr \left\{ L' \left[ L \left( X' V^{-1} X \right)^{-1} L \right]^{-1} L \left( X' V^{-1} X \right)^{-1} X' \frac{\partial V^{-1}}{\partial \theta_i} X \left( X' V^{-1} X \right)^{-1} \right\} \right), \]

\[ A_2 = \sum_{i=1}^{r} \sum_{j=1}^{r} \left( B_{ij} tr \left\{ L' \left[ L \left( X' V^{-1} X \right)^{-1} L \right]^{-1} L \left( X' V^{-1} X \right)^{-1} X' \frac{\partial V^{-1}}{\partial \theta_j} X \left( X' V^{-1} X \right)^{-1} \right\} \right), \]

\[ * tr \left\{ L' \left[ L \left( X' V^{-1} X \right)^{-1} L \right]^{-1} L \left( X' V^{-1} X \right)^{-1} X' \frac{\partial V^{-1}}{\partial \theta_i} X \left( X' V^{-1} X \right)^{-1} \right\}, \]

\[ (2-21) \]
\[ B = \frac{1}{2l} (A_1 + 6A_2), \]

\[ c_1 = \frac{g}{3l + 2(1-g)}, \]

\[ c_2 = \frac{l-g}{3l + 2(1-g)}, \]

\[ c_3 = \frac{l+2-g}{3l + 2(1-g)}, \]

\[ g = \frac{(l+1)A_1 - (l+4)A_2}{(l+2)A_2}, \]

and \( B_{ij} \) is the \( ij^{th} \) element of the matrix \( B(\theta) \).

The Kenward-Roger method is available in the PROC MIXED procedure in SAS\textsuperscript{®} as the DDFM = KR option in the model statement. This option performs the adjustment to the MSEP estimate and the hypothesis test as described above.

We now demonstrate that in the single-dimension case, the Kenward-Roger degrees of freedom approximation reduces to the Giesbrecht-Burns degrees of freedom approximation reviewed in Section 2.2.1 (Kenward and Roger 1997, p. 988). Recall that the Giesbrecht-Burns test statistic did not use an adjustment for unknown \( \theta \) in the MSEP. Thus, while the Kenward-Roger and Giesbrecht-Burns test statistics differ in the estimate of standard error, the approximated distributions are identical.

Consider again a hypothesis test for the single dimension linear combination \( H_0 : \lambda' \beta = 0 \), under the linear mixed model. The Kenward-Roger test statistic is

\[ t_{kr} = \frac{\lambda' \hat{\beta}}{VAR_{kr}} \]

(2-22)

where
\[
\hat{VAR}_{kl} = \lambda'(X'V^{-1}X)^{-1}\lambda + 2\lambda'(X'V^{-1}X)^{-1} \left[B_{ij} \left(\lambda' \frac{\partial V^{-1}}{\partial \theta_i} V \frac{\partial V^{-1}}{\partial \theta_j} \right) - X' \frac{\partial V^{-1}}{\partial \theta_i} X(X'V^{-1}X)^{-1} X' \frac{\partial V^{-1}}{\partial \theta_j} X - \frac{1}{4} X' \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} X \right] (X'V^{-1}X)^{-1} \lambda \right|_{v=\tilde{v}}.
\]

In this case

\[
A_i = \sum_{j=1}^{r} \sum_{j=1}^{r} \text{cov}(\hat{\theta}_i, \hat{\theta}_j) \text{tr} \left[ -\lambda' (X'V^{-1}X)^{-1} \lambda \right] \lambda' X'V^{-1}Z \frac{\partial V}{\partial \theta_i} Z'V^{-1}X(X'V^{-1}X)^{-1} \lambda
\]

\[
\ast \text{tr} \left[ -\lambda' (X'V^{-1}X)^{-1} \lambda \right] \lambda' X'V^{-1}Z \frac{\partial V}{\partial \theta_j} Z'V^{-1}X(X'V^{-1}X)^{-1} \lambda
\]

\[
= \left(\lambda' (X'V^{-1}X)^{-1} \lambda\right)^2 \sum_{j=1}^{r} \sum_{j=1}^{r} \text{cov}(\hat{\theta}_i, \hat{\theta}_j) \left[ \lambda' X'V^{-1}Z \frac{\partial V}{\partial \theta_j} Z'V^{-1}X(X'V^{-1}X)^{-1} \lambda \right]
\]

\[
\ast \left[ \lambda' X'V^{-1}Z \frac{\partial V}{\partial \theta_j} Z'V^{-1}X(X'V^{-1}X)^{-1} \lambda \right]
\]

\[
= A_2
\]

since the quantities inside the trace functions are scalars. Note also that

\[
B = \frac{1}{2} (A_1 + 6 A_2) = \frac{7}{2} A_2,
\]

\[
g = \frac{2 A_2 - 5 A_1}{3 A_2} = -1,
\]

\[
c_1 = \frac{-1}{3(1) + 2(1+1)} = -\frac{1}{7},
\]

\[
c_2 = \frac{2}{7},
\]

and

\[
c_3 = \frac{4}{7}.
\]
Thus,
\[ E\left[ F_{KR} \right] = (1 - A_2)^{-1}, \]
\[ V\left[ F_{KR} \right] = 2 \left[ \frac{1 + \left( \frac{1}{7} \right) \left( \frac{7}{2} \right) A_2}{\left[ 1 - \left( \frac{2}{7} \right) \left( \frac{7}{2} \right) A_2 \right]^2 \left[ 1 - \left( \frac{4}{7} \right) \left( \frac{7}{2} \right) A_2 \right]} \right] = \frac{2 - A_2}{(1 - A_2)^2 \left( 1 - 2A_2 \right)}, \]
and
\[ \rho = \frac{2 - A_2}{(1 - A_2)^2 \left( 1 - 2A_2 \right)} \left[ \frac{(1 - A_2)^2}{2} \right] = \frac{2 - A_2}{2(1 - 2A_2)}. \]

The degrees of freedom for the Kenward-Roger hypothesis test are then
\[ df_{KR} = 4 + \frac{3}{2 - A_2} = \frac{2}{A_2} \]
\[ = 2 \left( \lambda' \left( X' V^{-1} X \right)^{-1} \lambda \right)^2 \left\{ \sum_{i=1}^{r} \sum_{j=1}^{r} \text{cov} \left( \hat{\theta}_i, \hat{\theta}_j \right) \left[ \lambda' X' V^{-1} Z \frac{\partial V}{\partial \theta_i} Z' V^{-1} X (X' V^{-1} X)^{-1} \lambda \right] \right. \]
\[ \left. \ast \left[ \lambda' X' V^{-1} Z \frac{\partial V}{\partial \theta_j} Z' V^{-1} X (X' V^{-1} X)^{-1} \lambda \right] \right\}. \] (2-23)

Note that the denominator of Equation (2-23) is the Taylor series expression for
\[ \text{var} \left[ \lambda' \left( X' V^{-1} X \right)^{-1} \lambda \right]. \] This is the same as the degrees of freedom estimate from the Giesbrecht-Burns test in Equations (2-16) and (2-17); however, recall that the Giesbrecht-Burns test statistic is
\[ t_{GB} = \frac{\lambda' \hat{\beta}}{\sqrt{\lambda' \left( X' V^{-1} X \right)^{-1} \lambda}}. \]

Thus, \( df_{KR} = df_{GB} \) but \( |t_{KR}| \leq |t_{GB}| \). Often, the discrepancy in these test statistics may not be enough to make a significant difference in the test results; however, there are cases when the
discrepancy changes the conclusion of the test. The question becomes, which test is better? The accuracy of the estimating distribution must also be questioned in either test. Note that because the estimated distributions are equal, we can isolate the impact of modifying the estimated variance of the EBLUP on the hypothesis test. We explore that situation further though a simulation study in Chapter 4.

2.3.2 Fai-Cornelius Multiple Dimension Hypothesis Testing Methods

Fai and Cornelius (1996) also use a moment matching method to determine the distribution for their multiple dimension test statistics. The differences between the Kenward-Roger method and the Fai-Cornelius method come in the particular MSEP estimate used in the test statistic, in the number of moments evaluated, and in the techniques used to evaluate the moments of the test statistic.

Fai and Cornelius (1996) offer two alternatives for the test statistic: one uses the naïve MSEP estimate in matrix form, \( \hat{M}_1(\hat{\theta}) \), and the other uses the naïve estimate plus a correction term denoted by, \( \hat{M}^{FC}(\hat{\theta}) = \hat{M}_1(\hat{\theta}) + C^{FC}(\hat{\theta}) \), which will be discussed thoroughly in Chapter 3. Note that neither of these takes into account the bias of \( \hat{M}_1(\hat{\theta}) \) for \( \hat{M}_1(\theta) \), as in the Kenward-Roger procedure. The test statistics are

\[
F_{FC1} = \frac{1}{\ell} (\hat{\beta} - \beta)' L' \left( L M_1(\hat{\theta}) L' \right)^{-1} L (\hat{\beta} - \beta) 
\]

and

\[
F_{FC2} = \frac{1}{\ell} (\hat{\beta} - \beta)' L' \left( L M^{FC}(\hat{\theta}) L' \right)^{-1} L (\hat{\beta} - \beta). 
\]

For each of these test statistics, both one and two moment-matching methods are offered for estimating an \( F \)-distribution, resulting in four alternatives for conducting a multiple dimension hypothesis test. The method of finding the moments is the same for both test statistics, so for
simplicity let $\widehat{\text{MSEP}}$ denote the estimator of the MSEP in either Fai-Cornelius test statistic. To evaluate the moments of the test statistics, first a spectral decomposition is performed so that

$$\mathbf{P}'\mathbf{L}'(\widehat{\text{MSEP}})\mathbf{L}\mathbf{P} = \text{diag}(\lambda_k)$$

where $\mathbf{P}$ is an orthogonal matrix of eigenvectors and $\lambda_k$,

$k = 1, 2, \ldots, l$, are the eigenvalues of $\mathbf{L}'(\widehat{\text{MSEP}})\mathbf{L}$. The test statistic is rewritten so that

$$lF_{FC} = (\hat{\beta} - \beta)' \mathbf{P}\mathbf{P}'(\mathbf{L}(\widehat{\text{MSEP}})\mathbf{L}')^{-1}\mathbf{P}\mathbf{P}'(\hat{\beta} - \beta)$$

$$= (\mathbf{P}'\hat{\beta} - \mathbf{P}'\beta)' (\mathbf{P}'(\widehat{\text{MSEP}})\mathbf{L}\mathbf{P})^{-1} (\mathbf{P}'\hat{\beta} - \mathbf{P}'\beta)$$

$$= (\mathbf{P}'\hat{\beta} - \mathbf{P}'\beta)' \text{diag}(\lambda_k)^{-1} (\mathbf{P}'\hat{\beta} - \mathbf{P}'\beta)$$

$$= \sum_{k=1}^{l} \left( \frac{(\mathbf{P}'\hat{\beta} - \mathbf{P}'\beta)^2}{\lambda_k} \right)$$

$$= \sum_{k=1}^{l} t^2_{\nu_k}$$

$$= Q.$$ (2-26)

They then note that $t_{\nu_k}$ would have an independent standard normal distribution if the MSEP were known. It is reasonable then to approximate the distribution of $t_{\nu_k}$ with a $t$-distribution and estimate the degrees of freedom, $\nu_k$, with a Satterthwaite procedure, i.e.,

$$\nu_k = \frac{2\left[ \hat{\nu}\left(\mathbf{P}'\hat{\beta} - \mathbf{P}'\beta\right)_k \right]}{\hat{\nu}\left(\mathbf{P}'\hat{\beta} - \mathbf{P}'\beta\right)_k} = \frac{2\left(\widehat{\text{MSEP}}\right)_k}{\hat{\nu}\left(\mathbf{P}'\hat{\beta} - \mathbf{P}'\beta\right)_k}.$$ 

As in the Kenward-Roger procedure, we have the underlying assumption that $\widehat{\text{MSEP}}$ has a $\chi^2$ distribution. If $\nu_k > 2$ for all $k = 1, 2, \ldots, l$, then the moments of $Q$ can be approximated as follows:
\[ E(Q) = \frac{1}{l} \sum_{k=1}^{l} E(t_{v_k}^2) = \frac{1}{l} \sum_{k=1}^{l} \frac{v_k}{v_k - 2} \]

and

\[ V(Q) = \frac{1}{l} \sum_{k=1}^{l} V(t_{v_k}^2) = \frac{1}{l} \sum_{k=1}^{l} \frac{2v_k^2(v_k - 1)}{(v_k - 2)^2(v_k - 4)} \cdot \]

From Equation (2-26) we see that the moments of the Fai-Cornelius test statistics are

\[ E(F_{FC}) = \frac{1}{l} E(Q) \]

and

\[ V(F_{FC}) = \frac{1}{l^2} V(Q) \cdot \]

One or both of these moments are matched to the moments of the approximating \( F \)-distribution, as in the Kenward-Roger procedure, to produce an estimate of degrees of freedom (and an estimated coefficient in the two-moment case). The one-moment method with the test statistic in Equation (2-24) is available in the PROC MIXED procedure of SAS® as the DDFM = SATTERTH option in the model statement. Note that in the single dimension case, the Fai-Cornelius test statistic in Equation (2-24) reduces to the Giesbrecht-Burns test statistic in Equation (2-15) and that the Fai-Cornelius one-moment method of estimating the distribution also produces the same results as the Giesbrecht-Burns method. Thus we are avoiding the discrepancies that occur between the Kenward-Roger method and the Giesbrecht-Burns method described in Section 2.3.1.

We are interested in comparing the available hypothesis testing methods to determine which one provides the most accurate Type-I error rate. Some work has already been done to this end by Schaalje, McBride, and Fellingham (2002). Schaalje et al. (2002) compare the one moment Fai-Cornelius method with the test statistic in Equation (2-24) (using the DDFM = SATTERTH option in SAS®) with the Kenward-Roger method (using the DDFM = KR option in
SAS®). They performed simulations for four split-plot designs, two of which were balanced and two of which were unbalanced. The simulation is repeated for five covariance structures to study the effect of the complexity of the covariance structure. Hypothesis tests were computed for whole plot effect and sub-plot effect using the DDFM = KR and DDFM = SATTERTH methods. Schaalje et al. (2002) calculated simulated Type I error rates and lack-of-fit tests to determine if the simulated p-values followed a uniform (0,1) distribution. Both methods performed equally well for the simple Compound Symmetric covariance structure, regardless of the design. For the other four covariance structures the Kenward-Roger method outperforms the Fai-Cornelius method in the lack-of-fit test and the type I error rates; however, as the sample size becomes smaller or the covariance structure becomes more complex, neither method performs well. Type I error rates are severely inflated under both methods when complex covariance structures are coupled with small sample sizes.

Note that when random effects are considered, it is more appropriate to focus the investigation on prediction limits. We will shift the focus to prediction intervals and expand the comparison to include modified Fai-Cornelius methods with adjusted MSEP estimates in Chapter 4. Although we confine the comparison to a single dimension linear combination in the balanced one-way random effects model, this comparison will paint a more complete picture of the methods considered and show the need for further investigation into the proposed methods.

### 2.4 Negative Variance Component Estimates

Very little research has focused on the impact of negative variance parameter estimates on BLUP procedures, the MSEP estimation methods, and hypothesis testing. Stroup and Littell (2002) begin to explore the impact of the method of variance component estimation and negative variance parameter estimates on tests of fixed effects. Their exploration reveals that REML may not always be the best choice in terms of power and true Type I error values for tests of fixed
effects when negative variance parameter estimates are prevalent. The impact of negative variance parameter estimates on the EBLUP and MSEP estimation for linear combinations involving random effects has not been thoroughly investigated. In Chapter 5, we delve into the complexities of negative variance component estimates and the impact on EBLUP and the associated MSEP estimates. Smith and Murray (1984) develop a method for handling negative covariance estimates. They propose that the variance parameter in question be considered as a covariance parameter, thereby allowing negative estimates. The impact of this strategy on EBLUP methods is considered in Chapter 5, leading to an alternative model and analysis procedure.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Fixed or Random Effects</th>
<th>Correction Term</th>
<th>Bias Term</th>
<th>MSEP Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kackar and Harville 1984</td>
<td>Fixed and random</td>
<td>$\text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$</td>
<td>None</td>
<td>$M_1(\hat{\theta}) + \text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$</td>
</tr>
<tr>
<td>Giesbrecht and Burns 1985</td>
<td>Fixed</td>
<td>None</td>
<td>None</td>
<td>$\lambda' \left( X'X^-1 \right)^{-1} \lambda$</td>
</tr>
<tr>
<td>Jeske and Harville 1988</td>
<td>Fixed and random</td>
<td>$\text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$</td>
<td>None</td>
<td>$M_1(\hat{\theta}) + \text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$</td>
</tr>
<tr>
<td>Prasad and Rao 1990</td>
<td>Fixed and Random</td>
<td>$E \left[ \text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right] \right]$</td>
<td>For $q = #$ small areas, in the Fey-Herriot model</td>
<td>$M_1(\hat{\theta}) + 2\text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$</td>
</tr>
<tr>
<td>Harville and Jeske 1992</td>
<td>Fixed and random</td>
<td>$\text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$</td>
<td>$E \left[ M_1(\hat{\theta}) \right] = M_1(\hat{\theta}) + \frac{1}{2} \text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$</td>
<td>$M_1(\hat{\theta}) - \frac{1}{2} \text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right] + \text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$ or $M_1(\hat{\theta}) + 2\text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$</td>
</tr>
<tr>
<td>Fai and Cornelius 1996</td>
<td>Fixed</td>
<td>$\text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$ defined for multiple dimension linear combination, denoted $\mathbf{C}^\text{FC}(\hat{\theta})$; shown to be nn</td>
<td>None</td>
<td>$M_1(\hat{\theta})$ or $M_1(\hat{\theta}) + \mathbf{C}^\text{FC}(\hat{\theta})$</td>
</tr>
<tr>
<td>Kenward and Roger 1997</td>
<td>Fixed</td>
<td>$\text{tr} \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right]$ defined for multiple dimension linear combination; $\mathbf{C}^\text{EA}(\hat{\theta})$</td>
<td>Defined for multiple dimension linear combination, $\text{BIAS}^\text{EA}(\hat{\theta})$</td>
<td>$M_1(\hat{\theta})$ or $M_1(\hat{\theta}) + \mathbf{C}^\text{EA}(\hat{\theta}) + \text{BIAS}^\text{EA}(\hat{\theta})$</td>
</tr>
<tr>
<td>Paper</td>
<td>Hypothesis</td>
<td>Test Stat</td>
<td>Distribution of Test Stat</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>-----------</td>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td>Kackar and Harville 1984</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Giesbrecht and Burns 1985</td>
<td>$H_0: \lambda \beta = 0$</td>
<td>$t_{GB} = \frac{\hat{\lambda} \hat{\beta}}{\sqrt{\lambda'(X'X)^{-1}\lambda}}$</td>
<td>$t_{df_{GB}}$ where $df_{GB} = \frac{2}{\text{var}\left[\lambda'(X'X)^{-1}\lambda\right]}$ Evaluated with Satterthwaite-type method</td>
<td></td>
</tr>
<tr>
<td>Jeske and Harville 1988</td>
<td>$H_0: \lambda \beta + \delta' u = 0$</td>
<td>$t_{JO} = \frac{\hat{\gamma}}{\sqrt{M(\hat{\beta}) + tr[A(\hat{\beta})B(\hat{\beta})]}}$</td>
<td>$t_{df_{JO}}$ where $df_{JO} = \frac{2}{\text{var}\left[M(\hat{\beta}) + tr[A(\hat{\beta})B(\hat{\beta})]\right]}$ Evaluated with Satterthwaite-type method</td>
<td></td>
</tr>
<tr>
<td>Prasad and Rao 1990</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Harville and Jeske 1992</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Fai and Cornelius 1996</td>
<td>$H_0: L\beta = 0$</td>
<td>$F_{rc1} = \frac{1}{L'(\hat{\beta} - \beta)'[LM(\hat{\omega})L']^{-1}L(\hat{\beta} - \beta)}$</td>
<td>One and Two moment matching; Spectral decomposition, Taylor series and Satterthwaite techniques used to determine moments. Note that this yields two choices of distributions for each test statistic.</td>
<td></td>
</tr>
<tr>
<td>Kenward and Roger 1997</td>
<td>$H_0: L\beta = 0$</td>
<td>$F_{rs} = \frac{1}{L'(\hat{\beta} - \beta)'[LM^e(\hat{\omega})L']^{-1}L(\hat{\beta} - \beta)}$</td>
<td>Two moment matching; Taylor series and conditional expectations used to determine moments.</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 3
ESTIMATES OF MSEP FOR MULTIPLE DIMENSION LINEAR COMBINATIONS

3.1 Reconciling Kenward-Roger (1997) and Fai-Cornelius (1996) Definitions of the Correction Term Approximation

3.1.1 Linear Combinations of Fixed Effects

Kenward and Roger (1997) and Fai and Cornelius (1996) extend the methods for estimating the MSEP summarized in Section 2.1 to multiple dimension linear combinations. Both methods require expanding the definition of the Kackar-Harville approximation of the correction term, and in the Kenward-Roger case, the bias correction term, into a matrix form. Recall that the Kackar-Harville approximation of the correction term for single dimension linear combinations is

\[ C(\theta) \approx \text{tr} \left[ A(\theta)B(\theta) \right] \]

where \( A(\theta) = \text{var}(d(y;\theta)) = E \left[ d(y;\theta)d(y;\theta)' \right] \),

\( d(y;\theta) = \partial \tilde{r}(y;\theta)/\partial \theta \), and \( B(\theta) \) is either an approximation to or the exact MSE matrix,

\[ E \left[ (\hat{\theta} - \theta)(\hat{\theta} - \theta)' \right] \].

Extending this definition to multiple dimension linear combinations is not necessarily unique.

In both papers, only linear combinations of the vector of fixed parameters are considered. In Section 3.1.2 we extend their method to include linear combinations involving both fixed and random effects. Noting that

\[ \text{tr} \left[ A(\theta)B(\theta) \right] = \sum_{j=1}^{r} \sum_{i=1}^{r} b_{ij} \text{cov} \left( \frac{\partial \tilde{r}}{\partial \theta_i}, \frac{\partial \tilde{r}}{\partial \theta_j} \right) \]

where \( b_{ij} \) is the \( ij \)th element of \( B(\theta) \), Kenward and Roger expand this definition to the multiple dimension case by

\[ C(\theta) \approx C^{KR}(\theta) = \sum_{j=1}^{l} \sum_{i=1}^{l} b_{ij} \text{cov} \left( \frac{\partial L\tilde{\beta}}{\partial \theta_i}, \frac{\partial L\tilde{\beta}}{\partial \theta_j} \right) \]

where \( L \) is an \( l \)-dimension matrix of constants representing the linear combinations of interest. We will denote this approximation as \( C^{KR}(\theta) \).
Note that,
\[
\text{cov} \left( \frac{\partial \mathbf{L} \hat{\beta}}{\partial \theta_i}, \frac{\partial \mathbf{L} \hat{\beta}}{\partial \theta_j} \right) = \begin{bmatrix}
\text{cov} \left( \frac{\partial \mathbf{L}_1 \hat{\beta}}{\partial \theta_i}, \frac{\partial \mathbf{L}_1 \hat{\beta}}{\partial \theta_j} \right) & \text{cov} \left( \frac{\partial \mathbf{L}_1 \hat{\beta}}{\partial \theta_i}, \frac{\partial \mathbf{L}_2 \hat{\beta}}{\partial \theta_j} \right) & \cdots & \text{cov} \left( \frac{\partial \mathbf{L}_1 \hat{\beta}}{\partial \theta_i}, \frac{\partial \mathbf{L}_l \hat{\beta}}{\partial \theta_j} \right) \\
\text{cov} \left( \frac{\partial \mathbf{L}_2 \hat{\beta}}{\partial \theta_i}, \frac{\partial \mathbf{L}_1 \hat{\beta}}{\partial \theta_j} \right) & \text{cov} \left( \frac{\partial \mathbf{L}_2 \hat{\beta}}{\partial \theta_i}, \frac{\partial \mathbf{L}_2 \hat{\beta}}{\partial \theta_j} \right) & \cdots & \text{cov} \left( \frac{\partial \mathbf{L}_2 \hat{\beta}}{\partial \theta_i}, \frac{\partial \mathbf{L}_l \hat{\beta}}{\partial \theta_j} \right) \\
\vdots & \vdots & \ddots & \vdots \\
\text{cov} \left( \frac{\partial \mathbf{L}_l \hat{\beta}}{\partial \theta_i}, \frac{\partial \mathbf{L}_1 \hat{\beta}}{\partial \theta_j} \right) & \text{cov} \left( \frac{\partial \mathbf{L}_l \hat{\beta}}{\partial \theta_i}, \frac{\partial \mathbf{L}_2 \hat{\beta}}{\partial \theta_j} \right) & \cdots & \text{cov} \left( \frac{\partial \mathbf{L}_l \hat{\beta}}{\partial \theta_i}, \frac{\partial \mathbf{L}_l \hat{\beta}}{\partial \theta_j} \right)
\end{bmatrix},
\]

where \( \mathbf{L}_i \) is the \( i^{th} \) row of \( \mathbf{L} \). Thus we can write the \( km^{th} \) element of \( \mathbf{C}^{\text{KR}}(\theta) \) as
\[
\{ \mathbf{C}^{\text{KR}}_{km}(\theta) \} = b_{11} \text{cov} \left( \frac{\partial \mathbf{L}_1 \hat{\beta}}{\partial \theta_1}, \frac{\partial \mathbf{L}_m \hat{\beta}}{\partial \theta_1} \right) + b_{12} \text{cov} \left( \frac{\partial \mathbf{L}_1 \hat{\beta}}{\partial \theta_1}, \frac{\partial \mathbf{L}_2 \hat{\beta}}{\partial \theta_2} \right) + \cdots + b_{qq} \text{cov} \left( \frac{\partial \mathbf{L}_l \hat{\beta}}{\partial \theta_q}, \frac{\partial \mathbf{L}_m \hat{\beta}}{\partial \theta_q} \right).
\]

Fai and Cornelius (1996) define the multiple dimension approximation to the correction term differently and restrict their definition to a compound symmetric covariance structure with two covariance parameters. We now expand that definition to include any covariance structure with \( r \) parameters and show its equivalence to \( \mathbf{C}^{\text{KR}}(\theta) \). They also establish the non-negative definiteness of their correction term definition, which also applies in this more general setting.

Let \( \mathbf{C}^{\text{FC}}(\theta) \) denote the Fai and Cornelius multiple dimension approximation to the correction term. Then,
\[
\mathbf{C}^{\text{FC}}(\theta) = \begin{bmatrix}
\text{tr} \{ \mathbf{A}_{11}(\theta) \mathbf{B}(\theta) \} & \cdots & \text{tr} \{ \mathbf{A}_{1l}(\theta) \mathbf{B}(\theta) \} \\
\vdots & \ddots & \vdots \\
\text{tr} \{ \mathbf{A}_{l1}(\theta) \mathbf{B}(\theta) \} & \cdots & \text{tr} \{ \mathbf{A}_{ll}(\theta) \mathbf{B}(\theta) \}
\end{bmatrix},
\]

where
and $B(\theta)$ is as before. Then the $km^{th}$ element of $C_{FC}^{\theta}(\theta)$ is

$$
\{C_{FC}^{\theta}(\theta)\}_{km} = tr\{A_{km}(\theta)B(\theta)\} = b_{11}\text{cov}\left(\frac{\partial L_i \hat{\beta}}{\partial \theta_i}, \frac{\partial L_m \hat{\beta}}{\partial \theta_i}\right) + b_{12}\text{cov}\left(\frac{\partial L_i \hat{\beta}}{\partial \theta_i}, \frac{\partial L_m \hat{\beta}}{\partial \theta_2}\right) + \cdots + b_{qq}\text{cov}\left(\frac{\partial L_i \hat{\beta}}{\partial \theta_q}, \frac{\partial L_m \hat{\beta}}{\partial \theta_q}\right).
$$

Thus, clearly $C_{KR}^{\theta}(\theta) = C_{FC}^{\theta}(\theta)$.

Fai and Cornelius (1996) have shown that this definition of the correction term approximation is non-negative definite. This becomes an important factor when the MSEP estimate is to be used in a test statistic or confidence interval. We define the most general form of this approximation and prove that it is non-negative definite in the Section 3.1.2.

### 3.1.2 Linear Combinations of Fixed and Random Effects

We can now generalize the definition of the multiple-dimension correction term approximation to include random effects and establish that it is non-negative definite. Let $K = [L | M]$ denote an $l$-dimension matrix of constants. We now consider the linear combination, $K \begin{pmatrix} \beta \\ u \end{pmatrix} = L\beta + Mu$. To expand the definition to random effects, we need to redefine $A_{km}(\theta)$ as
\[ A_{km}(\theta) = \text{cov} \left( \frac{\partial (\mathbf{L}_k \vec{\beta} + \mathbf{M}_k \vec{u})}{\partial \theta}, \frac{\partial (\mathbf{L}_m \vec{\beta} + \mathbf{M}_m \vec{u})}{\partial \theta} \right) \]

\[
= \begin{bmatrix}
\text{cov} \left( \frac{\partial (\mathbf{L}_k \vec{\beta} + \mathbf{M}_k \vec{u})}{\partial \theta_1}, \frac{\partial (\mathbf{L}_m \vec{\beta} + \mathbf{M}_m \vec{u})}{\partial \theta_1} \right) & \ldots & \text{cov} \left( \frac{\partial (\mathbf{L}_k \vec{\beta} + \mathbf{M}_k \vec{u})}{\partial \theta_r}, \frac{\partial (\mathbf{L}_m \vec{\beta} + \mathbf{M}_m \vec{u})}{\partial \theta_r} \right) \\
\vdots & \ddots & \vdots \\
\text{cov} \left( \frac{\partial (\mathbf{L}_k \vec{\beta} + \mathbf{M}_k \vec{u})}{\partial \theta_r}, \frac{\partial (\mathbf{L}_m \vec{\beta} + \mathbf{M}_m \vec{u})}{\partial \theta_1} \right) & \ldots & \text{cov} \left( \frac{\partial (\mathbf{L}_k \vec{\beta} + \mathbf{M}_k \vec{u})}{\partial \theta_r}, \frac{\partial (\mathbf{L}_m \vec{\beta} + \mathbf{M}_m \vec{u})}{\partial \theta_r} \right)
\end{bmatrix}
\]

where \( \mathbf{L}_k \) is the \( k \)th row of the fixed effects portion of the matrix of constants, \( \mathbf{L} \), and \( \mathbf{M}_k \) is the \( k \)th row of the random effects portion of the matrix of constants, \( \mathbf{M} \).

We now follow a similar procedure as used in Fai and Cornelius (1996) to establish that this generalized definition of \( \mathbf{C}^{KR}(\theta) \) is non-negative definite and thus the MSEP estimator utilizing this correction term estimator is nonsingular. First, however, we need to broaden the definition of \( \mathbf{C}^{KR}(\theta) \) to incorporate linear combinations of fixed and random effects.

First, using the result that \( \frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{x}} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{x}} \mathbf{A}^{-1} \), we have

\[
\frac{\partial \mathbf{L}_k \vec{\beta}}{\partial \theta_i} = \frac{\partial \mathbf{L}_k (X'V^{-1}X)^{-1}X'V^{-1}y}{\partial \theta_i} = \mathbf{L}_k (X'V^{-1}X)^{-1}X' \frac{\partial V^{-1}}{\partial \theta_i} y + \mathbf{L}_k \frac{\partial (X'V^{-1}X)^{-1}}{\partial \theta_i} X'V^{-1}y
\]

\[= \mathbf{L}_k (X'V^{-1}X)^{-1}X' \frac{\partial V^{-1}}{\partial \theta_i} y - \mathbf{L}_k (X'V^{-1}X)^{-1}X' \frac{\partial V^{-1}}{\partial \theta_i} X(X'V^{-1}X)^{-1}X'V^{-1}y\]

\[= \mathbf{L}_k (X'V^{-1}X)^{-1}X' \frac{\partial V^{-1}}{\partial \theta_i} (I-P)y\]

\[= -\mathbf{L}_k (X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} (I-P)y\]

\[= \mathbf{C}_{ki} V^{-1} (I-P)y\]

where \( \mathbf{C}_{ki} = -\mathbf{L}_k (X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial V}{\partial \theta_i} \) and \( \mathbf{P} = X(X'V^{-1}X)^{-1}X'V^{-1} \)

and
\[
\begin{align*}
\frac{\partial \mathbf{M}_k\mathbf{\tilde{u}}}{\partial \theta_i} &= \mathbf{M}_k \frac{\partial \mathbf{G} \mathbf{Z}'\mathbf{V}^{-1} (I - P) y}{\partial \theta_i} + \mathbf{M}_k \mathbf{GZ}' \frac{\partial \mathbf{V}^{-1}}{\partial \theta_i} (I - P) y - \mathbf{M}_k \mathbf{GZ}' \mathbf{V}^{-1} \mathbf{X} \frac{\partial \mathbf{\tilde{\beta}}}{\partial \theta_i} \\
&= \mathbf{M}_k \frac{\partial \mathbf{G} \mathbf{Z}'\mathbf{V}^{-1} (I - P) y}{\partial \theta_i} - \mathbf{M}_k \mathbf{GZ}' \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i} (I - P) y \\
&\quad + \mathbf{M}_k \mathbf{GZ}' \mathbf{V}^{-1} \mathbf{X} \left( \mathbf{X}'\mathbf{V}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i} \mathbf{V}^{-1} (I - P) y \\
&= \mathbf{D}'_{ki} \mathbf{V}^{-1} (I - P) y + \mathbf{E}'_{ki} \mathbf{V}^{-1} (I - P) y + \mathbf{F}'_{ki} \mathbf{V}^{-1} (I - P) y \\
&= (\mathbf{D}'_{ki} + \mathbf{E}'_{ki} + \mathbf{F}'_{ki}) \mathbf{V}^{-1} (I - P) y
\end{align*}
\]

where

\[
\mathbf{D}'_{ki} = \mathbf{M}_k \frac{\partial \mathbf{G}}{\partial \theta_i} \mathbf{Z}', \quad \mathbf{E}'_{ki} = -\mathbf{M}_k \mathbf{GZ}' \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i}
\]

and

\[
\mathbf{F}'_{ki} = \mathbf{M}_k \mathbf{GZ}' \mathbf{V}^{-1} \mathbf{X} \left( \mathbf{X}'\mathbf{V}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i}.
\]

Thus,

\[
\frac{\partial (\mathbf{L}_k\mathbf{\tilde{\beta}} + \mathbf{M}_k\mathbf{\tilde{u}})}{\partial \theta_i} = (\mathbf{C}'_{ki} + \mathbf{D}'_{ki} + \mathbf{E}'_{ki} + \mathbf{F}'_{ki}) \mathbf{V}^{-1} (I - P) y.
\]

We can now assess the \(ij\)th element of \(\mathbf{A}_{km}(\theta)\) as

\[
\{\mathbf{A}_{km(ij)}(\theta)\} = \text{cov} \left[ \frac{\partial (\mathbf{L}_k\mathbf{\tilde{\beta}} + \mathbf{M}_k\mathbf{\tilde{u}})}{\partial \theta_i}, \frac{\partial (\mathbf{L}_m\mathbf{\tilde{\beta}} + \mathbf{M}_m\mathbf{\tilde{u}})}{\partial \theta_j} \right]
\]

\[
= \text{cov} \left[ (\mathbf{C}'_{ki} + \mathbf{D}'_{ki} + \mathbf{E}'_{ki} + \mathbf{F}'_{ki}) \mathbf{V}^{-1} (I - P) y, \ (\mathbf{C}'_{mj} + \mathbf{D}'_{mj} + \mathbf{E}'_{mj} + \mathbf{F}'_{mj}) \mathbf{V}^{-1} (I - P) y \right]
\]

\[
= (\mathbf{C}'_{ki} + \mathbf{D}'_{ki} + \mathbf{E}'_{ki} + \mathbf{F}'_{ki}) \mathbf{V}^{-1} (I - P) \mathbf{V} (I - P)' \mathbf{V}^{-1} \left( \mathbf{C}'_{mj} + \mathbf{D}'_{mj} + \mathbf{E}'_{mj} + \mathbf{F}'_{mj} \right)
\]

\[
= (\mathbf{C}'_{ki} + \mathbf{D}'_{ki} + \mathbf{E}'_{ki} + \mathbf{F}'_{ki}) \mathbf{V}^{-1} (I - P) \left( \mathbf{C}'_{mj} + \mathbf{D}'_{mj} + \mathbf{E}'_{mj} + \mathbf{F}'_{mj} \right)
\]

since
\[(I - P)V(I - P)'V^{-1} = \left(V - X(X'V^{-1}X)^{-1}X'\right)\left(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\right)\]
\[= (I - P).\]

We now have
\[A_{km}(\theta) = \begin{bmatrix}
  (C_{k1} + D_{k1} + E_{k1} + F_{k1})' \\
  \vdots \\
  (C_{kr} + D_{kr} + E_{kr} + F_{kr})'
\end{bmatrix} \times V^{-1}(I - P)\left[\begin{bmatrix}
  (C_{m1} + D_{m1} + E_{m1} + F_{m1}) \\
  \vdots \\
  (C_{mr} + D_{mr} + E_{mr} + F_{mr})
\end{bmatrix}\right].\]

We can now generalize the definition of the Kackar-Harville correction term approximation as follows.

**Definition 3.1.** The multiple-dimension Kackar-Harville approximation for a linear combination including fixed and random effects is

\[C^{KR}(\theta) = \left\{\text{tr}\left[A_{km}(\theta)B(\theta)\right]\right\},\]

where \(A_{km}(\theta)\) is defined in Equation (3-2).

We now prove that \(C^{KR}(\theta)\) is non-negative definite.

**Theorem 3.1.** The multiple-dimension Kackar-Harville approximation for the correction term, as defined in Equation (3-3), is non-negative definite. Thus, the MSEP approximation utilizing this correction, \(M_{I}(\theta) + C^{KR}(\theta)\), is nonsingular.

**Proof.**

We begin by examining the \(km^{th}\) term of \(C^{KR}(\theta)\). We can write \(V^{-1}(I - P) = Q^{1/2}Q^{1/2}\) and \(B(\theta) = R^{1/2}R^{1/2}\) where \(Q^{1/2}\) and \(R^{1/2}\) both exist since \(V^{-1}(I - P)\) is non-negative definite and \(B(\theta)\) is positive definite (e.g., Schott 1997, p.16).
Consider

\[
\begin{align*}
\{ C_{km}^{KR} (\theta) \} \\
\quad = tr \{ A_{km} (\theta) B (\theta) \} \\
\quad = tr \left\{ \left[ \begin{array}{c} (C_{k1} + D_{k1} + E_{k1} + F_{k1})' \\
\vdots \\
(C_{k\nu} + D_{k\nu} + E_{k\nu} + F_{k\nu})' \\
\end{array} \right] V^{-1} (1 - P) \left[ \begin{array}{ccc}
(C_{m1} + D_{m1} + E_{m1} + F_{m1}) & \cdots & (C_{m\nu} + D_{m\nu} + E_{m\nu} + F_{m\nu}) \\
\end{array} \right] B (\theta) \right\} \\
\quad = tr \left\{ \left[ \begin{array}{c} (C_{k1} + D_{k1} + E_{k1} + F_{k1})' \\
\vdots \\
(C_{k\nu} + D_{k\nu} + E_{k\nu} + F_{k\nu})' \\
\end{array} \right] Q^{1/2} Q^{1/2} \left[ \begin{array}{ccc}
(C_{m1} + D_{m1} + E_{m1} + F_{m1}) & \cdots & (C_{m\nu} + D_{m\nu} + E_{m\nu} + F_{m\nu}) \\
\end{array} \right] R^{1/2} R^{1/2} \right\} \\
\quad = tr \left\{ \left[ \begin{array}{c} (C_{m1} + D_{m1} + E_{m1} + F_{m1}) & \cdots & (C_{m\nu} + D_{m\nu} + E_{m\nu} + F_{m\nu}) \\
\end{array} \right] R^{1/2} \left[ \begin{array}{c} (C_{k1} + D_{k1} + E_{k1} + F_{k1})' \\
\vdots \\
(C_{k\nu} + D_{k\nu} + E_{k\nu} + F_{k\nu})' \\
\end{array} \right] \right\} \\
\quad = tr \left\{ \left[ \begin{array}{c} (Q^{1/2} (C_{k1} + D_{k1} + E_{k1} + F_{k1}) & \cdots & (C_{k\nu} + D_{k\nu} + E_{k\nu} + F_{k\nu}) R^{1/2})' \\
\end{array} \right] \right\} \\
\quad = \left\{ vec \left[ \left[ \begin{array}{c} (Q^{1/2} (C_{m1} + D_{m1} + E_{m1} + F_{m1}) & \cdots & (C_{m\nu} + D_{m\nu} + E_{m\nu} + F_{m\nu}) R^{1/2})' \\
\end{array} \right] \right] \right\}' \\
\quad = \delta_m' \delta_k = \delta_i' \delta_i \\
\end{align*}
\]

where \( \delta_i = vec \left[ \left[ \begin{array}{c} (Q^{1/2} ((C_{r1} + D_{r1} + E_{r1} + F_{r1}) & \cdots & (C_{r\nu} + D_{r\nu} + E_{r\nu} + F_{r\nu}) R^{1/2}) \right] \right] \).

Thus we can write \( C_{km}^{KR} (\theta) \) as

\[
C_{km}^{KR} (\theta) = \Delta' \Delta
\]

where

\[
\Delta = (\delta_1, \ldots, \delta_i).
\]

(3-4)
Writing $C^{KR}(\theta)$ in this form proves that it is non-negative definite (e.g., Schott 1997, p.16).

Since $M_1(\theta) = E\left[\left(\left(L\hat{\beta} + M\hat{\mu}\right) - (L\beta + Mu)\right)^T\left(\left(L\hat{\beta} + M\hat{\mu}\right) - (L\beta + Mu)\right)\right]$ is positive definite, we have now established that the general approximation of MSEP,

$$M_1(\theta) + C^{KR}(\theta)$$  \hspace{1cm} (3-5)

is positive definite, and hence nonsingular, and thus is a valid measure of prediction error. □

### 3.2 Multiple Dimension Bias Correction

Recall that it is often necessary to correct for the bias introduced into the estimate of MSEP by using $M_1(\hat{\theta})$ to estimate $M_1(\theta)$. Kenward and Roger (1997) define a bias correction term for multiple dimension linear combinations akin to the single dimension bias correction offered by Harville and Jeske (1992) found in Equation (2-11). We now rewrite this term to resemble the correction term in Equation (3-1), in order to maintain consistency in our notation and enable expansion to include linear combinations containing random effects. The definiteness properties of either the bias correction term or the estimate of MSEP containing the bias correction term has not previously been addressed. We will address that issue in Section 3.2.3. We first consider linear combinations containing only fixed effects.

#### 3.2.1 Bias Correction Estimator for Multiple Dimension Linear Combinations of Fixed Effects

We let $b_{ij}$ again be the $ij^{th}$ element of $B(\theta)$. Kenward and Roger (1997) state the bias correction term as

$$BIAS^{KR}(\theta) = -\frac{1}{2} \sum_{j=1}^{r} \sum_{i=1}^{r} b_{ij} \operatorname{cov}(L\hat{\beta},L\hat{\beta}) \frac{\partial^2}{\partial \theta_i \partial \theta_j},$$  \hspace{1cm} (3-6)

where
\[
\begin{bmatrix}
\frac{\partial^2 \text{var}(L, \tilde{\beta})}{\partial \theta_i \partial \theta_j} & \frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_i \partial \theta_j} & \cdots & \frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_i \partial \theta_j} \\
\vdots & \ddots & \vdots & \vdots \\
\frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_i \partial \theta_j} & \frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_i \partial \theta_j} & \cdots & \frac{\partial^2 \text{var}(L, \tilde{\beta})}{\partial \theta_i \partial \theta_j}
\end{bmatrix}
\]

Thus, the \(km^{th}\) element of \(\text{BIAS}^{KR}(\theta)\) is

\[
\{\text{BIAS}^{KR}_{km}(\theta)\} = -\frac{1}{2} \left( b_{11} \frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_i^2} + b_{12} \frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_i \partial \theta_2} + \cdots + b_{ij} \frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_i \partial \theta_j} + \cdots + b_{rr} \frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_r^2} \right) \quad (3-7)
\]

We now rewrite \(\text{BIAS}^{KR}(\theta)\) as

\[
\text{BIAS}^{KR}(\theta) = -\frac{1}{2} \begin{bmatrix}
\text{tr}\{\Lambda_{11}(\theta)B(\theta)\} & \cdots & \text{tr}\{\Lambda_{1l}(\theta)B(\theta)\} \\
\vdots & \ddots & \vdots \\
\text{tr}\{\Lambda_{l1}(\theta)B(\theta)\} & \cdots & \text{tr}\{\Lambda_{ll}(\theta)B(\theta)\}
\end{bmatrix} \quad (3-8)
\]

where

\[
\Lambda_{km}(\theta) = \frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta \partial \theta'} = \begin{bmatrix}
\frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_i^2} & \cdots & \frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_i \partial \theta_r} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_l \partial \theta_l} & \cdots & \frac{\partial^2 \text{cov}(L, \tilde{\beta}, L, \tilde{\beta})}{\partial \theta_l \partial \theta_l}
\end{bmatrix}
\]

To confirm the equality of Equations (3-8) and (3-6) note that the \(km^{th}\) element of Equation (3-8) is also
\[
\{ \text{BIAS}^{KR}_{km} (\theta) \} = tr \left\{ \Lambda_{lm} (\theta) B (\theta) \right\} = - \frac{1}{2} \left\{ b_{11} \frac{\partial^2 \text{cov} (L_1 \bar{\beta}, L_n \bar{\beta})}{\partial \theta_i^2} + b_{12} \frac{\partial^2 \text{cov} (L_1 \bar{\beta}, L_n \bar{\beta})}{\partial \theta_1 \partial \theta_2} + \cdots + b_{ij} \frac{\partial^2 \text{cov} (L_1 \bar{\beta}, L_n \bar{\beta})}{\partial \theta_j \partial \theta_i} + \cdots + b_{rr} \frac{\partial^2 \text{cov} (L_r \bar{\beta}, L_n \bar{\beta})}{\partial \theta_r^2} \right\}
\]

(3-9)

which is identical to Equation (3-7). It has been shown by Kenward and Roger (1997) that Equation (3-6) reduces to the multiple dimension correction term (\( C^{KR} (\theta) \)) when the covariance matrix, \( V \) is linear in \( \theta \), which gives \( M^{KR} (\hat{\theta}) = M_1 (\hat{\theta}) + 2C^{KR} (\hat{\theta}) \) as an estimator of the MSEP.

In the next section, we use Equations (3-8) and (3-9) to define the estimator for bias correction for a multiple-dimension linear combination involving fixed and random effects and confirm that the resulting bias correction term continues to reduce to the correction term in linear covariance cases.

### 3.2.2 Bias Correction Estimator for Multiple Dimension Linear Combinations of Fixed and Random Effects

The extension to include random effects is a straightforward but tedious task. However, it is important to explore fully the functions involved to determine when the term simplifies to the correction term approximation in Definition 3.1, and also to assess the effect on the definiteness properties of the MSEP estimator. The generalized bias correction term has the same form as in the fixed effect case:

\[
\text{BIAS}^{KR} (\theta) = - \frac{1}{2} \begin{bmatrix}
tr \left\{ \Lambda_{11} (\theta) B (\theta) \right\} & \cdots & tr \left\{ \Lambda_{ii} (\theta) B (\theta) \right\} \\
\vdots & \ddots & \vdots \\
tr \left\{ \Lambda_{nn} (\theta) B (\theta) \right\} & \cdots & tr \left\{ \Lambda_{nn} (\theta) B (\theta) \right\}
\end{bmatrix}
\]

(3-10)

but now,
\[ \Lambda_{km}(\theta) = \frac{\partial^2 \text{cov} \left( L_k \hat{\beta} + M_k \hat{\mu}, L_m \hat{\beta} + M_m \hat{\mu} \right)}{\partial \theta \partial \theta'} \]

\[ = \begin{bmatrix}
\frac{\partial^2 \text{cov} \left( L_k \hat{\beta} + M_k \hat{\mu}, L_m \hat{\beta} + M_m \hat{\mu} \right)}{\partial \theta_1 \partial \theta_1} & \cdots & \frac{\partial^2 \text{cov} \left( L_k \hat{\beta} + M_k \hat{\mu}, L_m \hat{\beta} + M_m \hat{\mu} \right)}{\partial \theta_r \partial \theta_r} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 \text{cov} \left( L_k \hat{\beta} + M_k \hat{\mu}, L_m \hat{\beta} + M_m \hat{\mu} \right)}{\partial \theta_1 \partial \theta_r} & \cdots & \frac{\partial^2 \text{cov} \left( L_k \hat{\beta} + M_k \hat{\mu}, L_m \hat{\beta} + M_m \hat{\mu} \right)}{\partial \theta_r \partial \theta_r}
\end{bmatrix} \]  

(3-11)

Now considering a single element of this matrix we see that

\[ \{ \Lambda_{km}(\theta) \} = \frac{\partial^2 \text{cov} \left( L_k \hat{\beta} + M_k \hat{\mu}, L_m \hat{\beta} + M_m \hat{\mu} \right)}{\partial \theta_i \partial \theta_j} \]

\[ = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \left[ \text{cov} \left( L_k \hat{\beta}, L_m \hat{\beta} \right) + \text{cov} \left( M_k \hat{\mu}, L_m \hat{\beta} \right) + \text{cov} \left( L_k \hat{\beta}, M_m \hat{\mu} \right) + \text{cov} \left( M_k \hat{\mu}, M_m \hat{\mu} \right) \right] \]

\[ = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \left[ \text{cov} \left( L_k \hat{\beta}, L_m \hat{\beta} \right) \right] + \frac{\partial^2}{\partial \theta_i \partial \theta_j} \left[ \text{cov} \left( M_k \hat{\mu}, L_m \hat{\beta} \right) \right] + \frac{\partial^2}{\partial \theta_i \partial \theta_j} \left[ \text{cov} \left( L_k \hat{\beta}, M_m \hat{\mu} \right) \right] + \frac{\partial^2}{\partial \theta_i \partial \theta_j} \left[ \text{cov} \left( M_k \hat{\mu}, M_m \hat{\mu} \right) \right]. \]

(3-12)

We now investigate each piece of Equation (3-12) using Littell et al. (2006) as a reference for the covariances involved. Consider Part (1) of Equation (3-12):
\[(1) = \frac{\partial^2}{\partial \theta \partial \theta} \text{cov} \left( L_k \hat{\beta}, L_m \hat{\beta} \right) = \frac{\partial^2}{\partial \theta \partial \theta} \text{cov} \left[ L_k \left( X'V'X \right)^d X'V'y, L_m \left( X'V'X \right)^d X'V'y \right] \]

\[= \frac{\partial^2}{\partial \theta \partial \theta} \left[ L_k \left( X'V'X \right)^d X'V'X \left( X'V'X \right)^d L'_m \right] - \frac{\partial^2}{\partial \theta \partial \theta} L_k \left( X'V'X \right)^d L'_m \]

\[= \frac{\partial}{\partial \theta} \left[ L_k \left( X'V'X \right)^d X'V' \frac{\partial V}{\partial \theta} V'X \left( X'V'X \right)^d L'_m \right] \]

\[= L_k \left( X'V'X \right)^d X'V' \frac{\partial V}{\partial \theta} V'X \left( X'V'X \right)^d L'_m + \]

\[L_k \left( X'V'X \right)^d X'V' \frac{\partial V}{\partial \theta} V'X \left( X'V'X \right)^d L'_m \]

\[-L_k \left( X'V'X \right)^d X'V' \frac{\partial V}{\partial \theta} V'X \left( X'V'X \right)^d L'_m \]

\[+ L_k \left( X'V'X \right)^d X'V' \frac{\partial V}{\partial \theta} V'X \left( X'V'X \right)^d L'_m \]

\[= C_{i_k} V'P_{m_k} - C_{i_j} V'P_{m_j} - C_{i_l} V'P_{m_l} + C_{i_r} V'P_{m_r} \]

\[+ L_k \left( X'V'X \right)^d X'V' \frac{\partial V}{\partial \theta} V'X \left( X'V'X \right)^d L'_m \]

\[= -C_{i_k} V' \left( I - P \right) C_{m_k} - C_{i_j} V' \left( I - P \right) C_{m_j} + L_k \left( X'V'X \right)^d X'V' \frac{\partial V}{\partial \theta} V'X \left( X'V'X \right)^d L'_m, \]  

where \( C' \) is as defined in Section 3.1.2.

Now moving on to part (2) of Equation (3-12), we have

\[(2) = \frac{\partial}{\partial \theta \partial \theta} \text{cov} \left( M_k \hat{u}, L_m \hat{\beta} \right) = \frac{\partial}{\partial \theta \partial \theta} \text{cov} \left[ M_k \left( Z'V'X \right)^d X'V'y, L_m \left( X'V'X \right)^d X'V'y \right] \]

\[= \frac{\partial}{\partial \theta} \left[ M_k \frac{\partial Z}{\partial \theta} V'X \left( X'V'X \right)^d L'_m \right] + M_k \frac{\partial Z}{\partial \theta} \left[ V'X \left( X'V'X \right)^d L'_m \right] \]

\[+ M_k \frac{\partial Z}{\partial \theta} \left[ V'X \left( X'V'X \right)^d L'_m \right] \]  

\[= (3-14) \]
\[ \begin{align*}
&= M_k \frac{\partial G}{\partial \theta_i} Z V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&- M_k \frac{\partial G}{\partial \theta_i} Z V^{-1} X (X V^{-1} X)^{-1} X V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&+ M_k \frac{\partial G}{\partial \theta_j} Z V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&- M_k G V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&+ M_k G V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&+ M_k G V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} X (X V^{-1} X)^{-1} X V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&- M_k \frac{\partial G}{\partial \theta_j} Z V^{-1} X (X V^{-1} X)^{-1} X V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&+ M_k G V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&- M_k G V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&+ M_k G V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} X (X V^{-1} X)^{-1} X V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&- M_k G V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&+ M_k G V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&- M_k \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} Z V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&- M_k G V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} X (X V^{-1} X)^{-1} L'_m \\
&- M_k G V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} X (X V^{-1} X)^{-1} L'_m
\end{align*} \]
\[ -D'_k V^{-1} C_{mj} + D'_k V^{-1} P C_{mj} - D'_k V^{-1} C_{mi} + D'_k V^{-1} P C_{mi} \]
\[ -E'_k V^{-1} C_{mj} + E'_k V^{-1} P C_{mj} - E'_k V^{-1} C_{mi} + E'_k V^{-1} P C_{mi} \]
\[ -F'_k V^{-1} C_{mj} + F'_k V^{-1} P C_{mj} - F'_k V^{-1} C_{mi} + F'_k V^{-1} P C_{mi} \]
\[ \text{(3-15)} \]
\[ -M_k \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} Z V^{-1} X (X' V^{-1} X)^{-1} L'_m - M_k G Z V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} X (X' V^{-1} X)^{-1} L'_m \]
\[ -M_k G Z V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} X (X' V^{-1} X)^{-1} L'_m \]
\[ = -D'_k V^{-1} (I - P) C_{mj} - D'_k V^{-1} (I - P) C_{mi} \]
\[ -E'_k V^{-1} (I - P) C_{mj} - E'_k V^{-1} (I - P) C_{mi} \]
\[ -F'_k V^{-1} (I - P) C_{mj} - F'_k V^{-1} (I - P) C_{mi} \]
\[ \text{(3-16)} \]
\[ -M_k \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} Z V^{-1} X (X' V^{-1} X)^{-1} L'_m - M_k G Z V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} X (X' V^{-1} X)^{-1} L'_m \]
\[ -M_k G Z V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} X (X' V^{-1} X)^{-1} L'_m \]

where \( D'_k, E'_k, \) and \( F'_k \) are as defined in Section 3.1.2. Due to the symmetry of parts (2) and (3) of Equation (3-12), we combine (2) and (3) as follows:

\[ (2) + (3) = -\left( D_{ki} + E_{kij} + F_{kij} \right) V^{-1} (I - P) C_{mj} \]
\[ -\left( D_{kj} + E_{kij} + F_{kij} \right) V^{-1} (I - P) C_{mi} \]
\[ -C'_k V^{-1} (I - P) \left( D_{kj} + E_{kij} + F_{kij} \right) \]
\[ -C'_k V^{-1} (I - P) \left( D_{kj} + E_{kij} + F_{kij} \right) \]
\[ \left( \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} \right) = g_1 \left( \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} \right) \]

where \( g_1 \left( \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} \right) \) are linear functions of \( \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} \) and \( \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} \), respectively.

Part (4) of Equation (3-12) tediously completes \( \{ \Lambda_{km(ij)}(\theta) \} \) as follows:
\begin{align}
(4) &= \frac{\partial^2}{\partial \theta_i \partial \theta_j} \left[ \text{cov} \left( M_k \tilde{u}, M_m \tilde{u} \right) \right] = \frac{\partial^2}{\partial \theta_i \partial \theta_j} M_k \left[ G^{-1} \right] \left[ GZV^{-1} (I - P) ZG \right] M_m' \\
&= M_k \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} - M_k \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} ZV^{-1} (I - P) ZG M_m' \\
&- M_k GZV^{-1} (I - P) Z \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} M_m' + M_k GZV^{-1} \frac{\partial^2 P}{\partial \theta_i \partial \theta_j} ZG M_m' \\
&- M_k GZ'V^{-1} (I - P) ZGM_m' \\
&- D'_{ki} V^{-1} (I - P) D_{mj} - D'_{kj} V^{-1} (I - P) D_{mi} \\
&- F'_{ki} V^{-1} (I - P) D_{mj} - F'_{kj} V^{-1} (I - P) D_{mi} \\
&- F'_{ki} V^{-1} (I - P) E_{mj} - F'_{kj} V^{-1} (I - P) E_{mi} \\
&+ E'_{ki} V^{-1} PE_{mj} + E'_{kj} V^{-1} PE_{mi} \\
&+ E'_{ki} V^{-1} PF_{mj} + E'_{kj} V^{-1} PF_{mi}, \\
\end{align}

Now note that
\begin{align}
\frac{\partial P}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i} \left[ X \left( X'V^{-1}X \right)^{-1} X'V^{-1} \right] \\
&= X \left( X'V^{-1}X \right)^{-1} X'V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} X \left( X'V^{-1}X \right)^{-1} X'V^{-1} - X \left( X'V^{-1}X \right)^{-1} X'V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} \\
\end{align}

which leads to
\[
\frac{\partial^2 P}{\partial \theta_i \partial \theta_j} = X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \\
-X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \\
+X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \\
-X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \\
+X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \\
-X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \\
+X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} \\
+X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} \\
-X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} .
\]

Now we have

\[
M_k G Z V^{-1} \frac{\partial^2 P}{\partial \theta_i \partial \theta_j} - Z G M'_m = -F'_{ji} V^{-1} (I - P) F_{mi} - F'_{ji} V^{-1} (I - P) F_{mi} - F'_{ji} V^{-1} (I - P) E_{mi} \\
-F'_{ki} V^{-1} (I - P) E_{mj} + M_k X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}M'_m \\
-M_k X(X'V^{-1}X)^{-1}X'V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1}M'_m.
\]

To complete Equation (3-18) we also must investigate \( \frac{\partial^2 V^{-1}}{\partial \theta_i \partial \theta_j} : \)

\[
\frac{\partial^2 V^{-1}}{\partial \theta_i \partial \theta_j} = -\frac{\partial}{\partial \theta_j} \left( V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} \right) = V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} + V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} - V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} .
\]
This gives us

\[
M_z G' Z^{-1} (I - P) Z G'_m = M_z G' V^{-1} \frac{\partial V}{\partial \theta_i} V^{-1} \frac{\partial V}{\partial \theta_j} (I - P) Z G'_m \\
+ M_z G' V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} \frac{\partial V}{\partial \theta_i} (I - P) Z G'_m - M_z G' V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} (I - P) Z G'_m
\]

\[
= \tilde{E}' V^{-1} E_{m_j} + \tilde{E}' V^{-1} E_{m_i} + \tilde{E}' V^{-1} F_{m_j} + \tilde{E}' V^{-1} F_{m_i} - M_z G' V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} (I - P) Z G'_m.
\]

Utilizing these expressions for \( \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} \) and \( \frac{\partial^2 P}{\partial \theta_i \partial \theta_j} \) in Equation (3-18) results in

\[
(4) = M_z \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} - M_z \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} Z V^{-1} (I - P) Z G'_m
\]

\[
-M_z G' V^{-1} (I - P) Z \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} - M_z G' V^{-1} \frac{\partial^2 P}{\partial \theta_i \partial \theta_j} Z G'_m
\]

\[
-M_z G' \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} (I - P) Z G'_m
\]

\[
-D'_{ki} V^{-1} (I - P) D_{m_j} - D'_{ki} V^{-1} (I - P) D_{m_i} - D'_{ki} V^{-1} (I - P) E_{m_j} - D'_{ki} V^{-1} (I - P) E_{m_i}
\]

\[
-F'_{ki} V^{-1} (I - P) D_{m_j} - F'_{ki} V^{-1} (I - P) D_{m_i} - F'_{ki} V^{-1} (I - P) E_{m_j} - F'_{ki} V^{-1} (I - P) E_{m_i}
\]

\[
-D'_{ki} V^{-1} (I - P) F_{m_j} - D'_{ki} V^{-1} (I - P) F_{m_i} - D'_{ki} V^{-1} (I - P) E_{m_j} - D'_{ki} V^{-1} (I - P) E_{m_i}
\]

\[
-E'_{ki} V^{-1} (I - P) F_{m_j} - E'_{ki} V^{-1} (I - P) F_{m_i} - E'_{ki} V^{-1} (I - P) E_{m_j} - E'_{ki} V^{-1} (I - P) E_{m_i}
\]

\[
-F'_{ki} V^{-1} (I - P) F_{m_j} + F'_{ki} V^{-1} (I - P) F_{m_i} + M_z G' V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} (I - P) Z G'_m
\]

\[
+M_z G' X (X' V^{-1} X)^{-1} X' V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1} Z G'_m
\]

\[
-M_z G' X (X' V^{-1} X)^{-1} X' V^{-1} \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} V^{-1} Z G'_m
\]

(3-19)

which reduces to

\[
(4) = -\left( D_{ki} + E_{ki} + F_{ki} \right) V^{-1} (I - P) \left( D_{m_j} + E_{m_j} + F_{m_j} \right)
\]

\[
-\left( D_{kj} + E_{kj} + F_{kj} \right) V^{-1} (I - P) \left( D_{ki} + E_{ki} + F_{ki} \right)
\]

\[
+ f_2 \left( \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} \right) - g_2 \left( \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} \right).
\]

(3-20)
Piecing together (1) – (4) in Equation (3-12) results in:

\[
\Lambda_{km(ij)}(\theta) = \frac{\partial^2 \text{cov} \left( L_i \tilde{\beta} + M_i \tilde{u}, L_m \tilde{\beta} + M_m \tilde{u} \right)}{\partial \theta_i \partial \theta_j}
\]

\[
= -\left( C_{ki} + D_{ki} + E_{ki} + F_{ki} \right)' V^{-1} (I - P) \left( C_{mj} + D_{mj} + E_{mj} + F_{mj} \right) \\
- \left( C_{kj} + D_{kj} + E_{kj} + F_{kj} \right)' V^{-1} (I - P) \left( C_{mi} + D_{mi} + E_{mi} + F_{mi} \right) \\
+ f \left( \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} \right) + g \left( \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} \right). \tag{3-21}
\]

When \( G \) and \( V \) are linear in \( \theta \) the functions of \( f \left( \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} \right) \) and \( g \left( \frac{\partial^2 V}{\partial \theta_i \partial \theta_j} \right) \) will be zero. In that case

\[
\Lambda_{km}(\theta) = -2 \left[ \left( C_{k1} + D_{k1} + E_{k1} + F_{k1} \right)' \right. \\
\vdots \\
\left. \left( C_{kr} + D_{kr} + E_{kr} + F_{kr} \right)' \right] \\
\times V^{-1} (I - P) \left[ \left( C_{m1} + D_{m1} + E_{m1} + F_{m1} \right) \cdots \left( C_{mr} + D_{mr} + E_{mr} + F_{mr} \right) \right]. \tag{3-22}
\]

Comparing Equation (3-22) to Equation (3-2) gives the desired result that in the most general case involving multiple dimension linear combinations of fixed and random effects

\[
\text{BIAS}^{KR}(\theta) = C^{KR}(\theta) \quad \text{when} \quad G \text{ and } V \text{ are linear in } \theta. \quad \text{Thus, just as in the single dimension fixed effects linear combinations, when } G \text{ and } V \text{ are linear in } \theta, \quad M^{KR}(\hat{\theta}) = M_1(\hat{\theta}) + 2C^{KR}(\hat{\theta}) \]

is an estimator of the MSEP of the linear combination of fixed and random effects that accounts for both the correction term and the bias correction term.

### 3.2.3 Definiteness Properties of the Estimator for MSEP for Multiple-Dimension Linear Combinations of Fixed and Random Effects

Now that all pieces of the estimator have been defined in the multiple dimension setting an estimator for the MSEP can be defined as
\[
M_{KR}^{\hat{\theta}}(\theta) = M_1(\hat{\theta}) + C_{KR}^{\hat{\theta}} + BIAS_{KR}^{\hat{\theta}}
\] 

(3-23)

It has not been proven that this MSEP estimator, either in single or multiple dimensions, is non-negative definite. Just as in the single dimension case, we have shown above that when the variance matrix for the model is linear in \(\theta\), the bias correction estimator, \(BIAS_{KR}^{\hat{\theta}}(\theta)\), reduces to the correction term, \(C_{KR}^{\hat{\theta}}(\theta)\), which we have shown is non-negative definite in its most general form. Hence the simplified MSEP estimator,

\[
M_{KR}^{\hat{\theta}}(\theta) = M_1(\hat{\theta}) + 2C_{KR}^{\hat{\theta}}
\]

(3-24)

is nonsingular and therefore a valid measure of prediction error. However, when the covariance matrix, \(V\), is not linear in \(\theta\), neither the bias correction term nor the MSEP estimator including the bias correction term has been previously shown to be non-negative definite. In Chapter 4, by exploring the special case of the balanced one-way random effects model, we provide a counter example to prove that the bias correction is, in fact, indefinite. Indeed, we show that in a single dimension case the estimate of MSEP given in Equation (3-23) can be smaller than the naïve estimate, \(M_1(\hat{\theta})\). From Equation (3-21) it is clear that the indefiniteness of \(BIAS_{KR}^{\hat{\theta}}(\theta)\) is caused by \(f\left(\frac{\partial^2 G}{\partial \theta_i \partial \theta_j}\right)\) and \(g\left(\frac{\partial^2 V}{\partial \theta_i \partial \theta_j}\right)\). Note that it is not necessary for the bias correction term to be non-negative definite for the overall estimator of the MSEP to be non-negative definite. In fact, it is reasonable for the bias correction term to be indefinite since it is estimating the bias of \(M_1(\hat{\theta})\), which is also indefinite. In Chapter 4, we provide a single dimension case where the negativity of the bias correction term causes the entire estimate of \(M_1(\theta)\) to be
negative, proving that $M_1(\hat{\theta}) + BIAS^{KR}(\hat{\theta})$, the overall MSEP estimator for $M_1(\theta)$, is indefinite and a poor choice for estimating the MSEP.

### 3.3 Transform Invariance of MSEP Estimator

Another important issue concerning these estimators is transform invariance. Consider two parameterizations of a model, $\theta$ and $\Phi = g(\theta)$. Kackar and Harville (1984) showed that the approximation for the correction term in the single dimension case, $tr\left[A(\theta)B(\theta)\right]$, is transform invariant under the conditions that $\hat{\Phi} = g(\hat{\theta})$ and that

$$B'(\theta) = \left[J(\theta)^{-1}B(g(\theta))J(\theta)^{-1}\right]^{-1}$$

where $B^*$ and $B$ are as in $tr\left[A(\theta)B(\theta)\right]$ and $J(\theta)$ is the Jacobian of the transformation. Both of these conditions are satisfied by using either REML or ANOVA variance component estimation and with $B(\theta) = I^{-1}(\theta)$, the inverse of the observed information matrix. This proof applies to the multiple dimension case, as we have shown that the multiple dimension correction term approximation is simply a matrix composed of the single dimension approximations as the components. Since each element of the matrix is transformation invariant, the entirety is transform invariant as well.

It is also easy to show that the naïve estimator of the MSEP, $M_1(\theta)$, is also transform invariant under the condition that $\hat{\Phi} = g(\hat{\theta})$. The invariance of the naïve estimator indicates that the true bias of $M_1(\hat{\theta})$, $E\left[M_1(\hat{\theta})\right] - M_1(\theta)$, will also be transform invariant (again with the condition, $\hat{\Phi} = g(\hat{\theta})$). Thus a desirable property for the estimator of the bias correction term is transformation invariance. Clearly, $BIAS^{KR}(\theta)$ is not transform invariant. In Chapter 4, we demonstrate the lack of transform invariance for a single dimension linear combination in the
one way random effects model. This is an undesirable quality of the bias correction term estimator. We will investigate its effects on the balanced one-way random effects model in Chapter 4.
CHAPTER 4
BALANCED ONE-WAY RANDOM EFFECTS MODEL

Many of the methods summarized in Chapter 2 were developed only for estimating and testing fixed effects (e.g., Giesbrecht and Burns 1985; Kenward and Roger 1997); however, the methods are often extended to include random effects without documented justification (SAS® 2003). In Chapter 3, the elements involved in extending these methods to the most general cases were derived. By looking at the simplest model including random effects, namely the balanced one-way random-effects model, we can examine if the use of these methods is as straightforward as previously claimed or if additional issues arise. The simplicity of this model often allows closed form expressions to be obtained for the approximations and estimators of the MSEP and their expected values, in addition to the true value of the MSEP being estimated. Peixoto and Harville (1986) derived closed form expressions for the true MSEP for a class of models including the balanced one-way random effects model. Similar methods are used here to enable direct comparisons of various MSEP estimators. By using this simple, balanced model, we can measure the result of including random effects on the MSEP estimates, as well as identify issues that may be more difficult to handle in more complex models.

Consider the model:

\[
y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n
\]

\[
e_{ij} \sim N(0, \sigma^2), \quad a_i \sim N(0, \sigma_a^2)
\]

where \(-\infty < \mu < \infty, \sigma^2 > 0, \sigma_a^2 \geq 0\). We use the parameterization \(\theta = (\sigma^2, \nu)\) where

\[
\nu = \frac{\sigma^2}{\sigma^2 + n\sigma_a^2}
\]

and consider the problem of predicting the mean of the \(i^{th}\) level of the random effect. Note that the variance matrix, \(V\), is not linear in this parameterization, nor is this the parameterization used by SAS®. This parameterization is chosen to demonstrate the impact of a
non-linear parameterization on the bias correction term and for consistency with past work on this topic. Comparisons will be made to results under the alternative parameterization, \( \Phi = (\sigma^2, \sigma_a^2) \) for which \( V \) is linear. The BLUP and the naïve MSEP, \( M_i(\theta) \), for \( t = a_i \) are straightforward to derive, as demonstrated in Kackar and Harville (1984, p. 857):

\[
\tilde{\tau}(\theta) = (1-\nu)(\bar{y}_i - \bar{y}), \text{ where } \bar{y}_i = \frac{\sum_{j=1}^{n} y_{ij}}{n}, \text{ and } \bar{y} = \frac{\sum_{i=1}^{k} \bar{y}_i}{k}, \tag{4-2}
\]

\[
M_i(\theta) = \frac{\sigma^2 (1-\nu)}{nk\nu} \left[ 1 + (k-1)\nu \right]. \tag{4-3}
\]

Noting that \( \tilde{\tau}(\theta) \) is linear in \( \theta \), the accuracy of the Kackar-Harville approximation of the correction term depends on four issues:

1) using the exact MSE of \( \hat{\theta} \) versus an approximation, such as \( I^{-1}(\theta) \), for \( B(\theta) \);
2) the appropriateness of the Taylor series approximation, depending in part on how close \( \theta \) is to the boundary of \( \Omega \);
3) whether \( \hat{\theta} \) is unbiased for \( \theta \);
4) the proximity of \( \text{cov}\left[ d(y;\theta)d(y;\theta)'(\hat{\theta}-\theta)(\hat{\theta}-\theta)'\right] \) to zero.

We can evaluate these four criteria to determine their effects on the accuracy of the Kackar-Harville approximation. We shall examine each of these criterion using both REML and ANOVA methods of variance component estimation.

4.1 **Kackar-Harville Approximation Based on REML Estimation Method**

Using REML methods to estimate the variance components results in \( \hat{\theta} = (\hat{\sigma}^2, \hat{\nu}) \) where
\[ \hat{\sigma}^2 = m_e \text{ if } m_a > m_e \]
\[ = \frac{(k-1)m_a + k(n-1)m_e}{nk-1} \text{ if } m_a \leq m_e \]
\[ \hat{\nu} = \min\left(\frac{m_e}{m_a}, 1\right) \tag{4-4} \]

where
\[ m_e = \frac{\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{k(n-1)} \text{ and } m_a = \frac{n \sum_{i=1}^k (\bar{y}_i - \bar{y})^2}{k-1} \]

(e.g., Kackar and Harville 1984; Searle, Casella and McCulloch 1992).

Note that these estimators remain within the parameter space, \( \Omega \), by definition. The EBLUP for \( t = a_i \) is
\[ \hat{\tau} = \tilde{\tau}\left(\hat{\Theta}\right) = (1 - \hat{\nu})(\bar{y}_i - \bar{y}) \tag{4-5} \]

and the naïve estimate of the MSEP of the EBLUP is
\[ M_1\left(\hat{\Theta}\right) = \frac{\hat{\sigma}^2 (1 - \hat{\nu})}{nk\hat{\nu}} \left[ 1 + (k-1)\hat{\nu} \right]. \tag{4-6} \]

The true correction term in the REML case, as derived by Kackar and Harville (1984, p. 731), is
\[ C_R\left(\hat{\Theta}\right) = E\left[ \hat{\tau} - \tilde{\tau} \right]^2 = \frac{\sigma^2 \nu (nk-1) \xi_i}{nk} \tag{4-7} \]

where
\[ \xi_i = \frac{\Psi_{v_e \nu^{-1}} (1 - \Psi)^{v_e \nu^{-1}}}{B(v_e, v_a) \nu} \left\{ \frac{2}{v_a} - \frac{(v_e + 1)}{v_e (v_a - 1)} - \frac{1}{\nu (v_a - 1)} \right\} \]
\[ + I_{v_e \nu} (v_e, v_a + 1) \left\{ \frac{v_a}{v_e + v_a} \left[ \frac{v_a (v_e + 1)}{v_e (v_a - 1)} - \frac{(1-\nu v)}{\nu^2} \right] + \frac{(1-\nu)^2}{\nu^2} \left[ \frac{v_a}{v_e + v_a} \right] \right\} \]
and
\[
B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad 1_x(a, b) = \left[ B(a, b) \right]^{-1} \int_0^x t^{a-1}(1-t)^{b-1} dt
\]
\[
v_e = \frac{k(n-1)}{2}, \quad v_a = \frac{k-1}{2}, \quad \Psi = \frac{v_e}{v_e + \nu v_a}.
\]

This results in a true MSEP, \( M_R(\theta) \), where
\[
M_R(\theta) = M_1(\theta) + C_R(\theta)
\]
\[
= M_1(\theta) + \frac{\sigma^2 \nu (nk-1) \xi_1}{nk}
\]
\[(4-8)\]

(Kackar and Harville 1984, p. 857).

In this model, the exact MSE matrix of \( \hat{\theta} \) is tractable when using REML estimates for the variance parameters. Thus assuming that \( \theta \) is sufficiently far from the boundary of \( \Omega \) and choosing \( B_R(\theta) \) to be the exact MSE matrix of \( \hat{\theta} \), we can assess the accuracy of
\[
tr\left[ A(\theta) B_R(\theta) \right]
\]
for \( C_R(\theta) \) by simply assessing the proximity of
\[
\text{cov}\left[ d(y; \theta) d(y; \theta)' , (\hat{\theta} - \theta)(\hat{\theta} - \theta)' \right]
\]
to zero.

To evaluate \( tr\left[ A(\theta) B_R(\theta) \right] \) first note that
\[
A(\theta) = \begin{bmatrix} 0 & 0 \\ 0 & a_{22} \end{bmatrix}
\]
\[(4-9)\]

and
\[
B_R(\theta) = \begin{bmatrix} E(\hat{\sigma}^2 - \sigma^2)^2 & E[(\hat{\nu} - \nu)(\hat{\sigma}^2 - \sigma^2)] \\ E[(\hat{\nu} - \nu)(\hat{\sigma}^2 - \sigma^2)] & E(\hat{\nu}^2) \end{bmatrix}
\]
\[(4-10)\]
where $a_{22} = \text{var}\left(\frac{\partial \hat{\xi}}{\partial \nu}\right) = \text{var}(\bar{y} - \bar{v}) = \frac{\sigma^2(k-1)}{nk\nu}$. As shown by Kackar and Harville (1984, p. 731), $E(\hat{v} - \nu)^2 = \nu^2 \xi_0$ where

$$
\xi_0 = \frac{\Psi^{v-1}(1-\Psi)^v}{B(v_e,v_a)\nu} \left\{ \frac{2}{(v_a-1)} - \frac{v_a(v_e+1)}{v_e(v_a-1)(v_a-2)} - \frac{1}{\nu(v_a-2)} \right\} \\
+ I_\Psi(v_e,v_a) \left\{ \frac{v_a^2(v_e+1)}{v_e(v_a-1)(v_a-2)} - \frac{2v_a}{(v_a-1)\nu^2} \right\} + \frac{(1-\nu)^2}{\nu^2}.
$$

As a result,

$$
\text{tr}\left[ A(\theta)B_R(\theta) \right] = a_{22} \nu^2 \xi_0 = \frac{\sigma^2\nu(k-1)\xi_0}{nk}.
$$

(4-11)

A Kackar-Harville approximation to the MSEP is then

$$
M_1(\theta) + \frac{\sigma^2\nu(k-1)\xi_0}{nk}
$$

(4-12)

(Kackar and Harville 1984, p. 857).

Now note that the only non-zero component of $\text{cov}\left[ d(y;\theta) d(y;\theta)' , (\hat{\theta} - \theta)(\hat{\theta} - \theta)' \right]$ is

$$
\text{cov}\left[ (\bar{y}_i - \bar{y})^2 , (\hat{v} - \nu)^2 \right] = E\left[ (\bar{y}_i - \bar{y})^2 (\hat{v} - \nu)^2 \right] - E\left[ (\bar{y}_i - \bar{y})^2 \right] E(\hat{v} - \nu)^2.
$$

Kackar and Harville (1984, p. 731) show that

$$
E\left[ (\bar{y}_i - \bar{y})^2 (\hat{v} - \nu)^2 \right] = \frac{\sigma^2\nu^2 \xi_1 (nk-1)}{nk} = C_R(\theta).
$$
It is also easy to verify that \( E \left[ (\bar{y}_i - \bar{y})^2 \right] = \frac{k-1}{nk} \left( \sigma^2 + n\sigma_a^2 \right) \), since \( \frac{(k-1)m_{\nu}}{\sigma^2 + n\sigma_a^2} \sim \chi^2_{k-1} \). Thus recalling that \( E \left( \hat{\nu} - \nu \right)^2 = \nu^2 \xi_0 \) and combining results, we have

\[
\text{cov} \left[ (\bar{y}_i - \bar{y})^2, (\hat{\nu} - \nu)^2 \right] = \frac{\sigma^2 \nu \xi_{\nu} (nk-1)}{nk} - \frac{\nu^2 \xi_0 (k-1)}{nk} \left( \sigma^2 + n\sigma_a^2 \right)
\]

\[
= \frac{\xi_{\nu} (nk-1)}{nk} - \frac{\nu^2 \xi_0 (k-1)}{nk}.
\]

(4-13)

Notice that this is simply the difference between the true correction term in the REML case, \( C_{\hat{\theta}}(\theta) \), and the approximation, \( tr \left[ A(\theta)B_{\hat{\theta}}(\theta) \right] \). This is only true when the exact MSE of \( \hat{\theta} \) is used for \( B_{\hat{\theta}}(\theta) \). Thus the condition for \( C_{\hat{\theta}}(\theta) = tr \left[ A(\theta)B_{\hat{\theta}}(\theta) \right] \), i.e., the condition for

\[
\text{cov} \left[ d(y;\theta)d(y;\theta)', (\hat{\theta} - \theta)(\hat{\theta} - \theta)' \right] = 0
\]

is

\[
\xi_{\nu} (nk-1) = \xi_0 (k-1).
\]

(4-14)

Figure 4-1 shows Equation (4-13) as a function of \( \nu \) for several values of \( k \), holding \( n = 6 \), with the condition in Equation (4-14) met when the functions cross the zero line. Note that throughout, without loss of generality, we hold \( \sigma^2 = 1 \). As \( k \) increases, the functions move closer to zero, indicating that the Kackar-Harville approximation for \( C_{\hat{\theta}}(\theta) \) improves as \( k \) increases. However, as \( k \) increases, the amount of information about the random effect increases, reducing the uncertainty about the covariance parameter estimates. Hence \( C_{\hat{\theta}}(\theta) \) tends toward zero and accounting for the correction term becomes less important, rendering the correction term approximation irrelevant. In the remainder of this chapter, the analysis will focus on a model with \( k = 6 \) levels of the random effect to demonstrate the impact of the different estimators for the MSEP.
In most models, the exact MSE of $\hat{\theta}$ is not tractable and an approximation of the MSE of $\hat{\theta}$ must be used. A common choice is $I^{-1}(\theta)$, the inverse of the observed information matrix, with $\theta$ replaced by $\hat{\theta}$ in practice. Consider the Kackar-Harville approximation of $C_r(\theta)$ where $B(\theta)$ is the asymptotic MSE of $\hat{\theta}$, $I^{-1}(\theta)$. Kackar and Harville (1984, p. 857) show that

$$tr[A(\theta)I^{-1}(\theta)] = \frac{2\sigma^2\nu kn - 1}{k^2 n(n - 1)}$$

(4-15)

where

$$I^{-1}(\theta) = \begin{bmatrix}
\frac{2(\sigma^2)^2}{k(n - 1)} & \frac{2\sigma^2\nu}{k(n - 1)} \\
\frac{2\sigma^2\nu}{k(n - 1)} & \frac{2\nu^2 kn - 1}{k(n - 1)(k - 1)}
\end{bmatrix}.$$  

(4-16)

Figure 4-2 compares the two Kackar-Harville approximations in Equations (4-11) and (4-15) with the correction term, $C_r(\theta)$, for $k = 6$ and $n = 6$. The success of the approximation is greatly influenced by the value of $\nu$. For small values of $\nu$, which correspond to large values of $\sigma_a^2$ relative to $\sigma^2$, $tr[A(\theta)I^{-1}(\theta)]$, is superior to $tr[A(\theta)B_r(\theta)]$ for approximating the true correction term, $C_r(\theta)$. However, as $\nu$ moves toward 1, which corresponds to the value of $\sigma_a^2$ growing relatively smaller, the roles reverse and $tr[A(\theta)B_r(\theta)]$ becomes the better approximation. In some cases, the choice of how to measure the variability of $\hat{\theta}$ may cause severe over-estimation of the correction term, in turn causing a greatly inflated value for the MSEP estimate. SAS® PROC MIXED uses $tr[A(\theta)I^{-1}(\theta)]$ which overestimates the true
correction term by over 100% for half of the parameter space. Using an inflated estimate for the MSEP could lead to overly conservative interpretation.

Because the correction term approximations in Equations (4-11) and (4-15) make use of the unknown variance parameters, estimates of the variance parameters must be substituted into these approximations. Note that there are now two levels of error being introduced into the correction term estimators: first, the true correction term is approximated by the Taylor series methods to yield the Kackar-Harville approximation; second, the variance parameters involved in the approximations are replaced by their estimators. In the case of $tr\left[A(\hat{\theta})I^1(\hat{\theta})\right]$ there is a third contribution to the error in that $I^1(\theta)$ is an approximation of the MSE of $\hat{\theta}$. Substituting the REML estimates of the variance components into Equations (4-11) and (4-15) gives two possible estimators for the correction term:

$$tr\left[A(\hat{\theta})B(\hat{\theta})\right] = \frac{\sigma^2\hat{\nu}(k-1)\hat{\xi}_0}{nk} \quad (4-17)$$

and

$$tr\left[A(\hat{\theta})I^1(\hat{\theta})\right] = \frac{2\sigma^2\hat{\nu}(kn-1)}{k^2n(n-1)}. \quad (4-18)$$

It is suggested by Prasad and Rao (1990) that the bias introduced by substituting variance parameters in these approximations is $o(k^{-1})$. To assess this assertion, we investigate the accuracy of Equations (4-17) and (4-18) by considering the expected values of both estimators. The expected value of Equation (4-17) is not available in closed form, so a simple Monte Carlo study was performed to evaluate the accuracy of the estimator for the one-way random effects model in Equation (4-1) with $k = 6$ and $n = 6$. Using the RANNOR function in SAS®, 10,000
independent sets of $e_{ij}$ and $a_i$ were generated from $N(0, \sigma^2 = 1)$ and $N\left(0, \sigma_i^2\right)$ for several values of $\sigma_i^2$, respectively. The values of $\sigma_i^2$ correspond to $\nu = \{0.05, 0.1, 0.15\ldots, 1.0\}$. Setting $
=0$, 10,000 sets of $\{y_{ij} : i=1\ldots6, j=1\ldots6\}$ were created from the model $y_{ij} = a_i + e_{ij}$. The Monte Carlo value of $E\left\{tr\left[A\left(\hat{\theta}\right)B_R\left(\hat{\theta}\right)\right]\right\} = E\left[\frac{\hat{\sigma}^2 \hat{\phi}(k-1)\hat{\xi}_{50}}{nk}\right]$ and the Monte Carlo standard error were obtained from the 10,000 replicates. Selected results of the study are contained in Table 4-1 (standard errors contained in parentheses) along with the values of $C_R\left(\theta\right)$ (Equation (4-7)) and $tr\left[A\left(\theta\right)B_R\left(\theta\right)\right]$ (Equation (4-11)).

Table 4-1. Accuracy of Kackar-Harville estimator (4-17) for $C_R\left(\theta\right)$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_R\left(\theta\right)$</td>
<td>.01537</td>
<td>.01890</td>
<td>.01758</td>
<td>.01660</td>
<td>.01850</td>
</tr>
<tr>
<td>$tr\left[A\left(\theta\right)B_R\left(\theta\right)\right]$</td>
<td>.04447</td>
<td>.03384</td>
<td>.01950</td>
<td>.01097</td>
<td>.00954</td>
</tr>
<tr>
<td>$E\left{tr\left[A\left(\hat{\theta}\right)B_R\left(\hat{\theta}\right)\right]\right}^{**}$</td>
<td>.03567</td>
<td>.02652</td>
<td>.01951</td>
<td>.01513</td>
<td>.00954</td>
</tr>
</tbody>
</table>

|       | (.00013) | (.00012) | (.00011) | (.00009) | (.00002) |

**Monte Carlo value. Standard error of simulation in parentheses.**

Figure 4-3 shows the results of the Monte Carlo study graphically. The accuracy of the estimator again largely depends on the value of $\nu$. For small values of $\nu$, which correspond to values of $\sigma_i^2$ that are large relative to $\sigma^2$, the estimator for the correction term is not accurate with expected values more than twice the function it is estimating. However, as $\nu$ approaches 1, the estimator for the correction term significantly underestimates the true correction term.
The expected value of Equation (4-18) can be expressed in closed form if we assume normal distributions. First define

\[ u_e = \frac{k(n-1)m_e}{\sigma^2}, \ u_a = \frac{(k-1)m_a}{\sigma^2 + n\sigma_a^2}, \ w = u_e + u_a, \ s = \frac{u_e}{u_e + u_a}. \]

It is easy to show that \( u_e \) and \( u_a \) are independent chi-square random variables with \( k(n-1) \) and \( (k-1) \) degrees of freedom, respectively (e.g., Searle 1971, p. 410). Also, \( w \) and \( s \) are independently distributed where \( w \) has a chi-square distribution with \( (kn-1) \) degrees of freedom, and \( s \) has a Beta\((\nu_e, \nu_a)\) distribution (e.g., Johnson and Kotz 1970, Sec. 24.2). Now note that if \( m_a > m_e \) then

\[ \frac{m_e}{m_a} < 1 \quad \text{so that} \quad \hat{\nu} = \min \left( \frac{m_e}{m_a}, 1 \right) = \frac{m_e}{m_a} \quad \text{but when} \quad m_a < m_e, \quad \frac{m_e}{m_a} > 1 \quad \text{so that} \quad \hat{\nu} = \min \left( \frac{m_e}{m_a}, 1 \right) = 1. \]

Defining \( \delta(m_a, m_e) = \begin{cases} 1 & \text{if } m_a > m_e \\ 0 & \text{if } m_a \leq m_e \end{cases} \) gives

\[
E \left[ \text{tr} \left( A(\hat{\theta}) I^{-1}(\hat{\theta}) \right) \right] = \frac{2(kn-1)}{kn(n-1)} E \left( \hat{\sigma}^2 \hat{\nu} \right)
\]

\[ = \frac{2(kn-1)}{kn(n-1)} \left[ E \left( \frac{m_e^2}{m_a} \delta(m_a, m_e) \right) + E \left( \frac{(k-1)m_a + k(n-1)m_e}{kn-1} (1 - \delta(m_a, m_e)) \right) \right]. \tag{4-19} \]

Noting that if \( m_a > m_e \) then \( s < \Psi \), so that \( \delta(s, \Psi) = \begin{cases} 1 & \text{if } s < \Psi \\ 0 & \text{if } s \geq \Psi \end{cases} \) and that \( \frac{m_e^2}{m_a} = \left( \frac{\nu \sigma^2 \nu_a}{k(n-1)\nu_e} \right) \frac{w s^2}{(1-s)} \), we see that
\[ (1) = E \left( \frac{m_e^2}{m_a} \delta(m_a, m_e) \right) = \frac{\nu \sigma^2 v_a}{k(n-1)v_e} E \left( \frac{ws^2}{1-s} \delta(s, \Psi) \right) \]
\[ = \frac{\nu \sigma^2 v_a}{k(n-1)v_e} E(w) E \left( \frac{s^2}{1-s} \delta(s, \Psi) \right) \]
\[ = \frac{\nu \sigma^2 v_a}{k(n-1)v_e} (kn-1) \int_0^\Psi \frac{s^2}{1-s} B(v_e, v_a) ds \]
\[ = \frac{\nu \sigma^2 v_a}{k(n-1)v_e} (kn-1) \left\{ \frac{B(v_e + 2, v_a - 1)}{B(v_e, v_a)} I_{\Psi}(v_e + 2, v_a - 1) \right\}. \tag{4-20} \]

Now note that \( u_e = sw \) so \( m_a = \frac{\sigma^2 sw}{k(n-1)} \) and \( m_a = \frac{(\sigma^2 + n\sigma_a^2)(1-s)w}{(k-1)} \) which leads to

\[ (2) = E \left( \frac{(k-1)m_a + k(n-1)m_e}{kn-1} (1 - \delta(m_a, m_e)) \right) = E \left( \frac{(k-1)m_a}{kn-1} (1 - \delta(m_a, m_e)) \right) \]
\[ + E \left( \frac{k(n-1)m_e}{kn-1} (1 - \delta(m_a, m_e)) \right) \]
\[ = \frac{\sigma^2 + n\sigma_a^2}{kn-1} E \left( w(1-s)(1 - \delta(s, \Psi)) + \frac{\sigma^2}{kn-1} E\left( ws(1 - \delta(s, \Psi)) \right) \right) \]
\[ = \left( \sigma^2 + n\sigma_a^2 \right) \int_0^1 (1-s) s^{v_e-1} (1-s)^{v_a-1} ds + \frac{\sigma^2}{\Psi} \int_0^1 s^{v_e-1} (1-s)^{v_a-1} ds \]
\[ = \left( \sigma^2 + n\sigma_a^2 \right) \frac{B(v_e, v_a + 1)}{B(v_e, v_a)} \left[ 1 - I_{\Psi}(v_e, v_a + 1) \right] + \frac{\sigma^2}{\Psi} \frac{B(v_e + 1, v_a)}{B(v_e, v_a)} \left[ 1 - I_{\Psi}(v_e + 1, v_a) \right] \]
\[ = \left( \sigma^2 + n\sigma_a^2 \right) \frac{v_a}{v_e + v_a} \left[ 1 - I_{\Psi}(v_e, v_a + 1) \right] + \frac{\sigma^2}{v_e + v_a} \frac{v_e}{v_e + v_a} \left[ 1 - I_{\Psi}(v_e + 1, v_a) \right]. \tag{4-21} \]

Substituting Equations (4-20) and (4-21) for (1) and (2) in Equation (4-19), respectively, we have

\[ E \left[ tr \left( A(\hat{\theta}) I(\hat{\theta}) \right) \right] = \frac{2(kn-1)}{k^2 n(n-1)} E(\hat{\sigma}^2 \hat{v}) \]
\[ = \frac{2(kn-1)}{k^2 n(n-1)} \left\{ \sigma^2 \nu (kn-1)v_a B(v_e + 2, v_a - 1) I_{\Psi}(v_e + 2, v_a - 1) \right\} \]
\[ + \sigma^2 \frac{v_e}{v_e + v_a} \left[ 1 - I_{\Psi}(v_e + 1, v_a) \right] + \sigma^2 \frac{v_a}{v_e + v_a} \left[ 1 - I_{\Psi}(v_e, v_a + 1) \right]. \tag{4-22} \]
Figure 4-4 shows the accuracy of the Kackar-Harville estimator (Equation (4-22)) for the Kackar-Harville approximation (Equation (4-15)) and the correction term (Equation (4-7)) holding \( k = 6 \) and \( n = 6 \). Just as for the approximation, the accuracy of \( E \left[ tr \left( A (\hat{\theta}) I^{-1} (\hat{\theta}) \right) \right] \) depends largely on the value of \( \nu \). While \( tr \left( A (\hat{\theta}) I^{-1} (\hat{\theta}) \right) \) is a fairly accurate estimator for \( tr \left[ A (\theta) I^{-1} (\theta) \right] \) over the entire parameter space, it may severely over-estimate \( C_R (\theta) \) for larger values of \( \nu \). We wish to assess the assertion of Prasad and Rao (1990) that

\[
E \left[ tr \left( A (\hat{\theta}) I^{-1} (\hat{\theta}) \right) \right] = tr \left[ A (\theta) I^{-1} (\theta) \right] + o \left( k^{-1} \right) \text{ for this model.}
\]

Noting that \( \lim_{k \to \infty} \Psi = 1 \) yields the result that \( \lim_{k \to \infty} I_{\Psi} = 1 \) so that

\[
E \left[ tr \left( A (\hat{\theta}) I^{-1} (\hat{\theta}) \right) \right] - tr \left[ A (\theta) I^{-1} (\theta) \right] = \frac{2 (kn-1)}{k^2 n (n-1)} \sigma^2 \nu \left\{ \frac{(kn-1) V_c B (v_e + 2, v_a - 1) I_{\Psi} (v_e + 2, v_a - 1)}{k(n-1) V_c B (v_e, v_a)} - 1 \right. \\
+ \sigma^2 \nu^{-1} \frac{V_c}{V_c + V_a} \left[ 1 - I_{\Psi} (v_e + 1, v_a) \right] + \sigma^2 \nu^{-2} \frac{V_a}{V_c + V_a} \left[ 1 - I_{\Psi} (v_e, v_a + 1) \right] \right\} = o \left( k^{-1} \right)
\]

confirming, for this example, the assertion of Prasad and Rao (1990).

These results also show the importance of studying the effect of substituting variance parameter estimates for the unknown values in the approximations for the correction term, the impact of which has been typically overlooked. In the case of \( tr \left[ A (\theta) B_R (\theta) \right] \) substituting variance parameter estimates improves the accuracy for \( C_R (\theta) \) (in the long run). However, in the case of \( tr \left[ A (\theta) I^{-1} (\theta) \right] \) the substitution creates more inaccuracy for \( C_R (\theta) \) for most values of \( \nu \).
Figure 4-5 compares the accuracy of Equations (4-17) and (4-18) for the correction term. The plot shows that $tr\left[A\left(\hat{\theta}\right)B_r\left(\hat{\theta}\right)\right]$ is more accurate (in the long-run) than $tr\left[A\left(\hat{\theta}\right)I^1\left(\hat{\theta}\right)\right]$ for $\upsilon > 0.30$; however, for $\upsilon < 0.30$, the reverse is true. This is the same phenomenon that we noticed with the approximations in Figure 4-2. Note that SAS® PROC MIXED by default uses $I^1\left(\theta\right)$ with REML estimates for $\hat{\theta}$ in the calculation of the MSEP estimator.

4.2 Kackar-Harville Approximation Based on ANOVA Estimation Methods

The best choice of method for estimating $\theta$ may depend on the model of interest, the level of balance in the design, and the goals of the analysis. Stroup and Littell (2002) demonstrate the effect of different estimation methods on the power and control of type I error for hypothesis tests on the fixed effects in analyzing an unbalanced, multi-location experiment. Their study indicates that REML and ML are not always the best choices when negative estimates of variance components are likely, and may lead to inflated Type I error rates and low power. The impact of the choice of variance component estimation method and negative variance component estimates on the Kackar-Harville approximation has not been investigated thoroughly. In Chapter 5 we elaborate on the impact of negative variance component estimates. To begin, however, we develop the correction term, several approximations, and the expected values of the estimators under ANOVA estimation for the balanced one-way random effects model.

The ANOVA estimators of $\theta = \left(\sigma^2, \upsilon\right)$ are $\hat{\theta} = \left(\hat{\sigma}^2, \hat{\upsilon}\right)$ where
$$\hat{\sigma}^2 = m_e = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} (y_{ij} - \bar{y}_i)^2}{k(n-1)},$$

$$\tilde{\nu} = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + n\hat{\sigma}_a^2} = \frac{m_e}{m_a},$$

$$\hat{\sigma}_a^2 = \frac{m_a - m_e}{n}, m_a = \frac{n \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2}{k-1}.$$ (4-23)

These estimates are obtained under the METHOD = TYPE 3 option in SAS® PROC MIXED.

The true correction term, $C_A(\theta)$, is derived under normality assumptions as follows:

$$C_A(\theta) = E \left[ (\bar{y}_i - \bar{y})^2 (\tilde{\nu} - \nu)^2 \right] = (kn)^{-1} E \left[ n \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2 (\tilde{\nu} - \nu)^2 \right]$$

$$= \frac{\sigma^2 + n\sigma_a^2}{kn} E \left[ (k-1)m_a \frac{1}{\sigma^2 + n\sigma_a^2} (\tilde{\nu} - \nu)^2 \right] = \frac{\sigma^2 + n\sigma_a^2}{kn} E \left[ u_a (\tilde{\nu} - \nu)^2 \right]$$

$$= \frac{\sigma^2 + n\sigma_a^2}{kn} E \left[ w(1-s) (\tilde{\nu} - \nu)^2 \right] = \frac{\sigma^2 + n\sigma_a^2}{kn} \nu^2 E \left[ w(1-s) \left( \frac{\tilde{\nu}}{\nu} - 1 \right)^2 \right]$$

$$= \frac{\sigma^2 + n\sigma_a^2}{kn} \nu^2 E \left[ w(1-s) \left( \frac{V_s}{V_e (1-s)} - 1 \right) \right] = \frac{\sigma^2 + n\sigma_a^2}{kn} \nu^2 E \left[ w(1-s) \left( \frac{V_s}{V_e (1-s)} - 1 \right)^2 \right]$$

$$= \frac{\sigma^2 + n\sigma_a^2}{kn} \nu^2 \left[ (1-s) \left( \frac{V_s}{V_e (1-s)} - 1 \right)^2 \right],$$

by independence of $w$ and $s$, and noting that $w \sim \chi^2_{kn-1}$. Recalling that $s \sim \text{Beta}(v_e, v_a)$, we find that

$$E \left[ (1-s) \left( \frac{V_s}{V_e (1-s)} - 1 \right)^2 \right] = E \left[ \frac{V_s^2}{V_e^2 (1-s)} - 2 \frac{V_s}{V_e} + (1-s) \right]$$

$$= \frac{V_e B(V_e + 2, V_a - 1)}{V_e B(V_e, V_a)} - 2 \left( \frac{V_a}{V_e + V_a} \right) + \frac{V_a}{V_e + V_a}$$

$$= \frac{V_e B(V_e + 2, V_a - 1)}{V_e B(V_e, V_a)} - \frac{V_a}{V_e + V_a}.$$

76
Thus the correction term is
\[
C_A(\theta) = \frac{\sigma^2 \nu (kn-1)}{kn} \left[ \frac{v_a^2 B(\nu_e, 2, \nu_a - 1) - \nu_a}{v_e B(\nu_e, \nu_a)} \right]
\]
\[
= \frac{\sigma^2 \nu (kn-1)}{kn} \left( \frac{\nu_a}{\nu_e (\nu_a - 1)} \right) \quad \text{(4-24)}
\]
\[
= \frac{2(kn-1)(k-1)}{k^2 n (n-1)(k-3)} \sigma^2 \nu.
\]

This gives the true MSEP for \( \hat{\tau} \),
\[
M_A(\theta) = M_1(\theta) + C_A(\theta)
\]
\[
= M_1(\theta) + \frac{2(kn-1)(k-1)}{k^2 n (n-1)(k-3)} \sigma^2 \nu. \quad \text{(4-25)}
\]

To evaluate the Kackar-Harville approximation for the correction term, we need to determine how to evaluate \( B(\theta) \). We consider three alternatives for \( B(\theta) \): the exact MSE matrix \( E[(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)'] \), the variance-covariance matrix of \( \tilde{\theta} \), and \( I^1(\theta) \). Recall from Equation (4-9) that due to the form of \( A(\theta) \) in this model we are only concerned with assessing the variability of \( \nu \) in order to evaluate the Kackar-Harville approximation to the correction term.

We begin with the exact MSE of \( \nu \), \( E[\nu - \nu]^2 \). Noting that \( \nu = \frac{m_c}{m_a} \sim F_{k(n-1),k-1} \) it is easily shown that \( E(\nu) = \left( \frac{k-1}{k-3} \right) \nu \), \( V(\nu) = \frac{2\nu^2 (k-1)^2 (kn-3)}{k(n-1)(k-3)^2 (k-5)} \), and
\[
E(\nu^2) = \frac{\nu^2 (k-1)^2 (k(n-1)+2)}{k(n-1)(k-3)(k-5)}. \quad \text{After algebraic simplification this leads to}
\]
\[ E[\tilde{\phi} - \nu]^2 = E(\tilde{\phi}^2) - 2\nu E(\tilde{\phi}) + \nu^2 \]
\[ = 2\nu^2 \left[ \frac{k(kn+3n-5)+1}{k(n-1)(k-3)(k-5)} \right]. \quad (4-26) \]

Letting \( B_A(\theta) = E\left[ (\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T \right] \) we have the following approximation for \( C_A(\theta) \):

\[ tr[ A(\theta) B_A(\theta) ] = \frac{2(k-1)(k(kn+3n-5)+1)}{k^2n(n-1)(k-3)(k-5)} \sigma^2 \nu. \quad (4-27) \]

Substituting the ANOVA estimates in Equation (4-23) for \( \theta \) yields this estimator for \( C_A(\theta) \):

\[ tr[ A(\tilde{\theta}) B_A(\tilde{\theta}) ] = \frac{2(k-1)(k(kn+3n-5)+1)}{k^2n(n-1)(k-3)(k-5)} \sigma^2 \tilde{\nu}. \quad (4-28) \]

To assess the accuracy of Equation (4-28) for estimating Equation (4-27) and \( C_A(\theta) \), we evaluate the expected value. Recalling the relationship \( \frac{m_e^2}{m_a} = \left( \frac{\nu\sigma^2v_a}{k(n-1)v_e} \right) \frac{w s^2}{(1-s)} \) where \( w \) and \( s \) are independently distributed with \( w \sim \chi^2_{kn-1} \) and \( s \sim Beta(v_e, v_a) \), we have

\[ E[\tilde{\sigma}^2\tilde{\nu}] = E\left[ \frac{m_e^2}{m_a} \right] = \left( \frac{\sigma^2\nu v_a}{k(n-1)v_e} \right) E[w] E\left[ \frac{s^2}{(1-s)} \right] \]
\[ = \left( \frac{\sigma^2\nu v_a}{k(n-1)v_e} \right) (kn-1) \left( \frac{B(v_e+2, v_a-1)}{B(v_e, v_a)} \right) \]
\[ = \cdots = \frac{(k-1)(k(n-1)+2)}{k(n-1)(k-3)} \sigma^2 \nu. \quad (4-29) \]

Thus,

\[ E\left[ tr[ A(\tilde{\theta}) B_A(\tilde{\theta}) ] \right] = \frac{2(k-1)^2(k(kn+3n-5)+1)(k(n-1)+2)}{k^3 n (n-1)^2 (k-3)^2 (k-5)} \sigma^2 \nu. \quad (4-30) \]
Again, we confirm the asymptotic assertion by Prasad and Rao (1990), noting that

$$E\left\{tr\left[A(\tilde{\theta})B_A(\tilde{\theta})\right]\right\} - tr\left[A(\theta)B_A(\theta)\right] = \frac{2(k-1)^2(k(kn+3n-5)+1)(k(n-1)+2)}{k^3n(n-1)^2(k-3)^2(k-5)} \sigma^2 \nu - \frac{2(k-1)(k(kn+3n-5)+1)}{k^2n(n-1)(k-3)(k-5)} \sigma^2 \nu$$

$$= \frac{2(k-1)(k(kn+3n-5)+1)}{k^2n(n-1)(k-3)(k-5)} \sigma^2 \nu \left\{ \frac{(k-1)(k(n-1)+2)}{k(n-1)(k-3)} - 1 \right\}$$

$$= o\left(k^{-1}\right).$$

The second alternative for $B(\theta)$ is the exact variance-covariance matrix of $\tilde{\theta}$, $B_A^*(\theta)$.

Because $\tilde{\nu}$ is not unbiased for $\nu$, this is not the same as the MSE matrix for $\tilde{\theta}$. As noted above,

$$V(\tilde{\nu}) = \frac{2\nu^2(k-1)^2(kn-3)}{k(n-1)(k-3)^2(k-5)},$$

which leads to the following approximation for $C_A(\theta)$:

$$tr\left[A(\theta)B_A^*(\theta)\right] = \frac{2(k-1)^3(kn-3)}{k^2n(n-1)(k-3)^2(k-5)} \sigma^2 \nu.$$  (4-31)

Again substituting the ANOVA estimates for $\theta$ in Equation (4-31) yields an estimator for $C_A(\theta)$:

$$tr\left[A(\tilde{\theta})B_A^*(\tilde{\theta})\right] = \frac{2(k-1)^3(kn-3)}{k^2n(n-1)(k-3)^2(k-5)} \tilde{\sigma}^2 \tilde{\nu}.$$  (4-32)

Using Equation (4-29) we have

$$E\left\{tr\left[A(\tilde{\theta})B_A^*(\tilde{\theta})\right]\right\} = \frac{2(k-1)^4(kn-3)(k(n-1)+2)}{k^3n(n-1)^2(k-3)^3(k-5)} \sigma^2 \nu.$$  (4-33)

Again we find that
\[
E\left\{tr\left[ A(\hat{\theta})B'_d(\hat{\theta})\right]\right\} - tr\left[ A(\theta)B'_d(\theta)\right] \\
= \frac{2(k-1)^3(kn-3)(k(n-1)+2)}{k^3n(n-1)^2(k-3)^3(k-5)}\sigma^2\nu - \frac{2(k-1)^3(kn-3)}{k^2n(n-1)(k-3)^2(k-5)}\sigma^2\nu \\
= \frac{2(k-1)^3(kn-3)}{k^2n(n-1)(k-3)^2(k-5)}\sigma^2\nu \left\{(k-1)(k(n-1)+2)\right\} \frac{1}{k(n-1)(k-3)} - 1 \\
= o(k^{-1}).
\]

The third alternative we consider for \( B(\theta) \) is to use the inverse of the observed information matrix, \( I^{-1}(\theta) \), as an approximation. Note that \( I^{-1}(\theta) \) is the same as the asymptotic variance of the REML estimators of \( \theta \), in Equation (4-16). Thus the Kackar-Harville approximation with ANOVA variance component estimates using \( I^{-1}(\theta) \) is the same as in the REML case in Equation (4-15), i.e.,

\[
tr\left[ A(\theta)I^1(\theta)\right] = \frac{2(kn-1)}{k^2n(n-1)}\sigma^2\nu, \quad (4-34)
\]

and the estimator formed by substituting ANOVA estimates for \( \theta \) is

\[
tr\left[ A(\hat{\theta})I^1(\theta)\right] = \frac{2(kn-1)}{k^2n(n-1)}\tilde{\sigma}^2\nu. \quad (4-35)
\]

Using Equation (4-29) we have

\[
E\left\{tr\left[ A(\hat{\theta})I^1(\hat{\theta})\right]\right\} = \frac{2(kn-1)}{k^2n(n-1)}E\left[ \tilde{\sigma}^2\nu \right] \\
= \frac{2(kn-1)(k-1)(k(n-1)+2)}{k^3n(n-1)^2(k-3)}\sigma^2\nu \quad (4-36)
\]

which again yields
\[ E\left[ tr\left[ A(\hat{\theta})I^{-1}(\hat{\theta})\right]\right] - tr\left[ A(\theta)I^{-1}(\theta)\right] \\
= \frac{2(kn-1)(k-1)(k(n-1)+2)}{k^2n(n-1)^2(k-3)}\sigma^2\nu - \frac{2(kn-1)}{k^2n(n-1)}\sigma^2\nu \\
= \frac{2(kn-1)}{k^2n(n-1)}\sigma^2\nu \left\{ \frac{(k-1)(k(n-1)+2)}{k(n-1)(k-3)} - 1 \right\} \\
= o\left(k^{-1}\right). \]

Figure 4-6 compares the three Kackar-Harville approximations in Equations (4-27), (4-31) and (4-34) to \( C_\theta(\theta) \) for \( k = 6 \) and \( n = 6 \). Clearly \( tr\left[ A(\theta)I^{-1}(\theta)\right] \) more closely approximates \( C_\theta(\theta) \) than the other two approximations, for all values of \( \nu \). Thus, although we would expect \( I^{-1}(\theta) \) to be less accurate than \( B_\theta(\theta) \) or \( B_\theta^*(\theta) \) in evaluating the MSE of \( \hat{\theta} \), especially for small values of \( k \), the approximation of \( C_\theta(\theta) \) utilizing \( I^{-1}(\theta) \) is more accurate than the either of the other choices across the entire parameter space. In fact, the Kackar-Harville approximation utilizing the exact MSE of \( \hat{\theta} \) is the least accurate of all three approximations. Note that this is in contrast to the results from REML variance component estimates in Figure 4-2, where the best approximation depended on the value of \( \nu \).

Figure 4-7 compares \( E\left\{ tr\left[ A(\hat{\theta})B_\theta(\hat{\theta})\right]\right\} \) (Equation (4-30)) to \( tr\left[ A(\theta)B_\theta(\theta)\right] \) and the correction term, \( C_\theta(\theta) \). The bias for the correction term introduced by replacing the unknown variance components with ANOVA estimates in the correction term approximation is significant and increases as \( \nu \) moves toward 1. Figure 4-8 demonstrates a similar result for the Kackar-Harville approximation and estimator utilizing the variance of \( \theta \). \( E\left\{ tr\left[ A(\hat{\theta})B_\theta^*(\hat{\theta})\right]\right\} \) (Equation (4-33)) is again inflated over the approximation and they both overestimate the true
correction term increasingly in \( \nu \). Figure 4-9 shows a much different result than the previous two figures since the accuracy of the estimator utilizing \( I^1(\hat{\theta}) \) (Equation (4-36)) is improved over that of the approximation (Equation (4-34)). While \( tr[A(\theta)I^1(\theta)] \) underestimates \( C_A(\theta) \) by up to 50\%, \( E\left[ tr[A(\hat{\theta})I^1(\hat{\theta})] \right] \) is much more accurate and consistently overestimates \( C_A(\theta) \) by only 6.67\% across the entire parameter space.

Figure 4-10 compares all three estimators (Equations (4-30), (4-33), and (4-36)) relative to \( C_A(\theta) \). Clearly the Kackar-Harville estimator utilizing \( I^1(\hat{\theta}) \) is far more accurate than either of the other two choices. This is counter-intuitive: the Kackar-Harville estimator using an approximation of the MSE of \( \hat{\theta} \) outperforms the other estimators using exact measures of the variation of \( \hat{\theta} \). Note again that this is a different result than seen with REML variance component estimates where the results were dependent on the value of \( \nu \). There, we saw that the Kackar-Harville estimator utilizing the exact MSE of \( \hat{\theta} \), \( B_R(\theta) \), often performed better than the one using an approximation of the MSE of \( \hat{\theta} \), \( I^1(\theta) \) (see Figure 4-5). The choice of which Kackar-Harville estimator for the correction term is best, therefore, depends, in part, on what method is used for variance component estimation.

The final step in this assessment is to compare the ANOVA results to the REML results. Figure 4-11 compares the accuracy of \( tr[A(\hat{\theta})I^1(\hat{\theta})] \) (Equation (4-36)) for \( C_A(\theta) \) with the accuracy of the two REML estimators, \( tr[A(\hat{\theta})B_R(\hat{\theta})] \) (Monte Carlo expected value) and \( tr[A(\hat{\theta})I^1(\hat{\theta})] \) (Equation (4-19)) for \( C_R(\theta) \), with \( k = 6 \) and \( n = 6 \). The ANOVA estimator
utilizing the information matrix, \( tr \left[ A(\hat{\theta})I^{-1}(\hat{\theta}) \right] \), appears to perform better than either REML based estimator, although the added variation to the MSEP of \( \hat{\tau} \) is larger for the EBLUP with ANOVA variance components, i.e., \( C_A(\theta) > C_R(\theta) \). We can assess the accuracy relative to the appropriate correction term to better compare the performance of the estimators. The relative bias of the two REML based estimators of the correction term and the ANOVA based estimator with \( I^{-1}(\hat{\theta}) \) is tabulated in Table 4-2 and graphically depicted in Figure 4-12. Relative bias was calculated as

\[
\frac{E\left\{ tr \left[ \hat{A}\hat{B} \right] \right\} - C(\theta)}{C(\theta)}.
\]

The ANOVA estimator of the correction term has consistently lower relative bias than either of the REML estimators. Thus, although the variation added to the MSEP by using ANOVA variance component estimates is larger than when using REML variance component estimates, a more consistently accurate estimate of the added variation is available in the ANOVA case. While the accuracy of both REML estimators depends largely on the value of \( \nu \), the ANOVA estimator is significantly more accurate across all values of \( \nu \).

This special case has demonstrated the importance of the choice of variance component estimation method, the choice of measure of variation of \( \hat{\theta} \), and the added variability of substituting parameter estimates for unknown variance components in the correction term estimates. While ANOVA methods add more variability to the EBLUP than REML estimation, the estimator for the MSEP of the EBLUP is more accurate with ANOVA methods than with REML methods. The best choice of how to measure the variability of \( \hat{\theta} \) was not consistent across variance component estimation methods. For ANOVA methods, a better choice for \( B(\theta) \)
was $I^t(\theta)$, while for REML estimation the better choice depended on the value of $\nu$. We also discovered that the effect of substituting parameter estimates depends, in part, on the choice for $B(\theta)$ and the variance component estimation method. The impact of this substitution is often considerable and could cause the MSEP estimate to be significantly inflated.

Table 4-2. Relative bias (%) of estimators of correction term

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>REML**</td>
<td>132.0</td>
<td>40.3</td>
<td>11.0</td>
<td>-8.8</td>
<td>-48.5</td>
</tr>
<tr>
<td>$E\left{tr\left[ A(\hat{\theta})B_x(\hat{\theta}) \right]\right}$</td>
<td>33.7</td>
<td>85.0</td>
<td>152.3</td>
<td>202.8</td>
<td>220.8</td>
</tr>
<tr>
<td>ANOVA</td>
<td>6.7</td>
<td>6.7</td>
<td>6.7</td>
<td>6.7</td>
<td>6.7</td>
</tr>
</tbody>
</table>

**Monte Carlo values used to compute relative bias.

4.3 Bias Correction Term for the Balanced One-Way Random Effects Model

4.3.1 Bias Correction Term Approximation under $\theta$ Parameterization

We now address the bias correction term approximation for the one-way random effects model. We will consider the case where $B(\theta) = I^t(\theta)$ so that the calculations for the approximation will be the same for the REML and ANOVA, and so that closed-form expressions are attainable. The difference comes in the expected values of the estimators of the bias correction term due to the truncation of the REML estimates. We will show that in either case the bias correction estimator can be negative, thus proving by counter-example that the bias
correction estimator is indefinite. We begin however, with demonstrating the lack of transform invariance.

Recall that the parameterization we are working with is \( \theta = (\sigma^2, \nu) \). If we were considering the parameterization \( \Phi = (\sigma^2, \sigma_z^2) \), wherein the variance matrix, \( V \), is linear in the parameters, the bias correction term would be identical to the correction term approximation, which is transform invariant, i.e., \( -\frac{1}{2}tr(A(\Phi)B(\Phi)) = tr(A(\Phi)B(\Phi)) = tr(A(\theta)B(\theta)) \), and non-negativity would not be an issue since \( tr(A(\theta)B(\theta)) \) is non-negative definite in its most general definition as proven in Chapter 3. This is not the case with \( \theta \) as the parameterization, as will be demonstrated in this chapter, since \( -\frac{1}{2}tr(A(\theta)B(\theta)) \neq tr(A(\theta)B(\theta)) \).

We now give a simple demonstration that the bias correction term approximation is not transform invariant, as discussed in Section 3.3, through the continuing example of the balanced one-way random effects model. The true bias of the naïve MSEP, \( E\left[M_1(\hat{\theta})\right] - M_1(\theta) \), is invariant to transformation as long as \( \tau(\hat{\theta}) = \tau(\hat{\Phi}) \). This follows from the definition of \( M_1(\theta) = E\left[\tau(\theta) - t\right]^2 \). Thus the lack of transform invariance of the bias estimate is a concern.

To calculate the bias correction term approximation under the \( \theta \) parameterization, we first compute the second derivatives of the naïve MSEP, \( M_1(\theta) \), with respect to the variance parameters:

\[
\frac{\partial M_1(\theta)}{\partial \sigma^2} = \frac{1 - 2\nu + k\nu - k\nu^2 + \nu^2}{nk\nu} \\
\Rightarrow \frac{\partial^2 M_1(\theta)}{\left(\partial \sigma^2\right)^2} = 0,
\]

85
and

$$\frac{\partial M_1(\theta)}{\partial \nu} = \frac{\sigma^2 (k - 2 + 2\nu(1-k))}{nk\nu} - \frac{\sigma^2 (1-2\nu+k\nu-k\nu^2+\nu^2)}{nk\nu^2}$$

$$= \frac{\sigma^2 (\nu^2(1-k)-1)}{nk\nu^2}$$

$$\Rightarrow \frac{\partial^2 M_1(\theta)}{(\partial \nu)^2} = \frac{2\sigma^2}{nk\nu^3},$$

and

$$\frac{\partial^2 M_1(\theta)}{\partial \sigma^2 \partial \nu} = \frac{(1-k)\nu^2-1}{nk\nu^2}.$$

So,

$$\Lambda(\theta) = \begin{bmatrix} 0 & \frac{(1-k)\nu^2-1}{nk\nu^2} \\ \frac{(1-k)\nu^2-1}{nk\nu^2} & \frac{2\sigma^2}{nk\nu^3} \end{bmatrix}. \quad (4-37)$$

Utilizing $I^1(\theta)$ from Equation (4-16), we have

$$\Lambda(\theta)I^1(\theta) = \begin{bmatrix} \frac{2\sigma^2 \left\{(1-k)\nu^2-1\right\}}{nk^2(n-1)\nu} & f(\sigma^2, \nu) \\ f(\sigma^2, \nu) & \frac{2\sigma^2 \left\{(1-k)\nu^2-1\right\}}{nk^2(n-1)\nu} + \frac{4\sigma^2 (kn-1)}{nk^2(n-1)(k-1)\nu} \end{bmatrix} \quad (4-38)$$

where $f(\sigma^2, \nu)$ is a function of $\theta = (\sigma^2, \nu)$ that is inconsequential in this calculation.

This gives

$$tr[\Lambda(\theta)I^1(\theta)] = 4\sigma^2\left[\frac{(1-k)\nu^2-1}{nk^2(n-1)\nu} + \frac{kn-1}{nk^2(n-1)(k-1)\nu}\right]. \quad (4-39)$$

Thus the bias correction term for the naïve estimator of the MSEP under the $\theta = (\sigma^2, \nu)$ parameterization is
\[
\text{BIAS}^k_R(\theta) = -\frac{1}{2} tr \left[ \Lambda(\theta) \Gamma^1(\theta) \right] = 2\sigma^2 \left[ \frac{(k-1)\nu^2 + 1}{nk^2(n-1)\nu} - \frac{kn-1}{nk^2(n-1)(k-1)\nu} \right] \tag{4-40} \]

which is obviously different from the bias correction term approximation used under the \( \Phi \) parameterization, which would be identical to Equation (4-15). Clearly the lack of transform invariance is attributable to the higher order derivatives in \( \Lambda(\theta) \). We will examine the impact of the parameterization by comparing each bias correction estimate to the true bias.

### 4.3.2 Expected Value of BIAS\(^k_R(\theta)\) with REML Covariance Parameter Estimation

We begin by calculating the expected value of the bias correction term estimator once the REML estimates are substituted for the unknown variance components.

Recall from Equations (4-19), (4-20), and (4-21) that

\[
E\left( \hat{\sigma}^2 \hat{\nu} \right) = \left\{ \frac{\sigma^2 \nu(kn-1)\nu_a B(\nu_c + 2, \nu_a - 1)I_\Psi(\nu_c + 2, \nu_a - 1)}{k(n-1)\nu B(\nu_c, \nu_a)} \right. \\
+ \left. \sigma^2 \frac{\nu_c}{\nu_c + \nu_a} \left[ 1 - I_\Psi(\nu_c + 1, \nu_a) \right] + \sigma^2 \nu^{-1} \frac{\nu_a}{\nu_c + \nu_a} \left[ 1 - I_\Psi(\nu_c, \nu_a + 1) \right] \right\} \tag{4-41} 
\]

We also need

\[
E\left( \hat{\sigma}^2 \right) = E\left[ m_{\delta}(m_a, m_c) \right] + E\left[ \frac{(k-1)m_a + k(n-1)m_c}{kn-1} \left( 1 - \delta(m_a, m_c) \right) \right] \tag{4-42} 
\]

where

\[
(1) = \frac{\sigma^2 + n\sigma^2}{k-1} E\left( w(1-s) \delta(s, \Psi) \right) \\
= \sigma^2 \nu^{-1} \frac{kn-1}{k-1} \int_0^w \frac{s^{\nu-1}(1-s)^{\nu-1}}{B(\nu_c, \nu_a)} ds \\
= \sigma^2 \nu^{-1} I_\Psi(\nu_c, \nu_a + 1)
\]

and from Equation (4-21),
\begin{equation}
(2) = (\sigma^2 + n\sigma^2_a) \frac{v_a}{v_e + v_a} [1 - I(\psi)(v_e + 1)] + \sigma^2 \frac{v_e}{v_e + v_a} [1 - I(\psi)(v_e + 1,v_a)].
\end{equation}

We can utilize Equations (4-41) and (4-42) to compute the expected value of \( \text{BIAS}^{KR}(\hat{\theta}) \) as follows:

\begin{align*}
E \left[ \text{BIAS}^{KR}(\hat{\theta}) \right] &= E \left\{ -\frac{1}{2} \text{tr} \left[ \Lambda(\hat{\theta})\Lambda^{-1}(\hat{\theta}) \right] \right\} \\
&= 2E \left\{ \sigma^2 \left[ \frac{(k-1)\hat{\theta}^2 + 1}{nk^2(n-1)\hat{\theta}} - \frac{kn-1}{nk^2(n-1)(k-1)\hat{\theta}} \right] \right\} \\
&= 2 \left( \frac{k-1}{nk^2(n-1)} \right) E(\hat{\sigma}^2\hat{\theta} - 2)E(\hat{\sigma}^2\hat{\theta}^{-1}) \\
&= 2 \left\{ \left( \frac{k-1}{nk^2(n-1)} \right) \left( \sigma^2 \nu(n-1)B(v_e + 2,v_a - 1)I(\psi)(v_e + 2,v_a - 1) \right) \right. \\
& \quad \quad + \sigma^2 \frac{v_e}{v_e + v_a} [1 - I(\psi)(v_e + 1,v_a)] + \sigma^2 \nu^{-1} \frac{v_e}{v_e + v_a} [1 - I(\psi)(v_e + 1,v_a)] \\
& \quad \quad + \left( \frac{1}{nk(k-1)} \right) \left( \sigma^2 \nu^{-1}I(\psi)(v_e,v_a + 1) + (\sigma^2 + n\sigma^2_a) \frac{v_e}{v_e + v_a} [1 - I(\psi)(v_e,v_a + 1)] \right) \\
& \quad \quad + \sigma^2 \frac{v_e}{v_e + v_a} [1 - I(\psi)(v_e + 1,v_a)] \right\}.
\end{align*}

\begin{equation}
(4-43)
\end{equation}

To evaluate the accuracy of the bias correction term estimator, the true bias correction term (which is the negative true bias) of \( M_1(\hat{\theta}) \) for \( M_1(\theta) \) can also be calculated as \( M_1(\theta) - E \left[ M_1(\hat{\theta}) \right] \), for REML variance component estimates of \( \theta \). Because \( M_1(\hat{\theta}) = 0 \) when \( m_a < m_e \), the expected value of \( M_1(\hat{\theta}) \) is
\[
E\left[M_1(\hat{\theta})\right] = E\left\{\frac{\hat{\sigma}^2(1-\hat{\phi})}{nk\hat{\phi}}\left[1 + (k-1)\hat{\phi}\right]\right\} \\
= E\left\{\frac{\hat{\sigma}^2(1-\hat{\phi})}{nk\hat{\phi}}\left[1 + (k-1)\hat{\phi}\right]\delta(m_a, m_e)\right\} \\
= E\left\{\left(\frac{m_a}{nk} + \frac{(k-2)m_e}{nk} - \frac{(k-1)m_e^2}{nkm_a}\right)\delta(m_a, m_e)\right\}.
\]

(4-44)

We establish that

\[
E\left[\frac{m_a}{nk}\delta(m_a, m_e)\right] = \frac{\sigma^2}{nk}I_q(v_e, v_a + 1),
\]

(4-45)

\[
E\left[\frac{(k-2)m_e}{nk}\delta(m_a, m_e)\right] = \frac{(k-2)\sigma^2}{nk^2(n-1)}E\left[(sw)\delta(s, \Psi)\right] \\
= \frac{(k-2)(kn-1)\sigma^2}{nk^2(n-1)}E\left[s\delta(s, \Psi)\right] \\
= \frac{(k-2)(kn-1)\sigma^2}{nk^2(n-1)}B(v_e + 1, v_a)I_q(v_e + 1, v_a) \\
= \frac{(k-2)\sigma^2}{nk}I_q(v_e + 1, v_a),
\]

(4-46)

and

\[
E\left[\frac{(k-1)m_e^2}{nkm_a}\delta(m_a, m_e)\right] = \left\{\frac{k-1}{nk}\left(\frac{\nu\sigma^2v_a}{k(n-1)v_e}\right)(kn-1)\right\} \\
\quad \times \left\{\frac{B(v_e + 2, v_a - 1)}{B(v_e, v_a)}I_q(v_e + 2, v_a - 1)\right\} \\
= \frac{(k-1)^2(kn-k+2)\sigma^2\nu}{nk^2(n-1)(k-3)}I_q(v_e + 2, v_a - 1).
\]

(4-47)

Combining Equations (4-45), (4-46), and (4-47) into Equation (4-44) gives us the expected value for \(M_1(\hat{\theta})\) with REML variance component estimates:
\[ E\left[M_1(\hat{\theta})\right] = \frac{\sigma^2}{nk} I_{\nu}(\nu,\nu_a+1) + \frac{(k-2)\sigma^2}{nk} I_{\nu}(\nu_c+1,\nu_a) \]
\[- \frac{(k-1)^2}{nk^2} \frac{(kn-k+2)\nu}{(n-1)(k-3)} I_{\nu}(\nu_c+2,\nu_a-1).\] (4-48)

Thus the true bias correction term of \( M_1(\hat{\theta}) \) for \( M_1(\theta) \) is

\[
M_1(\theta) - E\left[M_1(\hat{\theta})\right] = \frac{\sigma^2}{nk} \left\{ \frac{(1-\nu)}{\nu} \left[ 1 + (k-1)\nu \right] \right\} \\
+ \frac{(k-1)^2}{k(n-1)(k-3)} \frac{(kn-k+2)\nu}{I_{\nu}(\nu_c+2,\nu_a-1)} \]
\[- \nu^{-1} I_{\nu}(\nu_c,\nu_a+1) - (k-2) I_{\nu}(\nu_c+1,\nu_a-1).\] (4-49)

We can now make comparisons among the proposed bias correction term estimators of the naïve MSEP, \( M_1(\theta) \). Figure 4-13 pictorially demonstrates the bias of \( M_1(\hat{\theta}) \) for \( M_1(\theta) \) under REML methods. From this figure it is easy to see that \( M_1(\hat{\theta}) \) has both positive and negative bias depending on the value of \( \nu \). Figure 4-14 depicts the true bias correction term, \( M_1(\theta) - E\left[M_1(\hat{\theta})\right] \), from Equation (4-49), along with the bias correction term approximation,

\[
- \frac{1}{2} \text{tr}\left[ \Lambda(\theta) \Gamma^{-1}(\theta) \right] \] (Equation (4-40)) and its expected value under REML estimation (Equation (4-43)). First note that the effect of substituting variance parameter estimates in the bias correction approximation does not have much of an impact, unlike the results of the correction term approximation. Although \( - \frac{1}{2} \text{tr}\left[ \Lambda(\theta) \Gamma^{-1}(\theta) \right] \) and \( E\left\{ - \frac{1}{2} \text{tr}\left[ \Lambda(\hat{\theta}) \Gamma^{-1}(\hat{\theta}) \right] \right\} \) are nearly equal over the entire span of \( \nu \), both significantly underestimate the true bias correction term over most of the parameter space. In fact, the true bias correction term is positive for \( \nu < 0.75 \) reflecting that \( M_1(\hat{\theta}) \) is underestimating \( M_1(\theta) \). However, the bias approximation and
expected value are negative. Thus by using this bias correction term estimator, we would actually exacerbate the problem and significantly increase the amount by which \( M_1(\hat{\theta}) \) is already underestimating \( M_1(\theta) \). That is, while \( M_1(\hat{\theta}) \) tends to underestimate \( M_1(\theta) \) for \( \nu < 0.75 \), the “correction” by the bias correction term actually causes further underestimation of \( M_1(\theta) \), especially for small values of \( \nu \). The use of this form of the bias correction term would actually produce a poorer estimate for \( M_1(\theta) \) than using \( M_1(\hat{\theta}) \) alone. As \( \nu \) increases, we see that the true bias correction term becomes negative, reflecting that \( M_1(\hat{\theta}) \) is overestimating \( M_1(\theta) \); thus it is appropriate for the bias correction term estimator to be negative. As shown in Figure 4-14, the bias correction term approximation performs significantly better in this upper range of the parameter space than for \( \nu < .75 \).

Now we compare this result to the bias adjustment under the linear parameterization, \( \Phi = (\sigma^2, \sigma^2_\omega) \), recalling that

\[
-\frac{1}{2} tr \left( A(\Phi) I^{-1}(\Phi) \right) = tr \left( A(\Phi) I^{-1}(\Phi) \right) = tr \left( A(\theta) I^{-1}(\theta) \right),
\]

where \( tr \left( A(\theta) I^{-1}(\theta) \right) \) is in Equation (4-15).

Figure 4-15 shows \( M_1(\theta) - E \left[ M_1(\hat{\theta}) \right] \) along with \( tr \left[ A(\theta) I^{-1}(\theta) \right] \) (Equation (4-15)) and \( E \left[ tr \left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right] \right] \) (Equation (4-22)). We see a much different result for this bias correction term than we did for \( -\frac{1}{2} tr \left[ A(\theta) I^{-1}(\theta) \right] \). Now we see the bias correction term approximation and estimator significantly overestimating the true bias correction term, especially for large values of \( \nu \). For small values of \( \nu \), the estimator is fairly accurate. This is the exact opposite of
the previous result in Figure 4-14. Figure 4-16 compares the accuracy of \(-\frac{1}{2} tr\left[ \Lambda(\hat{\theta}) I^{-1}(\hat{\theta}) \right]\) and \(tr\left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right]\) for \(M_1(\theta) - E\left[ M_1(\tilde{\theta}) \right]\). For \(\nu < .75\), \(M_1(\theta) - E\left[ M_1(\tilde{\theta}) \right]\) is positive. In this range, \(tr\left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right]\) is a significantly more accurate estimator because

\[
E\left\{ -\frac{1}{2} tr\left[ \Lambda(\hat{\theta}) I^{-1}(\hat{\theta}) \right]\right\} \text{ is negative. For } \nu > .75, M_1(\theta) - E\left[ M_1(\tilde{\theta}) \right]\text{ is negative, and}
\]

\[
-\frac{1}{2} tr\left[ \Lambda(\hat{\theta}) I^{-1}(\hat{\theta}) \right]\text{ is significantly more accurate than } tr\left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right]. \text{ Thus the choice of which estimator is better depends on the value of } \nu \text{ and also on whether it is better to over- or under-estimate } M_1(\theta). \text{ Note that SAS® uses the } \Phi \text{ parameterization in the DDFM = KR option of PROC MIXED. Hence the results from Figure 4-15 are applicable to SAS® procedures.}

4.3.3 Expected Value of \(\text{BIAS}^{KR}(\theta)\) with ANOVA Covariance Parameter Estimation

We can do this same comparison for ANOVA variance parameter estimates. Recall that the bias correction approximation in Equation (4-40) is the same under ANOVA and REML variance parameters due to the choice of \(I^{-1}(\theta)\). However, because the ANOVA variance component estimates are not truncated, the expected values of the bias correction term estimator, \(E\left\{ -\frac{1}{2} tr\left[ \Lambda(\hat{\theta}) I^{-1}(\hat{\theta}) \right]\right\}\), and the naïve MSEP, \(E\left[ M_1(\tilde{\theta}) \right]\) will differ from the REML case.

First, recall from Equation (4-29) that

\[
E\left[ \tilde{\sigma}^2 \tilde{\nu} \right] = \frac{(k-1)(k(n-1)+2)}{k(n-1)(k-3)} \sigma^2 \nu \quad (4-50)
\]

and note that

\[
E\left[ \tilde{\sigma}^2 \nu^{-1} \right] = E(m_\nu) = E\left( \frac{\nu^{-1} u_x}{k-1} \right) = \sigma^2 \nu^{-1} \quad (4-51)
\]
Using Equations (4-50) and (4-51), the expected value for the bias correction term estimator is

\[
E \left[ \text{BIAS}^{KR} (\hat{\theta}) \right] = E \left\{ -\frac{1}{2} \text{tr} \left[ \Lambda (\hat{\theta}) I^1 (\hat{\theta}) \right] \right\} \\
= 2E \left\{ \hat{\sigma}^2 \left[ \frac{(k-1) \hat{\nu}^2 + 1}{nk^2 (n-1) \hat{\nu}} - \frac{kn-1}{nk^2 (n-1) (k-1) \hat{\nu}} \right] \right\} \\
= 2 \left\{ \frac{k-1}{nk^2 (n-1)} \right\} E \left( \hat{\sigma}^2 \hat{\nu} \right) - \frac{2}{nk} E \left( \hat{\sigma}^2 \hat{\nu}^{-1} \right) \\
= 2 \left\{ \frac{k-1}{nk^2 (n-1)} \right\} \frac{(k-1)(k(n-1)+2)}{k(n-1)(k-3)} \sigma^2 \nu \\
- \frac{2}{nk} \frac{\sigma^2 \nu^{-1}}{nk^2 (k-1)} \right\}. \quad (4-52)
\]

Note that by allowing the variance parameter estimates to take values outside of the parameter space, \( M_1 (\hat{\theta}) \) does not maintain the non-negative definite property. That is, in taking the expected value of \( M_1 (\hat{\theta}) \) with ANOVA variance parameter estimates, we are including negative values of \( M_1 (\hat{\theta}) \). Allowing variance parameter estimates outside of the parameter space in expected value calculations does not affect the definiteness properties of the bias correction term, \(-\frac{1}{2} \text{tr} \left[ \Lambda (\hat{\theta}) I^1 (\hat{\theta}) \right] \), since it is indefinite as will be discussed in Section 4.3.4. The calculation for the expected value of \( M_1 (\hat{\theta}) \) proceeds as follows:

\[
E \left[ M_1 (\hat{\theta}) \right] = E \left\{ \frac{\hat{\sigma}^2 (1-\hat{\nu})}{nk \hat{\nu}} \left[ 1 + (k-1) \hat{\nu} \right] \right\} \\
= E \left\{ \frac{m_a}{nk} + \frac{(k-2)m_e}{nk} - \frac{(k-1)m_e^2}{nk m_a} \right\} \\
= \frac{\sigma^2 \nu^{-1}}{nk} + \frac{(k-2)\sigma^2}{nk} - \frac{(k-1)^2 (kn-k+2) \sigma^2 \nu}{nk^2 (n-1)(k-3)}. \quad (4-53)
\]
Thus the true bias correction term (negative bias) of $M_i(\hat{\theta})$ for $M_i(\theta)$ is

$$M_i(\theta) - E\left[M_i(\hat{\theta})\right] = \frac{\sigma^2(1-\nu)}{nk\nu}\left[1+(k-1)\nu\right]$$

$$- \frac{\sigma^2\nu^{-1}}{nk} - \frac{(k-2)\sigma^2}{nk} + \frac{(k-1)^2(kn-k+2)\sigma^2\nu}{nk^2(n-1)(k-3)}$$

$$= \frac{2(k-1)(kn-1)\sigma^2\nu}{nk^2(n-1)(k-3)}.$$  \hspace{1cm} (4-54)

We can now compare this to the bias correction approximation, $-\frac{1}{2}tr\left[\Lambda(\theta)I^i(\theta)\right]$ and the expected value of the bias correction estimator, $E\left\{-\frac{1}{2}tr\left[\Lambda(\hat{\theta})I^i(\hat{\theta})\right]\right\}$, as well as $tr\left[A(\theta)I^i(\theta)\right]$ and $E\left\{tr\left[A(\hat{\theta})I^i(\hat{\theta})\right]\right\}$.

Figure 4-17 shows $M_i(\theta)$ and $E\left[M_i(\hat{\theta})\right]$. We see that $E\left[M_i(\hat{\theta})\right]$ consistently underestimates $M_i(\theta)$ over the entire parameter space, growing more severe as $\nu$ approaches 1.

Figure 4-18 compares the true bias correction term, $M_i(\theta) - E\left[M_i(\hat{\theta})\right]$ to $-\frac{1}{2}tr\left[\Lambda(\theta)I^i(\theta)\right]$ and $E\left\{-\frac{1}{2}tr\left[\Lambda(\hat{\theta})I^i(\hat{\theta})\right]\right\}$. We see a very similar result as in Figure 4-14 for REML methods. Again, while $M_i(\hat{\theta})$ underestimates $M_i(\theta)$, adding $-\frac{1}{2}tr\left[\Lambda(\theta)I^i(\theta)\right]$ as a bias correction would increase the amount by which $M_i(\theta)$ is underestimated. The bias correction term improves as $\nu$ increases but never becomes positive, and therefore never improves the unadjusted estimator, $M_i(\hat{\theta})$. 

94
Again we want to compare this to the bias correction under the linear parameterization $\Phi = (\sigma^2, \sigma^2_\nu)$. Figure 4-19 shows $M_1(\theta) - E[M_1(\tilde{\theta})]$ along with $tr[A(\tilde{\theta})I^{-1}(\tilde{\theta})]$ (Equation (4-15)) and $E\left[tr\left[A(\tilde{\theta})I^{-1}(\tilde{\theta})\right]\right]$ (Equation (4-36)). The result is much improved in this parameterization. We can see that $tr\left[A(\tilde{\theta})I^{-1}(\tilde{\theta})\right]$ is a highly accurate estimator for the true bias correction term, $M_1(\theta) - E[M_1(\tilde{\theta})]$ and using this estimator would improve the estimator for $M_1(\theta)$ over the entire parameter space. That is, $M_1(\tilde{\theta}) + tr\left[A(\tilde{\theta})I^{-1}(\tilde{\theta})\right]$ is a more accurate estimator of $M_1(\theta)$ than $M_1(\tilde{\theta})$. Figure 4-20 compares the accuracy of $tr\left[A(\tilde{\theta})I^{-1}(\tilde{\theta})\right]$ and $-\frac{1}{2}tr\left[A(\tilde{\theta})I^{-1}(\tilde{\theta})\right]$ for $M_1(\theta) - E[M_1(\tilde{\theta})]$. Because $M_1(\theta) - E[M_1(\tilde{\theta})]$ is positive over the entire parameter space, $-\frac{1}{2}tr\left[A(\tilde{\theta})I^{-1}(\tilde{\theta})\right]$ is an unacceptable estimator of $M_1(\theta) - E[M_1(\tilde{\theta})]$. However, $tr\left[A(\tilde{\theta})I^{-1}(\tilde{\theta})\right]$ is a highly accurate estimator for $M_1(\theta) - E[M_1(\tilde{\theta})]$ with nearly constant relative bias of 6.67%.

### 4.3.4 Impact of Negative Values of Bias Correction Term Approximation

As mentioned in Section 3.2.3 the bias correction approximation given by Equation (3-10) in indefinite. The balanced one-way random effects model under the 

$\theta = (\sigma^2, \nu)$ parameterization provides a sufficient example. Figure 4-15, shows that $-\frac{1}{2}tr\left[A(\tilde{\theta})I^{-1}(\tilde{\theta})\right]$ takes negative values over the entire range of $\nu$. We can determine from Equation (4-40) when the bias correction term estimator will be negative in terms of $k$ and $n$. First rewrite Equation (4-40) as
\[-\frac{1}{2} tr\left[ \Lambda(\theta) I^t(\theta) \right] = 2\sigma^2 \left[ \frac{(k-1)\nu^2 + 1}{nk^2(n-1)\nu} - \frac{kn-1}{nk^2(n-1)(k-1)\nu} \right] \]
\[= \frac{2\sigma^2}{nk^2(n-1)(k-1)\nu} \left[ (k-1)^2\nu^2 - k(n-1) \right].\]

Noting that the denominator is always positive when \( \nu \) is in its parameter space, we determine that the bias correction term estimator in Equation (4-40) is negative when
\[\nu < \frac{k(n-1)}{(k-1)^2}.\]

For \( k = 6 \) and \( n = 6 \), as in our example, this value is \( \nu < 1.09 \) which means that the bias correction term estimator is negative over the entire \((0, 1]\) parameter space for \( \nu \). This counterexample suffices to prove that the bias correction term approximation in its most general form,
\[\text{BIAS}^{KR}(\theta) = -\frac{1}{2} \left[ tr\{\Lambda_{11}(\theta) B(\theta)\} \cdots tr\{\Lambda_{1n}(\theta) B(\theta)\} \right] \]
\[\vdots \quad \vdots \]
\[\left[ tr\{\Lambda_{n1}(\theta) B(\theta)\} \cdots tr\{\Lambda_{nn}(\theta) B(\theta)\} \right],\]
from Equations (3-10) and (3-11), is indefinite. This in itself is not necessarily problematic since the true bias correction term, \( M_1(\theta) - E\left[ M_1(\hat{\theta}) \right] \), is also indefinite. However, if the magnitude of this piece is large enough to cause the entire MSEP estimator to be indefinite, then problems arise. It is beyond the scope of this dissertation to determine if the MSEP estimator utilizing the correction term estimator and the bias correction term estimator, \( M_1(\hat{\theta}) + C^{KR}(\hat{\theta}) + \text{BIAS}^{KR}(\hat{\theta}) \), is non-negative definite. However, it can be shown that as an estimator for \( M_1(\theta) \),
\[M_1(\hat{\theta}) + \text{BIAS}^{KR}(\hat{\theta}) \quad (4-55)\]
is indefinite by looking at the same special case of the balanced one-way random effects model.
The form of Equation (4-55) for this model is

\[
M_1(\hat{\theta}) + \text{BIAS}^{KR}(\hat{\theta}) = \frac{\hat{\sigma}^2(1-\hat{\nu})}{nk\hat{\nu}}[1+(k-1)\hat{\nu}]
\]

\[
+ 2\hat{\sigma}^2\left[\frac{(k-1)\nu^2+1}{nk^2(n-1)\nu} - \frac{kn-1}{nk^2(n-1)(k-1)\nu}\right]
\]

\[
= \frac{\sigma^2\left((k-1)^2(2-k(n-1))\nu^2+k(n-1)(k-1)(k-2)\nu+k^2(n-1)-3k(n-1)\right)}{nk^2(n-1)(k-1)}
\]

(4-56)

To determine when Equation (4-56) is negative, we find the roots of the numerator (noting that the denominator is greater than zero for \(k, n > 1\)) by solving the quadratic equation in \(\nu\). This yields the following result:

\[
M_1(\hat{\theta}) + \text{BIAS}^{KR}(\hat{\theta}) < 0
\]

where

\[
\hat{\nu} > \sqrt{\frac{3k(n-1)-k^2(n-1)}{(k-1)^2(2-k(n-1))} + \frac{(k(n-1)(k-1)(k-1))^2}{4((k-1)^2(2-k(n-1)))^2}} - \frac{k(n-1)(k-1)(k-2)}{2(k-1)^2(2-k(n-1))}
\]

(4-57)

or

\[
\hat{\nu} < \sqrt{\frac{3k(n-1)-k^2(n-1)}{(k-1)^2(2-k(n-1))} + \frac{(k(n-1)(k-1)(k-1))^2}{4((k-1)^2(2-k(n-1)))^2}} - \frac{k(n-1)(k-1)(k-2)}{2(k-1)^2(2-k(n-1))}.
\]

For \(k = 6\) and \(n = 6\) this gives \(\hat{\nu} > 0.987\) and \(\hat{\nu} < -0.130\). While the lower bound on \(\hat{\nu}\) is outside the parameter space, the upper bound is within the parameter space and a valid result for the parameter estimate. In the case of \(k = 3\) and \(n = 3\) the result is \(\hat{\nu} > 0.75\) and \(\hat{\nu} < 0\). Again, the lower bound is outside of the parameter space but the upper bound is well within the parameter space. This demonstrates the danger in using \(M_1(\hat{\theta}) + \text{BIAS}^{KR}(\hat{\theta})\) as an estimator for \(M_1(\theta)\) or \(M(\theta)\) and serves as a general warning of the possible effect of the indefiniteness of
The simplicity of this example raises even more cause for concern in more complex models.

4.4 Performance of Overall Estimators for the MSEP

Now that all the pieces of the MSEP estimators have been evaluated, we can combine these results to get an overall picture for how accurate each MSEP estimator is for the true MSEP. Recall that the true MSEP values will differ for REML and ANOVA variance components, since each variance component estimation method adds a different amount of uncertainty to the MSEP of the EBLUP and a different amount of bias to $M_1(\theta)$. For REML methods, we will look at MSEP estimators utilizing both $B_\theta(\hat{\theta})$ and $I^{-1}(\hat{\theta})$ in the correction term estimator since there was no clear preference (see Figure 4-5). However, for ANOVA methods, the estimators using $I^{-1}(\hat{\theta})$ were clearly superior to other choices of $B(\theta)$ and thus we will limit our comparison to $I^{-1}(\hat{\theta})$ for ANOVA methods. Thus the following equations list the possible estimators for the MSEP of the EBLUP of $t = a_i$. For REML variance parameter estimates we have three options:

$$M_1(\hat{\theta}) + 2tr\left[ A(\hat{\theta})B_\theta(\hat{\theta}) \right],$$  \hspace{1cm} (4-58)

$$M_1(\hat{\theta}) + 2tr\left[ A(\hat{\theta})I^{-1}(\hat{\theta}) \right],$$  \hspace{1cm} (4-59)

and

$$M_1(\hat{\theta}) + tr\left[ A(\hat{\theta})I^{-1}(\hat{\theta}) \right] - \frac{1}{2}tr\left[ A(\hat{\theta})I(\hat{\theta}) \right].$$  \hspace{1cm} (4-60)

For ANOVA variance parameter estimates we have 2 choices for MSEP estimator:
\[ M_1(\hat{\theta}) + 2tr\left[A(\hat{\theta})I^{-1}(\hat{\theta})\right], \]  
\begin{equation*}
\text{(4-61)}
\end{equation*}

and

\[ M_1(\hat{\theta}) + tr\left[A(\hat{\theta})I^{-1}(\hat{\theta})\right] - \frac{1}{2}tr\left[A(\hat{\theta})I^{-1}(\hat{\theta})\right]. \]  
\begin{equation*}
\text{(4-62)}
\end{equation*}

Note that while Equations (4-58), (4-59), and (4-61) are derived under the \( \Phi = (\sigma^2, \sigma^2_a) \) parameterization, due to the translation invariance of the elements involved, for simplicity we will use \( \theta \) notation.

We begin with the REML comparison. Figure 4-21 shows the true MSEP for the EBLUP of \( t = a_i \) using REML variance parameter estimates, \( M_R(\theta) \), along with the expected values of the three options in equations (4-58), (4-59), and (4-60). For completeness, \( E[M_1(\hat{\theta})] \) is included as well. As we have seen throughout with REML variance parameter estimates, the results are dependent on the values of \( \nu \). For small values of \( \nu \), Equation (4-59) is most accurate, almost matching the true MSEP exactly. As \( \nu \) grows, though, both Equation (4-58) and (4-60) become more accurate than Equation (4-59). Note that SAS® uses Equation (4-59) as the estimator of the MSEP in the DDFM = KR option of PROC MIXED, which is not the best choice over a large range of \( \nu \). It is especially problematic for \( \nu \) close to one, which again corresponds to values of \( \sigma^2_a \) small relative to \( \sigma^2 \).

Now looking at ANOVA comparisons in Figure 4-22, we see that Equation (4-61) is most accurate for the true MSEP, \( M_d(\theta) \), over the entire range of \( \nu \). It is clearly the best choice for MSEP estimator when ANOVA methods are used.

We also want to determine if REML or ANOVA methods are preferred when comparing MSEP estimators. To do so, we look at the relative bias of each of the five MSEP estimators.
listed above for their respective true MSEP values. Table 4-3 summarizes the results for select values of $\nu$ and Figure 4-23 illustrates this comparison. The table and figure clearly shows that the MSEP estimator with the smallest relative bias over the entire range of $\nu$ is

$$M_1(\hat{\theta}) + 2\text{tr} \left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right]$$

which uses ANOVA methods. Thus for accuracy of the MSEP estimator, ANOVA variance component estimation under the $\Phi$ parameterization with MSEP estimators utilizing $B(\Phi) = I^4(\Phi)$ are the best choice among the ones compared here.

Table 4-3. Relative bias (%) of MSEP estimators under REML and ANOVA estimation

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1(\hat{\theta})$</td>
<td>-13.28</td>
<td>-26.27</td>
<td>-30.88</td>
<td>-22.76</td>
<td>37.38</td>
</tr>
<tr>
<td>$M_1(\hat{\theta}) + 2\text{tr} \left[ A(\hat{\theta}) B(\hat{\theta}) \right]$**</td>
<td>16.75</td>
<td>10.59</td>
<td>11.71</td>
<td>36.21</td>
<td>140.50</td>
</tr>
<tr>
<td>$M_1(\hat{\theta}) + 2\text{tr} \left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right]$</td>
<td>4.03</td>
<td>22.33</td>
<td>65.91</td>
<td>173.14</td>
<td>620.57</td>
</tr>
<tr>
<td>$M_1(\hat{\theta}) + \text{tr} \left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right] - \frac{1}{2} \text{tr} \left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right]$</td>
<td>-26.84</td>
<td>-18.26</td>
<td>2.72</td>
<td>57.93</td>
<td>294.72</td>
</tr>
<tr>
<td>$M_1(\hat{\theta})$</td>
<td>-17.72</td>
<td>-51.38</td>
<td>-93.33</td>
<td>-142.68</td>
<td>-200.00</td>
</tr>
<tr>
<td>$M_1(\hat{\theta}) + 2\text{tr} \left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right]$</td>
<td>1.19</td>
<td>3.42</td>
<td>6.24</td>
<td>9.51</td>
<td>13.35</td>
</tr>
<tr>
<td>$M_1(\hat{\theta}) + \text{tr} \left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right] - \frac{1}{2} \text{tr} \left[ A(\hat{\theta}) I^{-1}(\hat{\theta}) \right]$</td>
<td>-29.70</td>
<td>-36.58</td>
<td>-49.77</td>
<td>-67.18</td>
<td>-88.37</td>
</tr>
</tbody>
</table>

**Monte Carlo values used to compute relative bias.

It is important to note that while this is the MSEP estimator used by SAS® in the DDFM = KR option when ANOVA methods are chosen with this model, the MSEP estimates are truncated at zero which changes the expected value of the MSEP estimator from the result given here. The disadvantage of using ANOVA methods is the likelihood of negative variance parameter estimates. However, by truncating the MSEP estimates, SAS® negates the advantage
of increased accuracy associated with ANOVA methods. The implication of allowing untruncated MSEP values is that negative estimates of the MSEP are possible. The impact of negative variance parameter estimates on EBLUP procedures in the balanced one-way random effects model are further explored in Chapter 5.

4.5 Comparison of Prediction Interval Methods

Several methods for conducting hypothesis tests in linear mixed models are reviewed in Chapter 2. Table 2-1 contains a summary of the methods, including how the distribution is determined and which standard error is used in the test statistic. We now want to examine these methods in terms of the balanced one-way random effects model. Because we are estimating the realized value of a random effect, it is more appropriate to examine the methods in terms of prediction intervals. The goal of this examination is to determine the accuracy of the prediction interval estimation methods and determine where further investigation is necessary.

Four prediction intervals will be compared in a Monte Carlo simulation study to determine which method has the most accurate coverage rate. All of the prediction intervals have a similar structure. The estimate of the realized value of the random effect, \( a_i \), is \( \hat{\tau} \), the EBLUP for \( a_i \), as in Equation (4-2). This is consistent in all four methods. The structure for the prediction intervals is

\[
\hat{\tau} \pm \left( \sqrt{MSEP} \right) t_{1-\alpha/2,df},
\]

(4-63)

The estimate of \( \sqrt{MSEP} \) and the degrees of freedom for the \( t \)-distribution depend on the method used to calculate the prediction interval. The methods used to determine the MSEP estimate and the degrees of freedom are summarized in Table 4-4.
### Table 4-4. Summary of prediction interval procedures.

<table>
<thead>
<tr>
<th>GB</th>
<th>KR</th>
<th>FC(Φ)</th>
<th>FC(θ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSEP</td>
<td>(M_1(\Phi))</td>
<td>(M_1(\Phi))</td>
<td>(M_1(\hat{\theta}))</td>
</tr>
<tr>
<td></td>
<td>(+2\text{tr}[A(\Phi)B(\Phi)])</td>
<td>(+2\text{tr}[A(\Phi)B(\Phi)])</td>
<td>(-\frac{1}{2}\text{tr}[A(\hat{\theta})B(\hat{\theta})])</td>
</tr>
<tr>
<td>DF</td>
<td>(df_{GB})</td>
<td>(df_{KR} = df_{GB})</td>
<td>(df_{FC(\Phi)})</td>
</tr>
<tr>
<td></td>
<td>(\frac{2 \left[ \text{MSEP}<em>{GB} \right]^2}{\text{var} \left[ \text{MSEP}</em>{GB} \right]})</td>
<td>(\frac{2 \left[ \text{MSEP}<em>{FC(\Phi)} \right]^2}{\text{var} \left[ \text{MSEP}</em>{FC(\Phi)} \right]})</td>
<td>(\frac{2 \left[ \text{MSEP}<em>{FC(\theta)} \right]^2}{\text{var} \left[ \text{MSEP}</em>{FC(\theta)} \right]})</td>
</tr>
</tbody>
</table>

Note: \(\text{var} \left[ \text{MSEP} \right] \approx \mathbf{h}(\cdot)^\prime \mathbf{B}(\cdot) \mathbf{h}(\cdot)\) where \(\mathbf{h}(\cdot) = \frac{\partial \left[ \text{MSEP}(\cdot) \right]}{\partial \cdot}\) and \(\mathbf{B}(\cdot) = \mathbf{I}^{-1}(\cdot)\).

Note that all of the methods use a “Satterthwaite-type” method for estimating the degrees of freedom. The degrees of freedom are calculated using

\[
df = \frac{2 \left[ \widehat{\text{MSEP}} \right]^2}{\text{var} \left[ \text{MSEP} \right]} \tag{4-64}
\]

where \(\text{var} \left[ \widehat{\text{MSEP}} \right] \approx \mathbf{h}(\cdot)^\prime \mathbf{B}(\cdot) \mathbf{h}(\cdot)\) for \(\mathbf{h}(\cdot) = \frac{\partial \left[ \text{MSEP}(\cdot) \right]}{\partial \cdot}\) and \(\mathbf{B}(\cdot) = \mathbf{I}^{-1}(\cdot)\). The disparity between the methods comes in the parameterization, and the estimate of the MSEP (or standard error) which subsequently impacts the degrees of freedom. We again study results under two parameterizations, \(\theta = (\sigma^2, \nu)\) and \(\Phi = (\sigma^2, \sigma_y^2)\). Note that due to the translation invariance of \(M_1(\cdot)\) and \(\text{tr}[A(\cdot)B(\cdot)]\), the figures and results demonstrated earlier in this chapter pertaining to \(M_1(\hat{\theta})\) and \(M_1(\hat{\theta}) + 2\text{tr}[A(\hat{\theta})B(\hat{\theta})]\) still apply to the standard error estimates used in the Giesbrecht-Burns, Kenward-Roger, and \(FC(\Phi)\) prediction interval procedures.
Two of the methods for formulating the prediction interval are used in SAS® PROC MIXED. The first, from the DDFM = SATTERTH option, is similar to the Giesbrecht-Burns (1985) procedure outlined in Chapter 2; however, in this application we are calculating prediction limits for the realized value of a random effect rather than a fixed effect, as in that paper. The method is also developed in Jeske and Harville (1988) and Fai and Cornelius (1996). We will refer to it as the Giesbrecht-Burns prediction interval method, since it uses the naïve MSEP estimate. The standard error is estimated by the naïve MSEP, \( \left( \hat{M}_1(\hat{\theta}) \right)^{1/2} \) where \( \hat{M}_1(\hat{\theta}) \) is in Equation (4-3).

The other method available in PROC MIXED through the DDFM = KR option is the Kenward-Roger (1997) approach. The standard error estimate includes a correction term and a bias correction term, as in Equation (4-59). Note that the Kenward-Roger prediction interval is determined under the \( \Phi \) parameterization. Recall that under the parameterization, \( \Phi = \left( \sigma^2, \sigma_a^2 \right) \), the bias correction term estimate is equal to the correction term estimate, thus the MSEP estimate used in the Kenward-Roger prediction interval is

\[
\hat{M}_1(\hat{\Phi}) + 2tr \left[ \mathbf{A}(\hat{\Phi}) \mathbf{B}(\hat{\Phi}) \right] = \hat{M}_1(\hat{\theta}) + 2tr \left[ \mathbf{A}(\hat{\theta}) \mathbf{B}(\hat{\theta}) \right].
\]

It is important to keep the parameterization in mind throughout, as the MSEP estimate and the degrees of freedom calculations depend on which parameterization is used. It is demonstrated in Chapter 2 that in a single-dimension linear combination, the Kenward-Roger method for determining the degrees of freedom reduces to the Giesbrecht-Burns method. Thus, although the Kenward-Roger and Giesbrecht-Burns methods for determining the prediction limits use different estimates for the standard error, the degrees of freedom for the \( t \)-distribution will be the same. Kenward-Roger
prediction limits adjust the MSEP estimate for estimating unknown variance components but do not adjust the \( t \)-distribution for the change in standard error.

The other two procedures for creating prediction intervals, \( FC(\Phi) \) and \( FC(\theta) \), are based on the Fai-Cornelius (1996) hypothesis testing methods. For the \( FC \) intervals, both the MSEP estimate and the degrees of freedom of the \( t \)-distribution account for the correction term and the bias correction term. Note that this is actually an extension of the methods developed by Fai and Cornelius (1996). The bias correction term was not considered as an adjustment to the MSEP estimate in the methods they proposed. The bias correction term is added here to be consistent with MSEP estimator used in the Kenward-Roger method. The FC methods differ in the parameterization. Under the \( \Phi = (\sigma^2, \sigma_a^2) \) parameterization, the MSEP estimate is the same as in the Kenward-Roger parameterization,

\[
M_i(\hat{\Phi}) + 2tr \left[ A(\hat{\Phi})B(\hat{\Phi}) \right] = M_i(\hat{\theta}) + 2tr \left[ A(\hat{\theta})B(\hat{\theta}) \right].
\]

Unlike the Kenward-Roger prediction interval, the \( FC(\Phi) \) procedure adjusts the degrees of freedom of the \( t \)-distribution for the adjustments to the MSEP estimate. The degrees of freedom are still calculated using a “Satterthwaite” method; however, now the corrected MSEP estimate,

\[
M_i(\hat{\Phi}) + 2tr \left[ A(\hat{\Phi})B(\hat{\Phi}) \right],
\]

is used in the calculation of the degrees of freedom, rather than the naïve MSEP estimate, \( M_i(\hat{\Phi}) \).

The final prediction limit method considered, \( FC(\theta) \), differs from the other three in the parameterization, \( \theta = (\sigma^2, \nu) \). Recall that the bias correction term estimate is not translation invariant, and thus by changing the parameterization, we also change the MSEP estimate. The
MSEP estimate for the $FC(\theta)$ procedure is $M_1(\hat{\theta}) + tr\left[ A(\hat{\theta})B(\hat{\theta}) \right] - \frac{1}{2} tr\left[ \Lambda(\hat{\theta})B(\hat{\theta}) \right]$. This is used in both the standard error estimate and the degrees of freedom calculation.

A Monte Carlo simulation study was conducted to compare the four prediction interval methods. The study is similar to the one conducted in Section 4.1 to evaluate the correction term estimators. Using the RANNOR function in SAS®, 10,000 independent sets of $e_{ij}$ and $a_i$ were generated from $N(0, \sigma^2 = 1)$ and $N(0, \sigma^2_a)$ for several values of $\sigma^2_a$, respectively. The values of $\sigma^2_a$ correspond to $\nu = \{0.05, 0.1, 0.15, 1.0\}$. Setting $\mu = 0$, 10,000 sets of $\{y_{ij}: i = 1, \ldots k, j = 1, \ldots 6\}$ were created from the model $y_{ij} = a_i + e_{ij}$. The number of levels of the random effect is varied to study the impact on the performance of the prediction interval procedures. The values for $k$ are $k = 3, 6, 15, or 30, while n is held constant at n = 6$. Simulating the data in this way allows the realized values generated for the $a_i$’s to be tracked. These are the realized values of the random effect for which the prediction intervals are constructed. By tracking the realized values of the random effect, we can determine how often the prediction intervals contain the true value of the realized random effect, providing the actual (simulated) coverage rate for a nominal 95% prediction interval.

The prediction limits for $a_i$ for the Giesbrecht-Burns and Kenward-Roger procedures are produced by the CL option in the ESTIMATE statement in PROC MIXED in SAS®, with the DDFM = SATTERTH or DDFM = KR options, respectively. To calculate the FC prediction limits for either parameterization, the degrees of freedom must be calculated directly and then used to determine the critical value from the $t$-distribution. For the $FC(\Phi)$ procedure, the standard error from the Kenward-Roger procedure is used to calculate the prediction limits. For
the \( FC(\theta) \) procedure, the standard error also has to be calculated directly in addition to the
degrees of freedom because of the change in parameterization. The SAS\textsuperscript{®} programming
statements used to perform the simulation study are contained in Appendix A along with the
derivatives needed to compute the degrees of freedom for the FC procedures.

The results of the simulation study are best demonstrated pictorially. Figures 4-24 through
4-27 contain the results of the simulation study for \( k = 3, 6, 15, \) and 30, respectively. It is clear
that the Kenward-Roger procedure produces the most inflated coverage rates consistently across
values of \( k \) and for the entire parameter space of \( \nu \). The prediction limits for the Kenward-
Roger procedure will always be larger than the Giesbrecht-Burns procedure since
\[
M_1(\hat{\theta}) + 2tr\left[A(\hat{\theta})B(\hat{\theta})\right] \geq M_1(\hat{\theta}).
\]
Adjusting the degrees of freedom in the \( FC(\Phi) \) procedure
for the adjusted MSEP estimate generally increases the degrees of freedom, causing the critical
value from the \( t \)-distribution to shrink and thus reducing the width of the prediction interval. The
coverage rates for the \( FC(\Phi) \) procedure are generally closer to 95% than the coverage rates for
the Kenward-Roger procedure.

The performance of the \( FC(\theta) \) prediction interval is more dependent on the value for \( k \)
and the parameter \( \nu \) than the other intervals. For \( k = 3 \), the coverage rate for small values of \( \nu \)
is exceedingly small. As \( \nu \) increases, the performance improves dramatically. This is due to the
inclusion of the bias correction term under the \( \theta = (\sigma^2, \nu) \) parameterization. Small values of \( k \)
combined with small values of \( \nu \) produce small MSEP estimates. Recall from Figure 4-23 that
the relative bias of \( M_1(\hat{\theta}) + tr\left[A(\hat{\theta})B(\hat{\theta})\right] - \frac{1}{2}tr\left[A(\hat{\theta})B(\hat{\theta})\right] \) approached -50% for small
values of \( \nu \). This is caused by the significant underestimation of the true bias correction term by
\[-\frac{1}{2} \text{tr}\left[ A(\hat{\theta})B(\hat{\theta}) \right]\] (see Figure 4-14). For larger values of $k$, the $FC(\theta)$ prediction interval procedure compares favorably with the other procedures and consistently improves on the performance of the Kenward-Roger prediction interval for larger values of $\nu$. For larger values of $\nu$ recall that $M_1(\hat{\theta}) + 2\text{tr} \left[ A(\hat{\theta})B(\hat{\theta}) \right]$ overestimated the true MSEP more so than $M_1(\hat{\theta}) + \text{tr} \left[ A(\hat{\theta})B(\hat{\theta}) \right] - \frac{1}{2} \text{tr} \left[ A(\hat{\theta})B(\hat{\theta}) \right]$. Thus we would expect a more accurate prediction interval with the $FC(\theta)$ method in this range of the parameter space.

To demonstrate the relationship between the prediction intervals more concretely, consider the hypothetical case where $k = 6, n = 6$, with data producing variance parameter estimates of $\hat{\sigma}^2 = 1$ and $\hat{\sigma}_a^2 = .5$. Table 4-5 contains the values for the standard error, degrees of freedom, critical $t$-value and the width of the prediction interval produced by each of the four procedures in this scenario. The width of the Kenward-Roger prediction interval increases over the width of the Giesbrecht-Burns interval due to the increase in the standard error estimate. The width of the $FC(\Phi)$ prediction interval is smaller than both the Giesbrecht-Burns and the Kenward-Roger intervals because of the increased degrees of freedom and subsequent decrease in the critical $t$-value. The width of the $FC(\theta)$ interval is by far the smallest due to the decrease in the standard error estimate and the critical $t$-value.

Finally, to demonstrate the effect of $k$ on the performance of the prediction interval procedures, Figure 4-28 shows the Kenward-Roger coverage rates for all values of $k$ considered in the simulation study. This is demonstrative of all the procedures in that as $k$ increases, the procedures perform better and the coverage rates move closer to the nominal value of 95%.
This study shows the importance of investigating the Fai-Cornelius methods for hypothesis tests and prediction intervals in more complex situations. These preliminary results indicate that the Fai-Cornelius methods often perform better than the Kenward-Roger and the Giesbrecht-Burns methods used in SAS®. The Kenward-Roger procedure produces an overly-conservative prediction interval because it does not adjust the degrees of freedom for the inflated MSEP estimate. The study also demonstrates the impact of the parameterization and the lack of transform invariance of the bias correction term on accuracy of the prediction interval methods. In more complex covariance structures, the parameterization may be vital to an accurate analysis.

Table 4-5. Effect of prediction interval procedures for $k = 6$, $n = 6$, $\hat{\sigma}^2 = 1$ and $\hat{\sigma}_u^2 = .5$

<table>
<thead>
<tr>
<th></th>
<th>GB</th>
<th>KR</th>
<th>$FC(\Phi)$</th>
<th>$FC(\theta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Error estimate</td>
<td>0.43301</td>
<td>0.46894</td>
<td>0.46894</td>
<td>0.40196</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>8.15</td>
<td>8.15</td>
<td>16.39</td>
<td>16.52</td>
</tr>
<tr>
<td>$t_{0.025,df}$</td>
<td>2.298</td>
<td>2.298</td>
<td>2.116</td>
<td>2.114</td>
</tr>
<tr>
<td>Width of CI</td>
<td>1.99</td>
<td>2.16</td>
<td>1.98</td>
<td>1.70</td>
</tr>
</tbody>
</table>
Figure 4-1. Equation (4-13) as a function of $\nu$ for several values of $k$, holding $n = 6$. Equation (4-14) holds when functions cross zero line.
Figure 4-2. Accuracy of approximations $\text{tr}\left[ A(\theta)B_R(\theta) \right]$ and $\text{tr}\left[ A(\theta)I^{-1}(\theta) \right]$ for the correction term, $C_R(\theta)$, under REML variance component estimation.
Figure 4-3. Accuracy of approximation $\text{tr}[\mathbf{A}(\theta) \mathbf{B}_R(\theta)]$ and estimator $E\left\{\text{tr}[\mathbf{A}(\hat{\theta}) \mathbf{B}_R(\hat{\theta})]\right\}$ for the correction term, $C_R(\theta)$, under REML variance component estimation.
Figure 4-4. Accuracy of approximation $\text{tr}[A(\theta)I^{-1}(\theta)]$ and estimator $E\left\{\text{tr}[A(\hat{\theta})I^{-1}(\hat{\theta})]\right\}$ for the correction term, $C_n(\theta)$, under REML variance component estimation.
Figure 4-5. Comparison of accuracy of estimators for the correction term, $C_R(\theta)$, under REML variance component estimation.
Figure 4-6. Comparison of accuracy of approximations $\text{tr} \left[ A(\theta) B_A(\theta) \right]$, $\text{tr} \left[ A(\theta) B_A^*(\theta) \right]$, and $\text{tr} \left[ A(\theta) B_{\theta}(\theta) \right]$ for the correction term, $C_A(\theta)$, under ANOVA variance component estimation.
Figure 4-7. Accuracy of approximation \( tr[A(\theta)B_A(\theta)] \) and estimator \( E\{tr[A(\hat{\theta})B_A(\hat{\theta})]\} \) for the correction term, \( C_A(\theta) \), under ANOVA variance component estimation.
Figure 4-8. Accuracy of approximation $\text{tr} \left[ A(\theta)B_A^*(\theta) \right]$ and estimator $E \left\{ \text{tr} \left[ A(\bar{\theta})B_A^*(\bar{\theta}) \right] \right\}$ for the correction term, $C_A(\theta)$, under ANOVA variance component estimation.
Figure 4-9. Accuracy of approximation $tr\left[A(\theta)I^{-1}(\theta)\right]$ and estimator $E\left\{tr\left[A(\bar{\theta})I^{-1}(\bar{\theta})\right]\right\}$ for the correction term, $C_A(\theta)$, under ANOVA variance component estimation.
Figure 4-10. Comparison of accuracy of estimators for the correction term, $C_{A}(\theta)$, under ANOVA variance component estimation.
Figure 4-11. Accuracy of estimators for the correction term under REML, $\hat{\theta}$, and ANOVA, $\hat{\theta}$, variance component estimation.
Figure 4-12. Relative bias of estimators for the correction term under REML, $\hat{\theta}$, or ANOVA, $\tilde{\theta}$, variance component estimation.
Figure 4-13. Bias produced by estimating \( M_1(\theta) \) with \( \hat{M}_1(\hat{\theta}) \).
Figure 4-14. Accuracy of approximation $-\frac{1}{2} \text{tr} \left[ \Lambda(\theta) \Gamma^{-1}(\hat{\theta}) \right]$ and estimator $E \left\{ -\frac{1}{2} \text{tr} \left[ \Lambda(\hat{\theta}) \Gamma^{-1}(\hat{\theta}) \right] \right\}$ for the true bias correction term, $M_1(\theta) - E \left[ M_1(\hat{\theta}) \right]$, under REML variance component estimation.
Figure 4-15. Accuracy of approximation \( tr\left[ A(\theta)I^{-1}(\theta) \right] \) and estimator \( E\left\{ tr\left[ A(\hat{\theta})I^{-1}(\hat{\theta}) \right] \right\} \) for the true bias correction term, \( M_1(\theta) - E\left[ M_1(\hat{\theta}) \right] \), under REML variance component estimation.
Figure 4-16. Comparing the accuracy of $E\left\{ \frac{1}{2} \text{tr} \left[ \Lambda(\hat{\theta}) \Gamma^{-1}(\hat{\theta}) \right] \right\}$ and $E\left\{ \text{tr} \left[ \Lambda(\hat{\theta}) \Gamma^{-1}(\hat{\theta}) \right] \right\}$ for the true bias correction term, $M_i(\theta) - E\left[ M_i(\hat{\theta}) \right]$. 
Figure 4-17. Bias produced by estimating $M_1(\theta)$ with $M_1(\hat{\theta})$. 
Figure 4-18. Accuracy of approximation $-\frac{1}{2} \text{tr}\left[ \Lambda(\theta) I^{-1}(\theta) \right]$ and estimator $E\left\{ -\frac{1}{2} \text{tr}\left[ \Lambda(\bar{\theta}) I^{-1}(\bar{\theta}) \right] \right\}$ for true bias correction term, $M_1(\theta) - E\left[ M_1(\bar{\theta}) \right]$, under ANOVA variance component estimation.
Figure 4-19. Accuracy of approximation, \( tr \left[ A(\theta) I^1(\theta) \right] \) and estimator \( E \left\{ tr \left[ A(\tilde{\theta}) I^1(\tilde{\theta}) \right] \right\} \) for true bias correction term, 
\( M_1(\theta) - E \left[ M_1(\tilde{\theta}) \right] \), under ANOVA variance component estimation.
Figure 4-20. Comparing the accuracy of $E\left\{ -\frac{1}{2} tr\left[ \Lambda(\bar{\theta}) I^{-1}(\bar{\theta}) \right] \right\}$ and $E\left\{ tr\left[ \Lambda(\bar{\theta}) I^{-1}(\bar{\theta}) \right] \right\}$ for the true bias correction term, $M_1(\theta) - E\left[ M_1(\tilde{\theta}) \right]$. 
Figure 4-21. Comparison of estimators of the MSEP for the EBLUP of $a_i$ utilizing REML variance component estimation.
Figure 4-22. Comparison of estimators of the MSEP for the EBLUP of $a_i$ utilizing ANOVA variance component estimation.
Figure 4-23. Relative bias of possible estimators for MSEP of EBLUP for $a_i$. 
Figure 4-24. Simulated coverage rates of prediction interval procedures for \( t = a_1 \) with \( k = 3 \).
Figure 4-25. Simulated coverage rates of prediction interval procedures for $t = a_i$ with $k = 6$. 
Figure 4-26. Simulated coverage rates of prediction interval procedures for $t = a_1$ with $k = 15$. 
Figure 4-27. Simulated coverage rates of prediction interval procedures for $t = a_1$ with $k = 30$. 
Figure 4-28. Simulated coverage rates for the Kenward-Roger prediction interval procedure.
CHAPTER 5  
THE EFFECT OF NEGATIVE VARIANCE COMPONENTS ON EBLUP

The effect of negative variance component estimates on EBLUP procedures is a topic that has been largely neglected (Stroup and Littell 2002). While the use of REML estimates, which are truncated at zero, would seem to negate this issue, there are still many circumstances when other variance parameter estimation methods are chosen and negative variance components may arise. For example, as shown in Stroup and Littell (2002) and in the results on the accuracy of the correction term and bias correction term estimates in Chapter 4, REML may not always be the preferred choice for variance component estimation. One strategy for handling negative variance component estimation is to allow that the component may be a covariance rather than a variance, thus alleviating the problem of truncating at zero. We begin to look at the impact of that strategy by again turning to the special case of the balanced one way random effects model.

5.1 Negative Variance Component Estimates and the Balanced One Way Random Effects Model

To demonstrate the current issues with negative variance component estimation, we first look at the current SAS® 9.1 output for PROC MIXED, both when the variance component estimates are truncated at zero and when they are not. Using the REML NOBOUND statement in the procedure statement allows the variance component estimates to take any value on the real line. Note that because we are in a balanced situation, the ANOVA variance component estimation would produce the same results as REML NOBOUND. We compare the effect of truncating using simulated data, generated from a normal distribution with $\sigma_a^2 = .01$ and $\sigma^2 = 1$. This distribution produces a high likelihood that the data generated will have larger variation within a block than between the blocks, hence often producing a negative estimate for $\sigma_a^2$. For demonstration purposes, the number of levels of the random effect is $k = 5$ and the number of
observations at each level is \( n = 6 \). A dataset that meets the above condition is given in Table 5-1. It is easy to spot that more variation occurs within the block levels than between them.

Table 5-1. Generated data that produces negative variance component estimate

<table>
<thead>
<tr>
<th>Replicate</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>2.34754</td>
<td>0.24672</td>
<td>-0.5585</td>
<td>0.60564</td>
<td>-0.9062</td>
<td>1.37627</td>
</tr>
<tr>
<td>Block 2</td>
<td>1.01248</td>
<td>-0.0493</td>
<td>0.49583</td>
<td>-1.3952</td>
<td>-0.8644</td>
<td>1.93825</td>
</tr>
<tr>
<td>Block 3</td>
<td>0.18826</td>
<td>0.91668</td>
<td>0.68925</td>
<td>0.96593</td>
<td>0.95929</td>
<td>-0.7838</td>
</tr>
<tr>
<td>Block 4</td>
<td>-0.4741</td>
<td>1.31312</td>
<td>-1.5036</td>
<td>0.06752</td>
<td>1.2383</td>
<td>-0.7888</td>
</tr>
<tr>
<td>Block 5</td>
<td>-1.2176</td>
<td>0.64999</td>
<td>0.32013</td>
<td>1.4081</td>
<td>-0.1774</td>
<td>1.77956</td>
</tr>
</tbody>
</table>

The results from the REML option in the MIXED procedure of SAS® are contained in Tables 5-2 and 5-3 and the results from the REML NOBOUND options in the MIXED procedure are contained in Tables 5-4 and 5-5. Also, for comparison, least square means (LSMEANS) generated by PROC GLM are given in Table 5-6. These give a starting point from which to compare the EBLUPS resulting from the REML and REML NOBOUND options. The following SAS® code generates the output in Tables 5-2 and 5-3:

```sas
proc mixed data = &sample method = reml ;
class blk trt;
model y = /ddfm = kenwardroger ;
random blk / s ;
```

Table 5-2. REML variance parameter estimates

<table>
<thead>
<tr>
<th>Covariance parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>0</td>
</tr>
<tr>
<td>Residual</td>
<td>1.0643</td>
</tr>
</tbody>
</table>

Table 5-3. Solution for random effects with REML variance parameter estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Block</th>
<th>Estimate</th>
<th>Std Err Pred</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>1</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Block</td>
<td>2</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Block</td>
<td>3</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Block</td>
<td>4</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Block</td>
<td>5</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
The NOBOUND option in the following SAS® code is the only difference from the previous code and generates the SAS® output in Tables 5-4 and 5-5:

```sas
proc mixed data = &sample method = reml nobound ;
class blk trt;
model y = /ddfm = kenwardroger ;
random blk / s ;
```

Table 5-4. REML NOBOUND variance parameter estimates

<table>
<thead>
<tr>
<th>Covariance Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blk</td>
<td>-0.1410</td>
</tr>
<tr>
<td>Residual</td>
<td>1.1810</td>
</tr>
</tbody>
</table>

Table 5-5. Solution for random effects with REML NOBOUND variance parameter estimates

| Effect | Block | Estimate | Std Err Pred | DF  | t Value | Pr > |t| |
|--------|-------|----------|--------------|-----|---------|------|---|
| Block  | 1     | -0.4846  | 0            | 29  | -Infty  | <.0001|
| Block  | 2     | 0.3461   | 0            | 29  | Infty   | <.0001|
| Block  | 3     | -0.4106  | 0            | 29  | -Infty  | <.0001|
| Block  | 4     | 0.8870   | 0            | 29  | Infty   | <.0001|
| Block  | 5     | -0.3379  | 0            | 29  | -Infty  | <.0001|

Table 5-6. LSMEANS from PROC GLM

<table>
<thead>
<tr>
<th>Block</th>
<th>LSMEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.51858487</td>
</tr>
<tr>
<td>2</td>
<td>0.18960919</td>
</tr>
<tr>
<td>3</td>
<td>0.48927203</td>
</tr>
<tr>
<td>4</td>
<td>-0.02458600</td>
</tr>
<tr>
<td>5</td>
<td>0.46046652</td>
</tr>
</tbody>
</table>

The results from the REML option, which truncates the variance component estimate for block effect at zero, are straightforward. The solution statement produces zero estimates for all blocks and missing values for the standard error, t-values, degrees of freedom, and p-values. The REML NOBOUND option is not quite so straightforward. When variance component estimates are positive, we can think of the BLUP as a “shrinkage” predictor. In other words, using the
variance component estimate as a weight, the block means are shrunk toward the overall mean. In addition, in a balanced model, the BLUPS maintain the rank order of the lsmeans. This can be easily seen in the structure of the BLUP for the balanced one-way random effects model in Equation (4-2).

Now, looking at the EBLUP estimates in Table 5-5, the first indication of a problem is that the standard errors are listed as zero and hence the t-values are listed as infinity. The calculations for standard error in this case are negative, causing SAS® to truncate the standard error at zero. Comparing the estimates in Table 5-5 to the LSMEANS from PROC GLM in Table 5-6 shows that the negative variance component estimate causes the EBLUP estimates to invert their rank order and expands the range of the estimates. The block with the largest LSMEAN (0.5186) now has the smallest EBLUP (-0.4846) and, likewise, the block with the smallest LSMEAN (-0.0246) now has the largest EBLUP (.8870). Also, the range of the LSMEANS is 0.5432 while the range of the EBLUP estimates is 1.3716. We normally expect the BLUP procedure to “shrink” the block means toward the overall mean according to the variation in the model; however, in this case, the estimates are inverted about the mean and expanded away from the overall mean. While the standard errors and t-values listed with the EBLUPs are a clear red flag, the unexpected results of the EBLUPS require further investigation. Note also that the standard errors may be positive even when $\sigma_u^2 < 0$; this would eliminate the red flag and leave more room for error in the interpretation of the results.

If the data analyst believes that a one-way random effects model is the true model for the research situation, then the recommendation from an EBLUP standpoint is to truncate at zero rather than proceed with a negative variance component estimate. Any use of EBLUPS while allowing the variance component to give a negative estimate is likely to yield misleading results.
The LSMEANS likely yield more reliable estimates than the EBLUPS from the REML NOBOUND option. Keep in mind that it is not always the best choice to truncate at zero when variance component estimates are negative if the analyst is interested in fixed effects (Stroup and Littell 2002).

5.2 Considering the Variance Parameter as a Covariance

One idea that surfaces when faced with negative variance component estimates is considering the parameter as a covariance rather than a variance, thus allowing the parameter estimate to take on values over the entire real line. Smith and Murray (1984) consider this option and the impact on estimation and hypothesis testing of the covariance component. For the balanced one way random effects model discussed in Chapter 4, the covariance structure would be modified from

$$\text{cov}(y_{ij}, y_{ij'}) = \sigma_a^2 + \sigma^2, \ i = i', \ j = j'$$
$$= \sigma_a^2, \ i = i', \ j \neq j'$$
$$= 0, \ i \neq i'$$

to

$$\text{cov}(y_{ij}, y_{ij'}) = \theta_a + \sigma^2, \ i = i', \ j = j'$$
$$= \theta_a, \ i = i', \ j \neq j'$$
$$= 0, \ i \neq i'$$

where

$$0 \leq |\theta_a| \leq \sigma^2 + \theta_a.$$
random sample from a population of cows, and $\sigma_a^2$, the variance associated with this population. When the estimate for $\sigma_a^2$ is negative, the definition of the parameter is changed from the population variance of the cows to the covariance between twins born to the same cow. Thus the population variance for the cows is ignored all together and the negative covariance estimate is interpreted as reflecting competition for nutrition between the twin calves. The negative value of the variance parameter estimate is likely reflecting the combination of the positive cow population variance and the negative correlation between twins born to the same cow. That is, there is more negative intraclass variability than positive interclass variability resulting in a negative parameter estimate.

The inversion of the LSMEANs that we see with the EBLUPs in PROC MIXED when the variance component estimate is negative occurs because the model is effectively changed to this Smith and Murray (1984) model. The EBLUP calculated by PROC MIXED for $a_i$ in this case is

$$\frac{n\hat{\theta}_a}{\hat{\sigma}^2 + n\hat{\theta}_a} (\bar{y}_i - \bar{y})$$

giving a negative coefficient when $-\frac{\hat{\sigma}^2}{n} < \hat{\theta}_a < 0$. Furthermore, $|\hat{\theta}_a|$ may be sufficiently large so that $\frac{n\hat{\theta}_a}{\hat{\sigma}^2 + n\hat{\theta}_a} < -1$. Thus rather than “shrinking” the means, as we expect EBLUP procedures to do, the EBLUP inverts and expands the range of the means. Clearly this is an unacceptable result. The question remains: what is an analyst to do when faced with this result. One possible solution is derived in the next section.
5.3 BLUP derivation for Random Effects Model with Correlated Errors

We now present a model that would account for both the population variance of the random effect and the correlation of replicates within a level of the random effect. Consider a model with a random effect, as well as correlated errors. The model and variance structure in Equation (4-1) would be modified to:

\[
y_{ij} = \mu + a_i + e_{ij}, i = 1, \ldots, k, j = 1, \ldots, n
\]

\[
a_i \sim N\left(0, \sigma^2_a\right)
\]

where \(e_{ij}\) are normally distributed with

\[
E(e_{ij}) = 0
\]

\[
\text{cov}(e_{ij}, e_{ij'}) = \begin{cases} 
\sigma^2 + \theta_a, & i = j', j = j' \\
\theta_a, & i = i', j \neq j' \\
0, & i \neq i' 
\end{cases}
\]

(5-1)

In mixed models notation, this equates to

\[
y = X\beta + Zu + e
\]

where

\[
X = 1_k, \quad \beta = \mu
\]

\[
Z = 1_k \otimes 1_n
\]

\[
u = [a_1, a_2, \ldots, a_k]' \sim N(0, G)
\]

where

\[
G = \sigma^2_a I_k
\]

and

\[
e = [e_{11}, e_{12}, \ldots, e_{in}]' \sim N(0, R)
\]

where

\[
R = I_k \otimes (\sigma^2 I_n + \theta_a J_n)
\]

so that
\[ V = ZGZ' + R = I_k \otimes \left\{ \sigma^2 I_n + \left( \sigma_a^2 + \theta_a \right) J_n \right\}. \]

From general matrix theory (e.g., Searle 1982), we have

\[ V^{-1} = I_k \otimes \left\{ \frac{1}{\sigma^2} \left( I_n - \frac{\sigma_a^2 + \theta_a}{\sigma^2 + n\left( \sigma_a^2 + \theta_a \right)} J_n \right) \right\} \]

which gives the BLUP for \( u \) as

\[ \hat{u} = GZV^{-1}(y - \hat{\mu}) = I_k \otimes \left[ \frac{\sigma_a^2}{\sigma^2 + n\left( \sigma_a^2 + \theta_a \right)} \right] I_n'(y - \bar{y}_n) \]

or the BLUP for \( a_i \) as

\[ \hat{a}_i = \frac{n \sigma_a^2}{\sigma^2 + n\left( \sigma_a^2 + \theta_a \right)} (\bar{y}_i - \bar{y}_n). \quad (5-2) \]

Note that for the model structure given in Equation (5-1) to be valid, \( V \) must be positive definite.

Therefore, the parameter space for which the model is valid is

\[ \sigma^2 > 0, \quad \sigma_a^2 > 0, \quad |\sigma_a^2 + \theta_a| < \frac{\sigma^2}{n}. \]

In this parameter space the coefficient for the BLUP, given in Equation (5-2), is non-negative. Thus the order of the LSMEANS for the blocks is preserved in the BLUP. One key consideration for this model is the problem of overparameterization. The parameters \( \sigma_a^2 \) and \( \theta_a \) are confounded and cannot be simultaneously estimated. One solution is to use an estimate for \( \sigma_a^2 \) from previous experience and consider the parameter as known. This leaves only \( \theta_a \) and \( \sigma^2 \) as unknown and the EBLUP analysis can be performed where the unknown variance parameters are estimated by an unbounded method, such as REML NOBOUND in PROC MIXED. There
are many situations, such as agriculture or education, where using a previous estimate as the parameter is realistic. Data contained in Table 5-5 are generated from a distribution as described in Equation (5-1) with \( \sigma^2 = .9, \theta_a = -.13 \) and \( \sigma_a^2 = .07 \).

<table>
<thead>
<tr>
<th>Replicate</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>1.77754</td>
<td>-0.72357</td>
<td>-0.79317</td>
<td>-0.79674</td>
<td>1.31455</td>
<td>-0.98605</td>
</tr>
<tr>
<td>Block 2</td>
<td>-0.62131</td>
<td>-1.15758</td>
<td>-1.21619</td>
<td>-0.24776</td>
<td>0.5867</td>
<td>0.12472</td>
</tr>
<tr>
<td>Block 3</td>
<td>-0.3412</td>
<td>-0.7861</td>
<td>0.41517</td>
<td>0.75414</td>
<td>1.34176</td>
<td>0.75806</td>
</tr>
<tr>
<td>Block 4</td>
<td>-0.05152</td>
<td>-0.9336</td>
<td>0.31271</td>
<td>-0.3097</td>
<td>1.04458</td>
<td>1.07845</td>
</tr>
<tr>
<td>Block 5</td>
<td>0.75952</td>
<td>-0.27854</td>
<td>0.28914</td>
<td>1.03234</td>
<td>-0.88724</td>
<td>-2.24931</td>
</tr>
</tbody>
</table>

The appropriate SAS® statements to analyze this data in the manner described in Equation (5-2) are

```sas
data gmat ;
input row col value;
cards;
1 1 .07
2 2 .07
3 3 .07
4 4 .07
5 5 .07 ;

proc mixed data = &sample method = reml ;
class blk trt ;
by rn ;
model y = /ddfm = kenwardroger ;
random blk / gdata=gmat g s ;
repeated trt / sub = blk type=cs ;
run ;
```

The `gdata` option after the random statement provides the known \( \mathbf{G} \) matrix which in this case is

\[
\mathbf{G} = .07 \mathbf{I}_n
\]

where \( \mathbf{I}_n \) is an \( n \)-dimensional identity matrix.

The output from this analysis of the data in Table 5-7 are contained in Tables 5-8 and 5-9. The EBLUPs for the levels of the random effect are generated by the `solutions` or `s` option.
in the random statement of the SAS® code, and are given in Table 5-9 under “Estimate.”

Comparison of these values to the LSMEANS generated by GLM in Table 5-10 show that the
function worked as expected, maintaining the rank order of the means and “shrinking” the
predictors toward the overall mean.

Table 5-8. Covariance parameter estimates under Model 1 (5-1)

<table>
<thead>
<tr>
<th>Covariance Parameter</th>
<th>Subject</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td>Block</td>
<td>0.9497</td>
</tr>
<tr>
<td>CS</td>
<td>Block</td>
<td>-0.1311</td>
</tr>
</tbody>
</table>

Table 5-9. Solution for random effects under Model (5-1)

| Effect | Block | Estimate | Std Err Pred | DF | t Value | Pr > |t| |
|--------|-------|----------|--------------|----|---------|-------|
| Block  | 1     | -0.00593 | 0.2646       | 2.16 | -0.02   | 0.9840 |
| Block  | 2     | -0.2849  | 0.2646       | 2.16 | -1.08   | 0.3868 |
| Block  | 3     | 0.2761   | 0.2646       | 2.16 | 1.04    | 0.3991 |
| Block  | 4     | 0.1560   | 0.2646       | 2.16 | 0.59    | 0.6113 |
| Block  | 5     | -0.1412  | 0.2646       | 2.16 | -0.53   | 0.6433 |

Table 5-10. LSMEANS from PROC GLM

<table>
<thead>
<tr>
<th>Block</th>
<th>y LSMEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.03457327</td>
</tr>
<tr>
<td>2</td>
<td>-0.42190439</td>
</tr>
<tr>
<td>3</td>
<td>0.35697247</td>
</tr>
<tr>
<td>4</td>
<td>0.19015370</td>
</tr>
<tr>
<td>5</td>
<td>-0.22234738</td>
</tr>
</tbody>
</table>

Note that the DDFM = KR method was used here which uses an adjusted standard error. If the
DDFM = SATTERTH option is used, the estimates and degrees of freedom remain the same, but
the standard error is reduced to .1647 which would impact the prediction limits as shown in
Section 4.5. Given the conclusion in Chapter 4 regarding the inaccuracy of the adjustment under
certain circumstances, the accuracy of either standard error estimate and their effect on prediction intervals for the realized values of the random effect needs further study. This remains as future research.

For comparison, Tables 5-11 and 5-12 contain the results of the analysis of the data in Table 5-7 if the correlation were ignored and the model were run as a random effects model with the REML NOBOUND option. We see that the variance parameter estimate is negative and the EBLUPs are inverted and expanded as shown in Section 5.1.

Table 5-11. Variance parameter estimates under random effects model

<table>
<thead>
<tr>
<th>Covariance Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>-0.06110</td>
</tr>
<tr>
<td>Residual</td>
<td>0.9497</td>
</tr>
</tbody>
</table>

Table 5-12. Solution for random effects under random effects model

| Effect | Block | Estimate | Std Err Pred | DF  | t Value | Pr > |t| |
|--------|-------|----------|---------------|-----|---------|-------|
| Block  | 1     | 0.005186 | 0             | 29  | Infty   | <.0001|
| Block  | 2     | 0.2487   | 0             | 29  | Infty   | <.0001|
| Block  | 3     | -0.2410  | 0             | 29  | -Infty  | <.0001|
| Block  | 4     | -0.1361  | 0             | 29  | -Infty  | <.0001|
| Block  | 5     | 0.1232   | 0             | 29  | Infty   | <.0001|

One issue that occurs with analyzing data in this way is that it is prone to convergence failure. Of 10,000 simulated data sets from the distribution described in Equation (5-1) and analyzed with this procedure, 21.5% failed to converge when wither REML or REML NOBOUND procedures were used. Even with this complication, this parameterization remains one of the only viable options if predictors of the realized values of the random effects are desired in a case with negative variance component estimation.
CHAPTER 6
CONCLUSION AND FUTURE WORK

The primary goal of this dissertation was to examine the impact of estimating variance components on the estimators of the MSEP for the EBLUP. This was accomplished by investigating properties of generalized components of the MSEP estimator, examining the behavior of various procedures in the balanced one-way random effects model, and determining the impact of negative variance components on EBLUP procedures.

Following the work by Kackar and Haville (1984), Kenward and Roger (1997), Jeske and Harville (1988), and Fai and Cornelius (1996), estimators for the correction term and the bias correction term were generalized to accommodate multiple dimension linear combinations of fixed and random effects. Properties of these generalized estimators were investigated. We showed that while for linear covariance structures the bias correction term estimator, $\text{BIAS}^{KR}(\theta)$, continues to reduce to the correction term estimator, $C^{KR}(\theta)$, which is non-negative definite, $\text{BIAS}^{KR}(\theta)$, in general, is indefinite. The indefiniteness of $\text{BIAS}^{KR}(\theta)$ causes the estimator, $M_1(\hat{\theta}) + \text{BIAS}^{KR}(\hat{\theta})$, of the naïve MSEP, $M_1(\theta)$, to be indefinite. Future work includes the investigation of definiteness properties of the MSEP estimator $M_1(\theta) + C^{KR}(\theta) + \text{BIAS}^{KR}(\theta)$, which includes both the correction term and the bias correction term. The magnitude of $M_1(\theta) + C^{KR}(\theta)$ may outweigh that of $\text{BIAS}^{KR}(\theta)$ to produce a positive definite estimator. We also demonstrated that $\text{BIAS}^{KR}(\theta)$ is not transform invariant, and that the choice of parameterization can have a significant impact on the bias correction term, the MSEP estimate, and the coverage rates of prediction intervals.

The balanced one-way random effects model provided a simple setting in which the various components of MSEP estimators could be thoroughly explored. Closed form
expressions allowed for direct examination of the correction term and bias correction term estimators. Through this example, the impact of variance component estimation techniques, the values of the variance components, and the parameterization was demonstrated. We determined that the MSEP estimator may significantly over- or under-estimate the true MSEP depending on these factors. We showed that truncating variance parameter estimates under REML estimation can cause significant (long-run) overestimation of both the correction term and the bias correction term, over most of the parameter space. While these procedures have been considered suspect near the boundary of the parameter space, this study showed that the accuracy of the methods is questionable over a much greater range of the parameter space.

We also discovered that the choice of how to measure the MSE of the variance parameter estimates can significantly impact the accuracy of the correction term estimator. In REML situations, using the exact MSE of the variance parameter estimates in the correction term estimator resulted in a more accurate correction term estimator than using the asymptotic MSE over two-thirds of the parameter space. However, if variance parameter estimates are left untruncated, as in ANOVA variance component estimates, the asymptotic MSE provided a more accurate correction term estimator over the entire parameter space.

The importance of the parameterization of the model is apparent in the bias correction term estimator since it is not transform invariant. Under a linear parameterization, $\text{BIAS}^{KR}(\theta)$ slightly overestimates the true bias correction term; while under a nonlinear parameterization $\text{BIAS}^{KR}(\theta)$ significantly underestimates the true bias correction term for a large portion of the parameter space when using REML variance component estimates. Under ANOVA variance parameter estimation, the linear parameterization provides a significantly more accurate bias correction term estimator. Even with this simple model, we see how the interaction of a few
factors complicates the results. The question of which MSEP estimator is best depends on the parameterization, the true value of the variance parameter, and the choice of variance parameter estimation method. Each of these factors deserves investigation in more complex models.

These same factors impact the coverage rates of prediction intervals for a realized value of the random effect. We introduced a modification of the Fai-Cornelius methods for approximating the degrees of freedom that includes a bias correction term in the MSEP estimator. The results of a Monte Carlo simulation study showed that the Fai-Cornelius based methods compare favorably with the Giesbrecht-Burns and Kenward-Roger methods currently available in SAS®. The simulation study also demonstrated a shortcoming of the Kenward-Roger method in that the degrees of freedom are not adjusted for the increased MSEP estimate for a single dimension linear combination. In fact, of the four prediction interval procedures included in the simulation study, the Kenward-Roger method consistently performed the poorest, producing the most inflated coverage rates. Because the modified Fai-Cornelius method adjusts both the MSEP and the distribution for the variance component estimates, this method deserves more attention and may significantly improve the coverage rates or Type I error for hypothesis testing in many situations. Future work includes further investigation of these methods in more complex settings.

The impact of negative variance component estimates on the EBLUP and the MSEP estimators was studied. We show that the EBLUP is no longer a reliable or valid estimator of the random effect when negative variance component estimates are present. The proposed solution by Smith and Murray (1984) of regarding the variance parameter as a covariance is shown to be invalid for EBLUP procedures. An alternative solution of a random effects model with correlated errors was presented. Further study of this proposed method is indicated. Future
work also includes the investigation of the impact of negative variance parameter estimates on fixed effects estimators, their standard errors, and distribution estimation techniques.
APPENDIX
DETAILS FOR MONTE CARLO SIMULATION STUDY ON PREDICTION INTERVALS

The calculations necessary for the $FC(\Phi)$ and $FC(\theta)$ prediction intervals are detailed below, followed by an example of the SAS® code used for the Monte Carlo simulation study to determine coverage rates for the four prediction interval procedures in Section 4.5.

Recall from Equation (4-64) that the inverse of the observed information matrix, $B(\bullet) = I^{-1}(\bullet)$, and the vectors of derivatives of the MSEP, $h(\bullet)$, are necessary for the “Satterthwaite”-method of calculating degrees of freedom. First we look at the $FC(\Phi)$ method.

The inverse of the observed information matrix for the $\Phi = (\sigma^2, \sigma_a^2)$ is

$$
B(\Phi) = I^{-1}(\Phi) = \begin{bmatrix}
\frac{2(\sigma^2)^2}{k(n-1)} & -2(\sigma^2)^2 \\
-2(\sigma^2)^2 & \frac{k(n-1)}{kn(n-1)}
\end{bmatrix}.
$$

(A-1)

We also need $h(\Phi) = \begin{bmatrix}
\frac{\partial \left\{ M_1(\Phi) + 2tr\left[ A(\Phi)B(\Phi) \right] \right\}}{\partial \sigma^2} \\
\frac{\partial \left\{ M_1(\Phi) + 2tr\left[ A(\Phi)B(\Phi) \right] \right\}}{\partial \sigma_a^2}
\end{bmatrix}'$

where, due to translation invariance, $M_1(\Phi)$ is as in Equation (4-3) and $tr\left[ A(\Phi)B(\Phi) \right]$ is as in Equation (4-15). Thus,

$$
\frac{\partial \left\{ M_1(\Phi) + 2tr\left[ A(\Phi)B(\Phi) \right] \right\}}{\partial \sigma^2} = \frac{(1-\nu)[1+(k-1)\nu]}{nk\nu}
$$

(A-2)

and
\[
\frac{\partial \left[ M_1(\Phi) + 2tr\left[ A(\Phi)B(\Phi) \right] \right]}{\partial \sigma^2_a} = \frac{\sigma^2 \left( 2n\sigma^2_a + k\sigma^2 \right) + n^2 \left( \sigma^2 \right)^2}{k \left( \sigma^2 + n\sigma^2_a \right)^2} - \frac{2n(kn-1)\left( \sigma^2 \right)^2}{k^2n(n-1)\left( \sigma^2 + n\sigma^2_a \right)^2}.
\]

(A-3)

The degrees of freedom for the \( FC(\Phi) \) method are calculated as

\[
df_{FC(\Phi)} = \frac{2 \left[ M_1(\Phi) + 2tr\left[ A(\Phi)B(\Phi) \right] \right]^2}{\text{var} \left[ M_1(\hat{\Phi}) + 2tr\left[ A(\hat{\Phi})B(\hat{\Phi}) \right] \right]}
\]

where

\[
\text{var} \left[ M_1(\hat{\Phi}) + 2tr\left[ A(\hat{\Phi})B(\hat{\Phi}) \right] \right] \approx h(\Phi)' I^{-1}(\Phi) h(\Phi)
\]

with the elements of \( h(\Phi) \) from Equations (A-2) and (A-3) and \( I^{-1}(\Phi) \) from Equation (A-1).

For the \( FC(\theta) \) method, we turn to the \( \theta = (\sigma^2, \nu) \) parameterization. The inverse of the observed information matrix is now

\[
B(\theta) = I^{-1}(\theta)
\]

\[
= \begin{bmatrix}
2(\sigma^2)^2 & 2(\sigma^2)\nu \\
k(n-1) & k(n-1) \\
2(\sigma^2)\nu & 2\nu^2(kn-1) \\
k(n-1) & k(k-1)(n-1)
\end{bmatrix}.
\]

(A-4)

We use the MSEP estimator utilizing both the correction term and the bias correction term elements to be consistent with the Kenward-Roger and the \( FC(\Phi) \) methods. Because the bias correction term estimator is not transform invariant, the MSEP estimate is different than the one used in the \( \Phi \) parameterization. The MSEP estimator in the \( \theta = (\sigma^2, \nu) \) parameterization is given in Equation (4-60). Thus
where

\[
\frac{\partial}{\partial \sigma^2} \left\{ M_i(\theta) + tr \left[ A(\theta) B(\theta) \right] - \frac{1}{2} tr \left[ A(\theta) B(\theta) \right] \right\} = (1 - \nu) \left[ 1 + (k-1) \nu \right]
\]

\[
\frac{\partial}{\partial \nu} \left\{ M_i(\theta) + tr \left[ A(\theta) B(\theta) \right] - \frac{1}{2} tr \left[ A(\theta) B(\theta) \right] \right\} = 2 \sigma^2 \nu \left[ 1 - k \nu^2 \right]
\]

and

\[
\frac{\partial}{\partial \sigma^2} \left\{ M_i(\theta) + tr \left[ A(\theta) B(\theta) \right] - \frac{1}{2} tr \left[ A(\theta) B(\theta) \right] \right\} = \frac{nk\nu}{k^2(n-1)} - 4 \frac{k(n-1) + (k-1)^2 \nu^2}{nk^2(n-1)(k-1)\nu^2}
\]

\[
\frac{\partial}{\partial \nu} \left\{ M_i(\theta) + tr \left[ A(\theta) B(\theta) \right] - \frac{1}{2} tr \left[ A(\theta) B(\theta) \right] \right\} = \frac{2 \sigma^2 (k-1)^2 \nu - k(n-1) + (k-1)^2 \nu^2}{nk^2(n-1)(k-1)\nu^2}
\]

The degrees of freedom for the \( FC(\theta) \) method are calculated as

\[
df_{FC(\theta)} = \frac{2 \left[ M_i(\hat{\theta}) + tr \left[ A(\hat{\theta}) B(\hat{\theta}) \right] - \frac{1}{2} tr \left[ A(\hat{\theta}) B(\hat{\theta}) \right] \right]^2}{\text{var} \left[ M_i(\hat{\theta}) + tr \left[ A(\hat{\theta}) B(\hat{\theta}) \right] - \frac{1}{2} tr \left[ A(\hat{\theta}) B(\hat{\theta}) \right] \right]}
\]

where

\[
\text{var} \left[ M_i(\hat{\theta}) + tr \left[ A(\hat{\theta}) B(\hat{\theta}) \right] - \frac{1}{2} tr \left[ A(\hat{\theta}) B(\hat{\theta}) \right] \right] \approx h(\hat{\theta})^T B(\hat{\theta}) h(\hat{\theta})
\]

with the elements of \( h(\Phi) \) from Equations (A-5) and (A-6), and \( \text{I}'(\Phi) \) from Equation (A-4).
An example of the SAS® code used for the Monte Carlo simulation study to determine coverage rates for the four prediction interval procedures, utilizing the above derivations for the $FC(\Phi)$ and $FC(\theta)$ methods, follows.

/*Sample program for Monte Carlo simulation study of prediction intervals*/

/*Generating Data for random effect and error from normal distributions and creating $y_{ij}$, observations*/

/*k=3,n=6,sigma(a)_squared=10*/

data normal;
  do mc = 1 to 10000;
    do sigma_a = (10)**.5 ;
      do i = 1 to 3 ;
        stda = rannor(17) ;
        do j = 1 to 6 ;
          e_ij = rannor (18) ;
          a = sigma_a*stda ;
          y_ij = a + e_ij ;
          output ;
        end ;
      end ;
    end ;
  end ;
run ;

/*Creates Satterthwaite prediction limits on $a_1$*/

proc mixed data = normal method=REML noclprint noinfo noitprint;
  class i ;
  model y_ij = / ddfm=Satterth ;
  random i ;
  estimate 'effect 1' | i 1 0 0 /cl;
  by mc ;
  ods listing exclude estimates ;
  ods listing exclude fitstatistics ;
  ods listing exclude covparms ;
  ods output estimates=est1 ;
run ;

/*Creates Kenward-Roger prediction limits on $a_1$*/

/*Outputs Kenward-Roger Standard Error for use in FC(\Phi) prediction intervals*/
/*Outputs Covariance Parameter Estimates to use in FC(\theta) calculations*/

proc mixed data = normal method=REML noclprint noinfo noitprint;
  class i ;
  model y_ij = / ddfm=kenwardroger ;
  run ;
random i ;
estimate 'effect 1' | i 1 0 0 /cl;
by mc ;
ods listing exclude estimates ;
ods listing exclude fitstatistics ;
ods listing exclude covparms ;
ods output estimates=est2 ;
ods output covparms=cp ;
run ;
data est1 ;
set est1 ;
drop Label ;
rename tValue=Satt_tval ;
rename DF=Satt_df ;
rename Probt = Satt_pval ;
rename StdErr=seSATT ;
rename upper = satt_uppercl ;
rename lower = satt_lowercl ;
run ;
data est2 ;
set est2 ;
drop Label estimate;
rename StdErr=seKR ;
rename tValue=KR_tval ;
rename DF=KR_df ;
rename Probt = KR_pval ;
rename upper = KR_uppercl ;
rename lower = KR_lowercl ;
run ;
proc sort data=cp ;
by covparm ;
data siga resid out ;
set cp ;
if covparm='i' then output siga ;
else if covparm='Residual' then output resid ;
run ;
data siga ;
set siga ;
siga = estimate ;
drop covparm estimate ;
run ;
data resid ;
set resid ;
resid = estimate ;
drop covparm estimate ;
run ;
/*Isolates a1 for each data set*/
data a ;
set normal ;
data est;
  merge est1 est2 siga resid;
  by mc;
  k=3;
  n=6;
  upshat = resid/(resid + n*siga);
  m1 = resid*(1-upshat)*(1+((k-1)*upshat))/(n*k*upshat);
/*elements of inverse information matrix for $\Phi$/
  b11 = 2*resid*resid/(k*(n-1));
  bsigsiga = -2*resid*resid/(k*n*(n-1));
  bsiga = 2*resid*resid*(1/(n*n)) *((1/(upshat*upshat*(k-1)))+(1/(k*(n-1))));
/*derivatives of $M_1$, $\Phi$ parameterization*/
  d1 = n*(k-1)*siga*siga/(k*(resid+n*siga)*(resid+n*siga));
  d2 = (resid*2*n*siga + k*resid + n*n*siga*siga)/(k*(resid+n*siga)*(resid+n*siga));
/*derivatives of $2^*trab$, $\Phi$ parameterization*/
  d1_FC2 = d1 + 2*((2*(k*n-1)*resid*resid + 4*(k*n-1)*n*resid*siga)/(k*k*n*(n-1)*(resid+n*siga)*(resid+n*siga)));
  d2_FC2 = d2 - 2*(2*n*(k*n-1)*upshat*upshat*upshat*upshat)/(k*k*n*(n-1));
/*$df_{FC2}$ = $FC(\Phi)$ degrees of freedom*/
  dbdFC2 = d1_FC2*(d1_FC2*b11 + d2_FC2*bsigsiga) + d2_FC2*(d1_FC2*bsigsiga + d2_FC2*bsiga);
  dfFC2 = 2*(seKR**2)*(seKR**2)/dbdFC2;
/*critical t-value for alpha=.05, $FC(\Phi)$ degrees of freedom*/
  tFC025 = quantile('T',.975,dfFC2);
/*$FC(\Phi)$ prediction limits*/
  FC_phi_uppercl = estimate +tFC025*seKR;
  FC_Phi_lowercl = estimate - tFC025*seKR;
/*MSEP estimate for $\theta$ parameterization*/
  msekrbias = m1 + 2*resid*upshat*(k*n-1)/(k*k*n*(n-1)) - 2*resid*((k*(n-1))-(k-1)*upshat*upshat*)/(n*k*k*(n-1)*(k-1)*upshat);
  sqmsekrbias = msekrbias**(1/2);
/*elements of inverse information matrix for $\theta$ (along with b11)*/
  b12 = 2*resid*upshat/(k*(n-1));
  b22 = 2*upshat*upshat*(k*n - 1)/(k*(n-1)*(k-1));
/*derivatives of $M_1$, $\theta$ parameterization*/
\[
\text{d1satt} = (1 - \text{upshat}) \times (1 + (k-1) \times \text{upshat}) / (n \times k \times \text{upshat}) ; \\
\text{d2satt} = \text{resid} \times ((1-k) \times \text{upshat} \times \text{upshat} -1) / (n \times k \times \text{upshat} \times \text{upshat}) ;
\]

/* derivative of trab + bias term, \( \theta \) parameterization */

\[
\text{d1_FC3} = \text{d1satt} + 2 \times \text{upshat} \times (k \times n - 1) / (k \times k \times n \times (n-1)) - (2 \times k \times (n-1) - 2 \times (k-1) \times (k-1) \times \text{upshat} \times \text{upshat}) / (n \times k \times k \times (n-1) \times (k-1) \times \text{upshat} \times \text{upshat}) ; \\
\text{d2_FC3} = \text{d2satt} + 2 \times \text{resid} \times (k \times n - 1) / (k \times k \times n \times (n-1)) \\
+ 2 \times \text{resid} \times (2 \times (k-1) \times (k-1) \times \text{upshat} \times \text{upshat} + k \times (n-1) - (k-1) \times (k-1) \times \text{upshat} \times \text{upshat}) / (n \times k \times k \times (n-1) \times (k-1) \times \text{upshat} \times \text{upshat}) ;
\]

/* dfFC3 = FC(\( \theta \)) degrees of freedom */

\[
\text{dfbFC3} = \text{d1_FC3} \times (\text{d1_FC3} \times \text{b11} + \text{d2_FC3} \times \text{b12}) + \text{d2_FC3} \times (\text{d1_FC3} \times \text{b12} + \text{d2_FC3} \times \text{b22}) ; \\
\text{dfFC3} = 2 \times \text{msekrbias} \times \text{msekrbias} / \text{dfbFC3} ;
\]

/* critical t-value, alpha = .05, FC(\( \theta \)) degrees of freedom */

\[
\text{tFC025bias} = \text{quantile('T',.975,dfFC3)} ;
\]

/* FC(\( \theta \)) prediction limits */

\[
\text{FC_theta_uppercl} = \text{estimate} + \text{tFC025bias} \times \text{sqmsekrbias} ; \\
\text{FC_theta_lowercl} = \text{estimate} - \text{tFC025bias} \times \text{sqmsekrbias} ;
\]

/* Flags each prediction interval containing a1 */

\[
\text{if satt_uppercl} = . \text{or satt_lowercl} = . \text{then satt_cover} = . ; \\
\text{else if satt_uppercl} > a \text{ and satt_lowercl} < a \text{ then satt_cover} = 1 ; \\
\text{else satt_cover}=0 ;
\]

\[
\text{if KR_uppercl} = . \text{or KR_lowercl} = . \text{then KR_cover} = . ; \\
\text{else if KR_uppercl} > a \text{ and KR_lowercl} < a \text{ then KR_cover} = 1 ; \\
\text{else KR_cover}=0 ;
\]

\[
\text{if FC_phi_uppercl} = . \text{or FC_phi_lowercl} = . \text{then FC_phi_cover} = . ; \\
\text{else if FC_phi_uppercl} > a \text{ and FC_phi_lowercl} < a \text{ then FC_phi_cover} = 1 ; \\
\text{else FC_phi_cover}=0 ;
\]

\[
\text{if FC_theta_uppercl} = . \text{or FC_theta_lowercl} = . \text{then FC_theta_cover} = . ; \\
\text{else if FC_theta_uppercl} > a \text{ and FC_theta_lowercl} < a \text{ then FC_theta_cover} = 1 ; \\
\text{else FC_theta_cover}=0 ;
\]

run ;

/* Calculates True Coverage Rates for Each Prediction Interval Method */

proc means data = est n mean stderr lclm uclm;
   var satt_cover KR_cover FC_phi_cover FC_theta_cover ;
   title 'Sigmasq_a=10 k=3 n=6' ;
   title2 'Simulated true coverage rates' ;
   title3 'BLUP a1' ;
run ;
LIST OF REFERENCES


BIOGRAPHICAL SKETCH

Jamie McClave Baldwin is a native of Gainesville, Florida. She is the daughter of Dr. James T. McClave and Mary Jay McClave, both of whom hold graduate degrees from the University of Florida. She is married to Ian Baldwin and mother to Caryss and Zach Baldwin. She will deliver their third child in January, 2008. She is a graduate of Gainesville High School, Vanderbilt University (B. A., mathematics and economics, magna cum laude, 1997) in Nashville, Tennessee, and the University of Florida (M. Stat., statistics, 1999; Ph.D., December 2007).

Jamie began her career in statistics while still in high school, as a data entry assistant at Info Tech, Inc. She has held several internships with Info Tech, Inc., allowing her to gain understanding of each step of the statistical analysis process. While at U. F., she worked as a graduate teaching assistant in the Department of Statistics and as a graduate research assistant in IFAS, Statistics. After graduation, Jamie will remain in Gainesville as a Statistical Consultant at Info Tech, Inc.