DISSERTATION

EVALUATION OF THE METHOD R PROCEDURE FOR ONE-WAY RANDOM EFFECTS MODELS

Submitted by

Sean Mahabir

Department of Statistics

In partial fulfillment of the requirements

for the degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall, 2003

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 3114685

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

UMI Microform 3114685

Copyright 2004 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

COLORADO STATE UNIVERSITY

August 20, 2003

WE HEREBY RECOMMEND THAT THE DISSERTATION EVALUATION OF THE METHOD R PROCEDURE FOR ONE-WAY RANDOM EFFECTS MODELS PREPARED UNDER OUR SUPERVISION BY SEAN MAHABIR BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

David C Bowden
Committee Member

Committee Member

Hariharan Iyer
Co-adviser

Phillip L. Chapman
Adviser

Richard A. Davis
Department Head

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT OF DISSERTATION

EVALUATION OF THE METHOD R PROCEDURE FOR ONE-WAY RANDOM EFFECTS MODELS

Method R (MR) is procedure (Reverter et al., 1994) for variance component estimation. It is best applied in situations where the computation of the preferred methods of ML and REML is infeasible or impossible to implement. Computationally, MR requires only the empirical best linear unbiased predictors (EBLUP's) for the random effects from the mixed model equations (MME). A single MR estimate for a variance component is the value for which the slope of the linear regression of the random effects EBLUP's from all the data, on the random effects EBLUP's from a sub-sample of the data, equals its expected value. Typically, the median of MR values obtained from repeated sub-sampling of the data is used.

To date, properties of the MR estimator are still poorly understood. Our investigation of balanced and unbalanced one-way random effects models reveals MR estimators to be conditional REML estimates based on the whole and sub-sample means in the balanced case, but not for the unbalanced case. Simulations of MR for the one-way balanced random effects model demonstrate the robustness of the median estimator of MR estimates for multiple sub-samples. The large sample variance of the MR estimator for the balanced one-way random effects model is decomposed into two components, the true MR variance and variance due to sub-sampling. From the derived expressions, it is shown that for the intra-class correlation coefficient, the asymptotic efficiency MR relative to the MLE is 1. Our large sample results also

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

demonstrate that the sub-sampling variance for 50% sub-sampling, is not a suitable estimator for the true MR variance in general, contrary to the assumption used in Mallinckrodt et.al. (1997).

The Bivariate MR procedure proposed by Reverter (1994) is shown to be equivalent to performing a multivariate regression of EBLUP's from all the data on EBLUP's from a sub-sample of the data. For the bivariate one-way random effects model, the MR procedure provides a conditional REML estimator for matrix of regression coefficients.

Sean Mahabir
Department of Statistics
Colorado State University
Fort Collins, Colorado 80523
Fall, 2003

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CONTENTS

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF FIGURES

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF TABLES

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 1

BACKGROUND

1.1 Introduction

Understanding the sources of variability in data is one of the defining elements of the Science of Statistics. The class of linear models encompasses a very broad collection of statistical models in which variance component estimation is an important concern. In this dissertation, our focus will be the sub-class of mixed linear models. In Section 1.2 we present a very brief history of variance component estimation leading up to likelihood based estimation, which has in recent time, become the preferred method of estimation for variance component models. This directly coincides with the advent of high speed computers that facilitate the ability to perform the often difficult and complex calculations involved in obtaining likelihood based estimates.

Despite the considerable progress made in computing likelihood based variance component estimates, there exist applications, in particular for unbalanced data, where the number of observations and/or the number variance components become so large that the use of conventional algorithms becomes infeasible. This situation is not an uncommon occurrence in animal breeding, where mixed linear models are often used for genetic prediction models for commercial livestock, where records are of the order of several millions. To appreciate the effort involved in computing likelihood based estimates, we review a few of the more popular algorithms used. In Section 1.3 the formulas required for likelihood based estimation are presented. Section

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.4 summarizes algorithms requiring derivatives and Section 1.5 summarizes derivative free algorithms. Section 1.6 addresses the relative attributes of the algorithms presented and introduces the **Method R** (MR) variance component procedure (Reverter (1994b)) as a procedure that can be used to obtain variance components when conventional algorithms may fail. What MR is, and what it's properties are, will be primary focus of the chapters to follow.

1.2 An Abbreviated History of Variance Component Estimation

Early Work, Pre-1900

Good accounts of the early history of variance components can be found in Scheffé (1956) and later in Anderson(1978,1979). These reviews trace fundamental concepts in variance component estimation to Astronomy. In a discussion by Plackett (1972), the method of least squares can be attributed independently to Legendre (1806) and Gauss (1809). According to Scheffé (1956), the first evidence of a mathematical model using variance components was by the astronomer Airy (1861), who formulated a one-way random effects model. Chauvenet (1863) also used a one-way random effects model, and derived the variance for the overall sample mean for the balanced model.

1900-Present

The introduction of the terms "variance" and "analysis of variance" was due to Fisher (1918). Kempthorne (1977) refers to Fisher's 1918 paper as "the basic and seminal paper in the theory of quantitative genetics" Fisher's paper implicitly employs variance component models and defines ratios for total variance in a trait, relative to variance attribute to constituent causes (Anderson (1978)).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In his book *Statistical Methods for Research Workers*, Fisher (1925a) lays the foundation for variance component theory and several subsequent methods of estimation. The principle of estimation arose through the demonstration of using analysis of variance to obtain an estimate of the intraclass correlation coefficient for the one-way balanced random model. The procedure equated sums of squares from the analysis of variance to their expected values, from which a set of equations in terms of the unknown parameters, the variance components, was constructed. Fisher's book extended the analysis of variance from the one-way model with unbalanced data to the two-way crossed model; however, estimation of variance components was not addressed.

The balanced and unbalanced one-way ANOVA was made clearer by Tippett (1931). Urquhart, Weeks and Henderson (1973) note that Fisher did not use linear models to explain analysis of variance. Tippett, however, did use linear models to explain the concept of analysis of variance. Tippett (1931) also dealt with the problem of optimal sampling designs for the one-way random model. Yates & Zacopanay (1935) made significant contributions for higher order sampling designs in their designs in cereal experiments. Neyman, Iwaszkiewicz & Kolodziejczyk (1935) examined the comparative efficiency of randomized block designs and Latin square designs. They made extensive use of linear models, including mixed linear models.

Daniels (1939) appears to have been the first to use the term "components of variance". Daniels (1939) and Winsor and Clarke (1940) both use an operator for expectation, whereas Fisher (1925a) uses expectation, but he does not use a specially defined operator for it.

Strategies for dealing with unbalanced data did not appear until Cochran (1939) and Winsor and Clarke (1940). Winsor and Clarke specifically addressed the problem

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of estimation of variance components for the one-way model with unbalanced data, while Cochran's paper was directed at the usefulness of variance components in selecting optimal designs.

Ganguli (1941) and Crump (1946,1947) extended the Winsor-Clarke method to higher order designs. Both noted the deficiency in analysis of variance method of producing negative estimates and suggested that such estimates be replaced by zero.

Also of note during this period was the work by Satterthwaite (1940) who obtained approximate sampling distributions of variance component estimates, and Wald (1940,1941), who addressed interval estimation for ratios of variance components for the 1-way and 2-way analysis of variance models with unbalanced data.

The 1950's brought in the first books to contain in in depth treatment of the subject of variance components. First in Anderson and Bancroft (1952), then later on in Bennett and Franklin (1954).

Henderson (1953) dealt with the difficult problem of using unbalanced data to estimate variance components. The paper presents three ways of using unbalanced data from random or mixed models. All three methods are adaptations of the ANOVA method of equating sums of squares to their expected values. For unbalanced data, Method I uses sums of squares of data that are analogues of those used with balanced data. It is applicable only to random effects models; Method II adjusts the data for the fixed effects in the model, and then uses Method I on the adjusted data. It can be used for mixed models, but does not allow for interactions between fixed and random effects; Method III is based on sums of squares resulting from fitting the linear model and it's sub-models. All three methods yield unbiased estimators,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and in the balanced case reduce the ANOVA method. Henderson's Methods I,II and III have all been used extensively in a broad variety of applications.

For balanced data, Graybill (1954) studied the sampling variances of ANOVA estimates of variance components for k-fold nested and random effects models. He showed that in the class of quadratic functions of the data that are unbiased estimators of the variance components, ANOVA estimators have minimum variance and are thus *best quadratic unbiased estimators* (BQUE).With the added assumption of normality, Graybill and Wortham showed that for balanced random effects models, ANOVA estimators of the variance components are also uniformly *best unbiased* (BUE). Graybill and Hultquist (1961) extended Graybill (1954) for all balanced random effect models, without requiring distributional assumptions. Albert (1976) showed that these properties also hold for the class of balanced mixed linear models. Thus for balanced data, ANOVA estimators for variance component models are BQUE, and under normality are BUE.

In contrast to balanced data, variance component estimators that are uniformly best do not exist for unbalanced data. This is problem is well documented in Scheffé (1959, Sec. 7.2). Scheffé also states that for unbalanced data, even under normality, the distribution theory becomes much more complicated. He also points to the problem of the a "general lack of uniqueness" of the ANOVA method for unbalanced data. As a result, there is an absence of the ability to determine relative optimality of ANOVA estimators for unbalanced data. The shortcomings of ANOVA estimators for unbalanced data eventually led to the consideration of maximum likelihood estimation.

The method of maximum likelihood was derived by Fisher (1922,1925b), but received little attention until the landmark paper by Hartley & Rao (1967). The reason for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

this neglect is attributed primarily to the complexity of the computations associated with arriving at solutions for the likelihood equations. The earliest use of the method of maximum likelihood for variance components models was Crump (1947,1951), who considered balanced and unbalanced one-way random effects models. He also derived formulae for calculating large sample variances of the maximum likelihood estimators. Herbach (1959) derived explicit formulas for maximum likelihood (ML) estimators for certain balanced data models. MLE's for variance components for several balanced data models are summarized in Corbeile and Searle (1976). Miller (1973,1977) worked on ML estimation for the 2-way random effects model for both the balanced and unbalanced cases, with or without interaction. Miller provided equations for maximum likelihood estimation, for which analytical solutions are unavailable. He also looked at the asymptotic properties of the estimators. Searle (1970) derived an expression for the large-sample dispersion matrix of ML estimators in the general unbalanced case.

W.A. Thompson (1962) proposed the idea of maximizing the part of the likelihood that is invariant to the fixed effects. This procedure is sometimes called marginal likelihood estimation, but is more commonly known as restricted maximum likelihood (REML). The procedure was formally introduced in a broad basis in Patterson and Thompson (1971). As with maximum likelihood, widespread application of REML was not immediate due to difficulty associated with computing solutions.

Townsend (1968), Harville (1969) and Townsend and Searle (1971) made significant contributions toward finding minimum variance quadratic unbiased estimators of variance components. This was followed by the work on minimum variance estimation by LaMotte (1970,1971,1973a,b,1976) and the papers on minimum-norm quadratic unbiased equation (MINQUE) by C.R. Rao (1970,1971,a,b,1972). MINQUE gives estimators that are unbiased, but the minimality property applies

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

only to the *a priori* values used for the variance components. Iterative use of MINQUE, using the successive solutions as *a priori* values until convergence is reached, is called I-MINQUE, whose solution is the same as that of REML (Hocking and Kutner, 1975), and are normally distributed under large sample theory (Brown, 1976). Indeed, the MINQUE solution for the initial set of *a priori* values is equivalent to the first round iterate from REML. For this reason, MINQUE is not often recommended as practical means of estimating variance components. A comprehensive treatment of the subject of MINQUE can be found in the book by Rao and Kleffe (1988).

Hill (1965,1967) dealt with estimation of variance components using Bayesian principles for the balanced 1-way random effects model. Several other similar works are reviewed in Khuri and Sahai (1985 pp. 283-284). Searle, Cassella and McCulloch (1992, Ch.9) also address this topic.

Several papers have been offered that compare the various procedures available for unbalanced data, for example Townsend and Searle (1971), Corbeile and Searle (1976), Swallow and Monahan (1984) and Li and Kotz (1978). In general, the recommended methodology is ML or REML. REML is often preferred for several reasons, one in particular being that it accounts for the "loss in degrees of freedom" attributable to estimating the fixed effects. Publications such as Harville (1977) and Searle, Cassella and McCulloch (1992) have brought increased attention to REML. Also with the introduction of available software, ML and REML are quite often the method of choice for variance component estimation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.3 Likelihood Based Estimation

Likelihood based estimation procedures have come to the forefront of the procedures available for variance component estimation. Maximum likelihood (ML) and Restricted maximum likelihood (REML), once computationally prohibitive, have become fairly routine with modern high speed computers. The optimal properties of these estimators make them an obvious first choice for many researchers, and is the reason this class of estimators will remain the focus of this discussion.

Commonly used algorithms for computation of ML/REML estimates include Newton-Rhaphson (NR), method of scoring (FS), EM-algorithm (EM) (Dempster, Laird and Rubin, 1977) and more recently, Derivative free (DF) methods (Smith and Graser, 1986; Graser, Smith and Tier, 1987) and Average information (AI) methods (Johnson and Thompson, 1995).

In applications where extremely large data sets are involved, matrix inversion necessary for ML/REML computation can often prove to be difficult or infeasible. Many techniques are available for reducing the computation necessary for ML/REML (Harville, 1977; Jenrich and Sampson, 1976; Hemmerle and Hartley, 1973; Lindstrom and Bates, 1988), nearly all rely on exploiting on some aspect of the structure of the matrices to be manipulated in order to speed iterations. Matrix inversion and storage can be particularly costly when dealing large data sets.

1.3.1 The Mixed Linear Model

The mixed linear model is given by

$$\mathbf{y} = \mathbf{X}\gamma + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1.3.1}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where $\mathbf{y}$ is a $N \times 1$ random vector of observable data points; $\mathbf{X}$ is the $N \times p$ design matrix corresponding to the fixed effects; γ is the $p \times 1$ vector of fixed effects parameters; $\mathbf{Z}$ is the $N \times q$ design matrix corresponding to the random effects; $\mathbf{u}$ is the $q \times 1$ vector of random effects parameters, and $\mathbf{e}$ is the $N \times 1$ random vector of unobservable random errors.

To denote that a random vector $\mathbf{y}$ follows a distribution with a given mean and covariance, we will write $\mathbf{y} \sim (.,.)$. The first argument specifies the mean vector and the second argument, the covariance matrix. To denote the mean, or expected value, we will use $E(.)$. To denote variance and covariance, we will use $var(.)$ and $cov(.,.)$, respectively.

For the above mixed model we will assume that $\mathbf{u} \sim (\mathbf{0}, \mathbf{D})$, $\mathbf{e} \sim (\mathbf{0}, \mathbf{R})$ and $cov(\mathbf{u}, \mathbf{e}) = \mathbf{0}$, where $\mathbf{D}$ is the $q \times q$ covariance matrix for $\mathbf{u}$, and $\mathbf{R}$ is the $N \times N$ covariance matrix for the random vector $\mathbf{e}$.

From the above it follows that $E(\mathbf{y}) = \mathbf{X}\gamma$ and $var(\mathbf{y}) = \mathbf{ZDZ}' + \mathbf{R} = \mathbf{V}$. The covariance matrix $\mathbf{V}$ can be considered to be a function of an unobservable parameter vector $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_c)'$, with the ϕ_i's $(i = 1, \ldots, c)$ representing the covariance parameters that define the structure of the dispersion of the random vector $\mathbf{y}$. Under the assumption that $\mathbf{u}$ and $\mathbf{e}$ are multivariate normal, the likelihood function for the mixed linear model is

$$L = (2\pi)^{-\frac{N}{2}} |\mathbf{V}|^{-\frac{1}{2}} exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\gamma)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\gamma)\right]. \quad (1.3.2)$$

The log-likelihood function is

$$l = -\frac{1}{2}\left[Nlog(2\pi) + log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\gamma)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\gamma)\right]. \quad (1.3.3)$$

$\mathbf{V}^{-1}$ can sometimes be difficult to obtain, depending on the size and complexity of the model. In instances where $\mathbf{R}^{-1}$ and $\mathbf{D}^{-1}$ are both easier to compute than

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$\mathbf{V}^{-1}$, the result below, proved by Henderson *et al.* (1959), may be of assistance in computation.

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}' + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\mathbf{R}.$$

1.3.2 First And Second Order Partial Derivatives

The maximum likelihood estimates (MLE's) are values for the unknown parameters that maximize the likelihood function. Since the relationship between the likelihood function and the log-likelihood function is monotonic, the MLE's can be derived by maximizing the log-likelihood function. To this end, we will need the first and second order derivatives of the log-likelihood function (Searle, Cassella & McCulloch(1992), Sec. 6.3).

The *first order partial derivatives* for the vector of fixed effects and covariance parameters are

$$\frac{\partial l}{\partial \gamma} = \mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\gamma) \tag{1.3.4}$$

$$\frac{\partial l}{\partial \phi_i} = -\frac{1}{2}\left[tr(\mathbf{V}^{-1}\mathbf{V}_i) - (\mathbf{y} - \mathbf{X}\gamma)'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\gamma)\right], \quad i = 1, \ldots, c \tag{1.3.5}$$

where $\mathbf{V}_i = \frac{\partial \mathbf{V}}{\partial \phi_i}$.

The *second order partial derivatives* for the vector of fixed effects and covariance parameters are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\frac{\partial^2 l}{\partial \gamma^2} = -\mathbf{X}\mathbf{V}^{-1}\mathbf{X} \tag{1.3.6}$$

$$\frac{\partial^2 l}{\partial \gamma \partial \phi_i} = -\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\gamma) \tag{1.3.7}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \phi_i \partial \phi_j} &= \omega_{ij}/2 - (\mathbf{y} - \mathbf{X}\gamma)'(\mathbf{V}^{-1}\mathbf{V}_j\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1} \\ &\quad + \mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j\mathbf{V}^{-1})(\mathbf{y} - \mathbf{X}\gamma)/2, \\ &\quad i, j = 1, \ldots, c \end{aligned} \tag{1.3.8}$$

where $\omega_{ij} = tr(\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j)$.

The elements of the *information matrix* can then be obtained by taking expectation of the negative second order partial derivatives,

$$E\left(-\frac{\partial^2 l}{\partial \gamma^2}\right) = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \tag{1.3.9}$$

$$E\left(-\frac{\partial^2 l}{\partial \gamma \partial \phi_i}\right) = 0, \quad i = 1, \ldots, c \tag{1.3.10}$$

$$E\left(-\frac{\partial^2 l}{\partial \phi_i \partial \phi_j}\right) = \omega_{ij}/2, \quad i, j = 1, \ldots, c. \tag{1.3.11}$$

Except in certain special cases, solving for MLE's will typically require numerical computation of the first and second order partial derivatives. If asymptotic variance estimates are required, the elements of the information matrix will also be needed. All of these require computing and storing $\mathbf{V}^{-1}$ and $\mathbf{V}_i$'s at each iteration. Thus, depending on the size and complexity of the problem, ML estimation can be computationally expensive.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.3.3 Restricted Maximum Likelihood Estimation

Assume that the $rank(\mathbf{X}) = r$. Let $\mathbf{K}$ be a non-unique $(N-r) \times N$ matrix whose rows form an orthonormal basis for the orthogonal complement of the column space of $\mathbf{X}$. Then $\mathbf{KX} = \mathbf{0}$. Also, let $\mathbf{X}_r$ be a non-unique $N \times r$ matrix, whose columns form a basis for the column space of $\mathbf{X}$. Define the transform $\mathbf{z}$ as

$$\mathbf{z} = \begin{bmatrix} \mathbf{K} \\ \mathbf{X}_r'\mathbf{V}^{-1} \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{Ky} \\ \mathbf{X}_r'\mathbf{V}^{-1}\mathbf{y} \end{bmatrix},$$

with distribution

$$\mathbf{z} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{X}_r'\mathbf{V}^{-1}\mathbf{X}\gamma \end{pmatrix}, \begin{pmatrix} \mathbf{KVK}' & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_r'\mathbf{V}^{-1}\mathbf{X}_r \end{pmatrix}\right]$$

The random variables $\mathbf{Ky}$ and $\mathbf{X}_r'\mathbf{V}^{-1}\mathbf{y}$ are independent random random vectors. Their respective log-likelihood functions l_1 and l_2 are as follows,

$$l_1 = -\frac{1}{2}\left[(N-r)log(2\pi) - log|\mathbf{KVK}'| - \mathbf{y}'\mathbf{K}'(\mathbf{KVK}')^{-1}\mathbf{Ky}\right] \qquad (1.3.12)$$

and

$$l_2 = -\frac{1}{2}\Big[(N-r)log(2\pi) - log|\mathbf{X}_r'\mathbf{V}^{-1}\mathbf{X}_r'| - (\mathbf{y}-\mathbf{X}\gamma)'\mathbf{V}^{-1}\mathbf{X}_r(\mathbf{X}_r\mathbf{V}^{-1}\mathbf{X}_r)^{-1}\mathbf{X}_r'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\gamma)\Big]. \qquad (1.3.13)$$

The likelihood of $\mathbf{X}_r'\mathbf{V}^{-1}\mathbf{y}$ depends on both the fixed effects parameters and on the covariance parameters. The likelihood of $\mathbf{Ky}$ depends solely on the covariance parameters, and sometimes referred to as the REML log-likelihood function. The values of the covariance parameters that maximize the REML log-likelihood function are the REML estimates of the covariance parameters. When REML estimates of the covariance parameters are evaluated, REML estimates for the fixed effects are determined from the likelihood function of $\mathbf{X}_r'\mathbf{V}^{-1}\mathbf{y}$, treating the estimated covariance parameters as known quantities using their REML estimates.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Now $\mathbf{KK}' = \mathbf{I}_{N-r}$ and $\mathbf{K}'(\mathbf{KK}')^{-1}\mathbf{K} = \mathbf{K}'\mathbf{K} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$, the projection matrix on the column space of the orthogonal complement of $\mathbf{X}$. From this, it is easy to show that

$$\mathbf{K}'(\mathbf{KVK}')^{-1}\mathbf{K} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}. \tag{1.3.14}$$

Using the following identity from Searle (1979).

$$log|\mathbf{KVK}'| = log|\mathbf{V}| + log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|, \tag{1.3.15}$$

We are now able to write the REML log-likelihood as a function independent of $\mathbf{K}$, denoted by l_R, as follows.

$$l_R = -\frac{1}{2}\left[(N-r)log(2\pi) + log|\mathbf{V}| + log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{P}\mathbf{y}\right], \tag{1.3.16}$$

where $\mathbf{P} = \mathbf{K}'(\mathbf{KVK}')^{-1}\mathbf{K}$.

The derivations of the necessary first and second order partial derivatives given below, can be found in Searle, Cassella & McCulloch (1992), Sec. 6.6. The *first order partial derivative* for the i^{th} covariance parameter is

$$\frac{\partial l_R}{\partial \phi_i} = -\frac{1}{2}\left(tr(\mathbf{V}_i\mathbf{P}) - \mathbf{y}'\mathbf{P}\mathbf{V}_i\mathbf{P}\mathbf{y}\right), \quad i = 1, \ldots, c. \tag{1.3.17}$$

The *second order partial derivative* involving the i^{th} and j^{th} covariance parameters is

$$\frac{\partial^2 l_R}{\partial \phi_i \partial \phi_j} = \frac{1}{2}tr(\mathbf{V}_i\mathbf{P}\mathbf{V}_j\mathbf{P}) - \mathbf{y}'\mathbf{P}\mathbf{V}_i\mathbf{P}\mathbf{V}_j\mathbf{P}\mathbf{y}, \quad i, j = 1, \ldots, c. \tag{1.3.18}$$

The corresponding element of the *information matrix* is then

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$E\left(-\frac{\partial^2 l_R}{\partial\phi_i\partial\phi_j}\right) = -\frac{1}{2}tr(\mathbf{V}_i\mathbf{P}\mathbf{V}_j\mathbf{P}), \quad i,j=1,\ldots,c. \tag{1.3.19}$$

As with maximum likelihood, REML estimates must typically be evaluated through iterative computation. At each iteration $\mathbf{V}^{-1}$ and $\mathbf{V}_i$'s must be computed and stored. Once again, depending on the size and complexity of the problem, computation could be a formidable task.

1.3.4 Henderson's Mixed Model Equations

Henderson *et al.* (1959) developed a set of equations called the mixed model equations (MME), for which the solutions give simultaneously a best linear unbiased estimator (BLUE) for $\mathbf{X}\gamma$ and a best linear unbiased predictor (BLUP) for $\mathbf{u}$. The mixed model equations were derived by maximizing the joint density of $\mathbf{y}$ and $\mathbf{u}$ for known $\mathbf{D}$ and $\mathbf{R}$. An interesting aspect of the MME is that they can be used to set up iterative procedures for computing ML and REML estimates (Henderson (1973); Graser, Smith and Tier (1987); Gilmour, Thompson and Cullis (1995)).

Henderson's mixed model equations (MME) are given below.

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{D}^{-1} \end{pmatrix}\begin{pmatrix} \gamma \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}. \tag{1.3.20}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.4 Computational Methods for ML/REML: Derivative Based

In this section, we will review a few of the more commonly used algorithms for computing ML/REML estimates that utilize partial derivatives of the likelihood.These methods typically use an iteration strategy of the following form (Searle, Cassella and McCulloch, 1992).

$$\boldsymbol{\theta}^{[i+1]} = \boldsymbol{\theta}^{[i]} + s^{[i]}\mathbf{M}^{[i]}\frac{\partial l}{\partial \theta}|_{\theta=\theta^{[i]}} \tag{1.4.1}$$

Here $\boldsymbol{\theta}^{[i]}$ is the estimate of the parameter vector at the i^{th} iteration. The scalar $s^{[i]}$ the value of step size at the i^{th} iteration. The matrix $\mathbf{M}^{[i]}$ is the modifier matrix at the i^{th} iteration and $\frac{\partial l}{\partial \theta}$ is the vector of first derivatives of the log-likelihood evaluated at $\boldsymbol{\theta}^{[i]}$.

1.4.1 Newton-Raphson

In the Newton-Raphson (NR) algorithm $\mathbf{M}^{[i]}$ is the inverse of the matrix of second order partial derivatives for the log-likelihood. The scalar $s^{[i]}$ the value of step size at the i^{th} iteration is usually set at a value of 1. Thus, effort goes into computing and storing the first and second order derivatives, which requires at each iteration, the computation and storage of $\mathbf{V}^{-1}$, the $\mathbf{V}_i$'s, the partial derivatives of $\mathbf{V}$ with respect to the covariance parameters, and traces of matrices of the form $\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j$.

NR generally converges in a few iterations; however, the solution may not necessarily be a global maximum and it is possible to obtain estimates outside the parameter space. Final estimates may also be affected by poor starting values.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.4.2 Method of Scoring

Method of Scoring (FS) uses the inverse of the information matrix, $\mathbf{I}(\theta^{[\mathbf{i}]})$ for $\mathbf{M}^{[\mathbf{i}]}$ and $s^{[i]} = 1$. As with Newton-Raphson, effort goes into computing and storing first order derivatives and the expected values of the second order derivatives at each iteration, which requires at each iteration, the computation and storage of $\mathbf{V}^{-1}$, the $\mathbf{V}_i$'s, the partial derivatives of $\mathbf{V}$ with respect to the covariance parameters, and traces of matrices of the form $\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j$.

As with Newton-Raphson, method of scoring converges in a few iterations, but is more robust to poor starting values than Newton-Raphson (Jenrich and Sampson (1986)).

1.4.3 Average Information REML

Average Information REML (AIREML) is a more recent procedure that was first proposed by Johnson and Thompson (1995). The algorithm combines observed and expected information to update estimates. $\mathbf{M}^{[\mathbf{i}]}$ is replaced by the inverse of the average information (AI) matrix, which is obtained by taking the average of the matrix of second derivatives (observed information) and the information matrix (expected information).

$$\begin{aligned} AI(\phi_i, \phi_j) &= \frac{1}{2}\left[\frac{\partial^2 l_R}{\partial\phi_i \partial\phi_j} + E\left(-\frac{\partial^2 l_R}{\partial\phi_i \partial\phi_j}\right)\right] \\ &= \mathbf{y}'\mathbf{P}\mathbf{V}_i\mathbf{P}\mathbf{V}_j\mathbf{P}\mathbf{y}, \quad i, j = 1, \ldots, c. \end{aligned} \tag{1.4.2}$$

The steps for AIREML are the following.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

step 0 Decide on starting values for the elements of ϕ.

step 1 Solve the mixed model equations (MME) to get estimates of ϕ and $\boldsymbol{\gamma}$. Store the residuals $\mathbf{e} = \mathbf{Py}$.

step 2 Solve the MME replacing $\mathbf{y}$ with $[\mathbf{V}_i\mathbf{Py},\ i = 1, \ldots, c]$.

step 3 Compute the AI matrix and the elements of the inverse of the coefficient matrix necessary for computing $\frac{\partial l_R}{\partial \phi}$

step 4 Update estimate of ϕ.

step 5 Check for convergence. Repeat steps 1 through 5 until a solution has been reached.

As with Newton-Raphson and Fisher-Scoring algorithms, computation and storage of $\mathbf{V}^{-1}$, the $\mathbf{V}_i$'s, the partial derivatives of $\mathbf{V}$ with respect to the covariance parameters. However, the use of average information matrix eliminates the need for evaluating the traces of matrices of the form $\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j$. The algorithm involves more steps per iteration but like Newton-Raphson and Fisher-Scoring, it generally converges in a few iterations.

1.5 Computational Methods for ML/REML: Non-Derivative Based

In this section, we will review a few of the algorithms for computing ML/REML estimates that do not require the direct computation of derivatives.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.5.1 ML/REML Estimates from Mixed Model Equations

Henderson (1973) gives an algorithm using solutions the MME given below, for setting up an iterative procedure for computing ML and REML estimates.

$$\begin{pmatrix} \mathbf{X'R^{-1}X} & \mathbf{Z'R^{-1}X} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z + D^{-1}} \end{pmatrix} \begin{pmatrix} \gamma \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{pmatrix} \tag{1.5.1}$$

Here $\boldsymbol{\phi}$ represents the vector of variance components parameters. $\mathbf{Z}_i$ is the component of $\mathbf{Z}$ corresponding to the i^{th} random effect, $\mathbf{u}_i$, with q_i levels.

$$\phi_i^{[m+1]} = \left(\hat{\mathbf{u}}_\mathbf{i}^{[\mathbf{m}]'}\hat{\mathbf{u}}_\mathbf{i}^{[\mathbf{m}]} + \phi_i^{[m]} tr(\mathbf{T}_{ii}^*)\right)/q_i, i = 1, \ldots, c-1. \tag{1.5.2}$$

$$\phi_c^{[m+1]} = \mathbf{y}'(\mathbf{y} - \mathbf{X}\hat{\gamma}^{[\mathbf{m}]} - \mathbf{Z}\hat{\mathbf{u}}^{[\mathbf{m}]})/N \tag{1.5.3}$$

where $\mathbf{T}_{ii}^*$ is the lower diagonal element of $(\mathbf{I} + \mathbf{Z'R^{-1}ZD})^{-1}$ that corresponds to $\mathbf{u_i}$. Starting with an initial estimate for parameter vector $\boldsymbol{\phi}^{[0]}$ and iterating until convergence yields the maximum likelihood estimate for $\boldsymbol{\phi}$.

The corresponding equations for REML estimation are

$$\phi_i^{[m+1]} = \left(\hat{\mathbf{u}}_\mathbf{i}^{[\mathbf{m}]'}\hat{\mathbf{u}}_\mathbf{i}^{[\mathbf{m}]} + \phi_i^{[m]} tr(\mathbf{T}_{ii})\right)/q_i, i = 1, \ldots, c-1. \tag{1.5.4}$$

$$\phi_c^{[m+1]} = \mathbf{y}'(\mathbf{y} - \mathbf{X}\hat{\gamma}^{[\mathbf{m}]} - \mathbf{Z}\hat{\mathbf{u}}^{[\mathbf{m}]})/(N-r). \tag{1.5.5}$$

where $\mathbf{T}_{ii}$ is the lower diagonal element of $(\mathbf{I} + \mathbf{Z'SZD})^{-1}$ that corresponds to $\mathbf{u_i}$, and $\mathbf{S} = \mathbf{R^{-1}} - \mathbf{R^{-1}X(X'R^{-1}X)^{-}X'R^{-1}}$. REML estimate of $\boldsymbol{\phi}$ is obtained as above.

For both ML and REML estimation, for each iteration $\mathbf{R}^{-1}$ must be stored and computed as well as the inverse of a $q \times q$ matrix before computing the trace of $\mathbf{T}_{ii}$ or $\mathbf{T}_{ii}^*$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.5.2 EM Algorithm

The *expectation-maximization (EM)* algorithm obtains maximum likelihood estimators by alternating between computing conditional expected values and maximizing the likelihood function.

Again, take ϕ to be a vector of variance components parameters and $\mathbf{Z}_i, \mathbf{u}_i$ and q_i as defined previously. Consider the entire data set to be the observed $\mathbf{y}$ together with the unobservable random vector $\mathbf{u}$. For the augmented data $\begin{bmatrix} \mathbf{y} \\ \mathbf{u} \end{bmatrix}$, the maximum likelihood estimators under normality are

$$\hat{\phi}_i = \mathbf{u}_i'\mathbf{u}_i/q_i, \quad i = 1, \ldots, c. \tag{1.5.6}$$

From which

$$\mathbf{X}\hat{\gamma} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\left(\mathbf{y} - \sum_{i=1}^{c-1}\mathbf{Z}_i\mathbf{u}_i\right) \tag{1.5.7}$$

Starting with initial estimates for the parameters, we first compute the conditional expected values of the sufficient statistics of the complete data, $\mathbf{u}_i'\mathbf{u}_i$, given the incomplete, data $\mathbf{y}$ **(expectation)**. The conditional expected values computed are then used in place of the sufficient statistics to obtain improved estimates of the parameters (representing MLE's for the complete data) **(maximization)**. The new parameter estimates can then be used to compute new conditional expectations. This procedure is repeated until convergence.

For ML the steps for EM algorithm are the following.

step 0 Decide on starting values for the parameters, $\gamma^{[0]}$ and $\phi^{[0]}$. Set m=0.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

step 1 (E-step) compute conditional means

$$\begin{aligned}\hat{t}_i^{[m]} &= E(\mathbf{u_i'u_i}|\mathbf{y})\ |_{\gamma^{[m]},\phi^{[m]}} \\ &= \phi_i^{2[m]}(\mathbf{y}-\mathbf{X}\gamma^{[m]})'((\mathbf{V^{[m]}})^{-1}\mathbf{Z}_i\mathbf{Z}_i'((\mathbf{V^{[m]}})^{-1}(\mathbf{y}-\mathbf{X}\gamma^{[m]})\end{aligned} \quad (1.5.8)$$

$$\begin{aligned}\hat{\mathbf{s}}^{[m]} &= E(\mathbf{y}-\textstyle\sum_{i=1}^{c-1}\mathbf{Z}_i\mathbf{u}_i|\mathbf{y})\ |_{\gamma^{[m]},\phi^{[m]}} \\ &= \mathbf{X}\gamma^{[m]}+\phi_c^{[m]}(\mathbf{V^{[m]}})^{-1}(\mathbf{y}-\mathbf{X}\gamma^{[m]})\end{aligned} \quad (1.5.9)$$

step 2 (M-step) Maximize the likelihood of the complete data.

$$\begin{aligned}\phi_i^{[m+1]} &= \hat{t}_i^{[m]}/q_i, \quad i=1,\ldots,c \\ \mathbf{X}\gamma^{[m+1]} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\hat{\mathbf{s}}^{[m]}\end{aligned} \quad (1.5.10)$$

step 3 If convergence is reached, set $\hat{\boldsymbol{\phi}} = \boldsymbol{\phi}^{[m+1]}$ and $\hat{\gamma} = \gamma^{[m+1]}$. Otherwise increase m by 1 and return to step 1.

EM for REML uses same arguments, in which case the complete data is the vector $\begin{bmatrix}\mathbf{Ky}\\ \mathbf{Ku}\end{bmatrix}$.

For variance components, EM will always generate estimates that are within the parameter space. While the steps for EM are relatively simple, the algorithm is generally slow to converge and effort goes into computing and storing $\mathbf{V}^{-1}$ for each iteration (the E-step).

1.5.3 Derivative Free REML (DFREML)

The derivative free (DF) algorithm described by Graser, Smith and Tier (1987) finds REML estimates of the elements ϕ by doing a direct search over the log-likelihood surface.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

To compute the log-likelihood DFREML uses the following form of the REML log-likelihood equation.

$$l_R = -log|\mathbf{C}| + log|\mathbf{R}| + log|\mathbf{D}| + \mathbf{y}'\mathbf{P}\mathbf{y} \tag{1.5.11}$$

where **C** is the coefficient matrix from the MME.

The difficult parts of the log-likelihood to calculate are $log|\mathbf{C}|$ and $\mathbf{y}'\mathbf{P}\mathbf{y}$. Assembling C requires computing and storing $\mathbf{R}^{-1}$ and $\mathbf{D}^{-1}$ and computing **P** requires computing and sorting $\mathbf{V}^{-1}$. Algorithms implemented for DFREML achieve these computations by solving the following augmented MME.

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{y}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{C} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{y}'\mathbf{R}^{-1}\mathbf{W} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \tag{1.5.12}$$

where $\mathbf{W} = [\mathbf{X}\vdots\mathbf{Z}]$.

Gaussian elimination automatically produces a known multiple of $\mathbf{y}'\mathbf{P}\mathbf{y}$. The sum of the log of the non-zero pivots (using the same ordering in each evaluation) gives $log|\mathbf{C}|$.

DFREML works well for models with small number of covariance parameters. For multiple trait analyses the method may converge slowly.

Boldman and Van Vleck (1991) used subroutines in SPARSPAK to substantially decrease the time to calculate the log-likelihood. SPARSPAK (George et al.,1980; Chu et al.,1984)) is based on Cholesky factorization rather than Gaussian elimination.

DF solution has half the accuracy in significant digits of the likelihood function being maximized (Press, Flannery, Teukolsky and Veterling, 1989). As the the number of variance components in the model increases, the accuracy is likely to decrease.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.6 Summary

For small to moderate sized data sets NR and FS are preferred because of quick convergence and variance estimates are generated. For larger scale problems DFREML has become popular because of its computational feasibility. Gilmour, Thompson and Cullis (1995) report that computational requirements are similar for EM and AI and three times that for DF. They also report that convergence in AI appears to be similar to FS. Convergence for AI is faster than for both EM and DF. This makes for a good case for using AI as a computing strategy for likelihood based estimation.

For all of the algorithms discussed, the size and complexity of the problem may pose a formidable task with respect to computation and storage of large matrices. In cases where the expense of computation and storage is too great, alternative methods of variance component estimation must be used. One such procedure is Method R (MR) Reverter et al.(1994b). The MR procedure for variance component estimation uses an iterative strategy that is based on a simple linear regression of estimates for the random effects for all the data on estimates of the random effects from part of the data. At each iteration of MR, the empirical BLUP's (EBLUP's) for all the data and for a subset of the data are computed. This requires solving the MME, which is a system of p+q equations. $\mathbf{D}^{-1}$ and $\mathbf{R}^{-1}$ must be computed. The MME can then be solved by using a linear solver to obtain the EBLUP's. This circumvents the need to compute and store the inverse of the coefficient matrix. MR does not provide ML or REML estimates of the covariance parameters; however, situations such as in some applications in animal breeding, the number of covariance parameters c and value of q can be quite large. It is not unusual for these data sets to be so large that algorithms for computing ML/REML estimates, such as those in Sections 1.4 and 1.5, become infeasible. In these instances, the MR procedure

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

offers a means of computing reasonable estimates of the covariance parameters. To date, the properties of estimators derived from the MR procedure are still unclear. In the following chapters we will attempt to shed additional light on some of the characteristics of the MR procedure.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 2

METHOD R PROCEDURE

2.1 Introduction

Method-R (MR) is an iterative procedure for estimating variance components in mixed linear models. The algorithm is based on a simple linear regression of random effect estimates from all the data on random effect estimates from part of the data. The estimates of the random effects are empirical best linear unbiased predictors (EBLUP's) obtained by solving Henderson's mixed model equations (MME). The procedure was first proposed by Reverter (1994b) as a technique for estimating heritability for the single trait animal model. What makes MR attractive is the simplicity of it's implementation, but more importantly, the feasibility of it's computation in contrast to maximum likelihood estimation procedures and other more traditional variance component estimation procedures when extremely large data sets are involved.

In this chapter the MR procedure for estimating variance components is defined formally for the univariate mixed linear model. The procedure is derived from the result that the covariance between BLUP's from all the data (whole sample BLUP's) and the covariance of the BLUP's from a subset of the data (sub-sample BLUP's) are equal. It will also be established that the MR procedure can also be defined from the simple regression through the origin of whole sample BLUP's on sub-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sample BLUP's. An outline of an algorithm to compute MR estimates of variance components is presented toward the end of the chapter.

2.2 The Mixed Linear Model

Consider the mixed linear model given below.

$$\mathbf{y} = \mathbf{X}\gamma + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{2.2.1}$$

where $\mathbf{y}$ is a $N \times 1$ random vector of observable data points; $\mathbf{X}$ is the $N \times p$ design matrix corresponding to the fixed effects; γ is the $p \times 1$ vector of fixed effects parameters; $\mathbf{Z}$ is the $N \times q$ design matrix corresponding to the random effects; $\mathbf{u}$ is the $q \times 1$ vector of random effects parameters and $\mathbf{e}$ is the $N \times 1$ random vector of unobservable random errors. Assume that $\mathbf{u} \sim (\mathbf{0}, \mathbf{D})$ and is independent of $\mathbf{e} \sim (\mathbf{0}, \mathbf{R})$, where $\mathbf{D}$ is the $q \times q$ covariance matrix for $\mathbf{u}$, and $\mathbf{R}$ is the $N \times N$ covariance matrix for the random vector $\mathbf{e}$. The covariance matrices $\mathbf{D}$ and $\mathbf{R}$ are typically functions of a small number of covariance parameters. The the covariance matrix for $\mathbf{y}$ is then,

$$var(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}. \tag{2.2.2}$$

Assuming that $\mathbf{D}$ and $\mathbf{R}$ are non-singular matrices, Henderson's mixed model equations (MME), Henderson (1950), can be written as,

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \gamma \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \tag{2.2.3}$$

As mentioned in Chapter 1, in instances where $\mathbf{D}^{-1}$ and $\mathbf{R}^{-1}$ are less costly to obtain than $\mathbf{V}^{-1}$, the following result due to Henderson et al.(1959) can be helpful.

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}. \tag{2.2.4}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From Christensen (1987), theorem 12.3.1, for known $\mathbf{D}$ and $\mathbf{R}$, if $\begin{bmatrix} \hat{\gamma} \\ \hat{\mathbf{u}} \end{bmatrix}$ is a solution to the mixed model equations, then $\mathbf{X}\hat{\gamma}$ is a $BLUE$ of $\mathbf{X}\gamma$, and $\hat{\mathbf{u}}$ is a $BLUP$ of $\mathbf{u}$. Solving the MME, the $BLUE$ of γ is given by

$$\hat{\gamma} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{2.2.5}$$

and the $BLUP$ of $\mathbf{u}$ is

$$\hat{\mathbf{u}} = \mathbf{D}\mathbf{Z}'(\mathbf{I} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y}. \tag{2.2.6}$$

In practice, the covariance matrices $\mathbf{D}$ and $\mathbf{R}$ are typically unknown. Replacing the true covariance matrices with estimated covariance matrices in the above equations, gives an estimator for $\mathbf{X}\gamma$ that is not a $BLUE$ but an *empirical best linear unbiased estimator (EBLUE)*, and a predictor for $\mathbf{u}$ that is not a $BLUP$ but an *empirical best linear unbiased predictor (EBLUP)*. We will denote these estimators as $\begin{bmatrix} \dot{\hat{\gamma}} \\ \dot{\hat{\mathbf{u}}} \end{bmatrix}$, where the estimator of $\mathbf{D}$ and $\mathbf{R}$ used will be made clear from the context.

2.3 Definition of Method R

Let $\hat{\mathbf{u}}$ be the $BLUP$ for $\mathbf{u}$ based on all the data,$\mathbf{y}$, denoted by $BLUP(\mathbf{u}|\mathbf{y})$. Similary, for a sub-sample of the data, $\mathbf{y}_1$, of dimension $M(M < N)$, let $\tilde{\mathbf{u}}$ be the $BLUP$ for $\mathbf{u}$ based on $\mathbf{y}_1$, denoted by $BLUP(\mathbf{u}|\mathbf{y}_1)$.

The MR procedure relies on the the result from the following theorem from Reverter (1994a), included here for completeness. The proof is essentially that presented by Reverter (1994b), with minor corrections in notation, and adjusted for the univariate case. The theorem for the multivariate case will be presented in Chapter 6.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Theorem 2.1 *Let* $\hat{\mathbf{u}}$ *be the BLUP for* $\mathbf{u}$ *using all the data, and let* $\tilde{\mathbf{u}}$ *be the BLUP for* $\mathbf{u}$ *from a subset of the data. Then the following holds,*

$$cov(\hat{\mathbf{u}}, \tilde{\mathbf{u}}) = var(\tilde{\mathbf{u}}).$$

Proof. Let $\mathbf{y}$ be the entire $N \times 1$ observable random vector of responses and $\mathbf{u}$ the $q \times 1$ unobservable vector of random effects. Then without loss in generality, the elements of $\mathbf{y}$ can be re-ordered and written as a partition, $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$, with $\mathbf{y}_1$ an $M \times 1$ random vector, and $\mathbf{y}_2$ an $(N - M) \times 1$ random vector. Let $\hat{\mathbf{u}} = BLUP(\mathbf{u}|\mathbf{y})$ and let $\tilde{\mathbf{u}} = BLUP(\mathbf{u}|\mathbf{y}_1)$. Let matrices $\mathbf{X}$ and $\mathbf{Z}$ represent the design matrices corresponding to fixed and random effects, respectively, for the re-ordered data vector. Also, let the matrices $\mathbf{D}$, $\mathbf{R}$ and $\mathbf{V}$ represent the covariance matrices corresponding to the re-ordered random vectors $\mathbf{u}$, $\mathbf{e}$ and $\mathbf{y}$, respectively. Then the mean and variance of $\begin{bmatrix} \mathbf{y} \\ \mathbf{u} \end{bmatrix}$ are

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{u} \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{X}\gamma \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{V} & \mathbf{ZD} \\ \mathbf{DZ}' & \mathbf{D} \end{bmatrix} \right).$$

Let $\mathbf{C} = cov(\mathbf{y}, \mathbf{u}) = \mathbf{ZD}$. Then with respect to the partition on $\mathbf{y}$,

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} = \begin{bmatrix} cov(\mathbf{y}_1, \mathbf{u}) \\ cov(\mathbf{y}_2, \mathbf{u}) \end{bmatrix}.$$

Then the mean and variance of $\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{u} \end{bmatrix}$ are

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{u} \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{X}_1\gamma \\ \mathbf{X}_2\gamma \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \mathbf{C}_1 \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \mathbf{C}_2 \\ \mathbf{C}_1' & \mathbf{C}_2' & \mathbf{D} \end{bmatrix} \right),$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where $\mathbf{V}_{ij}$'s are the resulting partition on $\mathbf{V}$ for the re-ordered data vector, $\mathbf{y}$. From Schott (1997) theorem 4.3, there exists a triangular $N \times N$ matrix $\mathbf{T}$ having non-negative diagonal elements, such that, $\mathbf{V} = \mathbf{TT}'$. Further, since $\mathbf{V}$ is a positive definite matrix, $\mathbf{T}$ is unique with positive diagonal elements. Then $\mathbf{T}$ can be written with respect to the partition on $\mathbf{y}$ as

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_{11} & 0 \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix}.$$

Let $\mathbf{z} = \mathbf{T}^{-1}\mathbf{y}$. Then the model for the transform $\mathbf{z}$ is

$$\mathbf{z} = \mathbf{T}^{-1}\mathbf{X}\gamma + \mathbf{T}^{-1}\mathbf{Z}\mathbf{u} + \mathbf{T}^{-1}\mathbf{e}.$$

The partition on $\mathbf{z}$ corresponding to that of $\mathbf{y}$ is

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11} & 0 \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}.$$

Let $\mathbf{W} = \mathbf{T}^{-1}\mathbf{X}$ with partition

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11} & 0 \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix},$$

and let $\mathbf{B} = cov(\mathbf{z}, \mathbf{u}) = \mathbf{T}^{-1}\mathbf{ZD}$ with partition

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} = \begin{bmatrix} cov(\mathbf{z}_1, \mathbf{u}) \\ cov(\mathbf{z}_2, \mathbf{u}) \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11} & 0 \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix}.$$

Then the mean and variance of $\begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{u} \end{bmatrix}$ is

$$\begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{u} \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{W}_1\gamma \\ \mathbf{W}_2\gamma \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{B}_1 \\ \mathbf{0} & \mathbf{I}_{N-m} & \mathbf{B}_2 \\ \mathbf{B}_1' & \mathbf{B}_2' & \mathbf{D} \end{bmatrix} \right).$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Since $\hat{\mathbf{u}} = BLUP(\mathbf{u}|\mathbf{y}) = BLUP(\mathbf{u}|\mathbf{z})$, solving the MME for $\mathbf{u}$ with respect to the transformed data vector $\mathbf{z}$ gives

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{D}(\mathbf{T}^{-1}\mathbf{Z})'(\mathbf{I}_m - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}')\mathbf{z} \\ &= \mathbf{D}(\mathbf{B}\mathbf{D}^{-1})'(\mathbf{I}_m - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}')\mathbf{z} \\ &= \mathbf{B}'(\mathbf{I}_m - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}')\mathbf{z} \\ &= \mathbf{B}'\mathbf{P}\mathbf{z},\end{aligned} \tag{2.3.1}$$

where

$$\begin{aligned}\mathbf{P} &= \mathbf{I}_N - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}' \\ &= \begin{bmatrix} \mathbf{I}_m - \mathbf{W}_1(\mathbf{W}_1'\mathbf{W}_1)^{-}\mathbf{W}'_1 & -\mathbf{W}_1(\mathbf{W}'_1\mathbf{W}_2)^{-}\mathbf{W}'_2 \\ -\mathbf{W}_2(\mathbf{W}'_2\mathbf{W}_1)^{-}\mathbf{W}'_1 & \mathbf{I}_{N-m} - \mathbf{W}_2(\mathbf{W}_1'\mathbf{W}_2)^{-}\mathbf{W}'_2 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}.\end{aligned} \tag{2.3.2}$$

Similarly, if we let $\tilde{\mathbf{u}} = BLUP(\mathbf{u}|\mathbf{y_1}) = BLUP(\mathbf{u}|\mathbf{z_1})$, solving the MME for $\mathbf{u}$ with respect to $\mathbf{z}_1$ gives

$$\begin{aligned}\tilde{\mathbf{u}} &= \mathbf{D}((\mathbf{I}_m\mathbf{0})\mathbf{T}^{-1}\mathbf{Z})'(\mathbf{I}_m - \mathbf{W}_1(\mathbf{W}_1'\mathbf{W}_1)^{-}\mathbf{W}'_1)\mathbf{z}_1 \\ &= \mathbf{D}((\mathbf{I}_m\mathbf{0})\mathbf{B}\mathbf{D}^{-1})'(\mathbf{I}_m - \mathbf{W}_1(\mathbf{W}_1'\mathbf{W}_1)^{-}\mathbf{W}'_1)\mathbf{z}_1 \\ &= \mathbf{B}'\begin{pmatrix} \mathbf{I}_m \\ \mathbf{0} \end{pmatrix}(\mathbf{I}_m - \mathbf{W}_1(\mathbf{W}_1'\mathbf{W}_1)^{-}\mathbf{W}'_1)\mathbf{z}_1 \\ &= \mathbf{B}'_1(\mathbf{I}_m - \mathbf{W}_1(\mathbf{W}_1'\mathbf{W}_1)^{-}\mathbf{W}'_1)\mathbf{z}_1 \\ &= \mathbf{B}'_1\mathbf{P}_{11}\mathbf{z}_1.\end{aligned} \tag{2.3.3}$$

Therefore

$$\begin{aligned}var(\tilde{\mathbf{u}}) &= \mathbf{B}'_1\mathbf{P}_{11}Var(\mathbf{z}_1)\mathbf{P}_{11}\mathbf{B}_1 \\ &= \mathbf{B}'_1\mathbf{P}_{11}\mathbf{I}_m\mathbf{P}_{11}\mathbf{B}_1 \\ &= \mathbf{B}'_1\mathbf{P}_{11}\mathbf{B}_1,\end{aligned} \tag{2.3.4}$$

since $\mathbf{P}_{11}$ is symmetric and idempotent, and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\begin{aligned} cov(\hat{\mathbf{u}}, \tilde{\mathbf{u}}) &= \mathbf{B}'\mathbf{P}cov(\mathbf{z}, \mathbf{z}_1)\mathbf{P}_{11}\mathbf{B}_1 \\ &= \mathbf{B}'\mathbf{P}\begin{pmatrix} \mathbf{I} \\ \mathbf{0} \end{pmatrix}\mathbf{P}_{11}\mathbf{B}_1 \\ &= \mathbf{B}'\begin{pmatrix} \mathbf{P}_{11}\mathbf{P}_{11} \\ \mathbf{P}_{21}\mathbf{P}_{11} \end{pmatrix}\mathbf{B}_1 \\ &= \mathbf{B}'\begin{pmatrix} \mathbf{P}_{11} \\ \mathbf{0} \end{pmatrix}\mathbf{B}_1. \end{aligned} \tag{2.3.5}$$

Since $\mathbf{P}_{21}$ belongs to the row space of $\mathbf{W}_1$, and $\mathbf{P}_{11}$ is the projection matrix onto the null space of $\mathbf{W}_1$, it follows that $\mathbf{P}_{21}\mathbf{P}_{11} = \mathbf{0}$. Therefore

$$\begin{aligned} cov(\hat{\mathbf{u}}, \tilde{\mathbf{u}}) &= \mathbf{B}'_1\mathbf{P}_{11}\mathbf{B}_1 \\ &= var(\tilde{\mathbf{u}}). \end{aligned} \tag{2.3.6}$$

□

The above theorem will now be used to derive the condition on which MR estimation is defined.

From Theorem 2.1 it follows that

$$\begin{aligned} & tr[cov(\hat{\mathbf{u}}, \tilde{\mathbf{u}})] = tr[var(\tilde{\mathbf{u}})] \\ \Rightarrow & \sum_{i=1}^{q} cov(\hat{u}_i, \tilde{u}_i) = \sum_{i=1}^{q} var(\tilde{u}_i). \end{aligned}$$

Since $E(\hat{u}_i) = E(\tilde{u}_i) = 0$, it follows that

$$\begin{aligned} & \sum_{i=1}^{q} E(\hat{u}_i\tilde{u}_i) = \sum_{i=1}^{q} E(\tilde{u}_i^2) \\ \Rightarrow & E\left(\sum_{i=1}^{q} \hat{u}_i\tilde{u}_i\right) = E\left(\sum_{i=1}^{q} \tilde{u}_i^2\right). \end{aligned} \tag{2.3.7}$$

In matrix notation,

$$E(\hat{\mathbf{u}}'\tilde{\mathbf{u}}) = E(\tilde{\mathbf{u}}'\tilde{\mathbf{u}}). \tag{2.3.8}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Also note that if $\mathbf{D}^{-1}$ is used as a weighting matrix, with elements d_{ij}, for $i, j = 1, \ldots, q$. Then

$$
\begin{aligned}
E(\hat{\mathbf{u}}'\mathbf{D}^{-1}\tilde{\mathbf{u}}) &= E\left(\sum_{i,j=1}^{q} \hat{u}_i d_{ij} \tilde{u}_j\right) \\
&= E\left(\sum_{i=1}^{q} \hat{u}_i d_{ii} \tilde{u}_i\right) + E\left(\sum_{i \neq j}^{q} \hat{u}_i d_{ij} \tilde{u}_j\right) \\
&= \sum_{i=1}^{q} d_{ii} E\left(\hat{u}_i \tilde{u}_i\right) + \sum_{i \neq j}^{q} d_{ij} E\left(\hat{u}_i \tilde{u}_j\right) \\
&= \sum_{i=1}^{q} d_{ii} E\left(\tilde{u}_i \tilde{u}_i\right) + \sum_{i \neq j}^{q} d_{ij} E\left(\tilde{u}_i \tilde{u}_j\right) \\
&= E(\tilde{\mathbf{u}}'\mathbf{D}^{-1}\tilde{\mathbf{u}}) \qquad (2.3.9)
\end{aligned}
$$

In the case that $\mathbf{D} = \mathbf{A}\sigma_a^2$, equation 2.3.9 reduces to the identity in Reverter (1994a), $E(\hat{\mathbf{u}}'\mathbf{A}^{-1}\tilde{\mathbf{u}}) = E(\tilde{\mathbf{u}}'\mathbf{A}^{-1}\tilde{\mathbf{u}})$.

The identity in equation 2.3.7 is the motivation for the following definition of MR. *The MR estimates for the covariance parameters are defined to be the values of the covariance parameters for which the EBLUP's,* $\dot{\hat{\mathbf{u}}}$ *and* $\dot{\tilde{\mathbf{u}}}$, *satisfy the following equality.*

$$\dot{\hat{\mathbf{u}}}'\hat{\mathbf{D}}^{-1}\dot{\tilde{\mathbf{u}}} = \dot{\tilde{\mathbf{u}}}'\hat{\mathbf{D}}^{-1}\dot{\tilde{\mathbf{u}}}, \qquad (2.3.10)$$

where $\hat{\mathbf{D}}$ is estimate of $\mathbf{D}$ generated from the estimates of the covariance parameters.

Alternatively, MR can be defined through the regression function $\hat{\mathbf{u}}$ on $\tilde{\mathbf{u}}$. Using the general result for linear models from Rao (1973) section 8a.2, the regression function derived from the conditional expectation of $\hat{\mathbf{u}}$ given $\tilde{\mathbf{u}}$ for fixed $\mathbf{D}$ and $\mathbf{R}$ is,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\begin{aligned} E(\hat{\mathbf{u}}|\tilde{\mathbf{u}}) &= E(\hat{\mathbf{u}}) + cov(\hat{\mathbf{u}}, \tilde{\mathbf{u}})var(\tilde{\mathbf{u}})^{-}(\tilde{\mathbf{u}} - E(\tilde{\mathbf{u}})) \\ &= \mathbf{B'}_1\mathbf{P}_{11}\mathbf{B}_1(\mathbf{B'}_1\mathbf{P}_{11}\mathbf{B}_1)^{-}\tilde{\mathbf{u}} \\ &= (\mathbf{P}_{11}\mathbf{B}_1)'[(\mathbf{P}_{11}\mathbf{B}_1)'(\mathbf{P}_{11}\mathbf{B}_1)]^{-}(\mathbf{P}_{11}\mathbf{B}_1)'\mathbf{z}_1 \\ &= (\mathbf{P}_{11}\mathbf{B}_1)'\mathbf{z}_1 \\ &= \tilde{\mathbf{u}}. \end{aligned} \tag{2.3.11}$$

Thus the regression function for $\hat{\mathbf{u}}$ on $\tilde{\mathbf{u}}$ is the simple linear regression function through the origin with slope 1, given by,

$$\hat{\mathbf{u}} = \tilde{\mathbf{u}} + \epsilon, \tag{2.3.12}$$

where ϵ is a $q \times 1$ random vector with $\epsilon \sim (\mathbf{0}, var(\hat{\mathbf{u}}|\tilde{\mathbf{u}}))$. We will denote the slope of the regression function by $\beta_{\hat{u}|\tilde{u}} = 1$. If $\mathbf{D}$ and $\mathbf{R}$ are estimated, the least squares estimator for the slope of the regression of $\dot{\hat{\mathbf{u}}}$ on $\dot{\tilde{\mathbf{u}}}$, denoted by $\hat{\beta}_{\dot{\hat{u}}|\dot{\tilde{u}}}$ is

$$\hat{\beta}_{\dot{\hat{u}}|\dot{\tilde{u}}} = \frac{\dot{\hat{\mathbf{u}}}'\hat{\mathbf{D}}^{-1}\dot{\tilde{\mathbf{u}}}}{\dot{\tilde{\mathbf{u}}}'\hat{\mathbf{D}}^{-1}\dot{\tilde{\mathbf{u}}}}. \tag{2.3.13}$$

Thus, an alternate definition for MR is the following. *The MR estimates of the covariance parameters are defined to be the values for the covariance parameters for which* $\hat{\beta}_{\dot{\hat{u}}|\dot{\tilde{u}}} = 1$.

In Chapter 6, the two approaches to defining MR estimates of the covariance parameters for the bivariate mixed linear model will be presented, and the question of their equivalence will be addressed.

2.4 Computing Method-R Estimates

The single trait animal model used by Reverter (1994b) will be used to describe the steps in the algorithm for MR. Let $\mathbf{D} = \mathbf{A}\sigma_a^2$ where $\mathbf{A}$ is a known $q \times q$ matrix. $\mathbf{A}$ is commonly referred to as the "relationship matrix". The elements of $\mathbf{A}$ describe the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

genetic relationships among the q random effect components (for example, sires or individual animals). Here we will assume that $\mathbf{R} = \mathbf{I}_N\sigma_e^2$. The MME for this model are

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & (\lambda\mathbf{A}^{-1} + \mathbf{Z}'\mathbf{Z}) \end{bmatrix} \begin{bmatrix} \gamma \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}, \tag{2.4.1}$$

where $\lambda = \sigma_e^2/\sigma_a^2$. One parameter of interest for this model is the proportion of total variation for the trait ($\mathbf{y}$) that is inherited, denoted by

$$\theta = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} = \frac{1}{1+\lambda} \,. \tag{2.4.2}$$

The parameter θ is referred to as *heritability* in genetic prediction models. In more general mixed linear models θ is known as the *intraclass correlation coefficient.*

What follows is a description of the steps required for computing a MR estimate of θ based on a single random sub-sample of the data (Reverter, 1994a), denoted by $\hat{\theta}_{MR}$. The iterative procedure uses a simple bisection of the parameter space of θ, the interval $(0, 1)$. An outline of the steps is included at the end of this section.

STEP 1: Let (c, d) denote the interval to be bisected. Set the initial interval to be, $\left(c^{(0)}, d^{(0)}\right) = (0, 1)$. The initial value for θ, denoted by $\theta^{(0)}$, is set at $\theta^{(0)} = \frac{c^{(0)}+d^{(0)}}{2} = .5$. The inverse of the relationship matrix $\mathbf{A}$ is computed. $\mathbf{A}$ is typically a sparse matrix and may be large depending on the order of the problem (number of individual animals). An Algorithm has been developed for the efficient computation of $\mathbf{A}^{-1}$ (Henderson (1975); Quaas (1975)) and need only be computed one time.

STEP 2: The MME for the whole and sub-sample data are built using the current value of θ for iteration j, $\theta^{(j)}$. Solutions for $\dot{\hat{\mathbf{u}}}$ and $\dot{\tilde{\mathbf{u}}}$ are obtained (which we will now refer to as $\hat{\mathbf{u}}$ and $\tilde{\mathbf{u}}$ for the remaining discussion of the algorithm) for notational

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

simplicity. For large data sets this can be done efficiently by solving the linear system of equations using a Gauss-Siedel algorithm or some similar algorithm for solving a linear system of equations that does not require computation or storage of the inverse of the coefficient matrix. In contrast, many of the more traditional methods reviewed in Chapter 1 require computation and storage of inverses for potentially large matrices at each iteration.

STEP 3: The slope for the simple linear regression of $\hat{\mathbf{u}}$ on $\tilde{\mathbf{u}}$ without intercept is computed. For the MR solution, we want the θ-value that satisfies

$$E\left(\hat{\mathbf{u}}'\mathbf{A}^{-1}\tilde{\mathbf{u}}\right) = E\left(\tilde{\mathbf{u}}'\mathbf{A}^{-1}\tilde{\mathbf{u}}\right).$$

Replacing the above by empirical moments, the MR estimate of θ must satisfy

$$\hat{\mathbf{u}}^{(\mathrm{j})\prime}\mathbf{A}^{-1}\tilde{\mathbf{u}}^{(j)} = \tilde{\mathbf{u}}^{(j)\prime}\mathbf{A}^{-1}\tilde{\mathbf{u}}^{(j)},$$

i.e. $\hat{\beta}^{(j)}_{\hat{\mathbf{u}}|\tilde{\mathbf{u}}} = 1$. If $\hat{\beta}^{(j)}_{\hat{\mathbf{u}}|\tilde{\mathbf{u}}} \neq 1$, then θ is either over-estimated or under-estimated. If $\hat{\beta}^{(j)}_{\hat{\mathbf{u}}|\tilde{\mathbf{u}}} > 1$ then

$$\hat{\mathbf{u}}^{(j)\prime}\mathbf{A}^{-1}\tilde{\mathbf{u}}^{(\mathrm{j})} > \tilde{\mathbf{u}}^{(j)\prime}\mathbf{A}^{-1}\tilde{\mathbf{u}}^{(\mathrm{j})}, \tag{2.4.3}$$

which implies that θ is under-estimated. If $\hat{\beta}^{(j)}_{\hat{\mathbf{u}}|\tilde{\mathbf{u}}} < 1$ then

$$\hat{\mathbf{u}}^{(j)\prime}\mathbf{A}^{-1}\tilde{\mathbf{u}}^{(\mathrm{j})} < \tilde{\mathbf{u}}^{(j)\prime}\mathbf{A}^{-1}\tilde{\mathbf{u}}^{(j)}, \tag{2.4.4}$$

which implies that θ is over-estimated. The above is well established for the one-way random effects model in Chapter 3. Models with covariances carrying properties similar to that of the one-way random effects model, also satisfy the above inequalities for the same reasons as the one-way model.

STEP 4: The next estimate of θ, $\theta^{(j+1)}$, is adjusted according to whether θ was over-estimated or under-estimated using the bisection method. If θ was under-estimated,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\left(c^{(j+1)}, d^{(j+1)}\right) = \left(\frac{c^{(j)} + \theta^{(j)}}{2}, \theta^{(j)}\right). \tag{2.4.5}$$

If θ was over-estimated,

$$\left(c^{(j+1)}, d^{(j+1)}\right) = \left(\theta^{(j)}, \frac{\theta^{(j)} + d^{(j)}}{2}\right). \tag{2.4.6}$$

Steps 2 through 4 are iterated until the desired precision, ε, for computing $\hat{\theta}_{MR}$ is achieved, i.e. when $|\hat{\theta}^{(j)} - \hat{\theta}_{MR}| < \varepsilon$ is satisfied.

The following is an outline of the steps described above.

Outline For Computing a MR Estimate From a Single Sub-Sample

1. Choose a starting values for θ, $\theta^{(0)}$. Use initial end points, $\left(c^{(0)}, d^{(0)}\right) = (0, 1)$. Compute $\mathbf{A}^{-1}$.

2. Compute EBLUPS $\hat{\mathbf{u}}^{(j)}$ and $\tilde{\mathbf{u}}^{(j)}$ by solving the MME for $\theta = \theta^{(j)}$ at the j^{th} iteration.

3. Compute the regression slope $\hat{\beta}^{(j)}_{\hat{\mathbf{u}}|\tilde{\mathbf{u}}} = \frac{\hat{\mathbf{u}}^{(j')}\mathbf{A}^{-1}\tilde{\mathbf{u}}^{(j)}}{\tilde{\mathbf{u}}^{(j')}\mathbf{A}^{-1}\tilde{\mathbf{u}}^{(j)}}$

- if $\hat{\beta}^{(j)}_{\hat{\mathbf{u}}|\tilde{\mathbf{u}}} > 1$ then θ was under-estimated. i.e. $\theta^{(j)} < \hat{\theta}_{MR}$.
- if $\hat{\beta}^{(j)}_{\hat{\mathbf{u}}|\tilde{\mathbf{u}}} < 1$ then θ was over-estimated. i.e. $\theta^{(j)} > \hat{\theta}_{MR}$.

4. Update estimate of θ using bisection algorithm.

- if θ was under-estimated, $\left(c^{(j+1)}, d^{(j+1)}\right) = \left(\frac{c^{(j)}+\theta^{(j)}}{2}, \theta^{(j)}\right)$.
- if θ was over-estimated, $\left(c^{(j+1)}, d^{(j+1)}\right) = \left(\theta^{(j)}, \frac{\theta^{(j)}+d^{(j)}}{2}\right)$.

Iterate steps 2 through 4 until until $|\hat{\theta}^{(j)} - \hat{\theta}_{MR}| < \varepsilon$, where ε is the desired level of precision.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 3

METHOD R ESTIMATION FOR THE ONE-WAY RANDOM EFFECTS MODEL

3.1 Introduction

To better understand the properties of MR estimator, we will examine the one-way balanced random effects model in some detail. The simplicity of this model allows for convenient comparison of the MR estimator to the ML and REML estimators. In particular, we will focus on estimating the intraclass correlation coefficient, $\theta = \sigma_a^2/(\sigma_a^2 + \sigma_e^2)$, referred to as heritability in genetic prediction models.

The vectors for the whole-sample and sub-sample BLUP's for the one-way balanced random effects model are each factored into two parts. One part consisting of insufficient statistics depending only on the data, the second part, a function of the unknown variance components. Regressing the vector of whole-sample insufficient statistics on the vector of sub-sample insufficient statistics, we are able to show that the least squares estimator of the slope is a conditional MLE, given the sub-sample group means. In addition, since θ can be expressed as a function of the expected value of the slope of the regression function for the insufficient statistics, the MR estimator of θ can be derived from the least squares estimator of the slope, and by the invariance property of MLE's, is also a conditional MLE of θ, given the sub-sample group means.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From the unconditional distribution of the estimator of the slope of the regression function of the insufficient statistics, we derive the unconditional distribution of the MR estimator of θ, $\hat{\theta}_{MR}$, and show it is median unbiased.

Finally, for the unbalanced one-way random effects model, it is established via counter-example, that the MR estimator of θ is not a conditional MLE, given the sub-sample group means.

3.2 The Balanced One-Way Random Effects Model

In this section we begin by presenting the balanced one-way random effects model. The MME are constructed and solved to obtain expressions for the BLUP's for the random effects vector. We then introduce notation for defining the corresponding BLUP's for a $100p\%$ random sub-sample, obtained by taking random proportion, p, of the data in each group.

The balanced one-way random effects model is given by

$$y_{ij} = \mu + u_i + e_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n. \tag{3.2.1}$$

For the above linear model, the fixed effects parameter is the scalar, μ, and the design matrix is given by, $\mathbf{X} = \mathbf{1}_N$, where $N = kn$. The random effects vector, $\mathbf{u}$, has corresponding design matrix $\mathbf{Z} = \mathbf{I}_k \otimes \mathbf{1}_n$. The covariance matrices for $\mathbf{u}$ and $\mathbf{e}$ are $\mathbf{D} = \mathbf{I}_k \sigma_a^2$ and $\mathbf{R} = \mathbf{I}_N \sigma_e^2$, respectively, with $u_i \sim N(0, \sigma_a^2)$, independent of $e_{ij} \sim N(0, \sigma_e^2)$.

3.2.1 MME for the Balanced One-Way Random Effects Model

The MME for the balanced one-way random effects model are

$$\frac{1}{\sigma_e^2} \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & (n + \lambda)\mathbf{I}_k \end{bmatrix} \begin{bmatrix} \mu \\ \mathbf{u} \end{bmatrix} = \frac{1}{\sigma_e^2} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix},$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which simplify to

$$\begin{bmatrix} kn & n\mathbf{1}'_k \\ n\mathbf{1}_k & (n+\lambda)\mathbf{I}_k \end{bmatrix} \begin{bmatrix} \mu \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} y_{..} \\ \mathbf{y}_{i.} \end{bmatrix}, \tag{3.2.2}$$

where $\lambda = \sigma_e^2/\sigma_a^2$, $y_{..} = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}$ and $\mathbf{y_{i.}}$ is a $k \times 1$ vector of group totals with r^{th}element $y_{r.} = \sum_{j=1}^{n} y_{rj}$, for $r = 1, \ldots, k$.

From row 1 of the MME :

$$kn\mu + n\sum_{i=1}^{k} u_i = y_{...} \tag{3.2.3}$$

From the i^{th} row of the MME:

$$n\mu + (n+\lambda)u_i = y_{i.}, \quad i = 2, \ldots, k+1. \tag{3.2.4}$$

Summing rows 2 through $k+1$ of the MME gives

$$kn\mu + (n+\lambda)\sum_{i=1}^{k} u_i = y_{...} \tag{3.2.5}$$

Subtracting equation (3.2.3) from (3.2.5) we obtain the following.

$$\lambda \sum_{i=1}^{k} u_i = 0,$$

which implies that the BLUE for μ is

$$\hat{\mu} = \overline{y}_{..}.$$

Substituting $\hat{\mu}$ for μ in equation (3.2.4) and solving for u_i we get,

$$\hat{u}_i = \frac{y_{i.} - n\hat{\mu}}{n+\lambda} = \left(\frac{n}{n+\lambda}\right)(\overline{y}_{i.} - \overline{y}_{..}), \quad i = 1, \ldots, k. \tag{3.2.6}$$

In matrix notation, the $BLUP$ of $\mathbf{u}$ is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
\hat{\mathbf{u}}_{\mathbf{k}\times\mathbf{1}} &= \frac{n}{n+\lambda}\left(\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{nk}\mathbf{1}_k\mathbf{1}'_N\mathbf{y}\right) \\
&= \frac{1}{n+\lambda}\left(\mathbf{I}_k \otimes \mathbf{1}'_n - \frac{1}{k}\mathbf{1}_k(\mathbf{1}'_k \otimes \mathbf{1}'_n)\right)\mathbf{y} \\
&= \frac{1}{n+\lambda}\left[(\mathbf{I}_k - \frac{1}{k}\mathbf{J}_k) \otimes \mathbf{1}'_n\right]\mathbf{y}. \qquad (3.2.7)
\end{aligned}
$$

Let $\mathbf{P}_{k\times k} = (\mathbf{I}_k - \frac{1}{k}\mathbf{J}_k)$, and $\mathbf{B}_{k\times N} = \mathbf{P} \otimes \mathbf{1}'_n$. Then $\hat{\mathbf{u}} = \frac{1}{n+\lambda}\mathbf{B}\mathbf{y}$. We now define the $k \times 1$ vector of insufficient statistics of group mean deviations from the overall mean as

$$\hat{\mathbf{v}} = \mathbf{B}\mathbf{y}. \qquad (3.2.8)$$

3.2.2 100 p % Sub-sampling per group

Suppose m observations are randomly sampled from each group of size n. Let the resulting sampled response vector be $\mathbf{y}^*$.

$$\mathbf{y}^* = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1m} \\ \hline \vdots \\ \hline y_{k1} \\ \vdots \\ y_{km} \end{bmatrix}.$$

Here y_{ij}'s are relabeled for convenience, so that $\mathbf{y}^*$ contains the first m responses in each group. Then the *BLUP* of $\mathbf{u}$ given $\mathbf{y}^*$ can be written as

$$
\begin{aligned}
\tilde{\mathbf{u}} &= \tfrac{1}{m+\lambda}\left[(\mathbf{I}_k - \tfrac{1}{k}\mathbf{J}_k) \otimes \mathbf{1}'_m\right]\mathbf{y}^* \\
&= \tfrac{1}{m+\lambda}\begin{bmatrix} y^*_{1.} - \frac{1}{k}y^*_{..} \\ \vdots \\ y^*_{k.} - \frac{1}{k}y^*_{k.} \end{bmatrix}.
\end{aligned}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Define the matrix as $\mathbf{G} = \mathbf{I}_k \otimes \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times(n-m)} \\ \mathbf{0}_{(n-m)\times m} & \mathbf{0}_{(n-m)\times(n-m)} \end{bmatrix}$.

Then

$$\mathbf{G}\mathbf{y}_{N\times 1} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1m} \\ \mathbf{0}_{n-m} \\ \hline \vdots \\ \hline y_{k1} \\ \vdots \\ y_{km} \\ \mathbf{0}_{n-m} \end{bmatrix} \Rightarrow \mathbf{BGy} = \begin{bmatrix} y^*_{1.} - \frac{1}{k}y^*_{..} \\ \vdots \\ y^*_{k.} - \frac{1}{k}y^*_{k.} \end{bmatrix},$$

where $\mathbf{B}$ is defined as above. The sub-sample $BLUP$ for $\mathbf{u}$ can be written in terms of the whole data vector $\mathbf{y}$ as $\tilde{\mathbf{u}} = \frac{1}{m+\lambda}\mathbf{BGy}$. Analogously, we can define the statistic $\tilde{\mathbf{v}}_{k\times 1}$ for the sub-sample as

$$\tilde{\mathbf{v}} = \mathbf{BGy}. \tag{3.2.9}$$

3.3 Regression of the whole sample insufficient statistics, $\hat{\mathbf{v}}$, on the sub-sample insufficient statistics, $\tilde{\mathbf{v}}$

In this section we will derive the regression function of $\hat{\mathbf{v}}$ on $\tilde{\mathbf{v}}$. We begin with deriving the joint distribution of $\hat{\mathbf{v}}$ and $\tilde{\mathbf{v}}$. The regression function is then derived from the conditional distribution of $\hat{\mathbf{v}}$ given $\tilde{\mathbf{v}}$.

3.3.1 Joint distribution of $\hat{\mathbf{v}}$ and $\tilde{\mathbf{v}}$

The joint distribution of $\hat{\mathbf{v}}$ and $\tilde{\mathbf{v}}$ is

$$\begin{pmatrix} \hat{\mathbf{v}} \\ \tilde{\mathbf{v}} \end{pmatrix}_{2k\times 1} \sim MVN_{2k}\left[\mathbf{0}_{2k\times 1}; \begin{pmatrix} \mathbf{BVB}' & \mathbf{BV}(\mathbf{BG})' \\ (\mathbf{BG})\mathbf{VB}' & (\mathbf{BG})\mathbf{V}(\mathbf{BG})' \end{pmatrix}\right].$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Evaluating the elements of the the above covariance, we have the following.

$$\begin{aligned}
\mathbf{BV} &= (\mathbf{P}\otimes\mathbf{1}_n')[(\mathbf{I}_k\otimes\mathbf{J}_n)\lambda^{-1}+(\mathbf{I}_k\otimes\mathbf{I}_n)]\sigma_e^2 \\
&= [n\lambda^{-1}(\mathbf{P}\otimes\mathbf{1}_n')+(\mathbf{P}\otimes\mathbf{1}_n')]\sigma_e^2 \\
&= (n\lambda^{-1}+1)(\mathbf{P}\otimes\mathbf{1}_n')\sigma_e^2
\end{aligned}$$

$$\begin{aligned}
\mathbf{BVB}' &= (n\lambda^{-1}+1)(\mathbf{P}\otimes\mathbf{1}_n')(\mathbf{P}\otimes\mathbf{1}_n) \\
&= n(n\lambda^{-1}+1)\mathbf{P}\sigma_e^2
\end{aligned}$$

$$\begin{aligned}
\mathbf{BG} &= (\mathbf{P}\otimes\mathbf{1}_n')\left[\mathbf{I}_k\otimes\begin{pmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\right] \\
&= (\mathbf{P}\otimes\mathbf{1}_n^{*'}),
\end{aligned}$$

where $\mathbf{1}_n^{*'}=(\mathbf{1}_m' \mid \mathbf{0}_{(n-m)}')$.

$$\begin{aligned}
\mathbf{BV}(\mathbf{BG})' &= (n\lambda^{-1}+1)(\mathbf{P}\otimes\mathbf{1}_n')(\mathbf{P}\otimes\mathbf{1}_n^{*})\sigma_e^2 \\
&= m(n\lambda^{-1}+1)\mathbf{P}\sigma_e^2
\end{aligned}$$

$$\begin{aligned}
(\mathbf{BG})\mathbf{V} &= (\mathbf{P}\otimes\mathbf{1}_n^{*'})[(\mathbf{I}_k\otimes\mathbf{J}_n)\lambda^{-1}+(\mathbf{I}_k\otimes\mathbf{I}_n)]\sigma_e^2 \\
&= [m\lambda^{-1}(\mathbf{P}\otimes\mathbf{1}_n^{*'})+(\mathbf{P}\otimes\mathbf{1}_n^{*'})]\sigma_e^2 \\
&= (m\lambda^{-1}+1)(\mathbf{P}\otimes\mathbf{1}_n^{*'})\sigma_e^2
\end{aligned}$$

$$\begin{aligned}
(\mathbf{BG})\mathbf{V}(\mathbf{BG})' &= (m\lambda^{-1}+1)(\mathbf{P}\otimes\mathbf{1}_n^{*'})(\mathbf{P}\otimes\mathbf{1}_n^{*})\sigma_e^2 \\
&= m(m\lambda^{-1}+1)\mathbf{P}\sigma_e^2
\end{aligned}$$

Using the above results, we can express the covariance of $\hat{\mathbf{v}}$ and $\tilde{\mathbf{v}}$ as

$$\mathbf{cov}(\hat{\mathbf{v}},\tilde{\mathbf{v}})=\begin{bmatrix} n(n\lambda^{-1}+1) & m(n\lambda^{-1}+1)\\ m(n\lambda^{-1}+1) & m(m\lambda^{-1}+1)\end{bmatrix}\otimes\mathbf{P}\sigma_e^2. \tag{3.3.1}$$

3.3.2 Conditional distribution of $\hat{\mathbf{v}}$ given $\tilde{\mathbf{v}}$

Using the general result for linear models from Rao (1973) Section 8a.2, the conditional expectation of $\hat{\mathbf{v}}$ given $\tilde{\mathbf{v}}$ is

$$\begin{aligned}
E(\hat{\mathbf{v}}\mid\tilde{\mathbf{v}}) &= E(\hat{\mathbf{v}})+cov(\hat{\mathbf{v}},\tilde{\mathbf{v}})\,(var(\tilde{\mathbf{v}}))^{-}\,(\tilde{\mathbf{v}}-E(\tilde{\mathbf{v}})) \\
&= (m(n\lambda^{-1}+1)\mathbf{P}\sigma_e^2)\,(m^{-1}(m\lambda^{-1}+1)^{-1}\mathbf{P}^{-}\sigma_e^{-2})\,\tilde{\mathbf{v}} \\
&= (n\lambda^{-1}+1)(m\lambda^{-1}+1)^{-1}\tilde{\mathbf{v}} \\
&= \tfrac{n+\lambda}{m+\lambda}\tilde{\mathbf{v}},
\end{aligned} \tag{3.3.2}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and the conditional variance of $\hat{\mathbf{v}}$ given $\tilde{\mathbf{v}}$ is

$$\begin{aligned} var(\hat{\mathbf{v}} \mid \tilde{\mathbf{v}}) &= var(\hat{\mathbf{v}}) - cov(\hat{\mathbf{v}},\tilde{\mathbf{v}})\,(var(\tilde{\mathbf{v}}))^{-}\,cov(\tilde{\mathbf{v}},\hat{\mathbf{v}}) \\ &= [n(n\lambda^{-1}+1) - m^2(n\lambda^{-1}+1)^2m^{-1}(m\lambda^{-1}+1)^{-1}]\,\mathbf{P}\sigma_e^2 \\ &= (n\lambda^{-1}+1)[n - m(n\lambda^{-1}+1)(m\lambda^{-1}+1)^{-1}]\mathbf{P}\sigma_e^2 \\ &= (n\lambda^{-1}+1)(m\lambda^{-1}+1)^{-1}[mn\lambda^{-1}+n-mn\lambda^{-1}-m]\mathbf{P}\sigma_e^2 \\ &= (n\lambda^{-1}+1)(m\lambda^{-1}+1)^{-1}(n-m)\mathbf{P}\sigma_e^2 \\ &= \tfrac{n+\lambda}{m+\lambda}(n-m)\mathbf{P}\sigma_e^2. \end{aligned} \tag{3.3.3}$$

Thus, the conditional distribution of $\hat{\mathbf{v}}$ given $\tilde{\mathbf{v}}$ is

$$\hat{\mathbf{v}} \mid \tilde{\mathbf{v}} \sim N_k\left(\frac{n+\lambda}{m+\lambda}\tilde{\mathbf{v}}, \frac{n+\lambda}{m+\lambda}(n-m)\mathbf{P}\sigma_e^2\right),$$

and the linear regression function of $\hat{\mathbf{v}}$ on $\tilde{\mathbf{v}}$ is

$$\hat{\mathbf{v}} = \beta\tilde{\mathbf{v}} + \mathbf{e}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}, \tag{3.3.4}$$

where $\beta = \frac{n+\lambda}{m+\lambda}$ and $\mathbf{e}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}} \sim N\left(\mathbf{0}, \frac{n+\lambda}{m+\lambda}(n-m)\mathbf{P}\sigma_e^2\right)$.

Let $\sigma^2 = \frac{n+\lambda}{m+\lambda}(n-m)\sigma_e^2$ and $\Sigma_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}} = \sigma^2\mathbf{P}$. Then the weighted least squares estimator of $\beta_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ is

$$\begin{aligned} \hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}} &= \left(\tilde{\mathbf{v}}'\Sigma^{-}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\tilde{\mathbf{v}}\right)^{-1}\tilde{\mathbf{v}}'\Sigma^{-}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\hat{\mathbf{v}} \\ &= (\sigma^2\tilde{\mathbf{v}}'\mathbf{P}\tilde{\mathbf{v}})^{-1}(\sigma^2\tilde{\mathbf{v}}'\mathbf{P}\hat{\mathbf{v}}), \end{aligned}$$

since $\mathbf{P}$ is a projection onto the space orthogonal to $\mathbf{1}_k$, $\tilde{\mathbf{v}}'\mathbf{P}\tilde{\mathbf{v}} = \tilde{\mathbf{v}}'\tilde{\mathbf{v}}$ and $\tilde{\mathbf{v}}'\mathbf{P}\hat{\mathbf{v}} = \tilde{\mathbf{v}}'\hat{\mathbf{v}}$. It then follows that,

$$\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}} = \frac{\tilde{\mathbf{v}}'\hat{\mathbf{v}}}{\tilde{\mathbf{v}}'\tilde{\mathbf{v}}}. \tag{3.3.5}$$

Thus, the weighted least squares estimator and the ordinary least squares estimator of β are identical.

Note that as a result of $\mathbf{P}$ being a singular matrix, the conditional likelihood function for $\hat{\mathbf{v}} \mid \tilde{\mathbf{v}}$ does not exist. Since $\mathbf{P}$ is symmetric idempotent, from Graybill(1976) Theorem 1.2.50, there exists an orthogonal $k \times k$ matrix $\mathbf{C}$ such that $\mathbf{P} = \mathbf{CDC}'$,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where $\mathbf{D}$ is the diagonal matrix of eigenvalues for $\mathbf{P}$, i.e. $\mathbf{D}_{k\times k} = Diag(\mathbf{1}_{k-1}, 0)$, from Graybill(1976) Theorem 1.7.2. Let $\mathbf{C}^*_{(k-1)\times k}$ matrix resulting from deleting the last row of $\mathbf{C}$. Applying $\mathbf{C}^*$ as a linear transform on $\hat{\mathbf{v}}$, the conditional distribution of $\mathbf{C}^*\hat{\mathbf{v}}$ is $N_{k-1}(\beta\mathbf{C}^*\tilde{\mathbf{v}}, \sigma^2\mathbf{I}_{k-1})$, since $\sigma^2\mathbf{C}^*\mathbf{P}\mathbf{C}^{*'} = \sigma^2\mathbf{C}^*\mathbf{C}'\mathbf{D}\mathbf{C}\mathbf{C}^{*'} = \sigma^2\mathbf{I}_{k-1}$. The corresponding least squares estimator of β based on the regression of $\mathbf{C}^*\hat{\mathbf{v}}$ on $\mathbf{C}^*\tilde{\mathbf{v}}$ is

$$\begin{aligned}\frac{(\mathbf{C}^*\tilde{\mathbf{v}})'(\mathbf{C}^*\hat{\mathbf{v}})}{(\mathbf{C}^*\tilde{\mathbf{v}})'(\mathbf{C}^*\tilde{\mathbf{v}})} &= \frac{\tilde{\mathbf{v}}'\mathbf{C}^{*'}\mathbf{C}^*\hat{\mathbf{v}}}{\tilde{\mathbf{v}}'\mathbf{C}^{*'}\mathbf{C}^*\tilde{\mathbf{v}}} \\ &= \frac{\tilde{\mathbf{v}}'\hat{\mathbf{v}}}{\tilde{\mathbf{v}}'\tilde{\mathbf{v}}} = \hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}},\end{aligned}$$

since $\mathbf{C}^{*'}\mathbf{C}^* = \mathbf{P}$ and $\mathbf{P}\hat{\mathbf{v}} = \hat{\mathbf{v}}$. Under the assumption that $\mathbf{C}^*\hat{\mathbf{e}}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ is multivariate normal, it follows from Result 7.4 in Johnson and Wichern (1992) that $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ *is the MLE for* β *based on the whole and sub-sample insufficient statistics.* Further, $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}} \sim N(\beta, \sigma^2(\tilde{\mathbf{v}}'\tilde{\mathbf{v}})^{-1})$ and is independent of the MLE of σ^2,

$$\begin{aligned}\hat{\sigma}^2 &= (\mathbf{C}^*\hat{\mathbf{e}}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}})'(\mathbf{C}^*\hat{\mathbf{e}}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}})/(k-1) \\ &= \hat{\mathbf{e}}'_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\mathbf{C}^{*'}\mathbf{C}^*\hat{\mathbf{e}}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}/(k-1) \\ &= \hat{\mathbf{e}}'_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\mathbf{P}\hat{\mathbf{e}}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}/(k-1) \\ &= \hat{\mathbf{e}}'_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\hat{\mathbf{e}}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}/(k-1),\end{aligned}$$

where $\hat{\mathbf{e}}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}} = \hat{\mathbf{v}} - \hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\tilde{\mathbf{v}}$. Also, recall that $\hat{\mathbf{v}} = (\mathbf{P}\otimes\mathbf{1}'_n)\mathbf{y} = k\mathbf{P}\overline{\mathbf{y}}$, where $\overline{\mathbf{y}}$ is the vector of whole-sample group means. Therefore $k\mathbf{C}^*\hat{\mathbf{v}} = k\mathbf{C}^*\mathbf{P}\overline{\mathbf{y}}$. The matrix $k\mathbf{C}^*\mathbf{P}$ has full row rank, equal to the dimension of the error space for the design matrix of the fixed effects , $\mathbf{X} = \mathbf{1}_N$. Thus, from Section 1.3.3, $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ is the *conditional REML estimator for* β*, based on the whole and sub-sample group means.*

It should be noted that since σ^2 is a function of β and σ_e^2, an estimator for σ_e^2 is

$$\hat{\sigma}_e^2 = \frac{\hat{\sigma}^2}{(n-m)\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}}.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

It follows from the invariance property of MLE's that the above expression gives a conditional MLE for σ_e^2, only if $\hat{\beta}_{\hat{\mathbf{v}}|\bar{\mathbf{v}}} > 0$. Otherwise, $\hat{\sigma}_e^2$ falls outside its parameter space. As our main parameter of interest is θ, the problem of dealing with MR estimates outside the parameter space will be addressed with respect to estimating θ in the following section.

3.4 Method R estimator for θ

In this section we a derive the MR estimator for θ from $\hat{\beta}_{\hat{\mathbf{v}}|\bar{\mathbf{v}}}$, the MR estimator for β. Further, the MR estimator $\hat{\theta}$, for θ, is shown to be a conditional MLE given the sub-sample groups means.

Recall that the parameter $\theta = \sigma_a^2/(\sigma_e^2 + \sigma_a^2)$, which can be written in terms of $\lambda = \sigma_e^2/\sigma_a^2$ as,

$$\theta = \frac{1}{1+\lambda}. \tag{3.4.1}$$

From Section 3.3.2, $\beta = (n+\lambda)/(m+\lambda)$. Thus, λ as a function of β is

$$\lambda = \frac{n - m\beta}{\beta - 1}. \tag{3.4.2}$$

Substituting for λ in equation 3.4.2 into equation 3.4.1, θ as a function of β is given by

$$\begin{aligned} \theta &= \left(1 + \frac{n - m\beta}{\beta - 1}\right)^{-1} \\ &= \frac{\beta - 1}{\beta - 1 + n - m\beta} \\ &= \frac{\beta - 1}{(n-1) - (m-1)\beta}. \end{aligned} \tag{3.4.3}$$

Note: Since $\sigma_a^2 = \sigma_e^2/\lambda$, a MLE of σ_a^2 is available as a result of the invariance property of MLE's. This estimator has similar problems to that for σ_e^2 mentioned

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

at the end of Section 3.3, and for the same reason stated above, a resolution is addressed with respect to estimation of θ.

From equation 3.4.3, the MR estimator for θ can be obtained by replacing β by it's MR estimator, $\hat{\beta}_{\hat{v}|\tilde{v}}$. Figure 3.1 illustrates the functional relationship between θ and β from equation 3.4.3. From figure 3.1, we see that if β is outside the interval $(1, n/m)$, θ is outside it's parameter space, $(0, 1)$. While replacing β in equation 3.4.3 with $\hat{\beta}_{\hat{v}|\tilde{v}}$ would seem to be a logical means for estimating θ, the possibility of $\hat{\beta}_{\hat{v}|\tilde{v}}$ falling outside the interval $(1, n/m)$ could pose a problem. One solution would be to define the MR estimator for θ in terms of $\hat{\beta}_{\hat{v}|\tilde{v}}$ as follows.

$$\hat{\theta} = \begin{cases} 0, & \hat{\beta}_{\hat{v}|\tilde{v}} < 1 \\ \frac{\hat{\beta}_{\hat{v}|\tilde{v}} - 1}{n-1-(m-1)\hat{\beta}_{\hat{v}|\tilde{v}}}, & 1 < \hat{\beta}_{\hat{v}|\tilde{v}} < n/m \\ 1, & \hat{\beta}_{\hat{v}|\tilde{v}} > n/m \end{cases} \qquad (3.4.4)$$

Notice that if $\hat{\theta}$ were unrestricted, by the invariance property of MLE's, it would also be a conditional MLE given the insufficient statistics. From equation 3.4.4, it is not obvious that $\hat{\theta}$ would remain a conditional MLE, as the function in 3.4.4 is not $1-1$. What follows covers the three situations that could arise. Here $L(.)$ is the likelihood function for β, which under the assumption of normality, is unimodal with maximum at $\beta = \hat{\beta}_{\hat{v}|\tilde{v}}$. As a result, we have the following.

1. $\hat{\beta}_{\hat{v}|\tilde{v}} < 1$, then $L(\hat{\beta}_{\hat{v}|\tilde{v}}) > L(1) > \underset{\{\beta:\ 1<\beta<n/m\}}{L(\beta)} \Rightarrow \hat{\theta} = 0$ is the MLE for θ.
2. $1 < \hat{\beta}_{\hat{v}|\tilde{v}} < n/m$, then $\hat{\theta} = \frac{\hat{\beta}_{\hat{v}|\tilde{v}} - 1}{n-1-(m-1)\hat{\beta}_{\hat{v}|\tilde{v}}}$ is the MLE for θ.
3. $\hat{\beta}_{\hat{v}|\tilde{v}} > n/m$, then $\underset{\{\beta:\ 1<\beta<n/m\}}{L(\beta)} < L(n/m) < L(\hat{\beta}_{\hat{v}|\tilde{v}}) \Rightarrow \hat{\theta} = 1$ is the MLE for θ.

Thus, we have established that $\hat{\theta}$, the MR estimator for θ, as defined in expression 3.4.4 is a *conditional MLE of θ , based on the insufficient statistics* $\hat{v}$ *and* $\tilde{v}$ and a conditional *REML estimator of θ, based on the whole and sub-sample group means.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

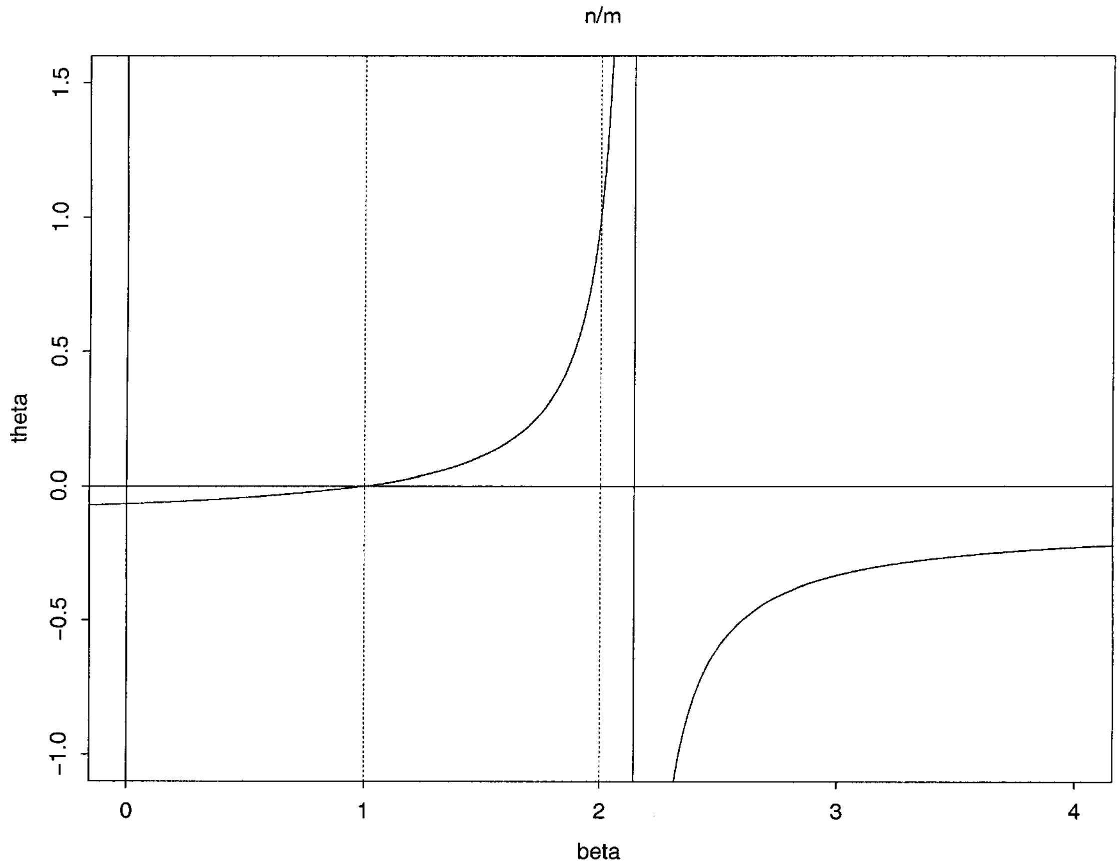

Figure 3.1: Graph of θ versus β. *θ falls outside it's parameter space,* $(0, 1)$*, when β falls outside* $(1, n/m)$.

In the following section, we wish to obtain the unconditional sampling distribution of $\hat{\theta}$. From the unconditional distribution of $\hat{\theta}$ we seek a characterization of the MR estimator of θ over all possible sub-samples. First, the unconditional distribution of the MR estimator of β is obtained. The resulting random variable will be referred to simply as $\hat{\beta}$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.5 Unconditional Distribution of $\hat{\beta}$

Under the assumption of multivariate normality for the regression residuals in Section 3.3, the distribution of $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ is multivariate normal with mean,

$$\begin{aligned} E(\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}) &= E\left(\frac{\tilde{\mathbf{v}}'\hat{\mathbf{v}}}{\tilde{\mathbf{v}}'\tilde{\mathbf{v}}}\right) \\ &= (\tilde{\mathbf{v}}'\tilde{\mathbf{v}})^{-1}\tilde{\mathbf{v}}'E(\hat{\mathbf{v}} \mid \tilde{\mathbf{v}}) \\ &= (\tilde{\mathbf{v}}'\tilde{\mathbf{v}})^{-1}\tilde{\mathbf{v}}'\tilde{\mathbf{v}}, \frac{n+\lambda}{m+\lambda}, \\ &= \beta, \end{aligned} \tag{3.5.1}$$

and variance,

$$\begin{aligned} var(\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}) &= \left(\tilde{\mathbf{v}}'\Sigma^{-}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\tilde{\mathbf{v}}\right)^{-1} \tilde{\mathbf{v}}'\Sigma^{-}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\Sigma_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\Sigma^{-}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\tilde{\mathbf{v}} \left(\tilde{\mathbf{v}}'\Sigma^{-}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\tilde{\mathbf{v}}\right)^{-1} \\ &= \left(\tilde{\mathbf{v}}'\Sigma^{-}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}\tilde{\mathbf{v}}\right)^{-1} \\ &= \frac{(n+\lambda)}{(m+\lambda)}\frac{(n-m)}{\tilde{\mathbf{v}}'\tilde{\mathbf{v}}}\sigma_e^2. \end{aligned} \tag{3.5.2}$$

Recall from equation 3.3.1, $\tilde{\mathbf{v}} \sim N(0, m(m\lambda^{-1}+1)\mathbf{P}\sigma_e^2)$. Let $\phi^2 = m(m\lambda^{-1}+1)\sigma_e^2$. Let $\mathbf{z} = \phi^{-1}\tilde{\mathbf{v}}$. Then $\mathbf{z} \sim \mathbf{N}_k(0, \mathbf{P})$. Since $\mathbf{P}$ is a symmetric idempotent matrix with rank $= k-1$, it follows (Graybill, theorem 4.7.1 pg. 140) that $\mathbf{z}'\mathbf{z} \sim \chi^2_{k-1}$. Let $Q = \tilde{\mathbf{v}}'\tilde{\mathbf{v}}$. Then $\mathbf{Q} \sim \phi^2\chi^2_{k-1}$. Let $\sigma^2 = \frac{n+\lambda}{m+\lambda}(n-m)\sigma_e^2$. Then we can write simply that $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}} \sim N(\beta, \sigma^2/Q)$.

For the remaining of this chapter we will refer to the distribution of $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ as the distribution of $\hat{\beta}|\tilde{\mathbf{v}}$.

Now $\hat{\beta}|\tilde{\mathbf{v}} \stackrel{d}{=} \hat{\beta}|Q \sim N(\beta, \sigma^2/Q)$. Let $\hat{\beta}' = Q^{1/2}(\hat{\beta}-\beta)/\sigma$. Then $\hat{\beta}'|Q \sim N(0,1)$. Since the conditional distribution of $\hat{\beta}'$ given Q does not depend on Q, it follows that $\hat{\beta}'$ and Q are independent. Therefore unconditionally $\hat{\beta}' \sim N(0,1)$, from which we have the following result.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\frac{\hat{\beta}'}{(\phi^{-1}Q^{1/2})/\sqrt{k-1}} = \frac{Q^{1/2}(\hat{\beta}-\beta)/\sigma}{(\phi^{-1}Q^{1/2})/\sqrt{k-1}} = \frac{\phi}{\sigma}\sqrt{k-1}(\hat{\beta}-\beta) \sim t_{k-1}. \tag{3.5.3}$$

Let $T \sim t_{k-1}$, then

$$\hat{\beta} \sim \frac{\sigma}{\phi}(k-1)^{-1/2}T + \beta.$$

Thus, the unconditional distribution of $\hat{\beta}$ is a scaled central t-distribution with $k-1$ degrees of freedom, shifted by β i.e. $\hat{\beta}$ has a location scale t-distribution, with location parameter β and scale parameter $\frac{\sigma^2}{\phi^2}(k-1)^{-1}$, denoted as

$$\hat{\beta} \sim t_{k-1}\left(\beta, \frac{\sigma^2}{\phi^2}(k-1)^{-1}\right).$$

The unconditional density of $\hat{\beta}$ is given by

$$\begin{aligned} f_{\hat{\beta}}(x) &= \frac{1}{\sqrt{\pi}}(k-1)^{-1/2}\left(\frac{\sigma^2}{\phi^2}(k-1)^{-1}\right)^{-1/2} \frac{\Gamma(k/2)}{\Gamma((k-1)/2)}\left[1 + \frac{(k-1)^{-1}(x-\beta)^2}{\left(\frac{\sigma^2}{\phi^2}(k-1)^{-1}\right)}\right]^{-k/2} \\ &= \frac{1}{\sqrt{\pi}}\left(\frac{\phi}{\sigma}\right)\frac{\Gamma(k/2)}{\Gamma((k-1)/2)}\left[\left(\frac{\phi}{\sigma}\right)^2(x-\beta)^2+1\right]^{-k/2}. \end{aligned}$$

The unconditional mean of $\hat{\beta}$ is $E(\hat{\beta}) = \beta$ and the unconditional variance of $\hat{\beta}$ is

$$\begin{aligned} var(\hat{\beta}) &= \frac{\sigma^2}{\phi^2}(k-1)^{-1}\frac{k-1}{(k-1)-2} \\ &= \frac{n+\lambda}{m+\lambda}(n-m)\sigma_e^2\left[m(m\lambda^{-1}+1)\sigma_e^2\right]^{-1}\frac{1}{k-3} \\ &= \frac{\lambda(n+\lambda)(n-m)}{m(m+\lambda)^2(k-3)}. \end{aligned} \tag{3.5.4}$$

Further, since the distribution for $\hat{\beta}$ is unimodal and symmetric about β, it is both mean and median unbiased. Let $\hat{\beta}_{.5}$ denote the median of the sampling distribution

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of $\hat{\beta}$ and $\hat{\theta}_{.5}$ denote the median of the sampling distribution of $\hat{\theta}$. Then there are three possible cases to consider.

1. $\hat{\beta}_{.5} \in (1, n/m)$: in which case there is a 1-1 relationship between $\hat{\beta}$ and $\hat{\theta}$, and $\hat{\beta}_{.5} = \beta \;\Rightarrow\; \hat{\theta}_{.5} = \theta$.

2. $\hat{\beta}_{.5} < 1$: in which case $\beta < 1 \;\Rightarrow\; \hat{\theta}_{.5} = 0$ and $\theta = 0$.

3. $\hat{\beta}_{.5} > n/m$: in which case $\beta > n/m \;\Rightarrow\; \hat{\theta}_{.5} = 1$ and $\theta = 1$.

Thus, it follows from the above arguments that $\hat{\theta}_{.5} = \theta$. Therefore $\hat{\theta}$ is a median unbiased estimator for θ. Also, since the distribution of $\hat{\theta}$ is asymmetric in general, $\hat{\theta}$ will not in general be mean unbiased.

3.6 Unconditional Distribution of $\hat{\theta}$

The unconditional distribution of $\hat{\theta}$ and it's moments will now be derived from unconditional distribution of $\hat{\beta}$. Recall that

$$\hat{\theta} = \begin{cases} 0 & ,\hat{\beta} < 1 \\ \frac{\hat{\beta}-1}{n-1-(m-1)\hat{\beta}} & ,1 < \hat{\beta} < n/m \\ 1 & ,\hat{\beta} > n/m \end{cases}$$

Therefore the density for $\hat{\theta}$ in terms of the density of $\hat{\beta}$ is

$$\begin{aligned} f_{\theta}(x) &= P_{\hat{\theta}}(0) + f_{\hat{\theta}}(x) \cdot I_{(0,1)}(x) + P_{\hat{\theta}}(1) \\ &= \textstyle\int_{-\infty}^{1} f_{\hat{\beta}}(u)\delta u \cdot I_{(0)}(x) + \left|\frac{\delta\hat{\beta}}{\delta\hat{\theta}}\right| f_{\hat{\beta}}(\hat{\theta}) \cdot I_{(0,1)}(x) + \int_{n/m}^{\infty} f_{\hat{\beta}}(u)\delta u \cdot I_{(1)}(x). \end{aligned}$$

The first and second moments about the origin are ,

$$E_{\hat{\theta}}(X) = 0 \cdot P_{\hat{\theta}}(0) + \int_0^1 x f_{\hat{\theta}}(x)\delta x + 1 \cdot P_{\hat{\theta}}(1), \tag{3.6.1}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and

$$E_{\hat{\theta}}(X^2) = 0^2 \cdot P_{\hat{\theta}}(0) + \int_0^1 x^2 f_{\hat{\theta}}(x)\delta x + 1^2 \cdot P_{\hat{\theta}}(1)$$

Therefore the variance of $\hat{\theta}$ is

$$var_{\hat{\theta}}(x) = E_{\hat{\theta}}(X^2) - [E_{\hat{\theta}}(X)]^2 \tag{3.6.2}$$

Figure 3.2 displays the densities for $\hat{\theta}$ when $n = 8$; $k = 8, 20, 50$ and $\theta = .0588, .5, .9488$. Notice that when the true value of θ is close to 0, the distribution of $\hat{\theta}$ is skewed right, with noticeable point mass at $\hat{\theta} = 0$ and small probability mass at $\hat{\theta} = 1$. When the true value of θ is 0.5, the distribution of $\hat{\theta}$ becomes more unimodal and less asymmetrical. When the true value of θ is close to 1, the distribution of $\hat{\theta}$ is skewed left, with noticeable point mass at $\hat{\theta} = 1$ and small probability mass at $\hat{\theta} = 0$. We can also see from figure 3.2 that as the value of k increases, the point masses at the endpoints of the parameter space are gradually reduced, and the mean-biasedness is also reduced in all cases.

3.7 Large Sample Properties of $\hat{\theta}$

In this section we will give attention to the large sample properties of MR for a single sub-sample. From equation 3.4.3, θ can be expressed as a differentiable function of β. First and second order partial derivatives with respect to β are

$$f'(\beta) = \frac{\partial\theta}{\partial\beta} = \frac{n-m}{[(n-1)-(m-1)\beta]^2}, \tag{3.7.1}$$

$$f''(\beta) = \frac{\delta^2\theta}{\partial\beta^2} = \frac{2(n-m)(m-1)}{[(n-1)-(m-1)\beta]^3}. \tag{3.7.2}$$

From Serfling (1980), Taylor's theorem is as follows.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

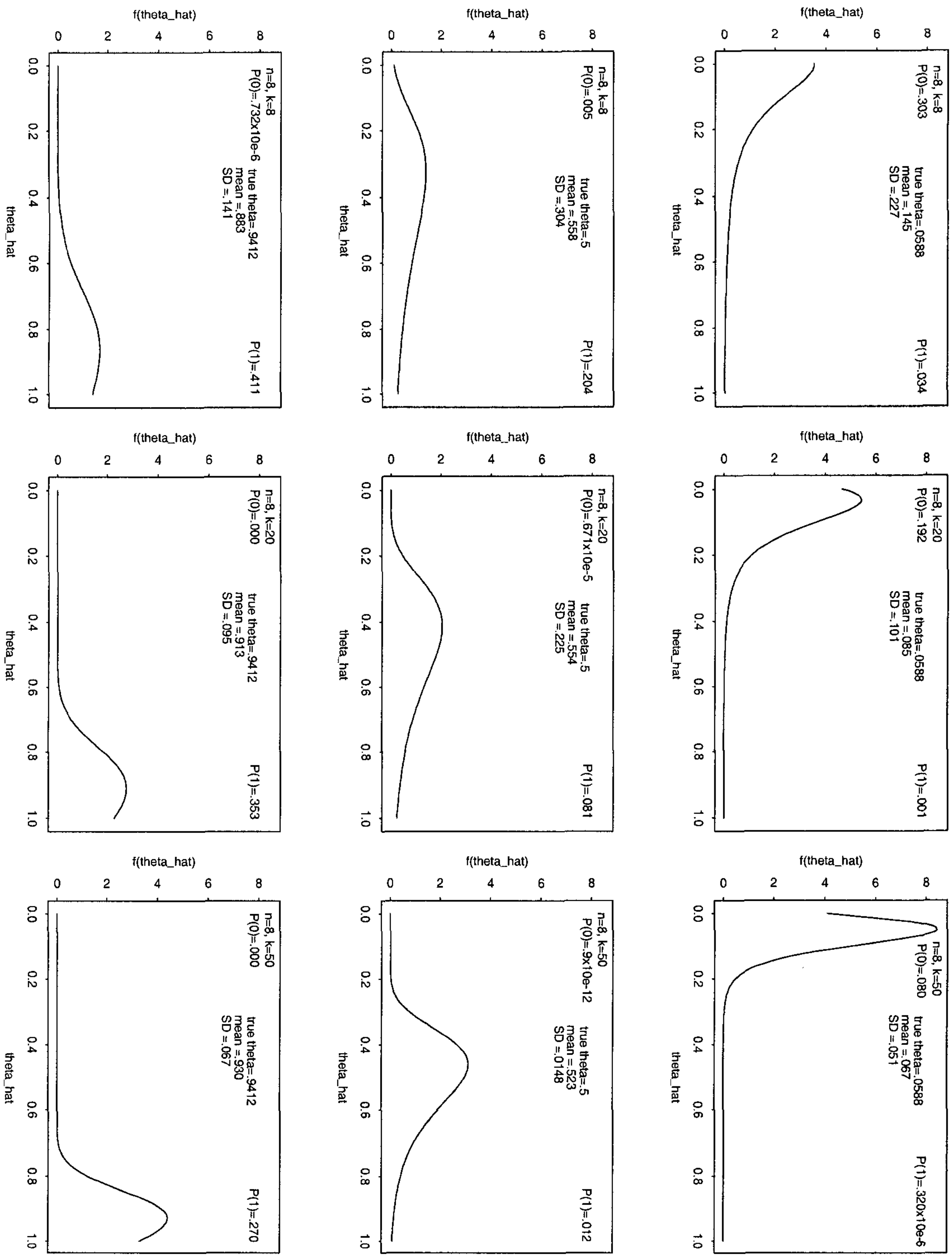

Figure 3.2: *Densities for* $\hat{\theta}$ *for* $n = 8$*;* $k = 8, 20, 50$ *and* $\theta = .0588, .5, .9412$*.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Theorem 3.1 *Let the function f have finite nth derivative $f^{(n)}$ everywhere in the open interval (a, b) and (n-1)th $f^{(n-1)}$ continuous in the closed interval $[a, b]$. Let $x \in [a, b]$. For each point $y \in [a, b]$, $y \neq x$, there exists a point z interior to the interval joining x and y such that*

$$f(y) = f(x) + \sum_{k=1}^{n-1} \frac{f^{(k)}(x)}{k!}(y-x)^k + \sum_{k=1}^{n} \frac{f^{(n)}(x)}{n!}(y-x)^n.$$

Applying Theorem 3.1, an approximation for the large sample mean for $\hat{\theta}$ will now be derived.

$$\begin{aligned} E(\hat{\theta}) &= E[f(\hat{\beta})] \\ &\doteq f[E(\hat{\beta})] + \frac{var(\hat{\beta})}{2} f''[E(\hat{\beta})] \\ &= \theta + \frac{1}{2}\frac{\lambda(n+\lambda)(n-m)}{m(m+\lambda)^2(k-3)} \frac{2(n-m)(m-1)}{\left[(n-1)-(m-1)\frac{n+\lambda}{m+\lambda}\right]^3} \\ &= \theta + \frac{\lambda(n+\lambda)(n-m)^2(m-1)}{m(m+\lambda)^2(k-3)} \frac{(m+\lambda)^3}{(n-m)^3(1+\lambda)^3} \\ &= \theta + \frac{\lambda(n+\lambda)(m+\lambda)(m-1)}{m(n-m)(1+\lambda)^3(k-3)}. \end{aligned} \tag{3.7.3}$$

Once again, applying Taylor's theorem, an approximation for the large sample variance for $\hat{\theta}$ is

$$\begin{aligned} var(\hat{\theta}) &= var[f(\hat{\beta})] \\ &\doteq var(\hat{\beta}) f'[E(\hat{\beta})]^2 \\ &= \frac{\lambda(n+\lambda)(n-m)}{m(m+\lambda)^2(k-3)} \frac{(n-m)^2}{\left[(n-1)-(m-1)\frac{n+\lambda}{m+\lambda}\right]^4} \\ &= \frac{\lambda(n+\lambda)(n-m)^3}{m(m+\lambda)^2(k-3)} \frac{(m+\lambda)^4}{(n-m)^4(1+\lambda)^4} \\ &= \frac{\lambda(n+\lambda)(m+\lambda)^2}{m(n-m)(1+\lambda)^4(k-3)}. \end{aligned} \tag{3.7.4}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For $m = 1$ or as $k \longrightarrow \infty$, the first order bias term in equation 3.7.3 vanishes. For fixed n and k, the large sample approximation for $var(\hat{\theta}_{MR})$ in equation 3.7.4 is minimized for $m = \frac{n\lambda}{n+2\lambda}$. Also, as $\lambda \longrightarrow 0, m \longrightarrow 0$, and as $\lambda \longrightarrow \infty, m \longrightarrow n/2$.

3.7.1 Asymptotic Relative Efficiency of MR for single sub-sample to MLE

From from Donner and Koval (1980), the large sample variance for the MLE of θ is

$$var_{asy}(\hat{\theta}_{MLE}) = \frac{2\lambda^2(n+\lambda)^2}{nk(1+\lambda)^4(n-1)}. \tag{3.7.5}$$

The ARE of $\hat{\theta}_{MR}$ with respect to $\hat{\theta}_{MLE}$ as $k \longrightarrow \infty$ is

$$\begin{aligned} ARE_{MR,MLE} &= \frac{2\lambda^2(n+\lambda)^2}{nk(1+\lambda)^4(n-1)}\left[\frac{\lambda(n+\lambda)(m+\lambda)^2}{m(n-m)(1+\lambda)^4(k-3)}\right]^{-1} \\ &= \frac{2m\lambda(n+\lambda)(n-m)}{n(n-1)(m+\lambda)^2}. \end{aligned} \tag{3.7.6}$$

Figure 3.3 displays ARE as a function of θ for $n = 4, 8, 20, 100$ and $m = 1, n/2, n-1$. For all of the observed examples, the $ARE_{MR,ML} < 1$. This reflects the loss in information in MR, as a result of only using a single sub-sample of the data. However, by taking multiple sub-samples we can increase the amount of information about θ. The mean and the median values of $\hat{\theta}$ -values for multiple sub-samples are natural choices as estimators. These new MR based estimators will be used in simulations presented in Chapter 4.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

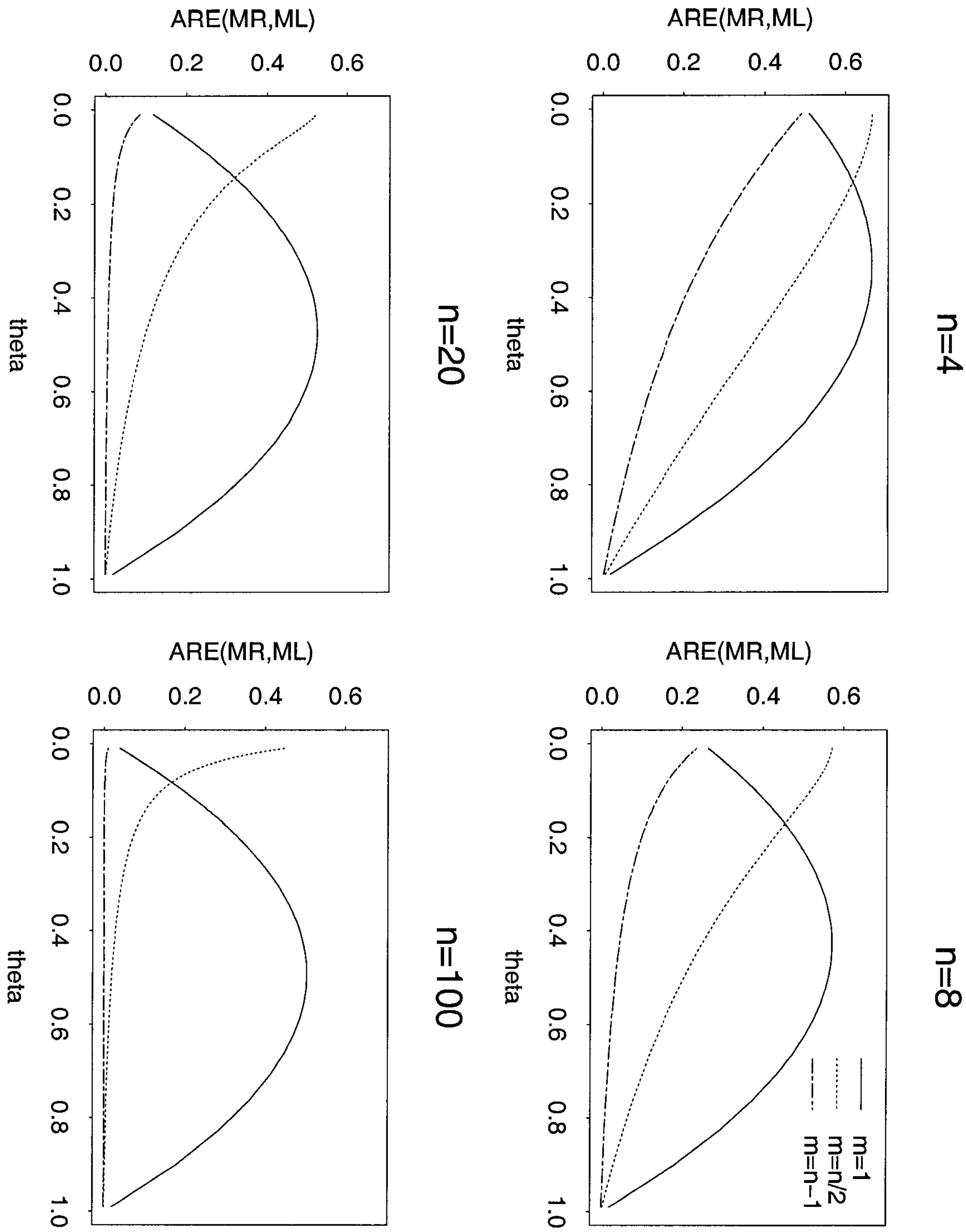

Figure 3.3: Graph of $ARE(MR, ML)$ versus θ, for $n = 4, 8, 20, 100$ and $m = 1, n/2, n-1$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.8 Unbalanced One-Way Random Effects Model

3.8.1 Introduction

In the previous section it was established that for a single sub-sample, the MR estimator of θ is a conditional MLE given the sub-sample group means. In this section, it will be shown through counter example, that this property does not hold for the unbalanced one-way random effects model.

The model for the unbalanced random effects model is given by

$$y_{ij} = \mu + u_i + e_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i. \tag{3.8.1}$$

Reverting to linear algebraic notation, let the response data vector be $\mathbf{y}$. Let design matrix for the fixed effects parameter be $\mathbf{X} = \mathbf{1}_N$, where $\sum_{i=1}^{k} n_i = N$. Let $\mathbf{u}$ be the random effects vector with the design matrix $\mathbf{Z} = \bigoplus_{i=1}^{k} \mathbf{1}_{n_i}$. And let the covariance matrices for $\mathbf{u}$ and $\mathbf{e}$ be, $\mathbf{D} = \mathbf{I}_k\sigma_a^2$ and $\mathbf{R} = \mathbf{I}_N\sigma_e^2$, respectively.

Then the MME for the unbalanced one-way random effects model reduced to

$$\frac{1}{\sigma_e^2}\begin{bmatrix} \sum_{i=1}^{k} n_i & (n_1, n_2, \ldots, n_k) \\ (n_1, n_2, \ldots, n_k)' & \mathbf{I_k}\lambda + diag(n_1, n_2, \ldots, n_k) \end{bmatrix}\begin{bmatrix} \mu \\ \mathbf{u} \end{bmatrix} = \frac{1}{\sigma_e^2}\begin{bmatrix} y_{..} \\ y_{i.} \end{bmatrix}, \tag{3.8.2}$$

where $\lambda = \sigma_e^2/\sigma_a^2$.

From row 1 of the MME :

$$\mu\sum_{i=1}^{k} n_i + \sum_{i=1}^{k} n_i u_i = y_{..}.$$

From the $(i+1)^{th}$ row of the MME:

$$n_i\mu + (n_i + \lambda)u_i = y_{i.}, \quad i = 2, \ldots, k+1. \tag{3.8.3}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Summing rows 2 through $k+1$ of the MME gives

$$\mu\sum_{i=1}^{k}n_i + \sum_{i=1}^{k}(n_i+\lambda)u_i = \sum_{i=1}^{k}y_{i.} = y_{..}. \tag{3.8.4}$$

Solving (1.8.3) for u_i gives the $BLUP$ of u_i, $\hat{u}_i$ as

$$\hat{u}_i = \frac{y_{i.} - n_i\hat{\mu}}{n_i+\lambda} = \frac{n_i}{n_i+\lambda}(\overline{y}_{i.} - \hat{\mu}). \tag{3.8.5}$$

Solving for μ in (1.8.4) gives the $BLUE$ for μ, $\hat{\mu}$ as

$$\hat{\mu} = \frac{y_{..} - \sum_{i=1}^{k} n_i u_i}{\sum_{i=1}^{k} n_i} = \overline{y}_{..} - \sum_{i=1}^{k}\frac{n_i}{N}u_i. \tag{3.8.6}$$

Substituting $\hat{u}_i$ from (1.8.5) for u_i in (1.8.6) we have,

$$\begin{aligned}\hat{\mu} &= \overline{y}_{..} - \sum_{i=1}^{k}\frac{n_i}{N}\frac{n_i}{n_i+\lambda}(\overline{y}_{i.} - \hat{\mu}) \\ &= \overline{y}_{..} - \sum_{i=1}^{k}\frac{n_i}{N}\frac{n_i}{n_i+\lambda}\overline{y}_{i.} + \sum_{i=1}^{k}\frac{n_i}{N}\frac{n_i}{n_i+\lambda}\hat{\mu}.\end{aligned}$$

Solving for $\hat{\mu}$ in terms of $\hat{u}_i$'s,

$$\begin{aligned}\hat{\mu} &= \frac{\overline{y}_{..} - \sum_{i=1}^{k}\frac{n_i}{N}\frac{n_i}{n_i+\lambda}\overline{y}_{i.}}{1 - \sum_{i=1}^{k}\frac{n_i}{N}\frac{n_i}{n_i+\lambda}} \\ &= \frac{\frac{1}{\sum_{i=1}^{k}n_i}\left[\sum_{i=1}^{k}n_i\overline{y}_{i.} - \sum_{i=1}^{k}\frac{n_i^2}{n_i+\lambda}\overline{y}_{i.}\right]}{\frac{1}{\sum_{i=1}^{k}n_i}\left[\sum_{i=1}^{k}n_i - \sum_{i=1}^{k}\frac{n_i^2}{n_i+\lambda}\right]} \\ &= \frac{\sum_{i=1}^{k}n_i\frac{\lambda}{n_i+\lambda}\overline{y}_{i.}}{\sum_{i=1}^{k}n_i\frac{\lambda}{n_i+\lambda}} \\ &= \frac{\sum_{i=1}^{k}\frac{y_{i.}}{n_i+\lambda}}{\sum_{i=1}^{k}\frac{n_i}{n_i+\lambda}}.\end{aligned} \tag{3.8.7}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Notice that the equations for $\hat{\mu}$ and $\hat{u}_i$'s depend on the unknown parameter λ, which is a function of the unknown variances, σ_e^2 and σ_a^2. Unlike the balanced model, where $\hat{\theta}_{MR}$ was computed from a closed form solution, $\hat{\theta}_{MR}$ for the unbalanced model must be computed using the iterative methods. Closed form solutions for estimates of θ do not exist for either ML or REML.

To show that for the unbalanced random effects model, $\hat{\theta}_{MR}$ is not a conditional MLE, we will compute $\hat{\theta}_{MR}$ and compute the conditional MLE of θ when regressing whole sample group means on sub-sample group means. Before presenting the results, we will give the equations for the conditional mean and covariance, and likelihood function, for the whole sample means given the sub-sample means.

3.8.2 Conditional Likelihood for Whole Sample Means Given Sub-Sample Means

Assuming that the distribution of the whole data vector $\mathbf{y}$ is multivariate normal, with mean $\mathbf{X}\mu$ and covariance $\mathbf{V}_y$. In terms of the variance components, the covariance of $\mathbf{y}$ is

$$\begin{aligned}
\mathbf{V}_y &= \mathbf{ZDZ}' + \mathbf{R} \\
&= \left[\bigoplus_{i=1}^{k} \mathbf{1}_{n_i}\right] \mathbf{I}_k \sigma_a^2 \left[\bigoplus_{i=1}^{k} \mathbf{1}'_{n_i}\right] + \mathbf{I}_N \sigma_e^2 \\
&= \sigma_a^2 \bigoplus_{i=1}^{k} \mathbf{J}_{n_i} + \sigma_e^2 \bigoplus_{i=1}^{k} \mathbf{I}_{n_i} \\
&= \sigma_e^2 \left[\bigoplus_{i=1}^{k} (\mathbf{I}_{n_i} + \lambda^{-1} \mathbf{J}_{n_i})\right] .
\end{aligned}$$

Let the vector of whole sample means be denoted by $\overline{\mathbf{y}}$, with

$$\overline{\mathbf{y}} = \begin{bmatrix} \overline{y}_1 \\ \vdots \\ \overline{y}_k \end{bmatrix} = diag(1/n_i)_{i=1}^{k} \mathbf{Z}'\mathbf{y} .$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Let the selection matrix for the sub-sample data be $\mathbf{G} = \bigoplus_{i=1}^{k} \begin{pmatrix} \mathbf{I}_{m_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$. Then the vector of sub-sample means, denoted by $\tilde{\mathbf{y}}$, can be written in terms of $\mathbf{y}$ as

$$\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_k \end{bmatrix} = diag(1/m_i)_{i=1}^{k} \mathbf{Z}'\mathbf{G}\mathbf{y} \ .$$

Computing the elements of $var\begin{pmatrix} \overline{\mathbf{y}} \\ \tilde{\mathbf{y}} \end{pmatrix}$,

$$\begin{aligned} var(\overline{\mathbf{y}}) &= diag(1/n_i)_{i=1}^{k}\ \mathbf{Z}'\mathbf{V}\mathbf{Z}\ diag(1/n_i)_{i=1}^{k}, \\ var(\tilde{\mathbf{y}}) &= diag(1/m_i)_{i=1}^{k}\ \mathbf{Z}'\mathbf{G}\mathbf{V}\mathbf{G}\mathbf{Z}\ diag(1/m_i)_{i=1}^{k}, \\ cov(\overline{\mathbf{y}}, \tilde{\mathbf{y}}) &= diag(1/n_i)_{i=1}^{k}\ \mathbf{Z}'\mathbf{V}\mathbf{G}\mathbf{Z}\ diag(1/m_i)_{i=1}^{k}. \end{aligned}$$

The conditional mean for $\overline{\mathbf{y}}$ given $\tilde{\mathbf{y}}$ is

$$\begin{aligned} E(\overline{\mathbf{y}} \mid \tilde{\mathbf{y}}) &= E(\overline{\mathbf{y}}) + cov(\overline{\mathbf{y}}, \tilde{\mathbf{y}}) var(\tilde{\mathbf{y}})^{-}(\tilde{\mathbf{y}} - E(\tilde{\mathbf{y}})) \\ &= diag(1/n_i)_{i=1}^{k}\ \mathbf{Z}'\mathbf{X}\mu + diag(1/n_i)_{i=1}^{k}\ \mathbf{Z}'\mathbf{V}\mathbf{G}\mathbf{Z}\ diag(1/m_i)_{i=1}^{k} \\ &\quad diag(m_i)_{i=1}^{k}\ (\mathbf{Z}'\mathbf{G}\mathbf{V}\mathbf{G}\mathbf{Z})^{-}\ diag(m_i)_{i=1}^{k} \\ &\quad (\tilde{\mathbf{y}} - diag(1/m_i)_{i=1}^{k}\ \mathbf{Z}'\mathbf{G}\mathbf{X}\mu) \\ &= diag(1/n_i)_{i=1}^{k}(n_i)_{i=1}^{k}\mu + diag(1/n_i)_{i=1}^{k} \left[\sigma_e^2 diag\left(m_i\left(\frac{n_i+\lambda}{\lambda}\right)\right)_{i=1}^{k}\right] \\ &\quad \left[\sigma_e^2 diag\left(m_i\left(\frac{m_i+\lambda}{\lambda}\right)\right)_{i=1}^{k}\right]^{-1} diag(m_i)_{i=1}^{k}(\tilde{\mathbf{y}} - diag(1/m_i)_{i=1}^{k}\ (m_i)_{i=1}^{k}\mathbf{X}\mu) \\ &= \mathbf{1}_k\mu + diag\left(\frac{m_i}{n_i}\frac{n_i+\lambda}{m_i+\lambda}\right)_{i=1}^{k} (\tilde{\mathbf{y}} - \mathbf{1}_k\mu) \\ &= diag\left(\frac{\lambda(n_i-m_i)}{n_i(m_i+\lambda)}\right)_{i=1}^{k} \mathbf{1}_k\mu + diag\left(\frac{m_i}{n_i}\frac{n_i+\lambda}{m_i+\lambda}\right)_{i=1}^{k} \tilde{\mathbf{y}}. \end{aligned} \tag{3.8.8}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The conditional variance for $\overline{\mathbf{y}}$ given $\tilde{\mathbf{y}}$ is

$$
\begin{aligned}
var(\overline{\mathbf{y}} \mid \tilde{\mathbf{y}}) &= var(\overline{\mathbf{y}}) - cov(\overline{\mathbf{y}}, \tilde{\mathbf{y}}) var(\tilde{\mathbf{y}})^{-} cov(\tilde{\mathbf{y}}, \overline{\mathbf{y}}) \\
&= diag(1/n_i)_{i=1}^{k} diag\left(n_i \frac{n_i+\lambda}{\lambda}\right)_{i=1}^{k} diag(1/n_i)_{i=1}^{k} \sigma_e^2 \\
&\quad - diag(1/n_i)_{i=1}^{k} diag\left(m_i \frac{n_i+\lambda}{\lambda}\right)_{i=1}^{k} diag(1/m_i)_{i=1}^{k} \sigma_e^2 \\
&\quad diag(m_i)_{i=1}^{k} diag\left(m_i \frac{\lambda}{m_i+\lambda}\right)_{i=1}^{k} diag(m_i)_{i=1}^{k} \sigma_e^{-2} \\
&\quad diag(1/m_i)_{i=1}^{k} diag\left(m_i \frac{n_i+\lambda}{\lambda}\right)_{i=1}^{k} diag(1/n_i)_{i=1}^{k} \sigma_e^2
\end{aligned}
$$

(Continues on the following page.)

$$
\begin{aligned}
var(\overline{\mathbf{y}} \mid \tilde{\mathbf{y}}) &= diag\left(\frac{n_i+\lambda}{n_i\lambda}\right)_{i=1}^{k} \sigma_e^2 - diag\left(\frac{m_i}{n_i^2} \frac{(n_i+\lambda)^2}{m_i+\lambda} \frac{1}{\lambda}\right)_{i=1}^{k} \sigma_e^2 \\
&= \sigma_e^2 \, diag\left[\frac{n_i+\lambda}{n_i\lambda}\left(1 - \frac{m_i}{n_i} \frac{n_i+\lambda}{m_i+\lambda}\right)\right]_{i=1}^{k} \\
&= \sigma_e^2 \, diag\left[\frac{n_i+\lambda}{n_i\lambda} \frac{\lambda(n_i-m_i)}{n_i(m_i+\lambda)}\right]_{i=1}^{k} \\
&= \sigma_e^2 \, diag\left[\frac{(n_i-m_i)(n_i+\lambda)}{n_i^2(m_i+\lambda)}\right]_{i=1}^{k}.
\end{aligned}
\qquad (3.8.9)
$$

As a side note, for the balanced random effects model, the conditional mean and variance of $\overline{\mathbf{y}}$ given $\tilde{\mathbf{y}}$ are

$$
E(\overline{\mathbf{y}} \mid \tilde{\mathbf{y}}) = \mu \frac{\lambda(n-m)}{n(m+\lambda)} \mathbf{1}_k + \frac{m}{n} \frac{(n+\lambda)}{(m+\lambda)} \tilde{\mathbf{y}} \qquad (3.8.10)
$$

and

$$
var(\overline{\mathbf{y}} \mid \tilde{\mathbf{y}}) = \frac{(n-m)(n+\lambda)}{n^2(m+\lambda)} \sigma_e^2 \mathbf{I}_k. \qquad (3.8.11)
$$

Therefore the simple linear regression function of $\overline{\mathbf{y}}$ on $\tilde{\mathbf{y}}$ is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\bar{\mathbf{y}} = \mathbf{1}_k\beta_0 + \tilde{\mathbf{y}}\beta_1 + \epsilon_{\bar{\mathbf{y}}|\tilde{\mathbf{y}}},$$

where $\beta_0 = \mu\frac{\lambda(n-m)}{n(m+\lambda)}$ and $\beta_1 = \frac{m}{n}\frac{(n+\lambda)}{(m+\lambda)}$ and $\epsilon_{\bar{\mathbf{y}}|\tilde{\mathbf{y}}} \sim N_k(\mathbf{0}, \frac{(n-m)(n+\lambda)}{n^2(m+\lambda)}\sigma_e^2\mathbf{I}_k)$. Notice that $\beta_1 = \frac{m}{n}\beta_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$.

Thus, the conditional maximum likelihood estimator of θ based on the regression of $\bar{\mathbf{y}}$ on $\tilde{\mathbf{y}}$ is derived by maximizing the conditional likelihood of $\bar{\mathbf{y}}$ given $\tilde{\mathbf{y}}$,

$$l = -\frac{1}{2}\left[klog2\pi + log|\mathbf{V}_{\bar{\mathbf{y}}|\tilde{\mathbf{y}}}| + (\bar{\mathbf{y}} - \mu_{\bar{\mathbf{y}}|\tilde{\mathbf{y}}})'\mathbf{V}_{\bar{\mathbf{y}}|\tilde{\mathbf{y}}}^{-1}(\bar{\mathbf{y}} - \mu_{\bar{\mathbf{y}}|\tilde{\mathbf{y}}})\right],$$

or equivalently, minimizing

$$log|\mathbf{V}_{\bar{\mathbf{y}}|\tilde{\mathbf{y}}}| + (\bar{\mathbf{y}} - \mu_{\bar{\mathbf{y}}|\tilde{\mathbf{y}}})'\mathbf{V}_{\bar{\mathbf{y}}|\tilde{\mathbf{y}}}^{-1}(\bar{\mathbf{y}} - \mu_{\bar{\mathbf{y}}|\tilde{\mathbf{y}}}). \qquad (3.8.12)$$

3.8.3 Counterexample

A small data set is used to illustrate that for the unbalanced random effects model, the MR estimate of θ for a single sub-sample is not the same as the conditional MLE of θ based on the conditional distribution of whole sample means given the sub-sample means. It is also shown for this data set, that by increasing the degree of unbalancedness, the discrepancy between the two estimators also increases.

The algorithm outlined in Section 2.4 was used to compute the MR estimate of θ. The conditional likelihood function of $\bar{\mathbf{y}}$ given $\tilde{\mathbf{y}}$, equation 3.8.12 was maximized using Maple V, with respect to μ and λ. $\hat{\theta}$ was then evaluated from $\hat{\lambda}$, using equation 3.4.1.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Maple V was used to compute both the MR estimates and conditional ML estimates of θ. The sub-sample for each case, were generated by taking a 50% random sub-samples from each group.

The first data set used was balanced, which served as a check for the results in Section 3.4. Unbalanced data sets of were then generated from the first by simply eliminating observations. For simplicity, group sizes kept at even values. Table 3.1 summarizes the results for the 4 data sets of increasing degree of unbalancedness.

Table 3.1: Method R and ML estimates of θ for data sets of increasing unbalancedness.

Whole Group Sizes	$\hat{\theta}_{MR}$	$\hat{\theta}_{ML}$	$\lvert\hat{\theta}_{MR} - \hat{\theta}_{ML}\rvert$
8,8,8,8	0.6819744368	0.6819744470	0.0000000102
8,8,8,6	0.6795280231	0.6811937246	0.0016657015
8,8,6,6	0.6470206082	0.6511951541	0.0041745459
8,8,6,4	0.8209108412	0.7986876067	0.0222232345

From table 3.1, we see that for the balanced design, estimates are equal to order 10^-7. For all the unbalanced designs, the two estimators are clearly not the same. As the degree of unbalancedness increases, it appears that so too, does the absolute error between the two estimators.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 4

SIMULATIONS FOR THE ONE-WAY MODEL FOR SMALL TO MODERATE SAMPLE SIZE

4.1 Introduction

In this chapter simulated data sets are used to evaluate the performance of MR in comparison to ML and REML for small to moderate sized samples. Data sets were generated for the one-way random effects model for several values of $\theta = \frac{\sigma_a^2}{\sigma_a^2+\sigma_e^2}$, and for six different error distributions. In Section 3.7.1 the motivation for multiple sub-sample based MR estimation was pointed out. The mean and median of MR estimates from repeated sub-samples were used as MR based estimators. For simulated data sets the comparative attributes of the two MR estimators, ML and REML, are examined through their derived empirical sampling distributions. In the latter part of this chapter MR based interval estimation for the one-way random effects model is discussed and illustrative results for bootstrap confidence intervals for the one-way model are presented.

4.2 Data Simulation

For the one-way balanced random effects model, data sets were simulated for each combination of number of groups $k = 4, 8, 20$; group size $n = 4, 8, 20$; standard deviation pairs $(\sigma_a, \sigma_e) = (1,4), (1,3), (1,2), (1,1), (2,1), (3,1), (4,1)$. For all simulations, the random effects vector was simulated so that $\mathbf{u}_{k\times 1} \sim N(\mathbf{0}_{k\times 1}, \mathbf{I}_k\sigma_a^2)$. The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

residual vector $\mathbf{e}_{N\times 1}$ was simulated so that $e_{ij} = \sigma_e(w_{ij} - E(w_{ij}))/\sqrt{var(w_{ij})}$, where w_{ij}'s are IID random variables from one of the following distributions.

- *Normal*

  $w_{ij} \sim N(0,1)$ for $i = 1, \ldots, k; j = 1, \ldots, n$; $E(w_{ij}) = 0$ and $var(w_{ij}) = 1$.

- *Student's-t, df=3*

  For $df = \nu$, $w_{ij} \sim t_\nu$ for $i = 1, \ldots, k; j = 1, \ldots, n$; $E(t_\nu) = 0$ and $var(t_\nu) = \nu/(\nu - 2)$.

- *Chi-square, df=3 and 5*

  For $df = \nu$, $w_{ij} \sim \chi^2_\nu$ for $i = 1, \ldots, k; j = 1, \ldots, n; E(w_{ij}) = \nu$ and $var(w_{ij}) = 2\nu$.

- *Contaminated normal* $(p = .1, c = 3)$ *and* $(p = .5, c = 3)$

  Let $x_{ij} \sim binomial(N, p)$, $z1_{ij} \sim N(0,1)$ and $z2_{ij} \sim N(0,1)$, withx_{ij}, $z1_{ij}$ and $z2_{ij}$ mutually independent random variables. Then

  $$w_{ij} = x_{ij} z1_{ij} + c(1 - x_{ij}) z2_{ij}.$$

  has a *contaminated normal distribution with parameters p and c*, with $E(w_{ij}) = 0$ and $var(w_{ij}) = c^2 - (c^2 - 1)p$.

1000 data sets were generated for all combinations of n, k, θ and the error distributions above. ML and REML estimates for θ, denoted by $\hat{\theta}_{ML}$ and $\hat{\theta}_{REML}$, respectively, were computed under the normal likelihood for each data set. For the MR estimator, 1000 sub-samples of size $kn/2$ ($m = n/2$ from each group) were obtained for each data set. The mean and median of the $\hat{\theta}_{MR}$-values generated for 1000 sub-samples, denoted by $\hat{\theta}_{MR-AVG}$ and $\hat{\theta}_{MR-MED}$, were used as the MR based estimators for θ.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.3 Simulation Results

In this section, the results for the simulations described in Section 4.2 are presented. The first set of figures included illustrate how the four estimators perform with respect to median bias and mean bias. Next we examine performance with respect to measures of dispersion, through the standard error and root mean square error for selected estimators. The last part of this section looks at the relationship between dispersion and sub-sample size.

4.3.1 Median Bias

Figures 4.1-4.6 display the medians of the empirical distributions for the four point estimators of θ plotted against θ for six different error distributions, $n = 4, 8, 20$ and $k = 4, 8, 20$. The error distributions used are normal, $t_{(3)}$, $\chi^2_{(3)}$, $\chi^2_{(5)}$, contaminated normal with $p = .1,\ c = 3$ and contaminated normal with $p = .5,\ c = 3$, respectively.

For all error distributions, the most striking attribute of these plots is the decrease in median bias as k increases. In contrast, the relationship between median bias and n does not appear to be as noticeable. Of the four estimators the $\hat{\theta}_{MR-AVG}$ appears to be the most erratic. $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{REML}$ perform consistently well in almost all all cases. $\hat{\theta}_{ML}$ appears to have median bias that is always greater in magnitude than $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{REML}$.

The behavior of median bias is relatively consistent over all the simulated error distributions, with the exception of the $t_{(3)}$ and the contaminated normal with $p = .1,\ c = 3$. For the $t_{(3)}$ there is an overall slight positive shift in median bias. In the case of the contaminated normal with $p = .1,\ c = 3$, the shift the median bias appears to be downward, and more dramatic than for any of the simulated error distributions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

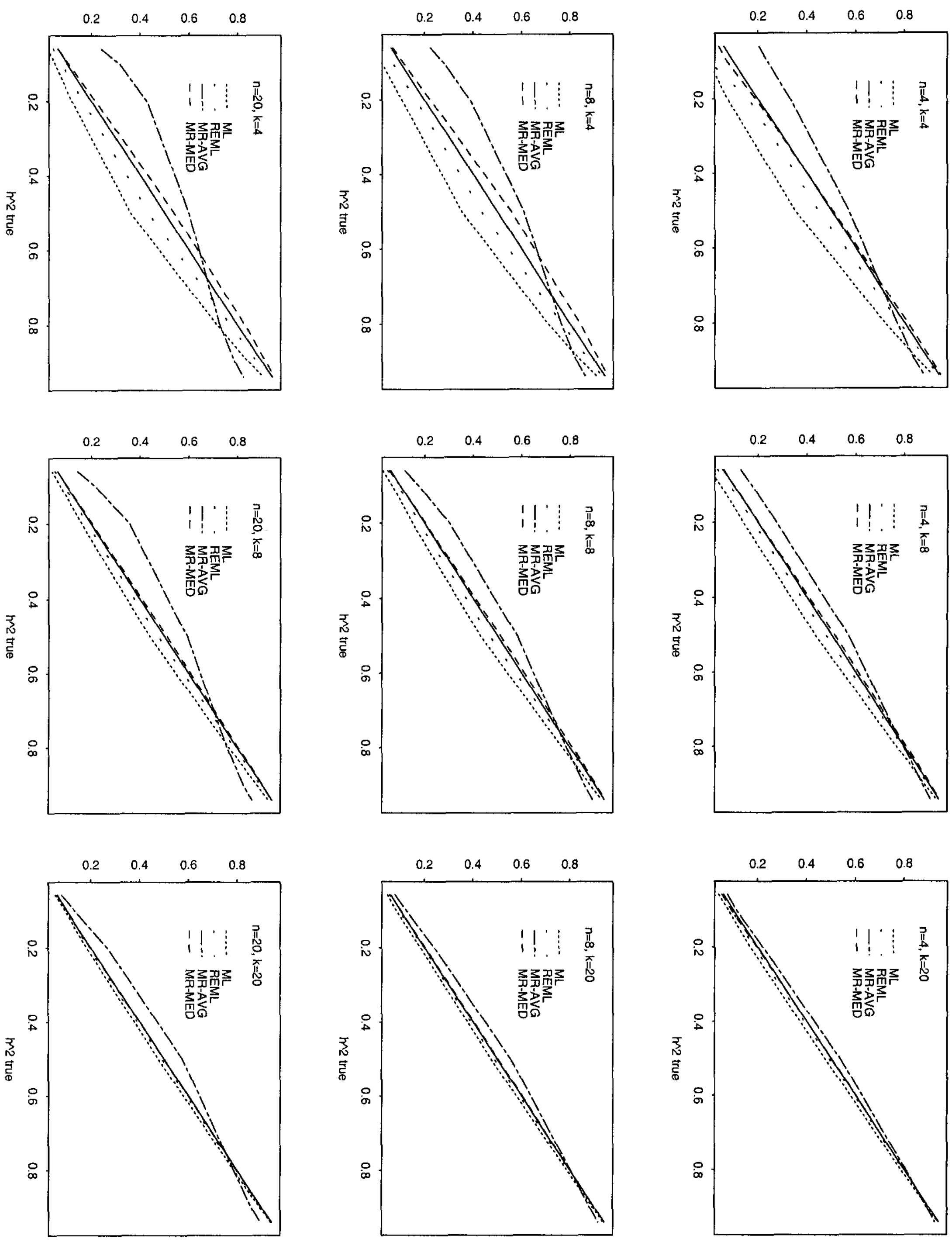

Figure 4.1: Normal random error simulation: median values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

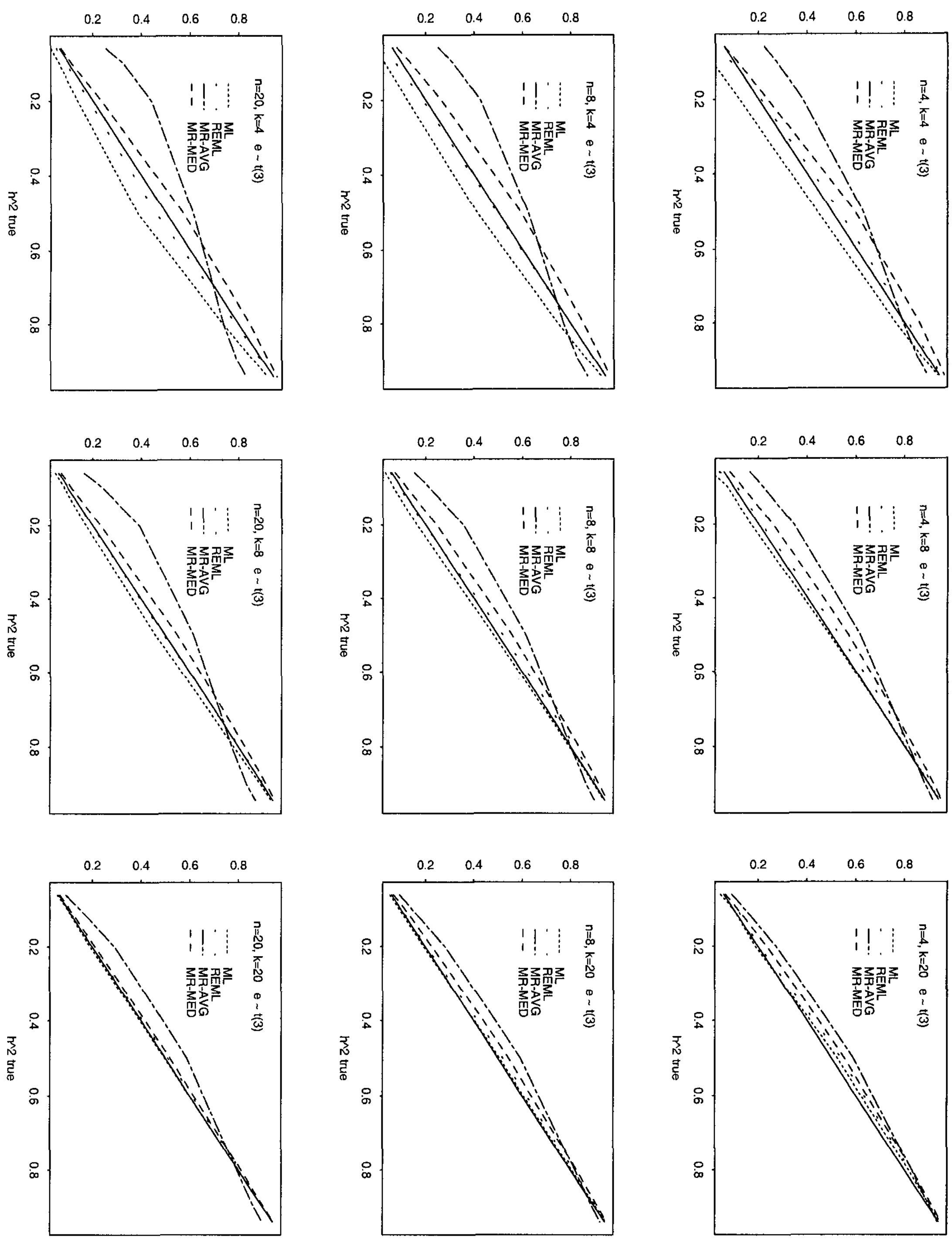

Figure 4.2: $t_{(3)}$ random error simulation: median values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

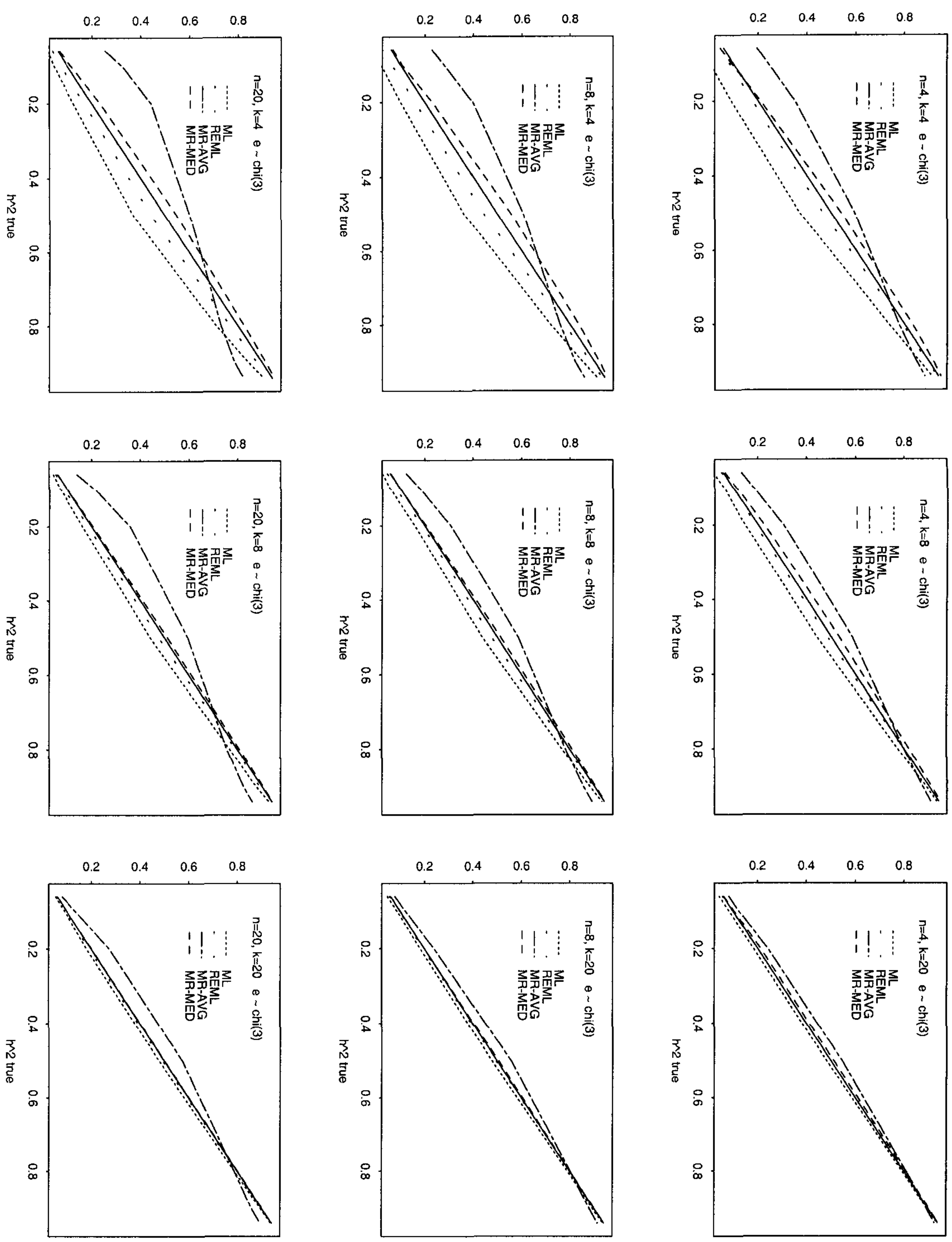

Figure 4.3: $\chi^2_{(3)}$ random error simulation: median values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

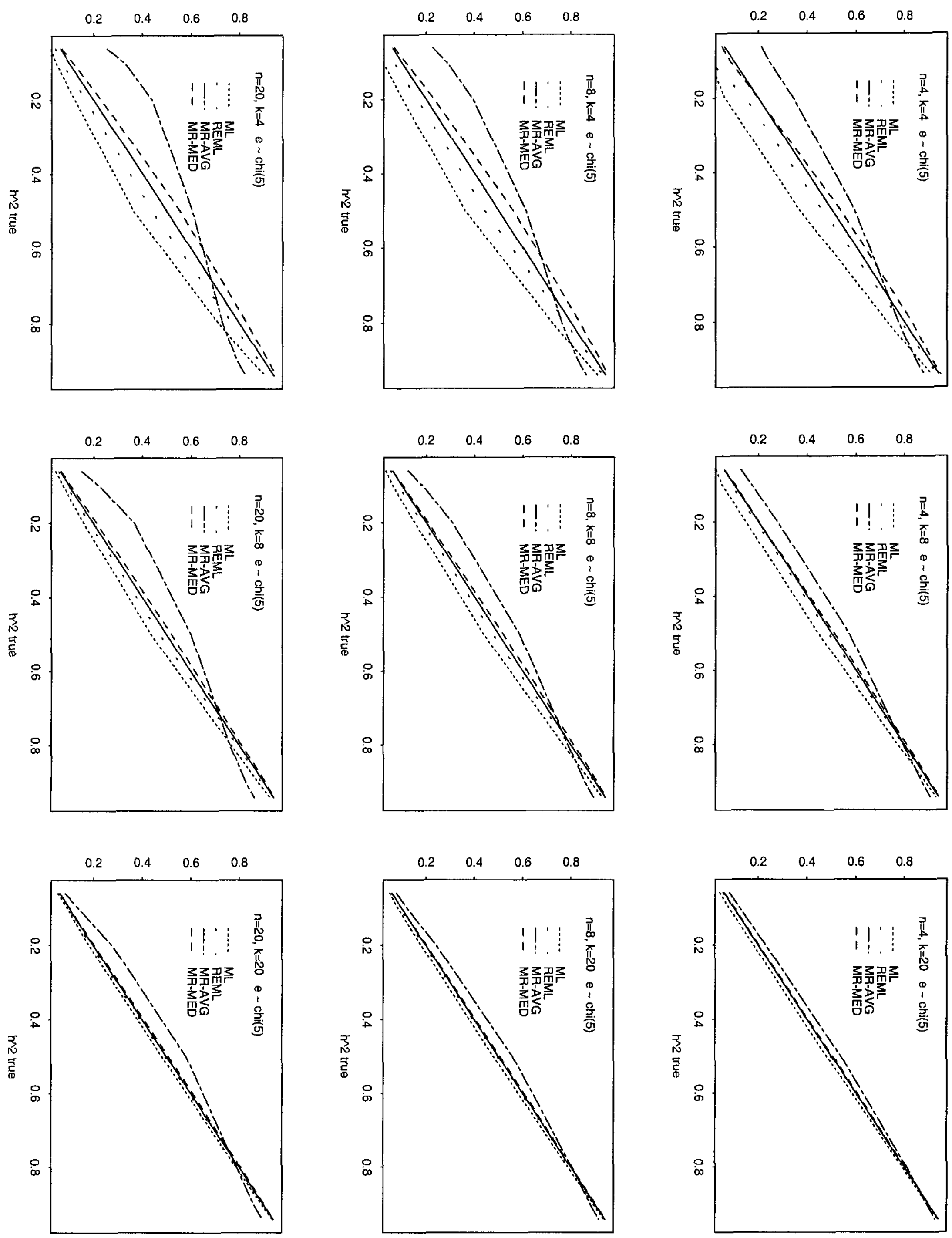

Figure 4.4: $\chi^2_{(5)}$ random error simulation: median values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

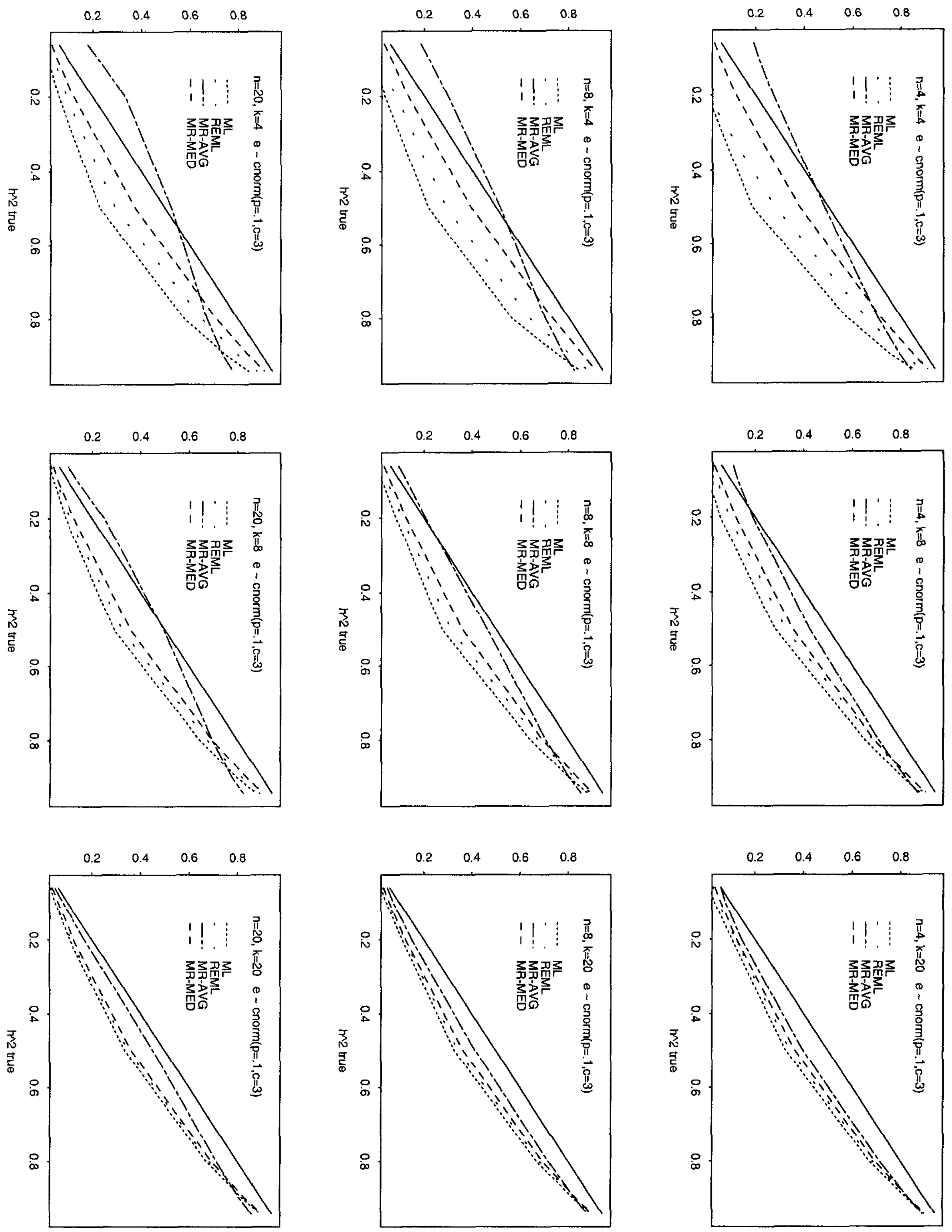

Figure 4.5: Contaminated normal ($p = .1, c = 3$) random error simulation: median values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

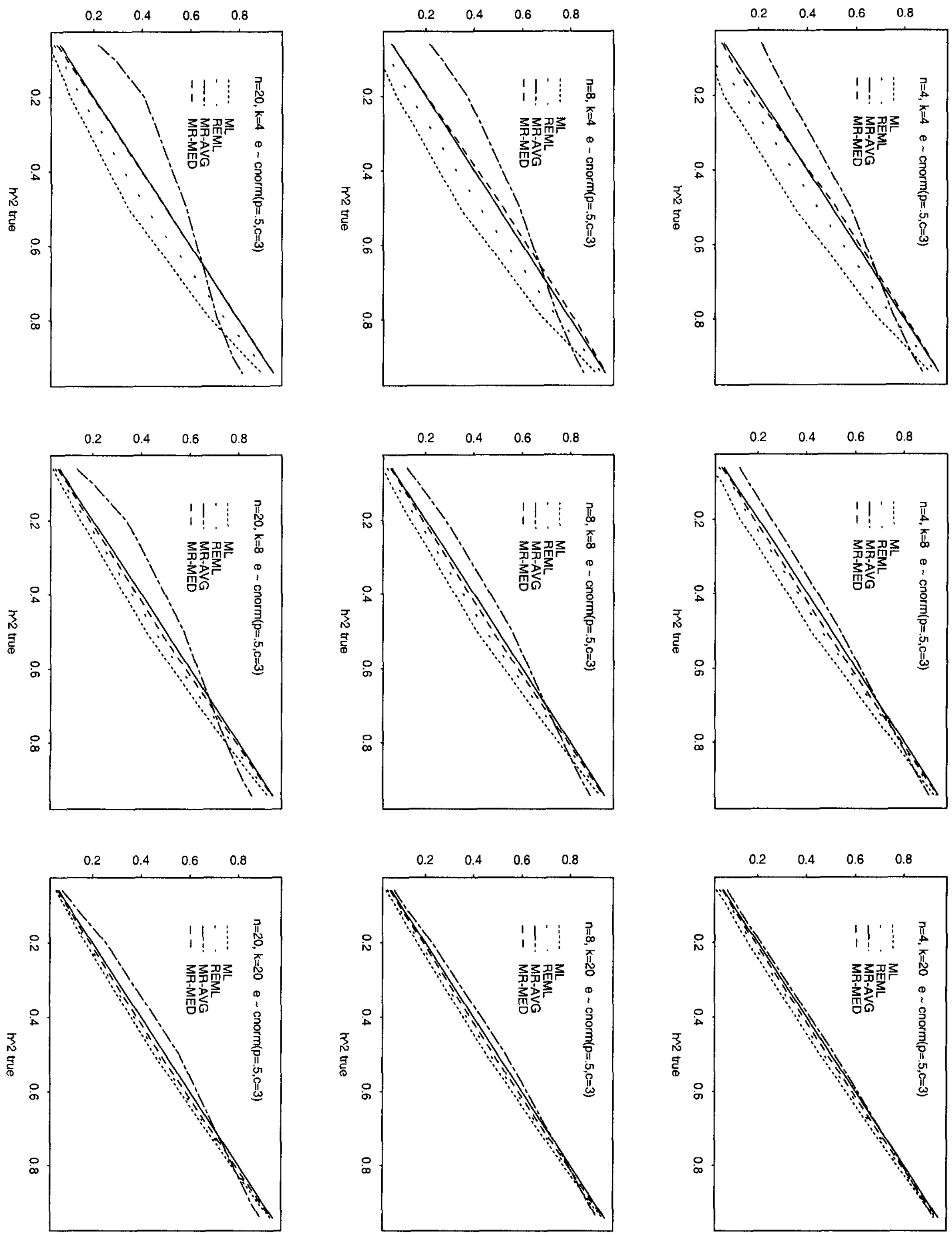

Figure 4.6: Contaminated normal ($p = .5, c = 3$) random error simulation: median values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.3.2 Mean Bias

Figures 4.7-4.12 display the means of the empirical distributions for the four point estimators of θ plotted against θ for six different error distributions, $n = 4, 8, 20$ and $k = 4, 8, 20$. The error distributions used are normal, $t_{(3)}$, $\chi^2_{(3)}$, $\chi^2_{(5)}$, contaminated normal with $p = .1,\ c = 3$ and contaminated normal with $p = .5,\ c = 3$, respectively. For all error distributions, as with median bias, mean bias decreases as k increases, but does not appear to be affected noticeably by n. Mean bias also appears to be influenced by θ. For small θ the mean bias appears to be to the right of that of the median bias, and to the left of median bias for large θ. This no doubt reflects the asymmetry in the empirical sampling distributions for these estimators. As observed with median bias, the error distributions that most effect mean bias are the $t_{(3)}$ and contaminated normal with $p = .1,\ c = 3$ distributions. A slight positive shift in mean bias is noted with the $t_{(3)}$ error distribution, and a more notable downward shift in mean bias with the contaminated normal with $p = .1,\ c = 3$ error distribution.

4.3.3 Standard Errors for REML and MR-MED

As $\hat{\theta}_{REML}$ and $\hat{\theta}_{MR-MED}$, are consistently the two best estimators of θ, with respect to median and mean bias, further comparisons will be confined to only these two estimators. Figures 4.13-4.18 display the standard errors of the empirical distributions of $\hat{\theta}_{REML}$ and $\hat{\theta}_{MR-MED}$ plotted against θ, for six different error distributions, $n = 4, 8, 20$ and $k = 4, 8, 20$. As before, the error distributions used are normal, $t_{(3)}$, $\chi^2_{(3)}$, $\chi^2_{(5)}$, contaminated normal with $p = .1,\ c = 3$ and contaminated normal with $p = .5,\ c = 3$, respectively.

For all error distributions, the standard errors of the two estimators decrease as both n and k increase. In fact they appear to converge to the same values at $k = 20$. For

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

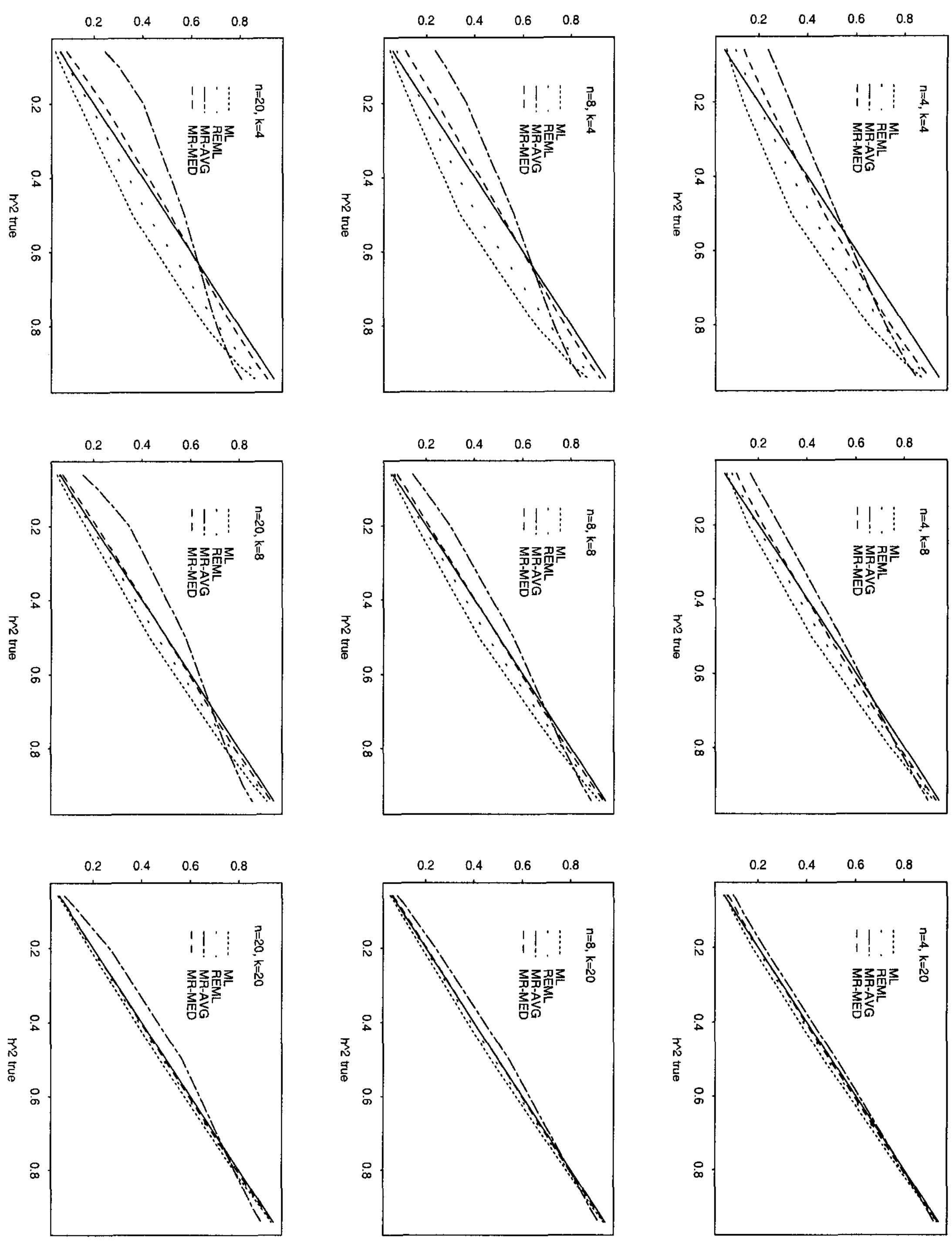

Figure 4.7: Normal random error simulation: mean values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

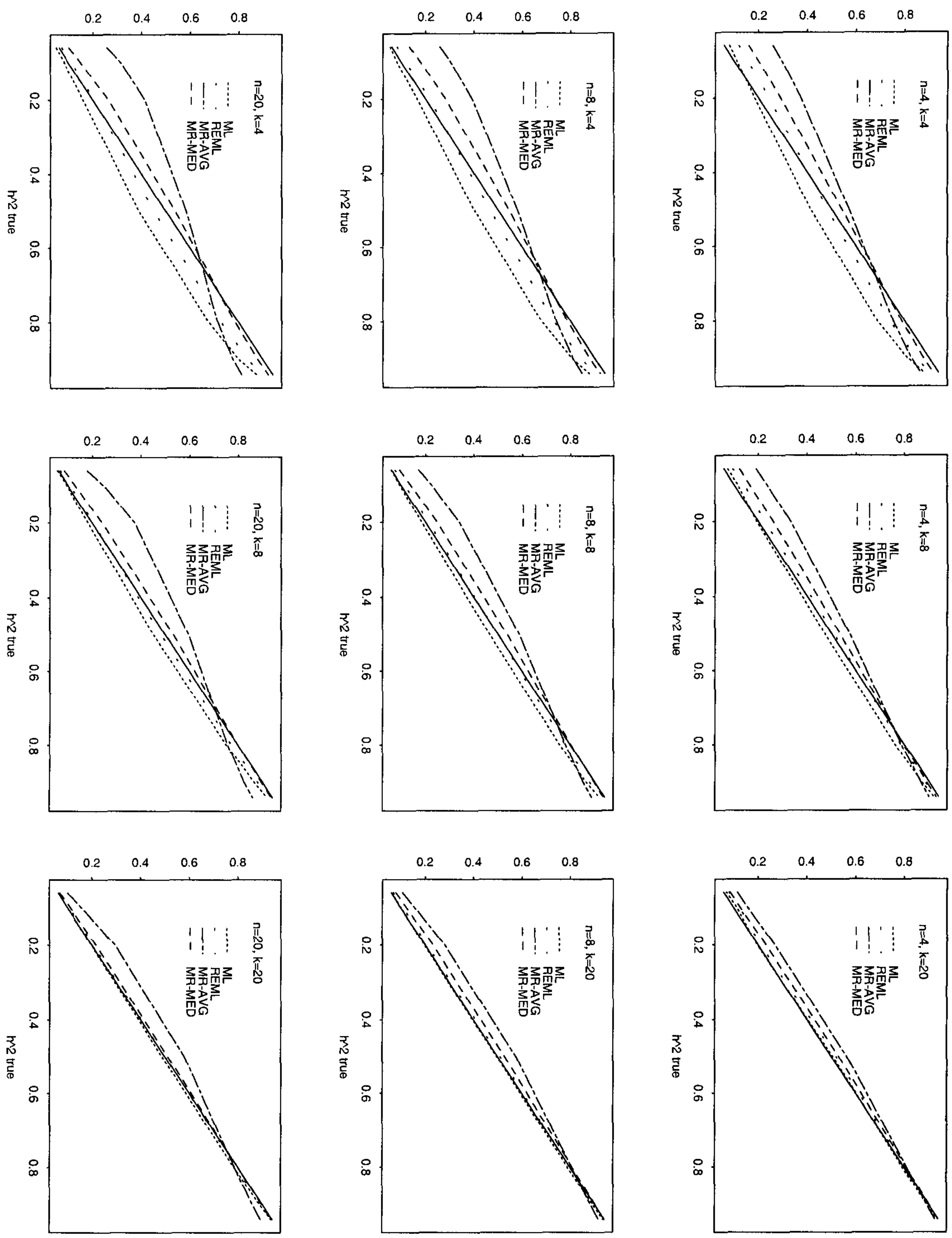

Figure 4.8: $t_{(3)}$ random error simulation: mean values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

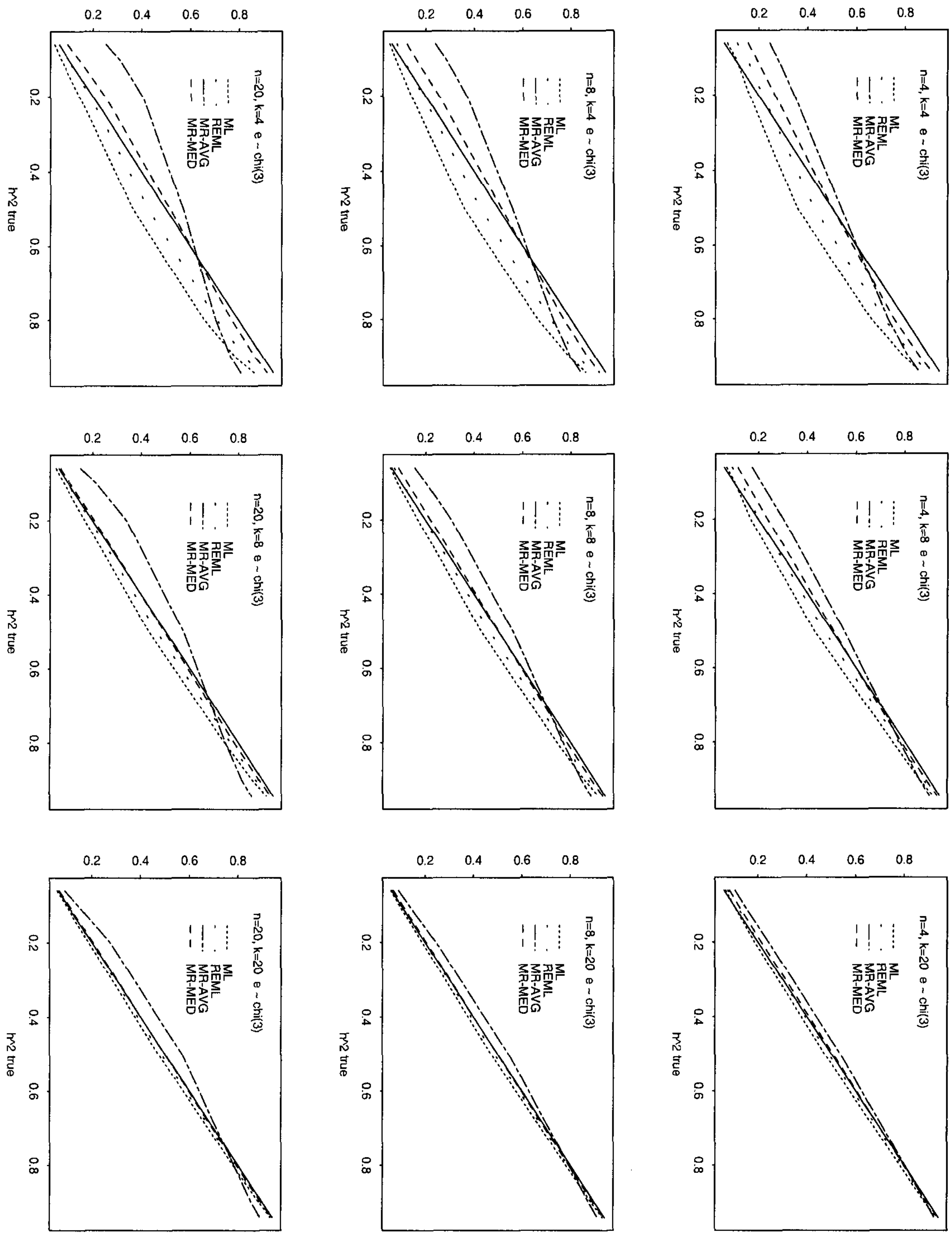

Figure 4.9: $\chi^2_{(3)}$ random error simulation: mean values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

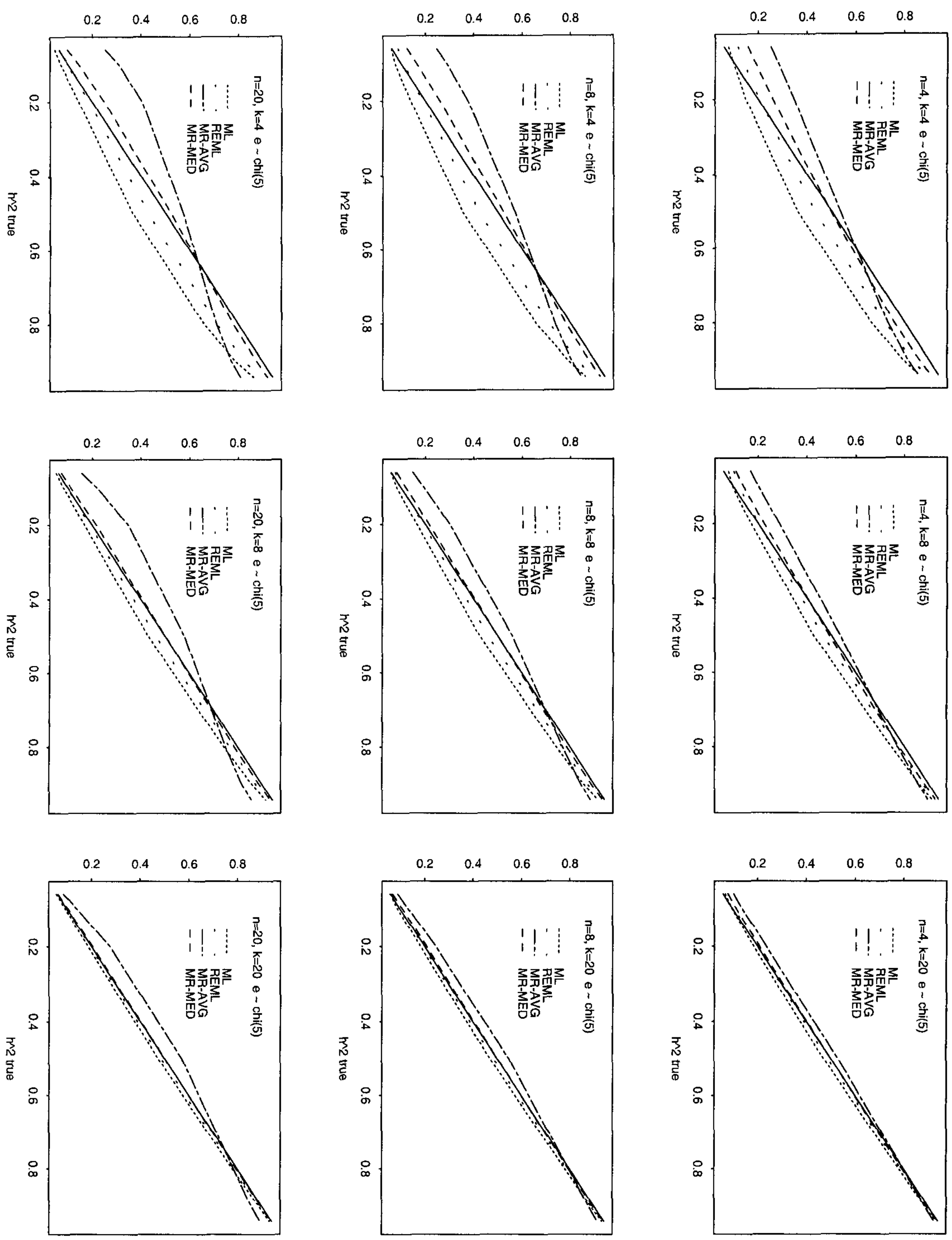

Figure 4.10: $\chi^2_{(5)}$ random error simulation: mean values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

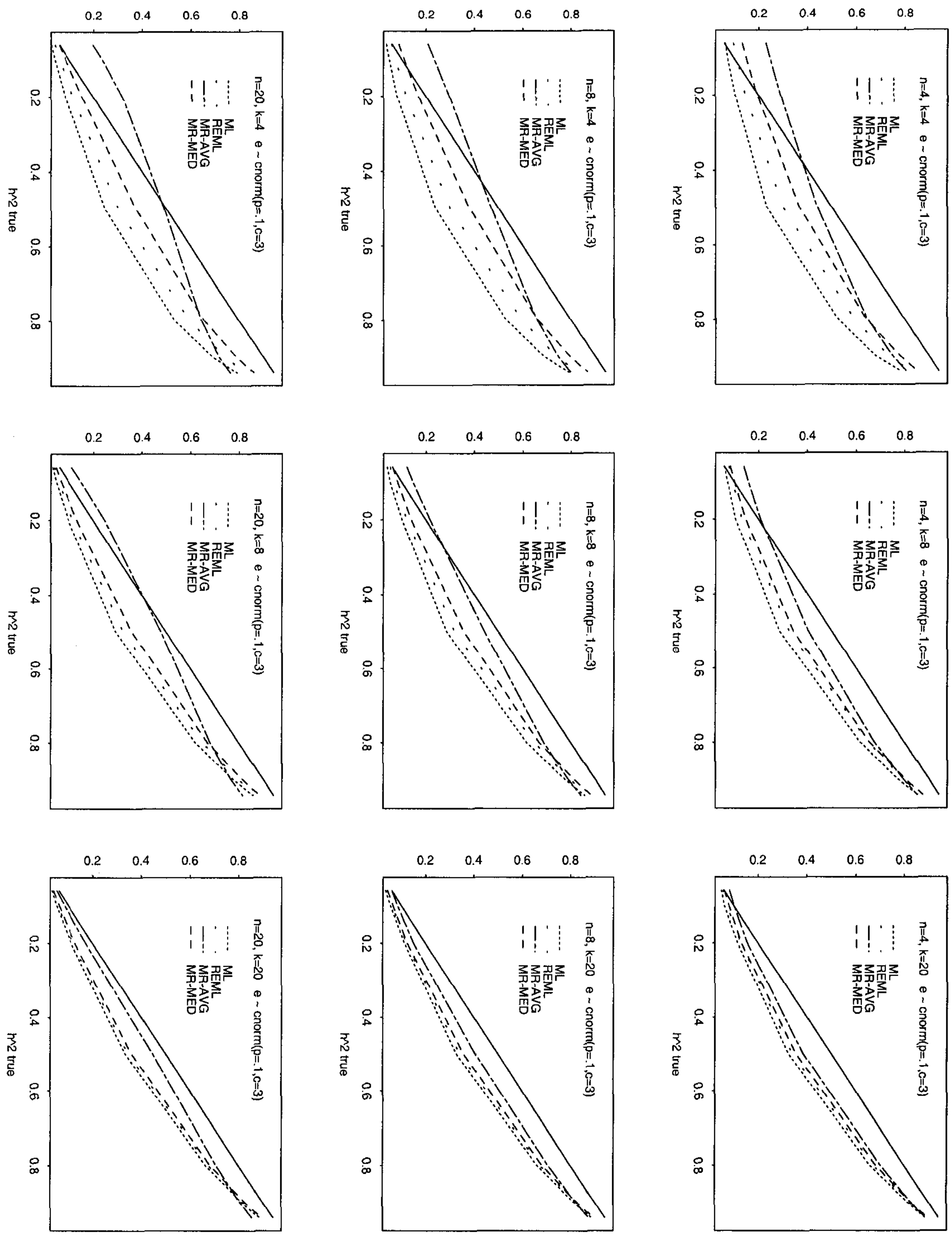

Figure 4.11: Contaminated normal ($p = .1, c = 3$) random error simulation: mean values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

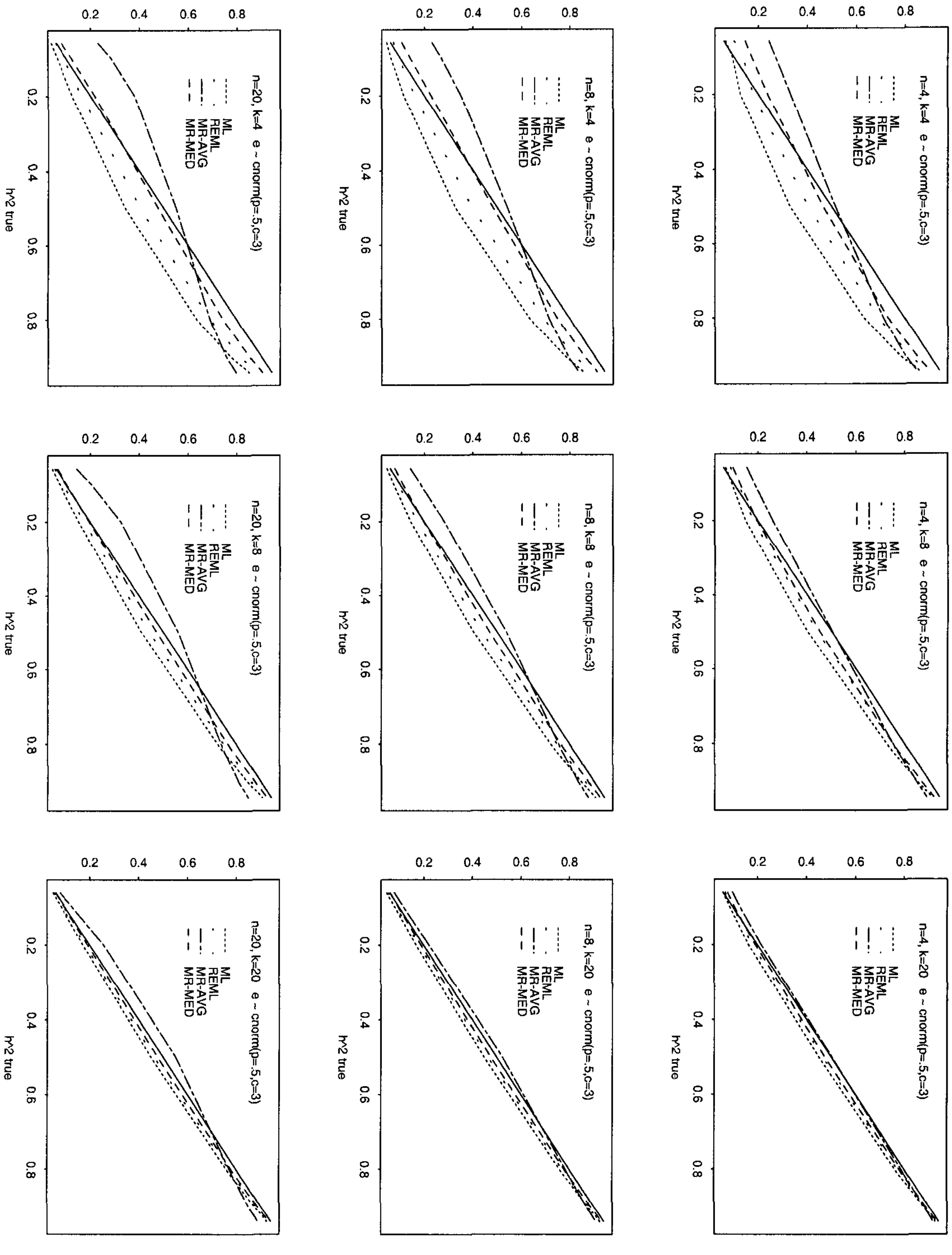

Figure 4.12: Contaminated normal ($p = .5, c = 3$) random error simulation: mean values for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

small n and k, the standard errors for REML appear to be consistently lower than that of MR-MED for small to moderate values of θ. However, at values of θ close to 1, the standard error for MR-MED is almost identical to that for REML, with the exception for the normal distribution, where the standard errors for REML are uniformly smaller than that for MR-MED for $n = 4$.

4.3.4 Root Mean Square Errors for REML and MR-MED

Figures 4.19-4.24 display the root mean square errors of the empirical distributions of $\hat{\theta}_{REML}$ and $\hat{\theta}_{MR-MED}$ plotted against θ, for six different error distributions, $n = 4, 8, 20$ and $k = 4, 8, 20$. As before, the error distributions used are normal, $t_{(3)}$, $\chi^2_{(3)}$, $\chi^2_{(5)}$, contaminated normal with $p = .1$, $c = 3$ and contaminated normal with $p = .5$, $c = 3$, respectively.

For all error distributions, as with the standard errors, the root mean square errors of the two estimators decrease as both n and k increase, and appear to converge to the same values at $k = 20$. For small n and k, the standard errors for REML appear to be consistently lower than that of MR-MED for small to moderate values of θ. However, at values of θ close to 1, the standard error for MR-MED is almost identical to that for REML, with the exception for the normal distribution, where the standard errors for REML are uniformly smaller than that for MR-MED for $n = 4$.

4.3.5 Sub-Sample Size Relationship With Dispersion

Figure 4.25 attempts to look at the effect that the sub-sample size, m, has on the precision of the MR based estimators, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$. Plotted are the values of the standard errors and the root mean square errors for the two estimators

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

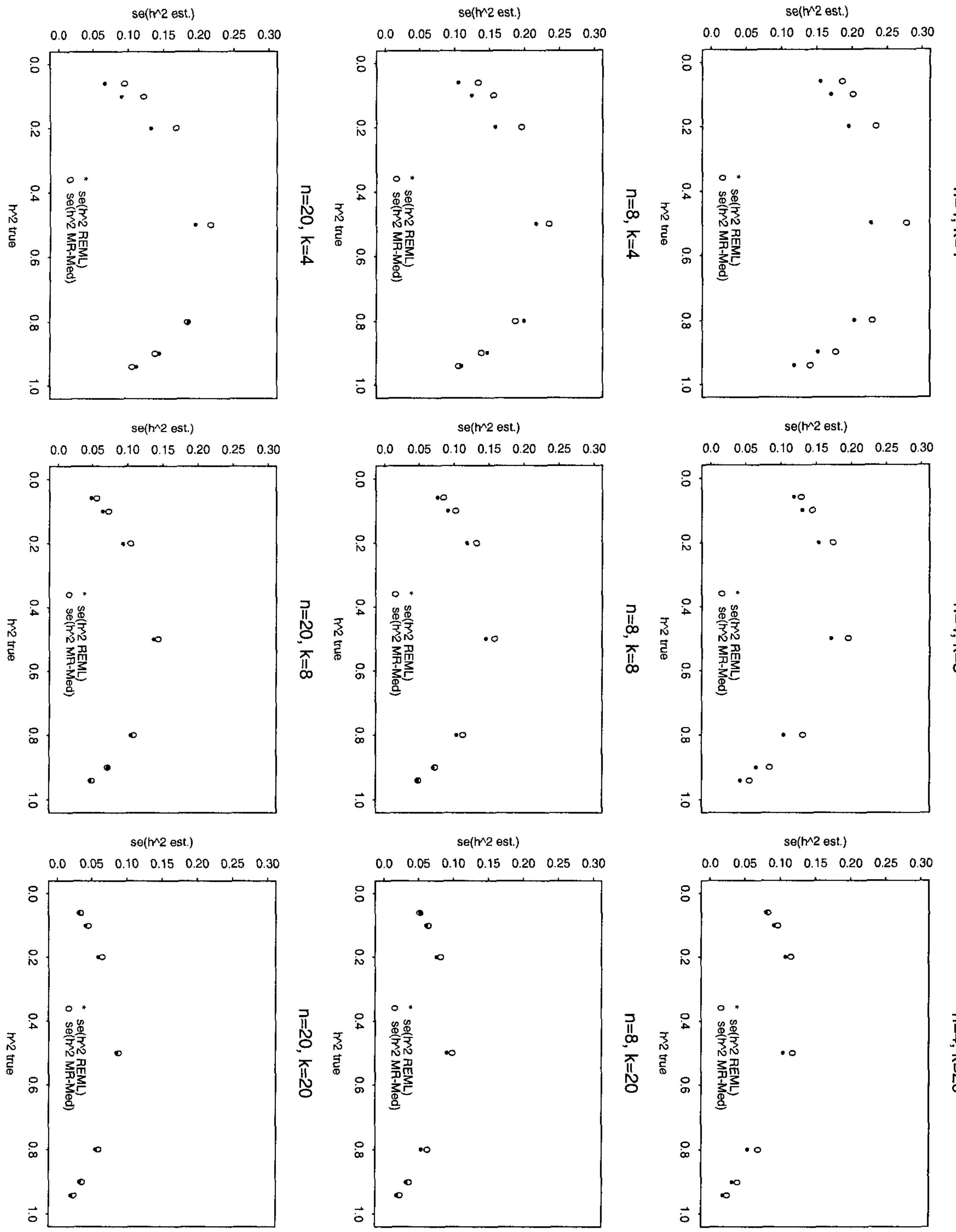

Figure 4.13: Normal random error simulation: standard errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

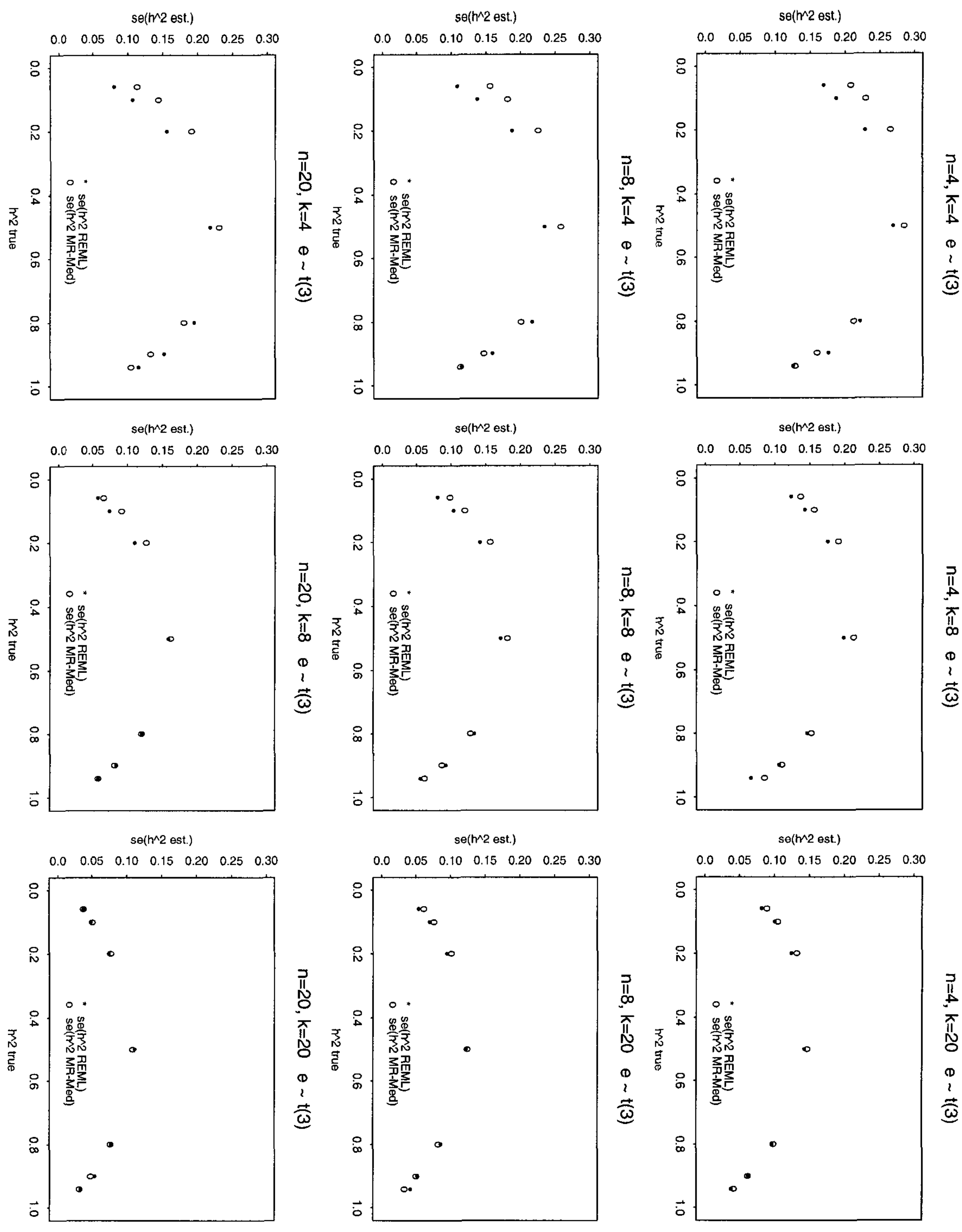

Figure 4.14: $t_{(3)}$ random error simulation: standard errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

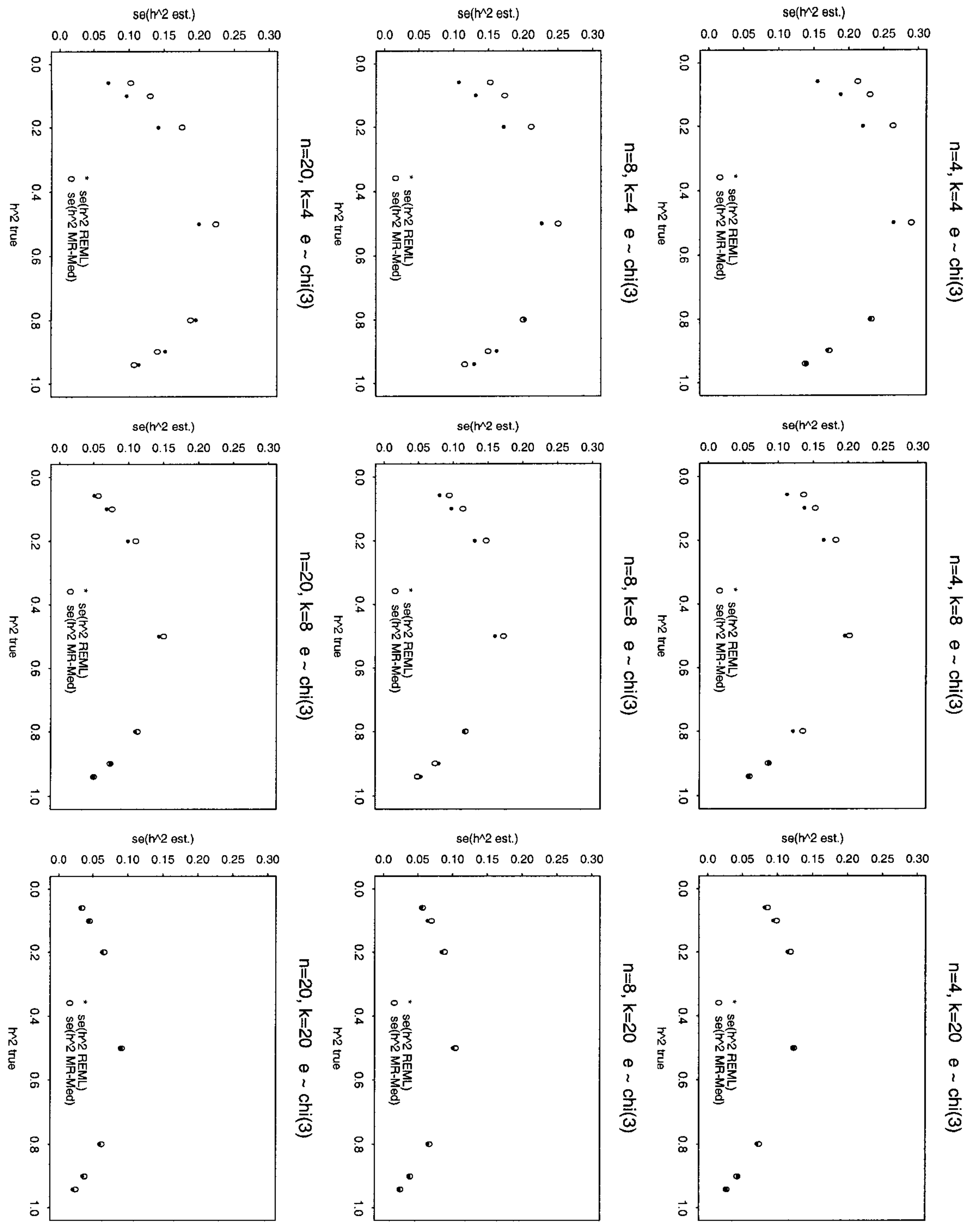

Figure 4.15: $\chi^2_{(3)}$ random error simulation: standard errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

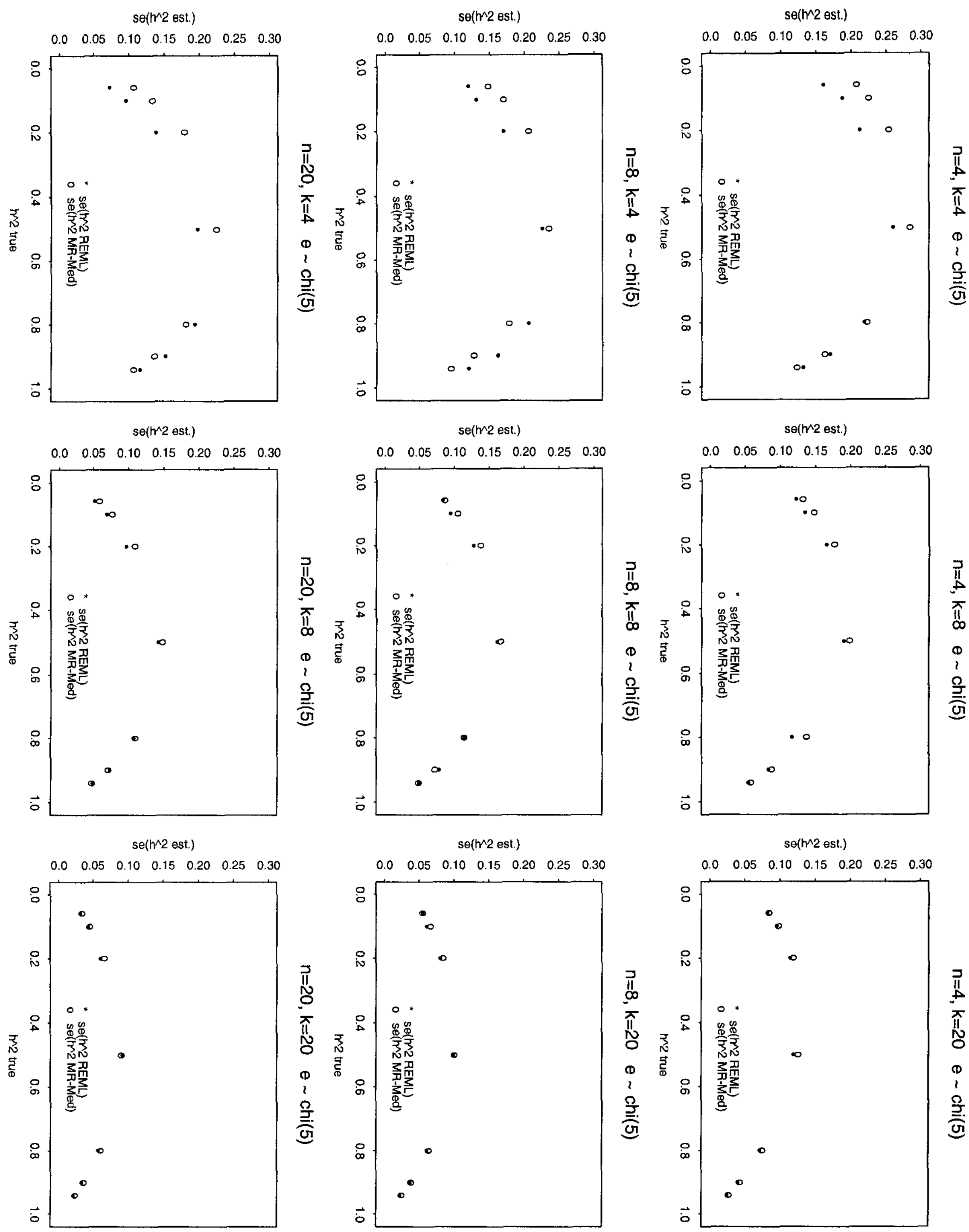

Figure 4.16: $\chi^2_{(5)}$ random error simulation: standard errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

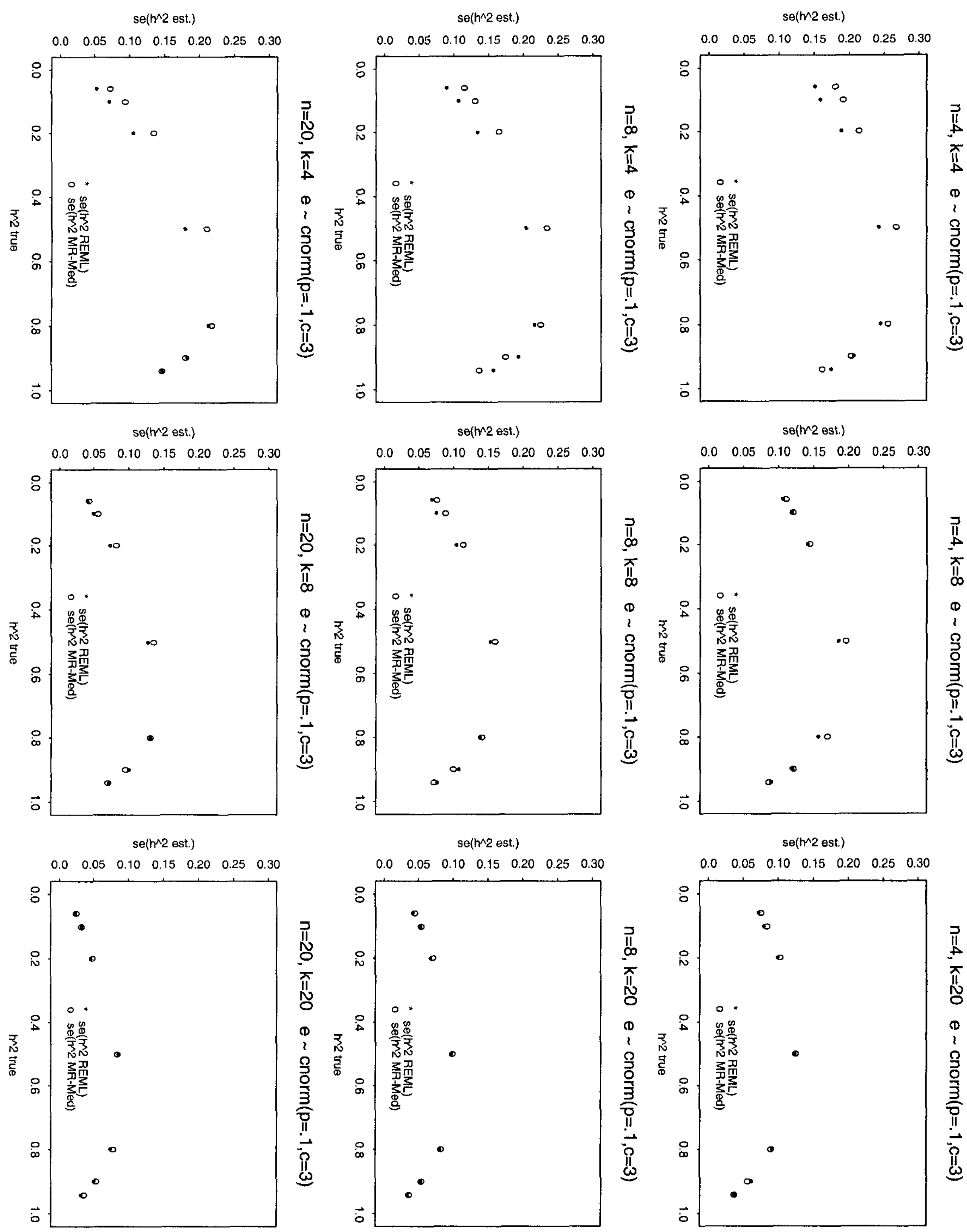

Figure 4.17: Contaminated normal ($p = .1, c = 3$) error simulation: standard errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

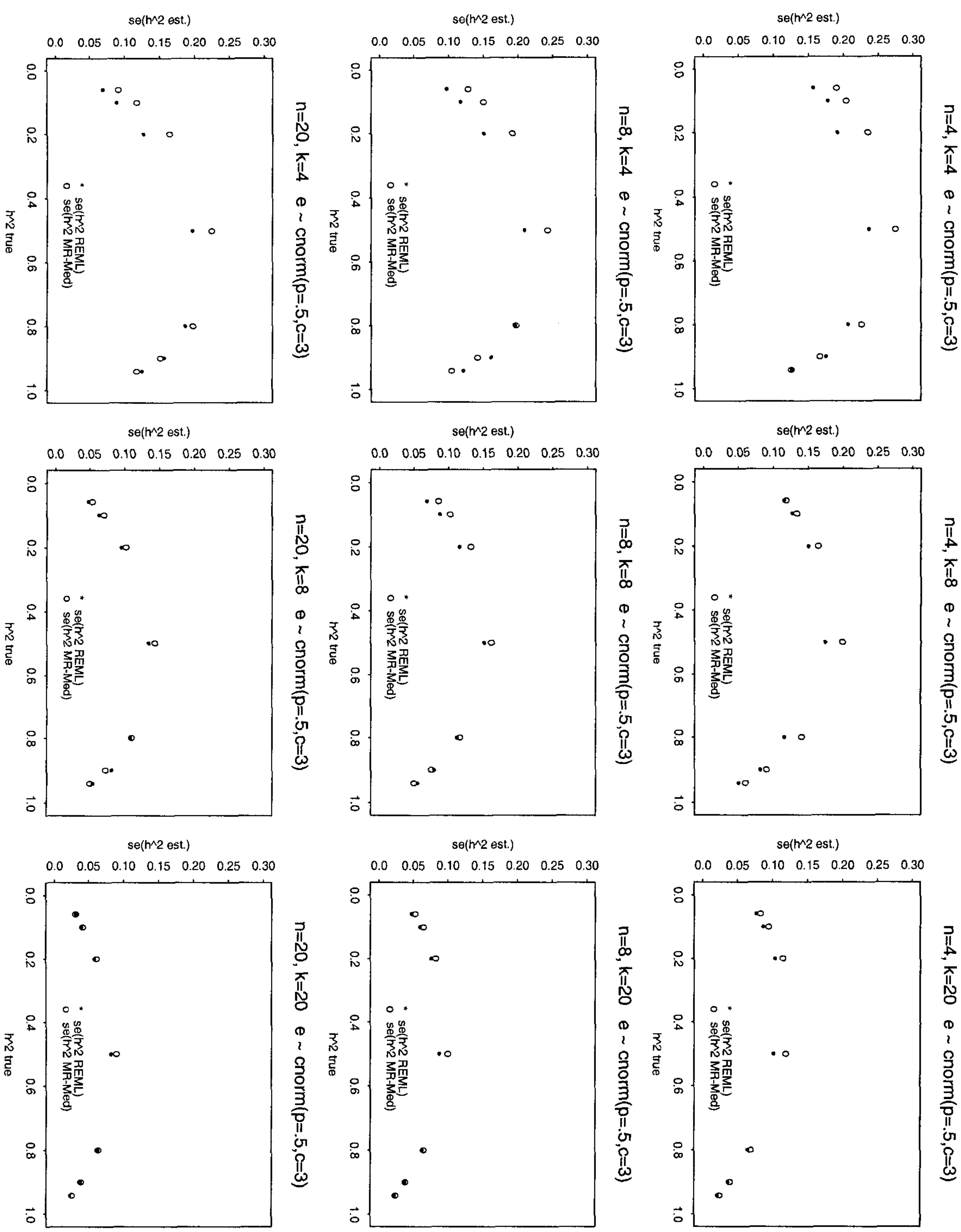

Figure 4.18: Contaminated normal ($p = .5, c = 3$) random error simulation: standard errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

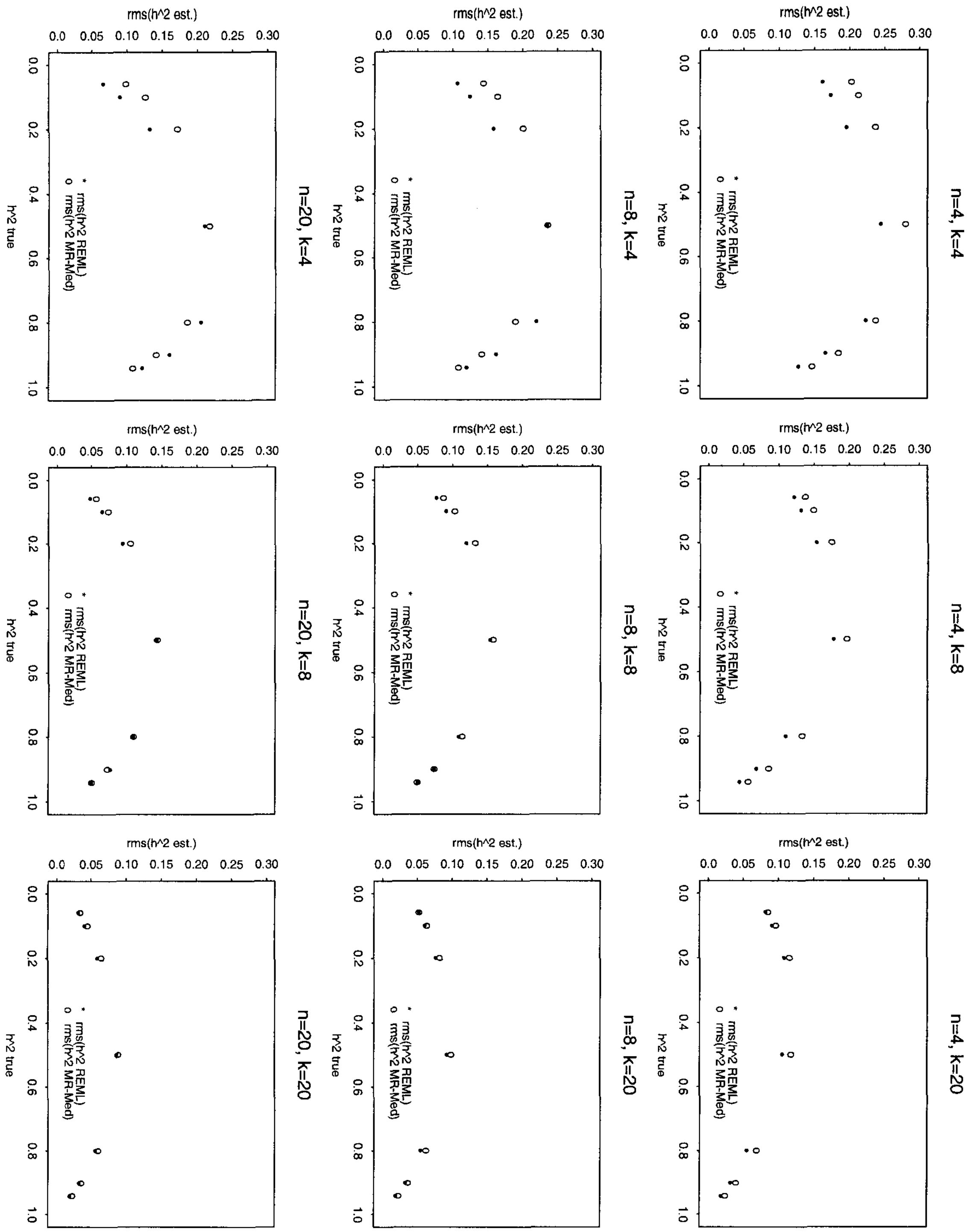

Figure 4.19: Normal random error simulation: root mean square errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

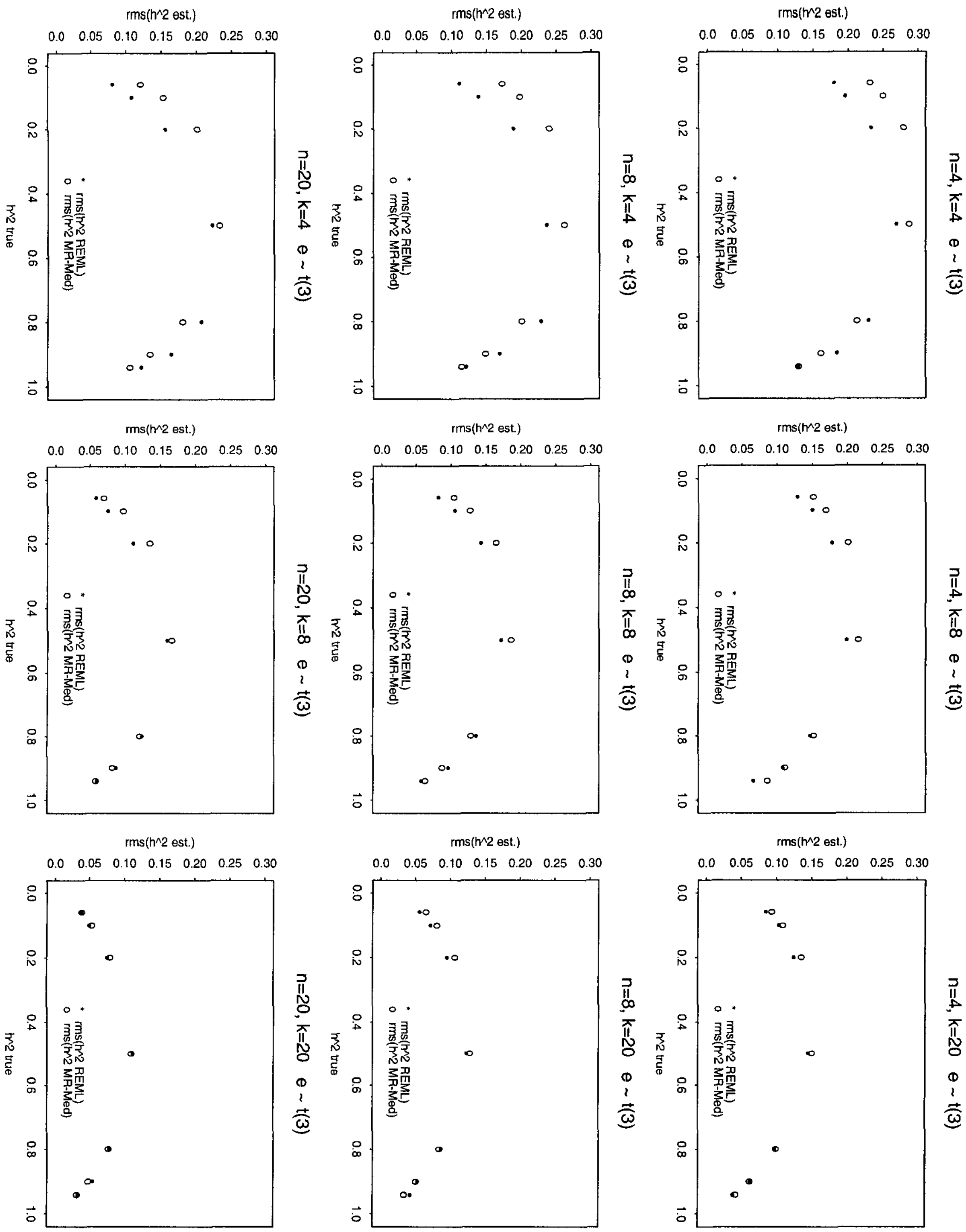

Figure 4.20: $t_{(3)}$ random error simulation: root mean square errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

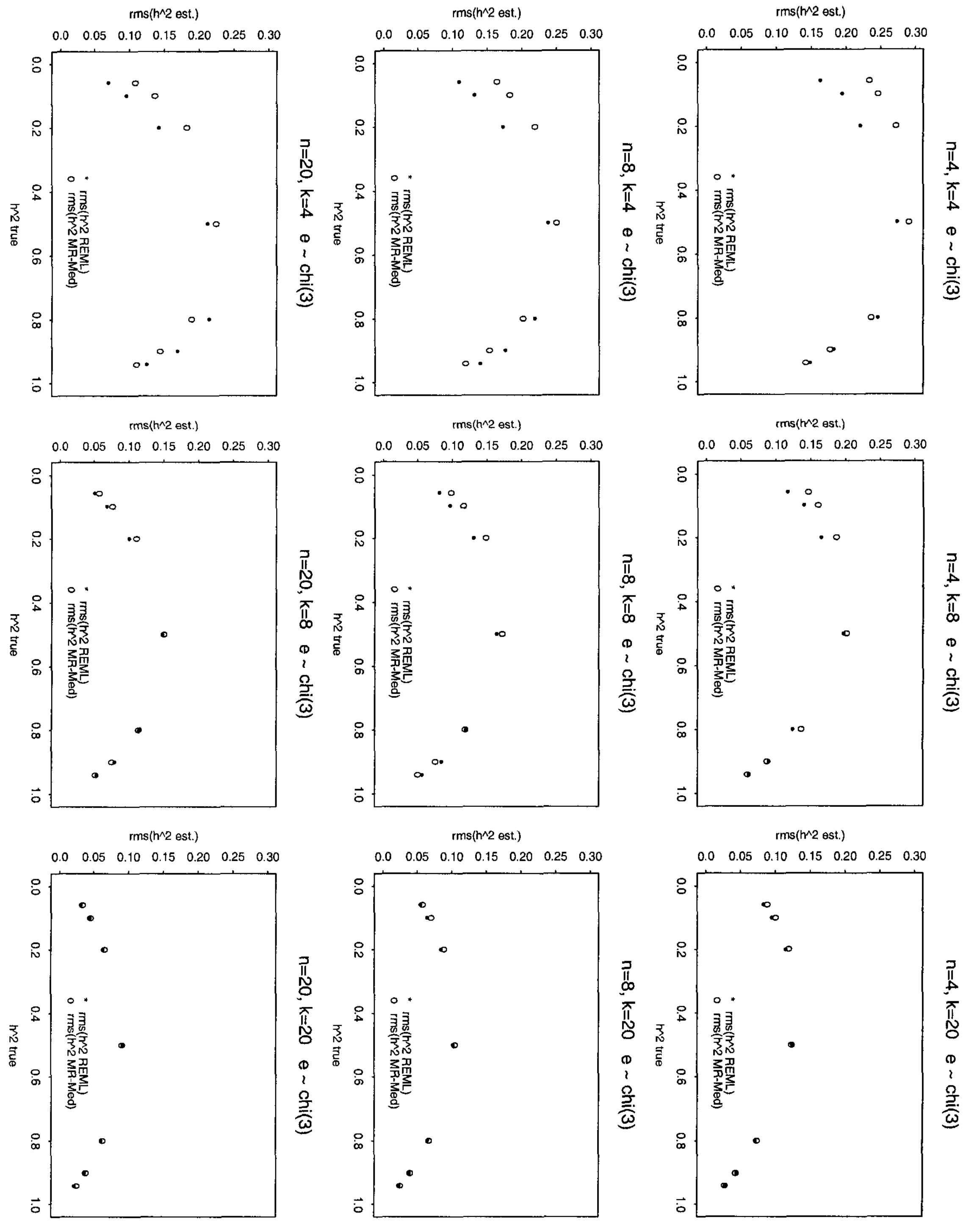

Figure 4.21: $\chi^2_{(3)}$ random error simulation: root mean square errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

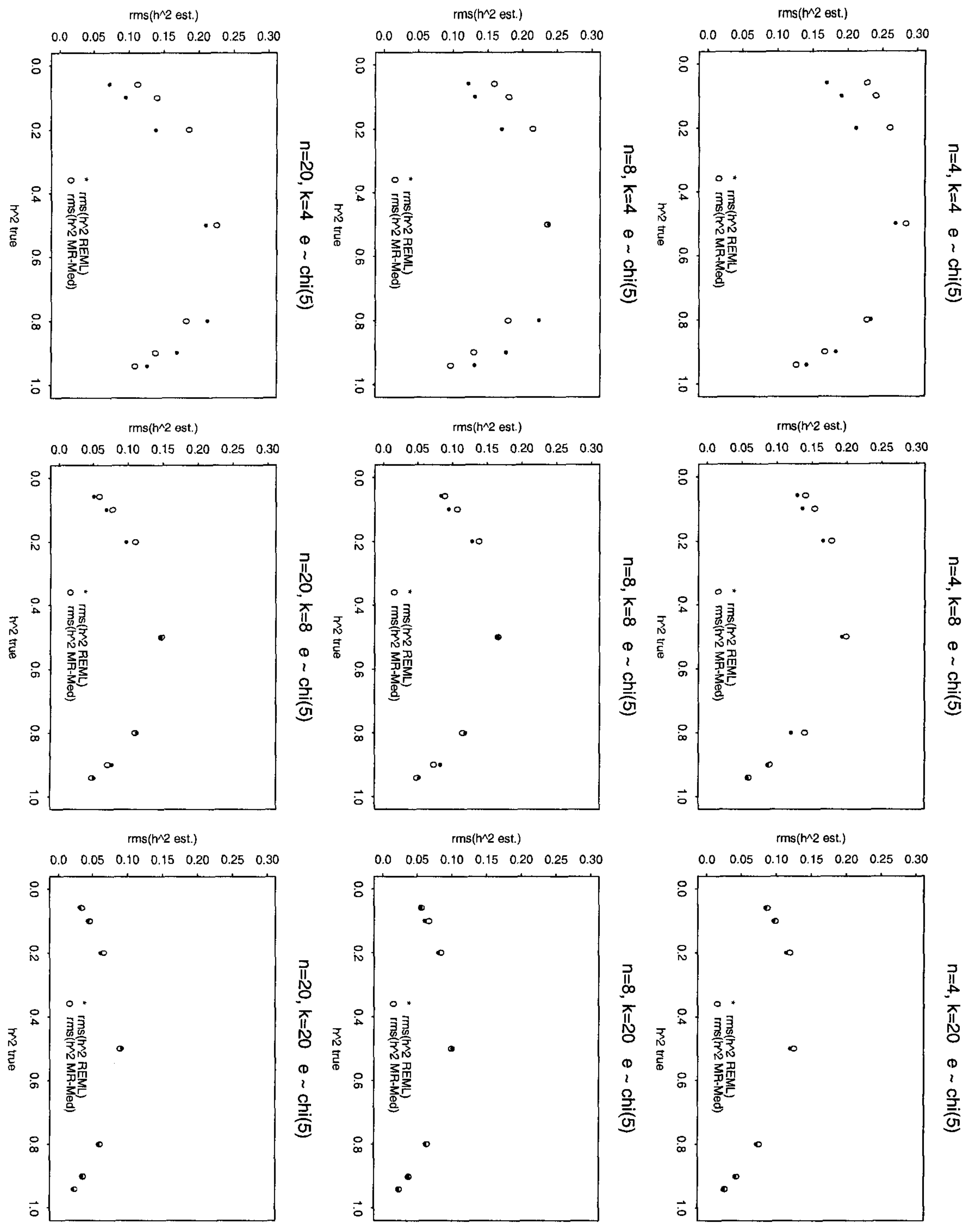

Figure 4.22: $\chi^2_{(5)}$ random error simulation: root mean square errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

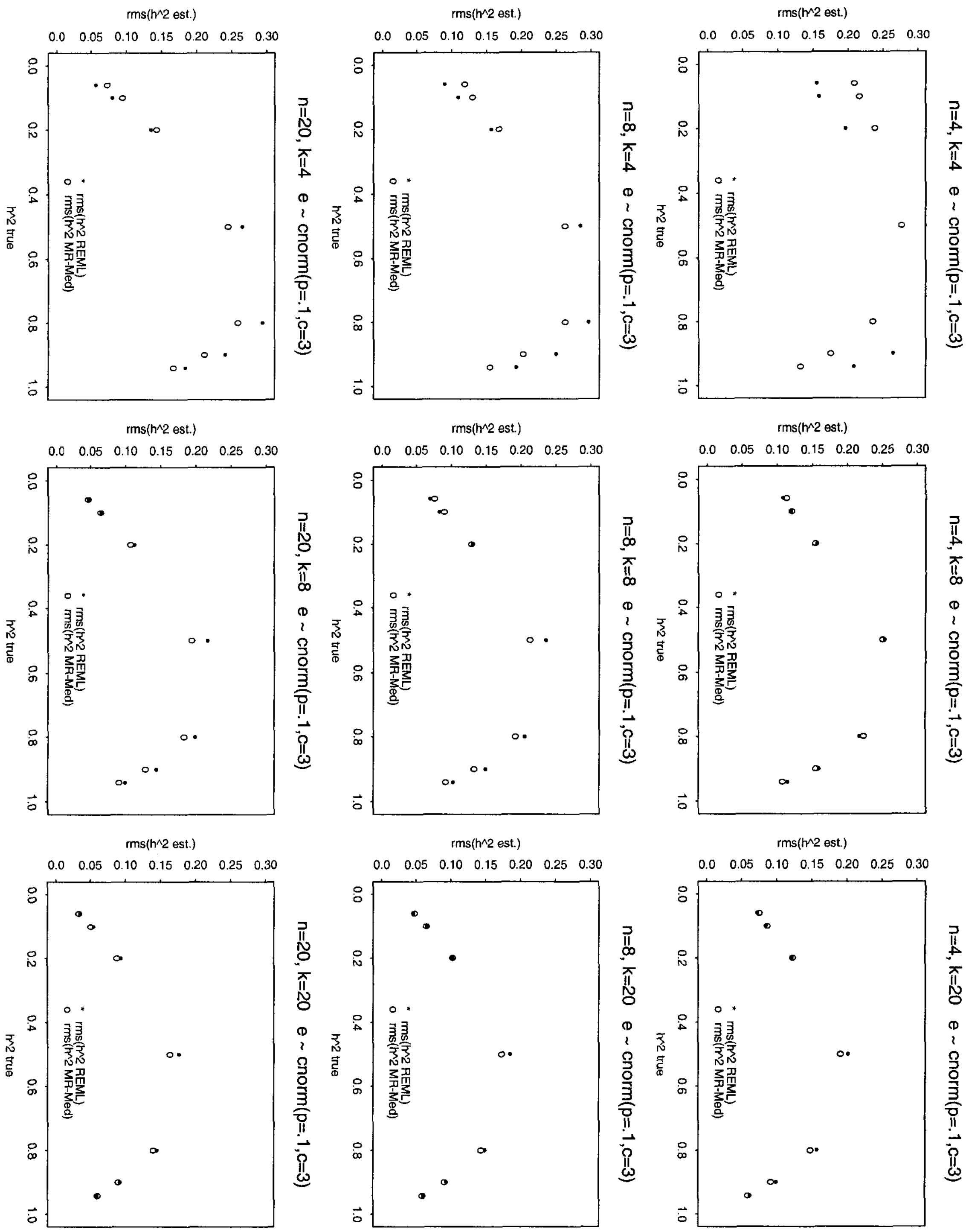

Figure 4.23: Contaminated normal ($p = .1, c = 3$) random error simulation: root mean square errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

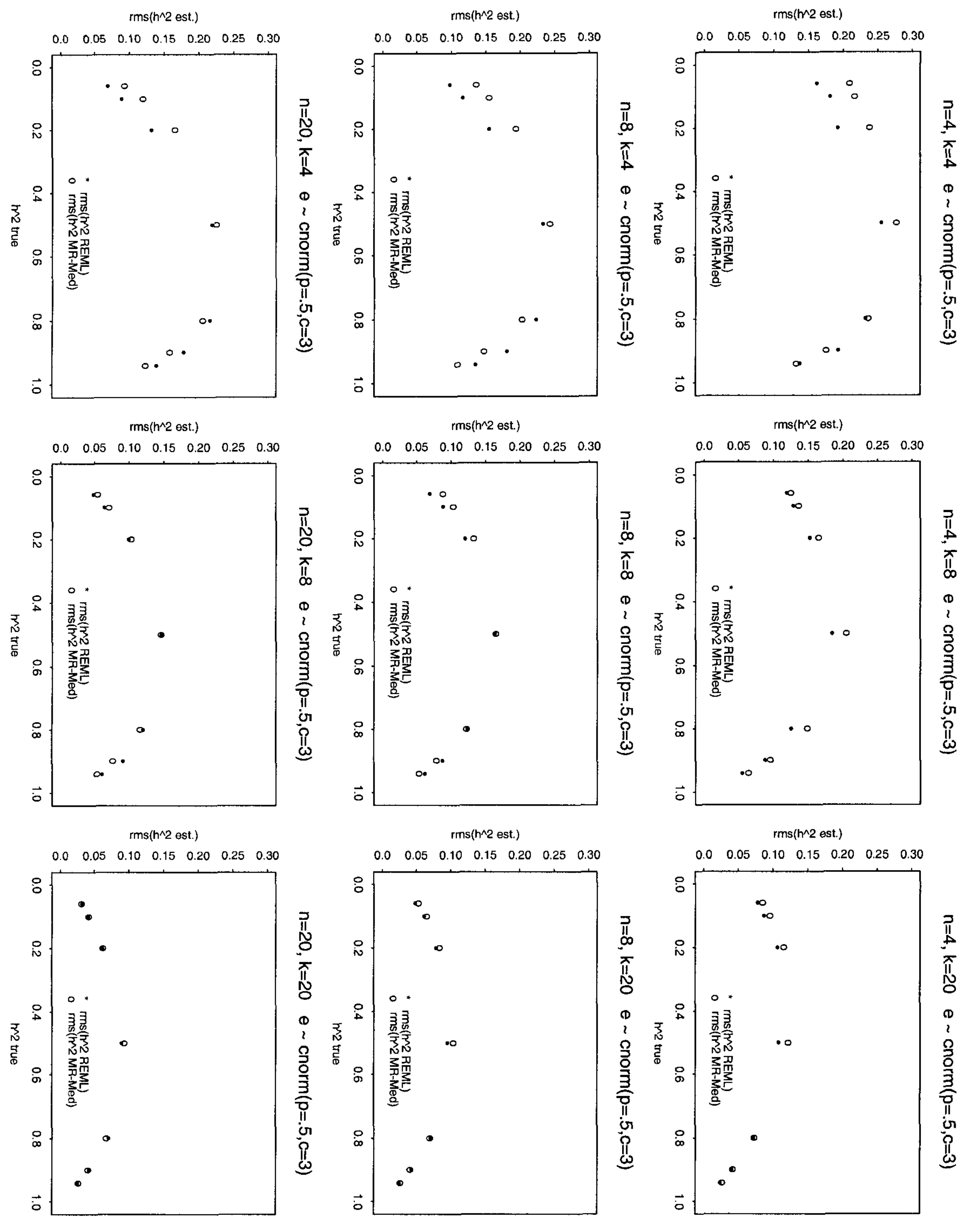

Figure 4.24: Contaminated normal ($p = .5, c = 3$) random error simulation: root mean square errors for $\hat{\theta}_{ML}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$, plotted against θ, for $n = 4, 8, 20$ and $k = 4, 8, 20$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

for $\theta = .9414, .5, .0588$, over values of m from 1 to 20, for fixed n of 20. For all θ values, the $\hat{\theta}_{MR-MED}$ appears to be the more stable of the two estimators with RMSE and SE that are consistently close in value, indicating low bias. Both the RMSE and SE of $\hat{\theta}_{MR-MED}$ decrease as θ decreases, and appear to increase only when m approaches n. In contrast, $\hat{\theta}_{MR-AVG}$ shows considerably high bias, that increases with m, when θ is close to 0 or 1. When θ is close to 0.5, $\hat{\theta}_{MR-AVG}$ has relatively stable bias, with increasing precision as m approaches n.

4.3.6 Summary for Balanced Random Effects model

Of the MR based estimators, $\hat{\theta}_{MR-MED}$ is to be recommended over $\hat{\theta}_{MR-AVG}$. Of the four estimators studied, $\hat{\theta}_{MR-AVG}$, is the most susceptible to extreme values, and thus the least stable. For our simulations of data of small to moderate sizes $\hat{\theta}_{MR-MED}$ performs well as a point estimator of θ when compared to $\hat{\theta}_{REML}$. For small n the RMSE for $\hat{\theta}_{MR-MED}$ compares well to that for $\hat{\theta}_{REML}$ for large values of θ, but is inferior to REML otherwise. For moderate to large n and k the $\hat{\theta}_{MR-MED}$ performs just as well $\hat{\theta}_{REML}$. Finally from the error distribution simulations, REML and MR-MED are both quite robust estimators for small to moderate n and k, but may have problems with bias in the case where the population is multimodal or heavy tailed.

In practice, the likelihood methods are to be preferred for the optimality properties and relative ease in computation, for small to moderate sample sizes. The simulations demonstrate that MR based estimators are a viable means of estimating parameters that compare well to likelihood based estimators, and even for small to moderate sized samples, these simulations have shown them to be robust estimators. These results are encouraging, as they demonstrate the credibility of MR

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

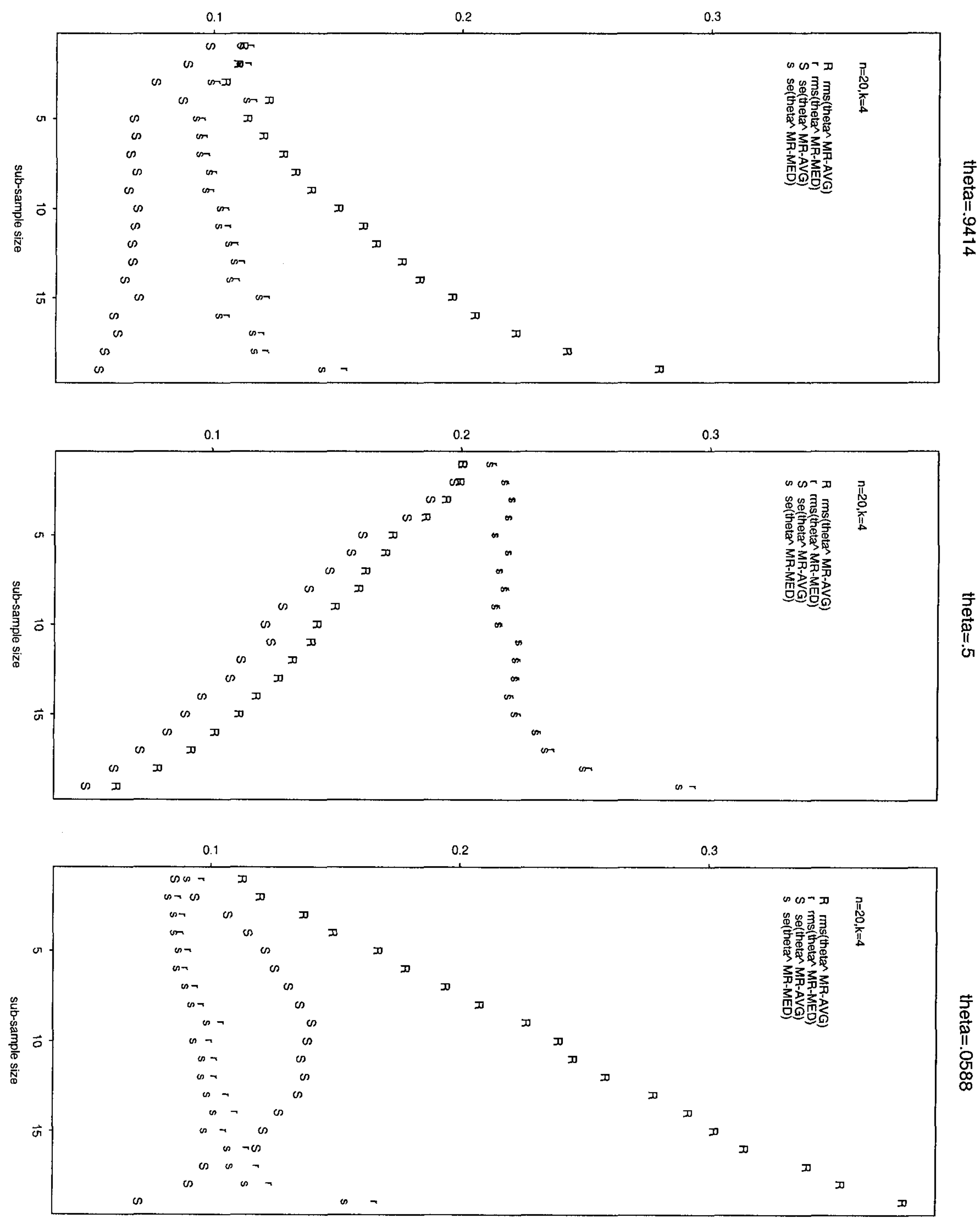

Figure 4.25: Normal random error simulation: standard errors and root mean square errors for $\hat{\theta}_{MR-MED}$ and $\hat{\theta}_{MR-AVG}$ plotted against sub-sample size m.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

based estimation and it's use in cases where likelihood based estimation is no longer feasible.

4.4 Method-R Based Interval Estimation for the Balanced One-Way Random Effects Model

4.4.1 Introduction

In this section we will look at MR based interval estimation for θ in the one-way balanced random effects model. In particular, bootstrap confidence intervals using MR, and the problems associated with them are discussed. Comments on a proposed method for bootstrap confidence intervals (Mallinckrodt, 1997) will also be included. The exact confidence interval for θ under normal distribution theory will be used a reference for the performance of the methods included in this section.

Normal Theory Exact Confidence Interval for θ

From Burdick and Graybill (1992), for the balanced one-way model under the assumption of normality, an exact confidence interval for $\theta = \frac{\sigma_a^2}{\sigma_e^2+\sigma_a^2}$ is given by

$$\left(\frac{S_1^2}{S_2^2 F_{1-\alpha/2:n_1,n_2}}, \frac{S_1^2}{S_2^2 F_{\alpha/2:n_1,n_2}} \right), \tag{4.4.1}$$

where S_1^2 and S_2^2 are the treatment and error mean squares, respectively, and $F_{1-\alpha/2:n_1,n_2}$ and $F_{1-\alpha/2:n_1,n_2}$ represent the quantile values for the central F-distribution with numerator degrees of freedom n_1 and denominator degrees of freedom n_2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4.4.2 Bootstrap Confidence Intervals for Method-R

Bootstrap methods of interval estimation provide a means of constructing good approximate confidence intervals even in situations where the sampling distribution of the estimator of interest is unknown or complex in nature (DiCiccio and Efron, 1996). This is precisely what makes bootstrap methods appealing, as a means of obtaining MR based confidence intervals. Recall that the MR estimator based on multiple sub-samples, involves statistics that are highly correlated. In Chapter 5, we will see that even for the simplest of models, deriving an expression for the variance of the MR estimator of θ quickly becomes a messy endevour. Efron and Tibshirani (1993) present several procedures for constructing bootstrap confidence intervals. Only two are presented here. The *bootstrap percentile method* and the *BCA method.*

Bootstrap Percentile Method

In order to compute the bootstrap percentile confidence interval for θ, we must first obtain a large number of bootstrap samples. A bootstrap sample is a random sample of observations from the data vector, $\mathbf{y}$, with replacement of size N. Let the number of bootstrap samples drawn be denoted by BS. Then we will denote the BS bootstrap samples by $\mathbf{y_1^*}, \mathbf{y_2^*}, \ldots, \mathbf{y_{BS}^*}$. The estimator of interest, say $\hat{\theta}_{MR-MED}$, is then computed for each of the bootstrap samples. These estimates will be denoted by, $\hat{\theta}^*_{1_{MR-MED}}, \hat{\theta}^*_{2_{MR-MED}}, \ldots, \hat{\theta}^*_{BS_{MR-MED}}$ make up the bootstrap distribution of $\hat{\theta}_{MR-MED}$. The 100 $(1-\alpha)$% percentile bootstrap confidence interval for θ is then given by,

$$(\hat{\theta}^*_{MR-MED,\alpha/2}, \hat{\theta}^*_{MR-MED,1-\alpha/2}), \qquad (4.4.2)$$

where $\hat{\theta}^*_{MR-MED,\alpha/2}$ and $\hat{\theta}^*_{MR-MED,1-\alpha/2}$, denote the $\alpha/2$ and $1-\alpha/2$ quantiles for the bootstrap distribution of $\hat{\theta}_{MR-MED}$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Bootstrap Bias Corrected Method (BC)

The bootstrap bias corrected (BC) method gives an adjustment for skewness 0 in the bootstrap distribution. The limits are constructed as in the percentile method, except that BC replaces the quantiles $\alpha/2$ and $1-\alpha/2$ with quantiles α_1 and α_2 given by,

$$\begin{aligned} \alpha_1 &= \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha/2)}}{1-(\hat{z}_0 + z^{(\alpha/2)})}\right) \\ \alpha_2 &= \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha/2)}}{1-(\hat{z}_0 + z^{(1-\alpha/2)})}\right), \end{aligned} \tag{4.4.3}$$

where $\hat{z}_0 = \Phi^{-1}\left(\frac{\#\hat{\theta}^*_{MR}<\hat{\theta}}{BS}\right)$.

4.4.3 Illustrative Calculations

Figures 4.26-4.28 show some preliminary results for percentile bootstrap confidence intervals for $\theta = .0588, .5, .9412$, respectively, based on MR-AVG and MR-MED for 20 data sets. Also included are the exact normal theory confidence limits. BCA bootstrap confidence intervals were attempted, but failed, due to undefined values for $\hat{z}_0$. The table below summarizes the observed coverage for the three methods. As the number of data sets used is small, the observed coverage is not expected to be an accurate estimate of coverage for these procedures, but is descriptive for this illustration.

Table 4.1: Percent Coverage for 95% confidence intervals for θ

θ	Exact	Percentile MR-AVG	Percentile MR-MED
.0588	95	100	100
.5	100	95	100
.9412	100	15	95

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From the above results we see that the percentile limits based on MR-AVG performs quite poorly for $\theta = .9412$, However coverage is adequate for small to moderate θ. In contrast percentile limits based on MR-MED performs just as well as the exact method for these preliminary results. In terms of the widths of the confidence intervals. for small θ, the precision of the exact intervals is consistently smallest, followed by that of the percentile limits based on MR-MED, then the percentile limits based on MR-AVG. In the case of $\theta = .5$, precision appears to be fairly similar among the three sets of intervals. For large θ, precision appears to be best for the exact intervals, while that of the two sets of percentile limits appear to be similar, with the percentile limits based on MR-MED being slightly more conservative.

4.4.4 Remarks

The percentile method works best when the shape of the bootstrap distribution is symmetric. In the instance that the bootstrap distribution is not symmetric, BC makes appropriate adjustments. While this method is appealing, computational problems arise due to the likelihood of drawing a value from either mass-point of the sampling distribution of $\hat{\theta}_{MR}$, for which "**" is undefined. Another drawback of bootstrap procedures in general, is that applied to MR, they require a nested series of sampling. First we have sub-sampling for single MR estimates, repeated r times to generate a single estimate of say, $\hat{\theta}_{MR-MED}$. Finally we have the sampling necessary to generate s bootstrap sample of estimates. Considering that both r and s are necessarily ¡large values (minimum of 1000), cumulatively, the effort involved can be non-trivial for even moderate sized data sets.

MR based bootstrap confidence intervals were proposed by Mallinckrodt et al. (1997). The procedure in Mallinckrodt et al. (1997) relies on the assumption that for 50% sub-sampling, the sample variance of MR estimates from repeated sub-samples,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

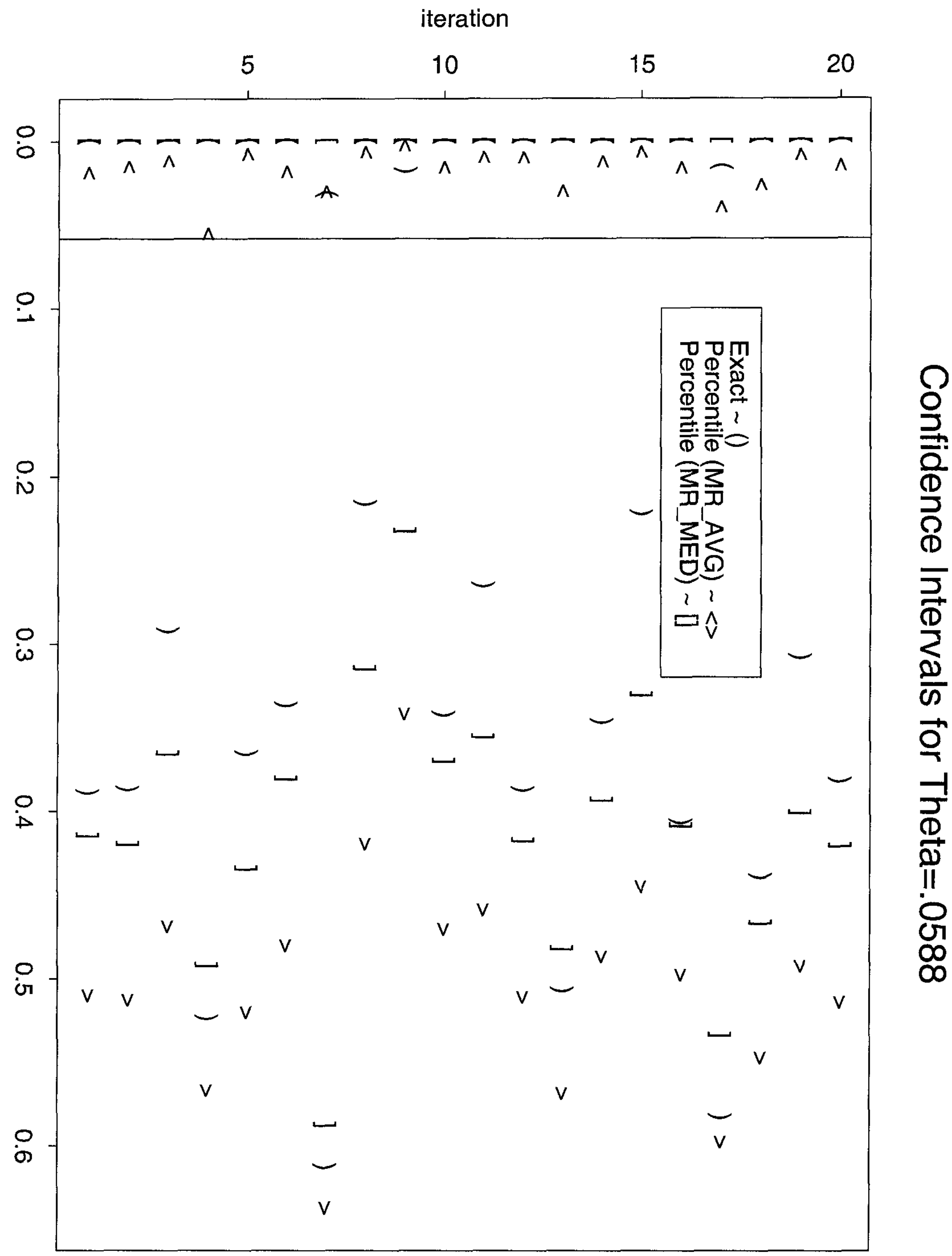

Figure 4.26: 95% Exact, percentile (MR-AVG), percentile (MR-MED) confidence intervals for $\theta = .0588$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

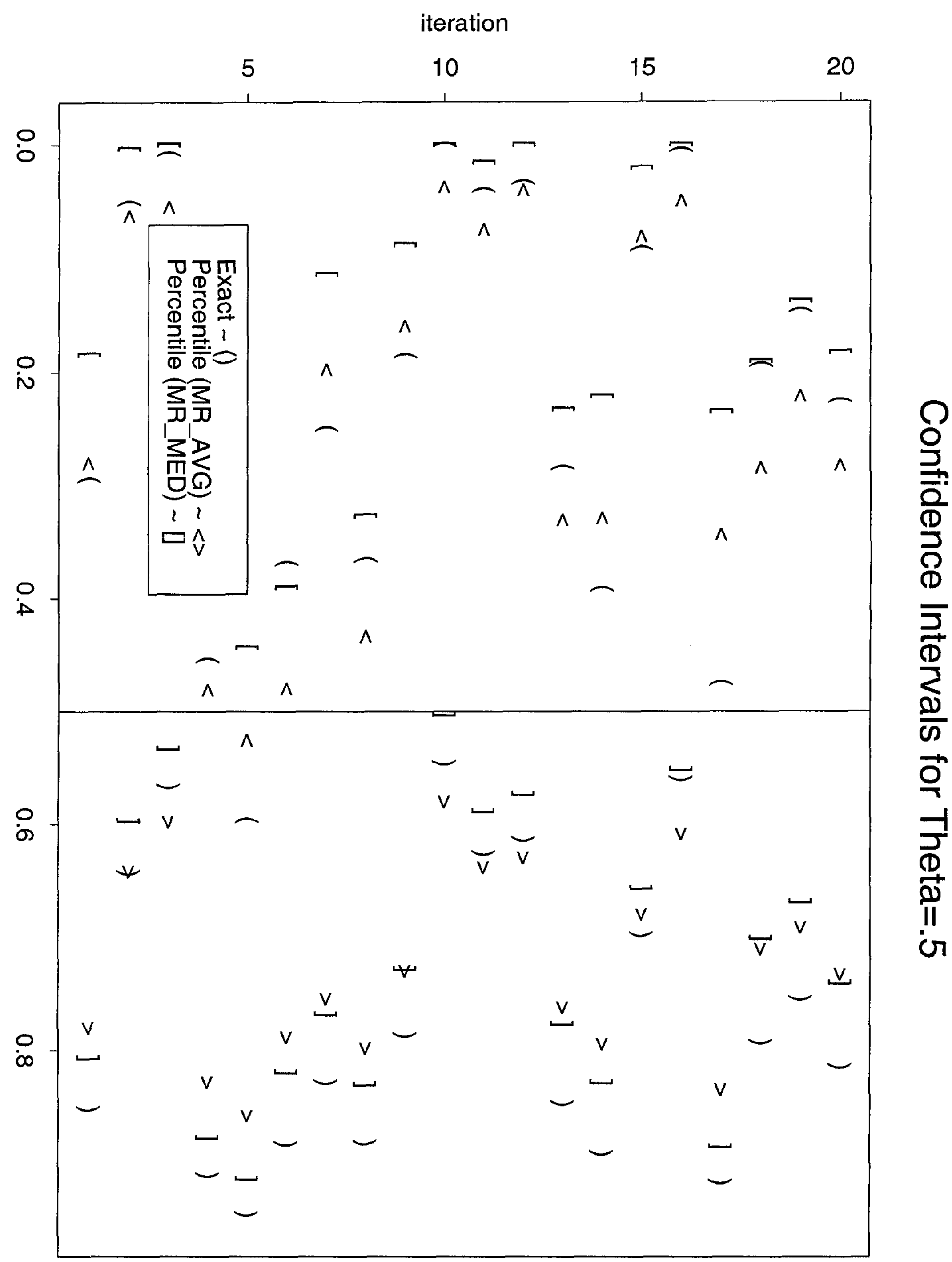

Figure 4.27: 95% Exact, percentile (MR-AVG), percentile (MR-MED) confidence intervals for $\theta = .5$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

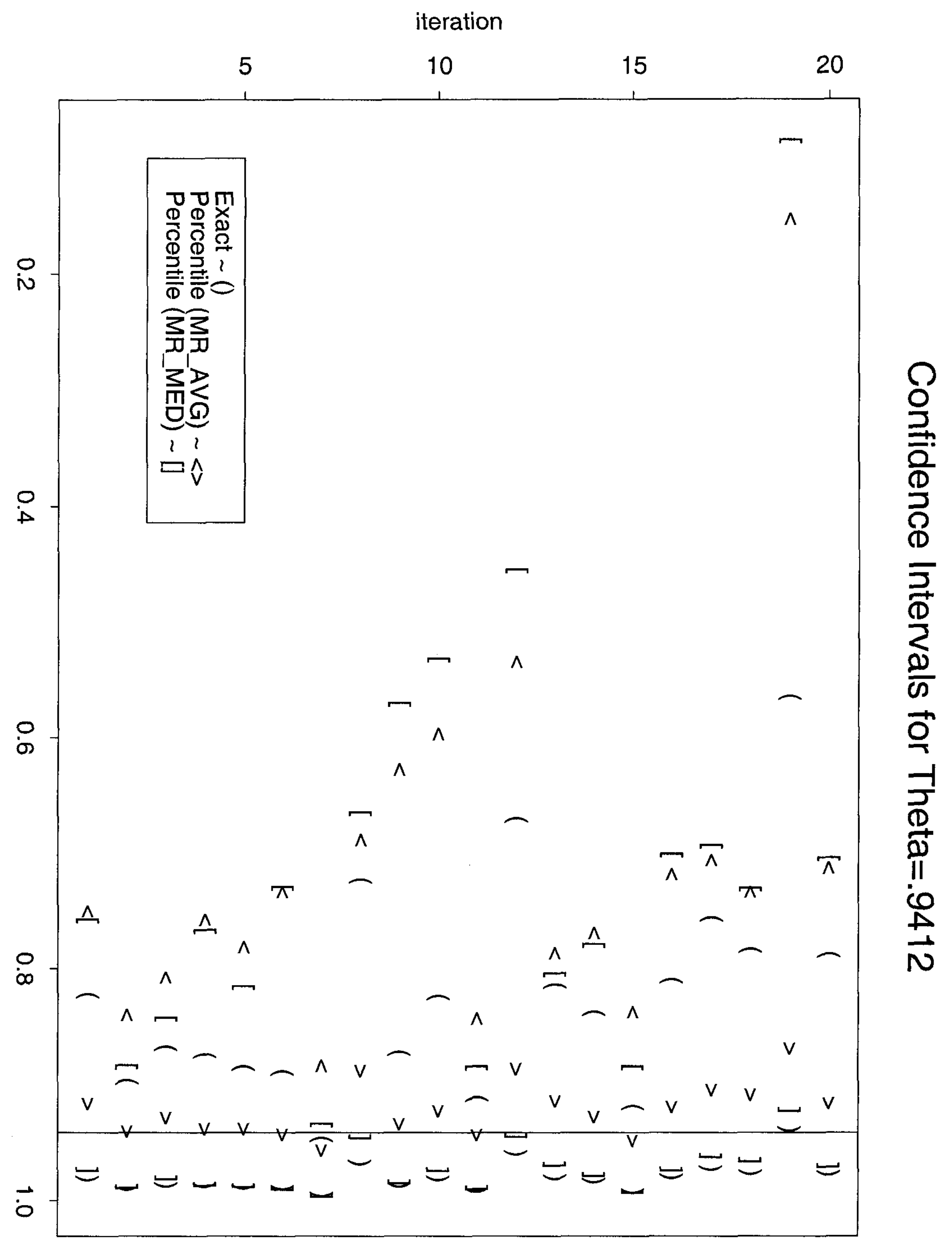

Figure 4.28: 95% Exact, percentile (MR-AVG), percentile (MR-MED) confidence intervals for $\theta = .9412$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is an adequate estimator of the bootstrap variance, as defined by Efron and Tibshirani (1996). Reverter et. al (1998a) present an approach for confidence regions for a vector of genetic parameters for a univariate full animal model. This procedure are similar to those in Mallinckrodt et al. (1997), in that they both assume that the bootstrap variance can adequately be approximated using the variance of MR estimates from repeated 50% sub-samples. Further discussion on this assumption and this approach to bootstrap confidence intervals will be addressed in Chapter 5.

In closing, for additional empirical works that compare MR to REML, see Cantet et.al.(2000), in which performance under selection is evaluated. Duangjinda et.al.(2001) examine performance under selection for four different animal breeding models. Both articles report the MSE for MR were higher than those for REML. Similar conclusions were drawn in Schenkel and Schaeffer (2000). Reverter(1998b) also reports on the optimality of MR.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 5

LARGE SAMPLE PROPERTIES OF THE METHOD R ESTIMATOR FOR THE BALANCED ONE-WAY MODEL

5.1 Introduction

The previous two chapters addressed the properties of the MR estimator in the balanced one-way ANOVA model for small to moderate samples sizes. In this chapter we focus on the MR estimator for large samples. The variance of the MR estimator based on a single sub-sample is decomposed into two parts, variation *between* independent data sets, and variance due to sub-sampling *within* a fixed data set. Expressions for the *between* and *within* components are derived. For large k, MR was shown to be asymptotically unbiased. Thus for large sample case, we will focus on the MR-AVG estimator. The asymptotics performed in this chapter involve letting the number of MR sub-samples go to infinity, thereby making the contribution of the within component of variance for the MR-AVG estimator go to zero. The between component of variance is studied for a fixed group size, n, as the number of groups, k, increases. The major result of this chapter is that the asymptotic variance of the MR estimator in the balanced one-way random effects ANOVA is the same as that of the variance as the asymptotic variance of the MLE given by Donner and Koval (1980). That is, the asymptotic efficiency of MR, relative to the MLE, in the balanced one-way model is 1, as the number of MR sub-samples and number of groups increase to infinity.

To obtain the variances, we first show that for large k, $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ can be approximated by $\hat{\beta}_{\hat{\mathbf{w}}|\tilde{\mathbf{w}}} = (\frac{1}{k}\hat{\mathbf{w}}\tilde{\mathbf{w}})/(\frac{1}{k}\tilde{\mathbf{w}}\tilde{\mathbf{w}})$, where $\hat{\mathbf{w}}$ and $\tilde{\mathbf{w}}$ are vectors of *non-centered* group totals

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

for the whole sample and sub-sample data, respectively. This provides a more mathematically tractable means for deriving the large sample approximation for the variance of $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$. The multivariate version of Taylor's theorem is applied to find first, the asymptotic variance of $\hat{\beta}_{\hat{\mathbf{w}}|\tilde{\mathbf{w}}}$, based on the moments of $\frac{1}{k}\hat{\mathbf{w}}\tilde{\mathbf{w}}$ and $\frac{1}{k}\tilde{\mathbf{w}}\tilde{\mathbf{w}}$. The univariate version of Taylor's theorem is then applied to obtain the asymptotic variance of heritability, $\hat{\theta}_{MR}$.

In the latter part of the chapter we examine the relationship of the between variance of $\hat{\theta}_{MR}$ to sub-sample size, and the relationship of the between variance of $\hat{\theta}_{MR}$ to the number of groups, k. These results provide evidence that for large k, the within component of variance for $\hat{\theta}_{MR}$ is not in general a good approximation for the between component of variance for $\hat{\theta}_{MR}$. This contradicts the claim by Mallinckrodt (1997), in which the within component of variance is used to estimate the true MR variance for constructing MR based bootstrap confidence intervals for $\hat{\theta}_{MR}$.

5.2 Outline for the Procedure used to Obtain Asymptotic Variances

What follows is an outline of the approach taken to obtain the asymptotic variances for $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ and $\hat{\theta}_{MR}$, and their respective between and within components.

1. Define the statistics $\hat{\mathbf{w}}$ and $\tilde{\mathbf{w}}$ as the whole and sub-sample totals, respectively. Express $\hat{\mathbf{w}}$ and $\tilde{\mathbf{w}}$ as functions of the raw data, the Y_{ij}'s, and indicator variables, the X_{ij}'s, used for sub-sample inclusion/exclusion. Show that $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ can be approximated by $\hat{\beta}_{\hat{\mathbf{w}}|\tilde{\mathbf{w}}}$ for large k. [Section 5.3]

2. Express $\hat{\beta}_{\hat{\mathbf{w}}|\tilde{\mathbf{w}}}$ as the ratio of the two statistics $\hat{\theta}_1 = \frac{1}{k}\hat{\mathbf{w}}'\tilde{\mathbf{w}}$ and $\hat{\theta}_2 = \frac{1}{k}\tilde{\mathbf{w}}'\tilde{\mathbf{w}}$. Apply the multivariate version of Taylor's theorem to obtain a large sample approximation to the mean and variance of $\hat{\beta}_{\hat{\mathbf{w}}|\tilde{\mathbf{w}}}$ in terms of the moments

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$\hat{\theta}_1$ and $\hat{\theta}_2$. Then apply Taylor's theorem once more to obtain a large sample approximation to the variance of $\hat{\theta}$ in terms of the moments $\hat{\theta}_1$ and $\hat{\theta}_2$. [Section 5.4]

3. Decompose the variance of $\hat{\theta}_{MR}$ over multiple sub-samples for repeated data collections into between and within components. The true MR variance is defined to be the between component of variance. Decompose the covariance of $\hat{\theta}_1$ and $\hat{\theta}_2$, $\Sigma_{\hat{\Theta}}$, into it's between and within components and use it to derive expressions for the between and within components of $var(\hat{\beta})$ in terms of the moments $\hat{\theta}_1$ and $\hat{\theta}_2$. Then use the components of $var(\hat{\beta})$ to derive the expressions for the between and within components of $var(\hat{\theta})$ in terms of the moments $\hat{\theta}_1$ and $\hat{\theta}_2$. [Section 5.5]

4. Present formulas for the elements of the covariance of $\hat{\Theta}$ expressed in terms of the unconditional moments for $\hat{\theta}_1$ and $\hat{\theta}_2$ and expectations with respect to the Y_{ij}'s of conditional expectations of $\hat{\theta}_1$ and $\hat{\theta}_2$ given the Y_{ij}'s. [Section 5.6]

5. Evaluate expectations of products of the X_{ij}'s and Y_{ij}'s required for evaluating the unconditional and conditional expectations of $\hat{\theta}_1$ and $\hat{\theta}_2$. [Section 5.7]

6. Evaluate expectations of products of $\hat{\theta}_1$ and $\hat{\theta}_2$ required for determining the elements of $\Sigma_{\hat{\Theta}}$ and its between and within components. [Section 5.8]

7. Evaluate the elements of $\Sigma_{\hat{\Theta}}$ and its between and within components using the equations in Section 5.6 and 5.8. [Section 5.9.1]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

8. Evaluate the formulas for the large sample approximations for the mean $\hat{\beta}$, variance of $\hat{\beta}$ and the between and within components of the variance of $\hat{\beta}$ using the results in Section 5.9.1. [Section 5.9.2]

9. Evaluate the formulas for the large sample approximations for the variance of $\hat{\theta}$ and the between and within components of the variance of $\hat{\theta}$ using the results in Section 5.9.1. [Section 5.9.3]

5.3 Large Sample Approximation for $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$

In this section we will show that for large k, $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ can be approximated by the simpler function, $\hat{\beta}_{\hat{\mathbf{w}}|\tilde{\mathbf{w}}} = (\frac{1}{k}\hat{\mathbf{w}}\tilde{\mathbf{w}})/(\frac{1}{k}\tilde{\mathbf{w}}\tilde{\mathbf{w}})$, where $\hat{\mathbf{w}}$ and $\tilde{\mathbf{w}}$ are vectors of non-centered group totals for the whole sample and sub-sample data, respectively. This simple approximation provides a more mathematical tractable means of deriving a large sample approximation for the variance of $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$. Ultimately, we seek a large sample approximation for the variance of $\hat{\theta}_{MR}$ through the large sample approximations for the moments of $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$.

Let X_{ij} be the indicator random variable that is 1 if Y_{ij} is included in the random sub-sample and 0 otherwise. For a sub-sample of size m, the probability mass function for X_{ij} is given below.

$$Pr(X_{ij} = x_{ij}) = \begin{cases} \frac{m}{n}, & x_{ij} = 1 \\ 1 - \frac{m}{n}, & x_{ij} = 0 \end{cases} \tag{5.3.1}$$

The statistics $\tilde{\mathbf{v}}'\tilde{\mathbf{v}}$ and $\tilde{\mathbf{v}}'\hat{\mathbf{v}}$ in terms of the x_{ij} and y_{ij} values, can be written as follows.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\tilde{\mathbf{v}}'\tilde{\mathbf{v}} = \sum_{i=1}^{k}\left(\sum_{j=1}^{n} x_{ij}y_{ij} - \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}y_{ij}\right)^2 \tag{5.3.2}$$

$$\tilde{\mathbf{v}}'\hat{\mathbf{v}} = \sum_{i=1}^{k}\left(\sum_{j=1}^{n} x_{ij}y_{ij} - \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}y_{ij}\right)\left(\sum_{j=1}^{n} y_{ij} - \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}\right) \tag{5.3.3}$$

From above, $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ is a non-linear function of the statistics $\tilde{\mathbf{v}}'\hat{\mathbf{v}}$ and $\tilde{\mathbf{v}}'\tilde{\mathbf{v}}$, which in turn are functions of the X_{ij}'s and Y_{ij}'s. An expression for the variance of $\hat{\theta}_{MR}$ can be derived from our knowledge of the X_{ij}'s and Y_{ij}'s. However, the complexity of the functional relationship between $\hat{\theta}_{MR}$ and the X_{ij}'s and Y_{ij}'s, makes for a formidable task. Large sample properties of equations 5.3.2 and 5.3.3 allow us to approximate $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ with a simpler function of the X_{ij}'s and Y_{ij}'s. The following arguments provide justification for using this simpler form.

Let $z_{ij} = x_{ij}y_{ij}$, $\hat{w}_i = \sum_{j=1}^{n} y_{ij}$ and $\tilde{w}_i = \sum_{j=1}^{n} x_{ij}y_{ij}$, and $\hat{\theta}_1 = \tilde{\mathbf{w}}'\hat{\mathbf{w}}/k$ and $\hat{\theta}_2 = \tilde{\mathbf{w}}'\tilde{\mathbf{w}}/k$. Treating n as fixed and w.l.o.g. we take $\mu = 0$.

$$\begin{aligned}
\frac{1}{k}\tilde{\mathbf{v}}'\tilde{\mathbf{v}} &= \frac{1}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{n} z_{ij} - \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n} z_{ij}\right)^2 \\
&= \frac{1}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{n} z_{ij} - \frac{n}{k}\sum_{i=1}^{k}\overline{z}_{i.}\right)^2 \\
&= \frac{1}{k}\sum_{i=1}^{k}\left[\left(\sum_{j=1}^{n} z_{ij}\right)^2 - \frac{2n}{k}\sum_{i=1}^{k}\overline{z}_{i.}\sum_{j=1}^{n} z_{ij} + \frac{n^2}{k^2}\left(\sum_{i=1}^{k}\overline{z}_{i.}\right)^2\right].
\end{aligned} \tag{5.3.4}$$

Since $\overline{z}_{i.}$'s are independent random variables with mean $\mu = 0$ and finite variance,

$$\frac{1}{k}\sum_{i=1}^{k}\overline{z}_{i.} = O_p(k^{-1/2}).$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

As $k \to \infty$,

$$\frac{1}{k}\sum_{i=1}^{k}\left[\left(\sum_{i=1}^{k}\frac{n^2}{k^2}\overline{z}_{i.}\right)^2\right] = \frac{1}{k}\sum_{i=1}^{k}n^2\left[O_p(k^{-1/2})\right]^2$$
$$= \sum_{i=1}^{k}\frac{n^2}{k}O_p(k^{-1})$$
$$= \sum_{i=1}^{k}O_p(k^{-2})$$
$$= O_p(k^{-1}),$$

and

$$\frac{1}{k}\sum_{i=1}^{k}\left[\frac{2n}{k}\sum_{i=1}^{k}\overline{z}_{i.}\sum_{j=1}^{n}z_{ij}\right] = \frac{1}{k}\sum_{i=1}^{k}2nO_p(k^{-1/2})O_p(1)$$
$$= \frac{1}{k}\sum_{i=1}^{k}O_p(k^{-1/2})$$
$$= O_p(k^{-1/2}).$$

Therefore

$$\frac{1}{k}\tilde{\mathbf{v}}'\tilde{\mathbf{v}} = \frac{1}{k}\tilde{\mathbf{w}}'\tilde{\mathbf{w}} + \left[O_p(k^{-1}) + O_p(k^{-1/2})\right]$$
$$= \frac{1}{k}\tilde{\mathbf{w}}'\tilde{\mathbf{w}} + O_p(k^{-1/2})$$
$$= \frac{1}{k}\tilde{\mathbf{w}}'\tilde{\mathbf{w}} + o_p(1).$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Similarly,

$$\begin{aligned}
\frac{1}{k}\tilde{\mathbf{v}}'\hat{\mathbf{v}} &= \frac{1}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{n} z_{ij} - \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n} z_{ij}\right)\left(\sum_{j=1}^{n} y_{ij} - \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}\right) \\
&= \frac{1}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{n} z_{ij} - \frac{n}{k}\sum_{i=1}^{k}\overline{z}_{i.}\right)\left(\sum_{j=1}^{n} y_{ij} - \frac{n}{k}\sum_{i=1}^{k}\overline{y}_{i.}\right) \\
&= \frac{1}{k}\sum_{i=1}^{k}\left[\sum_{j=1}^{n} z_{ij}\sum_{j=1}^{n} y_{ij} - \left(\frac{n}{k}\sum_{i=1}^{k}\overline{z}_{i.}\sum_{j=1}^{n} y_{ij} + \frac{n}{k}\sum_{i=1}^{k}\overline{y}_{i.}\sum_{j=1}^{n} z_{ij}\right)\right. \\
&\qquad \left. + \frac{n^2}{k^2}\sum_{i=1}^{k}\overline{z}_{i.}\sum_{i=1}^{k}\overline{y}_{i.}\right] \\
&= \frac{1}{k}\tilde{\mathbf{w}}'\hat{\mathbf{w}} + \frac{1}{k}\sum_{i=1}^{k}\left[n^2\left(\frac{1}{k}\sum_{i=1}^{k}\overline{z}_{i.}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\overline{y}_{i.}\right)\right. \\
&\qquad \left. - n\left(\frac{1}{k}\sum_{i=1}^{k}\overline{z}_{i.}\sum_{j=1}^{n} y_{ij} + \frac{1}{k}\sum_{i=1}^{k}\overline{y}_{i.}\sum_{j=1}^{n} z_{ij}\right)\right].
\end{aligned} \tag{5.3.5}$$

As $k \to \infty$,

$$\begin{aligned}
\frac{1}{k}\sum_{i=1}^{k}\left[n^2\left(\frac{1}{k}\sum_{i=1}^{k}\overline{z}_{i.}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\overline{y}_{i.}\right)\right] &= \frac{1}{k}\sum_{i=1}^{k} n^2[O_p(k^{-1/2})O_p(k^{-1/2})] \\
&= \sum_{i=1}^{k}\frac{n^2}{k}O_p(k^{-1}) \\
&= \sum_{i=1}^{k} O_p(k^{-2}) \\
&= O_p(k^{-1}),
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{k}\sum_{i=1}^{k}\left[n\left(\frac{1}{k}\sum_{i=1}^{k}\overline{z}_{i.}\sum_{j=1}^{n} y_{ij} + \frac{1}{k}\sum_{i=1}^{k}\overline{y}_{i.}\sum_{j=1}^{n} z_{ij}\right)\right] &= \frac{1}{k}\sum_{i=1}^{k}[n(O_p(k^{-1/2})O_p(1) \\
&\qquad + O_p(k^{-1/2})O_p(1))] \\
&= \frac{1}{k}\sum_{i=1}^{k} nO_p(k^{-1/2}) \\
&= \frac{1}{k}\sum_{i=1}^{k} O_p(k^{-1/2}) \\
&= O_p(k^{-1/2}).
\end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Therefore

$$\begin{aligned}\frac{1}{k}\tilde{\mathbf{v}}'\hat{\mathbf{v}} &= \frac{1}{k}\tilde{\mathbf{w}}'\hat{\mathbf{w}} + \left[O_p(k^{-1}) + O_p(k^{-1/2})\right] \\ &= \frac{1}{k}\tilde{\mathbf{w}}'\hat{\mathbf{w}} + O_p(k^{-1/2}) \\ &= \frac{1}{k}\tilde{\mathbf{w}}'\hat{\mathbf{w}} + o_p(1).\end{aligned}$$

For large k, the above arguments allow us to approximate $\frac{1}{k}\tilde{\mathbf{v}}'\tilde{\mathbf{v}}$ and $\frac{1}{k}\tilde{\mathbf{v}}'\hat{\mathbf{v}}$ with $\frac{1}{k}\tilde{\mathbf{w}}'\tilde{\mathbf{w}}$ and $\frac{1}{k}\tilde{\mathbf{w}}'\hat{\mathbf{w}}$, respectively. Thus, for large k, we can estimate $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}}$ with $\hat{\beta}_{\hat{\mathbf{w}}|\tilde{\mathbf{w}}} = \frac{\frac{1}{k}\tilde{\mathbf{w}}'\hat{\mathbf{w}}}{\frac{1}{k}\tilde{\mathbf{w}}'\tilde{\mathbf{w}}}$.

5.4 Large Sample Approximation for Moments of $\hat{\beta}$ and $\hat{\theta}$

In this section we will first derive expressions for large sample approximations to mean and the variance of $\hat{\beta}_{\hat{\mathbf{w}}|\tilde{\mathbf{w}}}$ by applying the multivariate version of Taylor's theorem to $\hat{\beta}_{\hat{\mathbf{w}}|\tilde{\mathbf{w}}}$ as a function of the statistics $\frac{1}{k}\tilde{\mathbf{w}}'\hat{\mathbf{w}}$ and $\frac{1}{k}\tilde{\mathbf{w}}'\tilde{\mathbf{w}}$. We then derive an expression for a large sample approximation for the variance of $\hat{\theta}_{MR}$ by applying univariate Taylor's theorem to $\hat{\theta}_{MR}$ as a function of $\hat{\beta}_{\hat{\mathbf{w}}|\tilde{\mathbf{w}}}$.

5.4.1 Taylor Series Approximations for $E(\hat{\beta})$ and $var(\hat{\beta})$ in terms of the moments of $\hat{\theta}_1$ and $\hat{\theta}_2$

Let $\hat{\theta}_1 = \frac{1}{k}\tilde{\mathbf{w}}'\hat{\mathbf{w}}$ and $\hat{\theta}_2 = \frac{1}{k}\tilde{\mathbf{w}}'\tilde{\mathbf{w}}$. Then from Section 5.3, for fixed n and large k, $\hat{\beta}_{\hat{\mathbf{v}}|\tilde{\mathbf{v}}} \doteq \frac{\hat{\theta}_1}{\hat{\theta}_2}$. For the remainder of this chapter, we will refer to $\frac{\hat{\theta}_1}{\hat{\theta}_2}$ as $\hat{\beta}$, and the corresponding estimate of heritability as $\hat{\theta} = (\hat{\beta} - 1)/[(n-1) - (m-1)\hat{\beta}]$. Also, let $\hat{\Theta} = (\hat{\theta}_1, \hat{\theta}_2)' \sim (\mu_{\hat{\Theta}}, \Sigma_{\hat{\Theta}})$, where

$$\mu_{\hat{\Theta}} = \begin{pmatrix} E(\hat{\theta}_1) \\ E(\hat{\theta}_2) \end{pmatrix} \text{ and } \Sigma_{\hat{\Theta}} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

From Serfling (1980), the multivariate version of Taylor's theorem is as follows.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Theorem 5.1 *Let the function f defined on $\mathbf{R}^m$ possess partial derivatives of order n at each point of an open set $\mathbf{S} \subset \mathbf{R}^m$. Let $\mathbf{x} \in \mathbf{S}$. For each point $\mathbf{y}$, $\mathbf{y} \neq \mathbf{x}$, such that the line segment $L(\mathbf{x}, \mathbf{y})$ joining $\mathbf{x}$ and $\mathbf{y}$ lies in $\mathbf{S}$, there exists a point $\mathbf{z}$ in the interior of $L(\mathbf{x}, \mathbf{y})$ such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \sum_{k=1}^{n-1} \frac{1}{k!} \sum_{i_1=1}^{m} \cdots \sum_{i_k=1}^{m} \frac{\partial^{(k)} f(t_1, \ldots, t_m)}{\partial t_{i_1} \ldots \partial t_{i_m}} \bigg|_{\mathbf{t}=\mathbf{x}} \prod_{j=1}^{k} (y_{i_j} - x_{i_j})$$
$$+ \frac{1}{n!} \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} \frac{\partial^{(n)} f(t_1, \ldots, t_m)}{\partial t_{t_1} \ldots \partial t_{i_n}} \bigg|_{\mathbf{t}=\mathbf{z}} \prod_{j=1}^{k} (y_{i_j} - x_{i_j}).$$

5.4.2 Large Sample Approximation for $E(\hat{\beta})$ and $var(\hat{\theta})$

Since $\hat{\beta}$ is a function of $\hat{\Theta}$, we can use Theorem 5.1 to derive a Taylor series approximation for $\hat{\beta}$ about the point $\hat{\Theta} = \mu_{\hat{\Theta}}$. Let $\hat{\beta}'(\hat{\Theta})$ be the vector of first order partial derivatives of $\hat{\beta}$ with respect to the elements of $\hat{\Theta}$ with $\hat{\beta}_i'(\hat{\Theta}) = \partial\hat{\beta}/\partial\hat{\theta}_i$ for $i = 1, 2$. And let $\hat{\beta}''(\hat{\Theta})$ be the matrix whose elements are the second order partial derivatives of $\hat{\beta}$ with respect to the elements of $\hat{\Theta}$, with $\hat{\beta}_{ij}''(\hat{\Theta}) = \partial^2\hat{\beta}/\partial\hat{\theta}_i\partial\hat{\theta}_j$ for $i, j = 1, 2$. The derivatives needed are the following,

$\hat{\beta}_1'(\hat{\Theta}) = \frac{1}{\hat{\theta}_2}$; $\hat{\beta}_2'(\hat{\Theta}) = -\frac{\hat{\theta}_1}{\hat{\theta}_2^2}$; $\hat{\beta}_{11}''(\hat{\Theta}) = 0$; $\hat{\beta}_{22}''(\hat{\Theta}) = \frac{2\hat{\theta}_1}{\hat{\theta}_2^3}$; $\hat{\beta}_{12}''(\hat{\Theta}) = -\frac{1}{\hat{\theta}_2^2}$.

Applying Theorem 5.1, a large sample approximation for $E(\hat{\beta})$ can be obtained by taking expectation of a second order Taylor series expansion of $\hat{\beta}$ about the point $\hat{\Theta} = \mu_{\hat{\Theta}}$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
E(\hat{\beta}) &\doteq E\Big\{\hat{\beta}(\hat{\Theta})|_{\hat{\Theta}=\mu_{\hat{\Theta}}} + \Big[\hat{\beta}'(\hat{\Theta})\Big]'|_{\hat{\Theta}=\mu_{\hat{\Theta}}}(\hat{\Theta}-\mu_{\hat{\Theta}}) \\
&\quad + \frac{1}{2!}(\hat{\Theta}-\mu_{\hat{\Theta}})'\hat{\beta}''(\hat{\Theta})|_{\hat{\Theta}=\mu_{\hat{\Theta}}}(\hat{\Theta}-\mu_{\hat{\Theta}})\Big\} \\
&= E\Big\{\hat{\beta}(\hat{\Theta})|_{\hat{\Theta}=\mu_{\hat{\Theta}}} + \sum_{i=1}^{2}(\hat{\theta}_i - E(\hat{\theta}_i))\hat{\beta}_i'(\hat{\Theta})|_{\hat{\Theta}=\mu_{\hat{\Theta}}} \\
&\quad + \frac{1}{2!}\sum_{i=1}^{2}(\hat{\theta}_i - E(\hat{\theta}_i))^2\hat{\beta}_{ii}''(\hat{\Theta})|_{\hat{\Theta}=\mu_{\hat{\Theta}}} \\
&\quad + (\hat{\theta}_1 - E(\hat{\theta}_1))(\hat{\theta}_2 - E(\hat{\theta}_2))\hat{\beta}_{12}''(\hat{\Theta})|_{\hat{\Theta}=\mu_{\hat{\Theta}}}\Big\} \\
&= \hat{\beta}(\mu_{\hat{\Theta}}) + \frac{1}{2}\sum_{i=1}^{2} 0.\hat{\beta}_i'(\mu_{\hat{\Theta}}) + \frac{1}{2}\sum_{i=1}^{2} var(\hat{\theta}_i)\hat{\beta}_{ii}''(\mu_{\hat{\Theta}}) + cov(\hat{\theta}_1,\hat{\theta}_2)\hat{\beta}_{12}''(\mu_{\hat{\Theta}}) \\
&= \frac{E(\hat{\theta}_1)}{E(\hat{\theta}_2)} + \frac{var(\hat{\theta}_1)}{2}.0 + \frac{var(\hat{\theta}_2)}{2}\frac{2E(\hat{\theta}_1)}{E(\hat{\theta}_2)^3} - cov(\hat{\theta}_1,\hat{\theta}_2)\frac{1}{E(\hat{\theta}_2)^2} \\
&= \frac{1}{E(\hat{\theta}_2)^3}\Big(E(\hat{\theta}_1)E(\hat{\theta}_2)^2 + var(\hat{\theta}_2)E(\hat{\theta}_1) - cov(\hat{\theta}_1,\hat{\theta}_2)E(\hat{\theta}_2)\Big)
\end{aligned} \tag{5.4.1}
$$

Once again, by applying Theorem 5.1, a large sample approximation for $var(\hat{\beta})$ can be obtained by computing the variance for the first order Taylor series expansion of $\hat{\beta}$ about $\hat{\Theta} = \mu_{\hat{\Theta}}$.

$$
\begin{aligned}
var(\hat{\beta}) &\doteq var\Big\{\hat{\beta}(\mu_{\hat{\Theta}}) + \Big[\hat{\beta}'(\mu_{\hat{\Theta}})\Big]'(\hat{\Theta}-\mu_{\hat{\Theta}})\Big\} \\
&= \Big[\hat{\beta}'(\mu_{\hat{\Theta}})\Big]' var(\hat{\Theta})\hat{\beta}'(\mu_{\hat{\Theta}}) \\
&= \sum_{i=1}^{2} var(\hat{\theta}_i)\Big[\hat{\beta}_i'(\hat{\Theta})_{\hat{\Theta}=\mu_{\hat{\Theta}}}\Big]^2 + 2cov(\hat{\theta}_1,\hat{\theta}_2)\Big[\hat{\beta}_1(\hat{\Theta})\hat{\beta}_2(\hat{\Theta})_{\hat{\Theta}=\mu_{\hat{\Theta}}}\Big], \text{ from 5.3.7} \\
&= \frac{1}{E(\hat{\theta}_2)^2}var(\hat{\theta}_1) + \left(-\frac{E(\hat{\theta}_1)}{E(\hat{\theta}_2)^2}\right)^2 var(\hat{\theta}_2) - 2\frac{E(\hat{\theta}_1)}{E(\hat{\theta}_2^3)}cov(\hat{\theta}_1,\hat{\theta}_2) \\
&= \frac{1}{E(\hat{\theta}_2)^4}\Big[E(\hat{\theta}_2)^2 var(\hat{\theta}_1) + E(\hat{\theta}_1)^2 var(\hat{\theta}_2) - 2E(\hat{\theta}_1)E(\hat{\theta}_2)cov(\hat{\theta}_1,\hat{\theta}_2)\Big].
\end{aligned} \tag{5.4.2}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5.4.3 Large Sample Approximation for $var(\hat{\theta})$

Recall that $\hat{\theta}$ as a function of $\hat{\beta}$ can be written as $\hat{\theta} = (\hat{\beta} - 1)/[(n-1) - (m-1)\hat{\beta}]$, with first order partial, $\partial\hat{\theta}/\partial\hat{\beta} = (n-m)/[(n-1) - (m-1)\hat{\beta}]^2$. The first order Taylor series approximation for $var(\hat{\theta}_{MR})$ about the point $\hat{\Theta} = E(\hat{\Theta})$ is then written as,

$$var(\hat{\theta}) \doteq \frac{var(\hat{\beta})}{2} \left(\frac{\partial\hat{\theta}}{\partial\hat{\beta}} \right)^2_{\hat{\Theta}=E(\hat{\Theta})} . \tag{5.4.3}$$

5.5 Between and Within Components of Variation for $\hat{\theta}_{MR}^{(r,s)}$

In this section the variance of the MR estimator based on a single sub-sample is decomposed into two parts, variation *between* independent data sets, and variance due to sub-sampling *within* a fixed data set. The *between* component is true variance of $\hat{\theta}_{MR}$, while the *within* component is the true measure of variability among MR estimates from the same data set. This is of particular interest, because the current practice (Reverter(1994b) and Mallinckrodt et al.(1997)) is to use the sample variance of $\hat{\theta}_{MR}$ for repeated sub-samples, as an estimator for the true variance of $\hat{\theta}_{MR}$ for a single sub-sample.

5.5.1 Decomposition for $var(\hat{\theta}_{MR})$ into Between and Within Components

Let the MR estimate from the s^{th} sub-sample of the r^{th} data collection, or experiment, be denoted by $\hat{\theta}_{MR}^{(r,s)}$, with $r = 1, \ldots, R$ and $s = 1, \ldots, S$. Here R represents the number of independent experiments, and S, total number of sub-samples drawn from each experiment. A one-way ANOVA model for $\hat{\theta}_{MR}^{(r,s)}$ is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\hat{\theta}_{MR}^{(r,s)} = \mu_{\hat{\theta}_{MR}} + D_r + \zeta_{rs}, \quad r = 1, \ldots, R; \; s = 1, \ldots, S. \tag{5.5.1}$$

The unconditional mean of $\hat{\theta}_{MR}$ is given by the parameter $\mu_{\hat{\theta}_{MR}}$. The random variable D_r represents the random effect corresponding to the r^{th} experiment, with $D_r \overset{IID}{\sim} (0, \sigma_D^2)$, $r = 1, \ldots, R$. The D_r's are independent of ζ's, which are dependent random variables with conditional variance, given the data for experiment r, as σ_ζ^2, and $\zeta_{rs} \sim (0, \sigma_\zeta^2)$, $s = 1, \ldots, S$. The variance of $\hat{\theta}_{MR}^{(r,s)}$ is given by

$$var(\hat{\theta}_{MR}^{(r,s)}) = \sigma_D^2 + \sigma_\zeta^2, \quad r = 1, \ldots, R; \; s = 1, \ldots, S. \tag{5.5.2}$$

The variance component σ_D^2 is the parameter representing the "between" experiment component variance of $\hat{\theta}_{MR}$, and is also the true variance of $\hat{\theta}_{MR}$ for a single sub-sample for a given data collection. The variance component σ_ζ^2 is the parameter for the "within" experiment component of variance of $\hat{\theta}_{MR}$. In practice, it is more common than not, for R to be 1, due to the cost involved in performing a single experiment. In contrast, the cost in obtaining $\hat{\theta}_{MR}$ for multiple sub-samples, will in general be less expensive, being dependent on the invested effort in computer programming required to obtain MR estimates for repeated sub-samples. In situations where R=1, it is not possible to estimate σ_D^2, however, an estimate for σ_ζ^2 is readily available through MR estimates obtained from repeated sub-samples.

In Section 5.4.3 we derived a large sample approximation to equation 5.5.2. Now we seek similar expressions for large sample approximations for the between and within components of variance of $\hat{\theta}_{MR}$. To accomplish this we first decompose $\Sigma_{\hat{\Theta}}$ into its respective within and between components. The large sample approximations for the components of variance for $var(\hat{\beta})$ are derived simply by replacing $\Sigma_{\hat{\Theta}}$ by it's partitioned form in equation 5.4.2. Similarly, the large sample approximations for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the components of variance for $var(\hat{\theta}_{MR})$ are derived simply by replacing $var(\hat{\beta})$ by it's resulting partitioned form in equation 5.4.3.

5.5.2 Decomposition for $\Sigma_{\hat{\Theta}}$ into Between and Within Components

Let $\hat{\Theta} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix}$ and let $\mu_{\hat{\Theta}}$ denote $E(\hat{\Theta}) = \begin{pmatrix} E(\hat{\theta}_1) \\ E(\hat{\theta}_2) \end{pmatrix}$. From equation 5.5.1, the corresponding model for $\hat{\Theta}$ can be written as follows.

$$\hat{\Theta}^{(r,s)} = \mu_{\hat{\Theta}} + \alpha_r + \eta_{rs}, \quad r = 1, \ldots, R; \; s = 1, \ldots, S. \tag{5.5.3}$$

For the above model, $\mu_{\hat{\Theta}}$ is true mean vector of $\hat{\Theta}^{(r,s)}$. The random effects vector corresponding to the r^{th} experiment is given by $\alpha_r = \begin{pmatrix} \alpha_{1,r} \\ \alpha_{2,r} \end{pmatrix} \overset{IID}{\sim} (\mathbf{0}, \Sigma_b)$, and the residual random vector is $\eta_r = \begin{pmatrix} \eta_{1,r} \\ \eta_{2,r} \end{pmatrix} \sim (\mathbf{0}, \Sigma_w)$, where Σ_w is the conditional covariance of η_r given the data for experiment r. The covariances Σ_b and Σ_w represent the between and within components for the covariance of $\hat{\Theta}^{(r,s)}$, with

$$\Sigma_b = \begin{pmatrix} \sigma_{1,b}^2 & \sigma_{12,b} \\ \sigma_{12,b} & \sigma_{2,b}^2 \end{pmatrix}; \quad \Sigma_w = \begin{pmatrix} \sigma_{1,w}^2 & \sigma_{12,w} \\ \sigma_{12,w} & \sigma_{2,w}^2 \end{pmatrix}.$$

From the above, it follows that $var(\hat{\Theta}^{(r,s)})$, which we will denote by

$$\Sigma_{\hat{\Theta}} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}, \tag{5.5.4}$$

and

$$\Sigma_{\hat{\Theta}} = \Sigma_b + \Sigma_w. \tag{5.5.5}$$

In the following sub-sections we will use the above decomposition for $\Sigma_{\hat{\Theta}}$ to derive the large sample approximations for the decompositions of $var(\hat{\beta})$ and $var(\hat{\theta}_{MR})$, respectively.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5.5.3 Decomposition for $var(\hat{\beta})$ into Between and Within Components

The large sample approximation for $var(\hat{\beta})$ in equation 5.4.2 can be expressed in terms of the partition of $\Sigma_{\hat{\Theta}}$ in equation 5.5.5 as follows.

$$\begin{aligned} var(\hat{\beta}) &\doteq \left[\hat{\beta}'(\mu_{\hat{\Theta}})\right]'(\Sigma_b + \Sigma_w)\hat{\beta}'(\mu_{\hat{\Theta}}) \\ &= \left[\hat{\beta}'(\mu_{\hat{\Theta}})\right]'\Sigma_b\hat{\beta}'(\mu_{\hat{\Theta}}) + \left[\hat{\beta}'(\mu_{\hat{\Theta}})\right]'\Sigma_w\hat{\beta}'(\mu_{\hat{\Theta}}) \\ &= var_{[b]}(\hat{\beta}) + var_{[w]}(\hat{\beta}). \end{aligned} \tag{5.5.6}$$

From the above, it follows that

$$var_{[w]}(\hat{\beta}) \doteq \frac{1}{E(\hat{\theta}_2)^4}\left[E(\hat{\theta}_2)^2 var_{[w]}(\hat{\theta}_1) + E(\hat{\theta}_1)^2 var_{[w]}(\hat{\theta}_2) - 2E(\hat{\theta}_1)E(\hat{\theta}_2)cov_{[w]}(\hat{\theta}_1, \hat{\theta}_2)\right], \tag{5.5.7}$$

and

$$var_{[b]}(\hat{\beta}) \doteq \frac{1}{E(\hat{\theta}_2)^4}\left[E(\hat{\theta}_2)^2 var_{[b]}(\hat{\theta}_1) + E(\hat{\theta}_1)^2 var_{[b]}(\hat{\theta}_2) - 2E(\hat{\theta}_1)E(\hat{\theta}_2)cov_{[b]}(\hat{\theta}_1, \hat{\theta}_2)\right]. \tag{5.5.8}$$

5.5.4 Decomposition for $var(\hat{\theta}_{MR})$ into Between and Within Components

Applying the partition on $var(\hat{\beta})$ in equation 5.5.6 to the equation for the large sample approximation for $var(\hat{\theta}_{MR})$ in equation 5.4.3, we have the following.

$$\begin{aligned} var(\hat{\theta}_{MR}) &\doteq \frac{var_{[w]}(\hat{\beta}) + var_{[b]}(\hat{\beta})}{2}\left(\frac{\partial\hat{\theta}_{MR}}{\partial\hat{\beta}}\right)^2_{\hat{\Theta}=E(\hat{\Theta})} \\ &= \frac{var_{[w]}(\hat{\beta})}{2}\left(\frac{\partial\hat{\theta}_{MR}}{\partial\hat{\beta}}\right)^2_{\hat{\Theta}=E(\hat{\Theta})} + \frac{var_{[b]}(\hat{\beta})}{2}\left(\frac{\partial\hat{\theta}_{MR}}{\partial\hat{\beta}}\right)^2_{\hat{\Theta}=E(\hat{\Theta})} \\ &= var_{[w]}(\hat{\theta}_{MR}) + var_{[b]}(\hat{\theta}_{MR}). \end{aligned} \tag{5.5.9}$$

Thus, it follows that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$var_{[w]}(\hat{\theta}_{MR}) = \frac{var_{[w]}(\hat{\beta})}{2}\left(\frac{\partial\hat{\theta}_{MR}}{\partial\hat{\beta}}\right)^2_{\hat{\Theta}=E(\hat{\Theta})} \tag{5.5.10}$$

and

$$var_{[b]}(\hat{\theta}_{MR}) = \frac{var_{[b]}(\hat{\beta})}{2}\left(\frac{\partial\hat{\theta}_{MR}}{\partial\hat{\beta}}\right)^2_{\hat{\Theta}=E(\hat{\Theta})} \tag{5.5.11}$$

Note that all of the large sample approximations for the $var(\hat{\beta})$ and $var(\hat{\theta}_{MR})$ and their respective between and within components are in terms of the moments for $\hat{\theta}_1$ and $\hat{\theta}_2$, which are still to be evaluated. In the following section, the functional relationship between $\hat{\theta}_1$ and $\hat{\theta}_2$ and the X_{ij}'s and the Y_{ij}'s, will be used to derive the moments for $\hat{\theta}_1$ and $\hat{\theta}_2$ needed to evaluate the large sample approximations for the variances given in Sections 5.4 and 5.5.

5.6 Between and Within Components of Covariance for the Elements of $\Sigma_{\hat{\Theta}}$

Let statistics $\hat{\theta}_1$ and $\hat{\theta}_2$ be defined as in Section 5.4, both of which are functions the x_{ij}'s and y_{ij}'s. Then the decompositions for the variances, $var(\hat{\theta}_i)$, $i = 1,2$ and $cov(\hat{\theta}_1, \hat{\theta}_2)$ can be expressed as follows

$$var(\hat{\theta}_i) = E[var(\hat{\theta}_i|\mathbf{Y})] + var[E(\hat{\theta}_i|\mathbf{Y})],\ i = 1, 2. \tag{5.6.1}$$

$$cov(\hat{\theta}_1, \hat{\theta}_2) = E[cov(\hat{\theta}_1, \hat{\theta}_2|\mathbf{Y})] + cov[E(\hat{\theta}_1|\mathbf{Y}), E(\hat{\theta}_2|\mathbf{Y})]. \tag{5.6.2}$$

The first term in equation 5.6.1,$E[var(\hat{\theta}_i|\mathbf{Y})]$, is the variance of the values of $\hat{\theta}_i$ for single data set, and will denoted as $var_{[w]}(\hat{\theta}_i) = \sigma^2_{i,w}$ from Section 5.5.2 . The second term in equation 5.6.1, $var[E(\hat{\theta}_i|\mathbf{Y})]$, is the variance of the mean $\hat{\theta}_i$ values for all datasets, and will be denoted by $var_{[b]}(\hat{\theta}_i) = \sigma^2_{i,b}$ from Section 5.5.2 . The two terms

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

comprising the decomposition of $cov(\hat{\theta}_1, \hat{\theta}_2)$ have similar interpretations. The first term in equation 5.6.2, $E[cov(\hat{\theta}_1, \hat{\theta}_2|\mathbf{Y})]$, is the covariance between $\hat{\theta}_1$ and $\hat{\theta}_2$ values for a single data set, and will be denoted as $cov_{[w]}(\hat{\theta}_1, \hat{\theta}_2) = \sigma_{12,w}$ from Section 5.5.2 . The second term in 5.6.2, $cov[E(\hat{\theta}_1|\mathbf{Y}), E(\hat{\theta}_2|\mathbf{Y})]$ is the covariance between mean $\hat{\theta}_1$ and $\hat{\theta}_2$ values for all datasets, and will be denoted as $cov_{[b]}(\hat{\theta}_1, \hat{\theta}_2) = \sigma_{12,b}$ from Section 5.5.2 .

Now,

$$\begin{aligned} var_{[w]}(\hat{\theta}_i) &= E[var(\hat{\theta}_i|\mathbf{Y})] \\ &= E[E(\hat{\theta}_i^2|\mathbf{Y})] - E\{[E(\hat{\theta}_i|\mathbf{Y})]^2\} \\ &= E(\hat{\theta}_i^2) - E\{[E(\hat{\theta}_i|\mathbf{Y})]^2\}, \text{ and} \end{aligned} \tag{5.6.3}$$

$$\begin{aligned} var_{[b]}(\hat{\theta}_i) &= var[E(\hat{\theta}_i|\mathbf{Y})] \\ &= E\{[E(\hat{\theta}_i|\mathbf{Y})]^2\} - \{E[E(\hat{\theta}_i|\mathbf{Y})]\}^2 \\ &= E\{[E(\hat{\theta}_i|\mathbf{Y})]^2\} - [E(\hat{\theta}_i)]^2. \end{aligned} \tag{5.6.4}$$

Also,

$$\begin{aligned} cov_{[w]}(\hat{\theta}_1, \hat{\theta}_2) &= E[cov(\hat{\theta}_1, \hat{\theta}_2|\mathbf{Y})] \\ &= E[E(\hat{\theta}_1\hat{\theta}_2|\mathbf{Y})] - E[E(\hat{\theta}_1|\mathbf{Y})E(\hat{\theta}_2|\mathbf{Y})] \\ &= E(\hat{\theta}_1\hat{\theta}_2) - E[E(\hat{\theta}_1|\mathbf{Y})E(\hat{\theta}_2|\mathbf{Y})], \end{aligned} \tag{5.6.5}$$

and

$$\begin{aligned} cov_{[b]}(\hat{\theta}_1, \hat{\theta}_2) &= cov[E(\hat{\theta}_1|\mathbf{Y}), E(\hat{\theta}_2|\mathbf{Y})] \\ &= E[E(\hat{\theta}_1|\mathbf{Y})E(\hat{\theta}_2|\mathbf{Y})] - E[E(\hat{\theta}_1|\mathbf{Y})]E[E(\hat{\theta}_2|\mathbf{Y})] \\ &= E[E(\hat{\theta}_1|\mathbf{Y})E(\hat{\theta}_2|\mathbf{Y})] - E(\hat{\theta}_1)E(\hat{\theta}_2) \end{aligned} \tag{5.6.6}$$

Therefore expressions for $var_{[w]}(\hat{\theta}_i)$ and $var_{[b]}(\hat{\theta}_i)$ can be found through $E(\hat{\theta}_i)$, $E(\hat{\theta}_i^2)$ and $E\{[E(\hat{\theta}_i|\mathbf{Y})]^2\}$ $(i = 1, 2)$. Similarly, $cov_{[w]}(\hat{\theta}_1, \hat{\theta}_2)$ and $cov_{[b]}(\hat{\theta}_1, \hat{\theta}_2)$ can be found through $E(\hat{\theta}_1)$, $E(\hat{\theta}_2)$, $E(\hat{\theta}_1\hat{\theta}_2)$ and $E[E(\hat{\theta}_1|\mathbf{Y})E(\hat{\theta}_2|\mathbf{Y})]$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In the next section we will evaluate the expectations for the functions of the X_{ij}'s and Y_{ij}'s, so that we can in turn evaluate the expectations of functions of $\hat{\theta}_1$ and $\hat{\theta}_2$, and then finally, we evaluate the large sample formulas for $var(\hat{\beta})$ and $var(\hat{\theta}_{MR})$ and their respective decompositions.

5.7 Preliminary Work for Evaluating Moments for $\hat{\theta}_1$ and $\hat{\theta}_2$

The statistics $\hat{\theta}_1$ and $\hat{\theta}_2$ involve products of sums of the $\mathbf{X}_{ij}$'s and $\mathbf{Y}_{ij}$'s. To reduce the complexity of the analytic arguments that follow, we will define a few new functions of the x_{ij}'s and y_{ij}'s. Let $A_i = \sum_{j=1}^{n} z_{ij}$ and $B_i = \sum_{j=1}^{n} y_{ij}$. Unless otherwise specified, $\sum_{i_r}$ will be used to denote $\sum\limits_{i_r=1}^{k}$, and $\sum_{j_r}$ will be used to denote $\sum\limits_{j_r=1}^{n}$, where $r = 1, \ldots, 4$. In similar fashion, the operand $\sum_{i_1 \neq i_2 \ldots \neq i_r}$ will be used to denote $\sum\limits^{k}\sum\limits^{k}\sum\limits^{k}_{i_1 \neq i_2 \ldots \neq i_r}$ and $\sum_{j_1 \neq j_2 \ldots \neq j_r}$ will be used to denote $\sum\limits^{n}\sum\limits^{n}\sum\limits^{n}_{j_1 \neq j_2 \ldots \neq j_r}$. Expectations of products of the A_i's and B_i's will be evaluated below.

5.7.1 Expectations of $\mathbf{X}_{ij}$ products

$\mathbf{X}_{ij}$ was defined to be the indicator random variable that has a value of 1 if observation y_{ij} is included in a random sub-sample, and 0 otherwise. The statistics $\hat{\theta}_1$ and $\hat{\theta}_2$ are functions of $\mathbf{X}_{ij}$'s of order 1 and 2, respectively. Thus, in order to evaluate first and second order central moments for $\hat{\theta}_1$ and $\hat{\theta}_2$, we need expectations of X_{ij} products for up to, and including order 4. In cases where the $\mathbf{X}_{ij}$'s are in the same group, sampling is with replacement. Therefore $E(\mathbf{X}_{ij_1}\mathbf{X}_{ij_2}) = \frac{m}{n}\frac{m-1}{n-1}$, if $j_1 \neq j_2$. In the instance that $j_1 = j_2$, $E(\mathbf{X}_{ij_1}\mathbf{X}_{ij_2}) = E(\mathbf{X}^2_{ij_1}) = \left(\frac{m}{n}\right)^2$. $\mathbf{X}_{ij}$'s in different groups are sampled independently, and as such, are independent variables. Therefore $E(\mathbf{X}_{i_1j_1}\mathbf{X}_{i_2j_2})$=$E(\mathbf{X}_{i_1j_1})E(\mathbf{X}_{i_2j_2})$, if $i_1 \neq i_2$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Listed below is a summary of the expectations of $\mathbf{X}_{ij}$ products of all types for up to, and including four terms.

$$E(\mathbf{X}_{ij}) = \frac{m}{n} \tag{5.7.1}$$

$$E(\mathbf{X}_{i_1j_1}\mathbf{X}_{i_2j_2}) = \begin{cases} \left(\frac{m}{n}\right)^2 & , i_1 \neq i_2 \\ \left(\frac{m}{n}\right)\frac{m-1}{n-1} & , i_1 = i_2, j_1 \neq j_2 \end{cases} \tag{5.7.2}$$

$$E(\mathbf{X}_{i_1j_1}\mathbf{X}_{i_2j_2}\mathbf{X}_{i_3j_3}) = \begin{cases} \left(\frac{m}{n}\right)^3 & , i_1 \neq i_2 \neq i_3 \\ \left(\frac{m}{n}\right)^2\frac{m-1}{n-1} & , i_1 = i_2 \neq i_3, j_1 \neq j_2 \\ \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\frac{m-2}{n-2} & , i_1 = i_2 = i_3, j_1 \neq j_2 \neq j_3 \end{cases} \tag{5.7.3}$$

$$\begin{aligned} &E(\mathbf{X}_{i_1j_1}\mathbf{X}_{i_2j_2}\mathbf{X}_{i_3j_3}\mathbf{X}_{i_4j_4}) \\ &= \begin{cases} \left(\frac{m}{n}\right)^4 & , i_1 \neq i_2 \neq i_3 \neq i_4 \\ \left(\frac{m}{n}\right)^3\frac{m-1}{n-1} & , i_1 = i_2 \neq i_3 \neq i_4, j_1 \neq j_2 \\ \left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)\frac{m-2}{n-2} & , i_1 = i_2 = i_3 \neq i_4, j_1 \neq j_2 \neq j_3 \\ \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\frac{m-2}{n-2}\right)\frac{m-3}{n-3} & , i_1 = i_2 = i_3 = i_4, j_1 \neq j_2 \neq j_3 \neq j_4 \end{cases} \end{aligned} \tag{5.7.4}$$

5.7.2 Expectations of products $\mathbf{Y}_{ij}$'s

We will now evaluate expectations of products of the $\mathbf{Y}_{ij}$'s, up to and including terms of order 4. Adopting the assumptions from Section 3.1, we have that u_i's are $u_i \overset{IID}{\sim} N(0, \sigma_a^2)$ for $i = 1, \ldots, k$, and independent of ϵ_{ij}'s, with $\epsilon_{ij} \overset{IID}{\sim} N(0, \sigma_e^2)$ for $i = 1, \ldots, k$ and $j = 1, \ldots, n$. From this, we have $E(u_i^r \epsilon_{jk}^s) = E(u_i^r)E(\epsilon_{jk}^s)$ for $r, s = 1, \ldots, 4$, where

$$E(u_i^r) = \begin{cases} 0, & \text{if r is odd} \\ \sigma_a^2, & \text{if r=2} \\ 3\sigma_a^4, & \text{if r=4} \end{cases} \qquad \text{and similarly, } E(\epsilon_{ij}^s) = \begin{cases} 0, & \text{if s is odd} \\ \sigma_a^2, & \text{if s=2} \\ 3\sigma_a^4, & \text{if s=4} \end{cases} .$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

As a result of the above, if r or s is odd, the product of the expectations is 0. For the following, let $E\left(\Pi_{k=1}^{k} y_{ik}^{r_k}\right)$ be denoted by $\mu_{r_1 r_2 \dots r_k}$.

$$\mu_1 = E\left(y_{11}\right) = E\left(u_1 + \epsilon_{11}\right) = 0. \tag{5.7.5}$$

$$\mu_2 = E\left(y_{11}^2\right) = var(y_{11}) = \sigma_a^2 + \sigma_e^2. \tag{5.7.6}$$

$$\begin{aligned}\mu_{11} = E\left(y_{11}y_{12}\right) &= E\left[\left(u_1 + \epsilon_{11}\right)\left(u_1 + \epsilon_{12}\right)\right] \\ &= E\left[u_1^2 + u_1\left(\epsilon_{11} + \epsilon_{12}\right) + \epsilon_{11}\epsilon_{12}\right] \\ &= E(u_1^2) \\ &= \sigma_a^2.\end{aligned} \tag{5.7.7}$$

$$\begin{aligned}\mu_3 = E\left(y_{11}^3\right) &= E\left[\left(u_1 + \epsilon_{11}\right)^3\right] \\ &= E\left(u_1^3 + 3u_1^2\epsilon_{11} + 3u_1\epsilon_{11}^2 + \epsilon_{11}^3\right) \\ &= E(u_1^3) \\ &= 0.\end{aligned} \tag{5.7.8}$$

$$\begin{aligned}\mu_{21} = E\left(y_{11}^2 y_{12}\right) &= E\left[\left(u_1 + \epsilon_{11}\right)^2\left(u_1 + \epsilon_{12}\right)\right] \\ &= E\left[\left(u_1^2 + 2u_1\epsilon_{11} + \epsilon_{11}^2\right)\left(u_1 + \epsilon_{12}\right)\right] \\ &= E\left(u_1^3 + 2u_1\epsilon_{11} + \epsilon_{11}^2 u_1 + u_1^2\epsilon_{12} + 2u_1\epsilon_{11}\epsilon_{12} + \epsilon_{11}^2\epsilon_{12}\right) \\ &= 0.\end{aligned} \tag{5.7.9}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
\mu_{111} = E\left(y_{11}y_{12}y_{13}\right) &= E\left[\left(u_1+\epsilon_{11}\right)\left(u_1+\epsilon_{12}\right)\left(u_1+\epsilon_{13}\right)\right] \\
&= E\left[\left(u_1^2+u_1(\epsilon_{11}+\epsilon_{12})+\epsilon_{11}\epsilon_{12}\right)\left(u_1+\epsilon_{13}\right)\right] \\
&= E\Big[u_1^3+u_1^2(\epsilon_{11}+\epsilon_{12}+\epsilon_{13})+u_1\epsilon_{11}\epsilon_{12} \\
&\qquad +u_1(\epsilon_{11}+\epsilon_{12}+\epsilon_{13})+\epsilon_{11}\epsilon_{12}\epsilon_{13}\Big] \\
&= 0.
\end{aligned} \tag{5.7.10}
$$

$$
\begin{aligned}
\mu_4 = E(y_{11}^4) &= E\left[\left(u_1+\epsilon_{11}\right)^4\right] \\
&= E\left(u_1^4+4u_1^3\epsilon_{11}+6u_1^2\epsilon_{11}^2+4u_1\epsilon_{11}^3+\epsilon_{11}^4\right) \\
&= E(u_1^4)+6E(u_1^2\epsilon_{11}^2)+E(\epsilon_{11}^4) \\
&= 3\sigma_a^4+6\sigma_a^2\sigma_e^2+3\sigma_e^4.
\end{aligned} \tag{5.7.11}
$$

$$
\begin{aligned}
\mu_{22} = E(y_{11}^2y_{12}^2) &= E\left[\left(u_1+\epsilon_{11}\right)^2\left(u_1+\epsilon_{12}\right)^2\right] \\
&= E(u_1^4+u_1^2\epsilon_{12}^2+u_1^2\epsilon_{11}^2+\epsilon_{11}^2\epsilon_{12}^2) \\
&= E(u_1^4)+E(u_1^2\epsilon_{12}^2)+E(u_1^2\epsilon_{11}^2)+E(\epsilon_{11}^2\epsilon_{12}^2) \\
&= 3\sigma_a^4+2\sigma_a^2\sigma_e^2+\sigma_e^4.
\end{aligned} \tag{5.7.12}
$$

$$
\begin{aligned}
\mu_{31} = E(y_{11}^3y_{12}) &= E\left[\left(u_1+\epsilon_{11}\right)^3\left(u_1+\epsilon_{12}\right)\right] \\
&= E\left[(u_1^3+3u_1^2\epsilon_{11}+3u_1\epsilon_{11}^2+\epsilon_{11}^3)(u_1+\epsilon_{12})\right] \\
&= E(u_1^4+3u_1^2\epsilon_{11}^2) \\
&= E(u_1^4)+3E(u_1^2\epsilon_{11}^2) \\
&= 3\sigma_a^4+3\sigma_a^2\sigma_e^2.
\end{aligned} \tag{5.7.13}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\mu_{211} = E(y_{11}^2 y_{12} y_{13}) = E\left[(u_1+\epsilon)^2(u_1+\epsilon_{12})(u_1+\epsilon_{13})\right]$$
$$= E(u_1^4 + u_1^2\epsilon_{11}^2)$$
$$= E(u_1^4) + E(u_1^2\epsilon_{11}^2) \tag{5.7.14}$$
$$= 3\sigma_a^4 + \sigma_a^2\sigma_e^2$$

$$\mu_{1111} = E(y_{11}y_{12}y_{13}y_{14}) = E\left[(u_1+\epsilon_{11})(u_1+\epsilon_{12})(u_1+\epsilon_{13})(u_1+\epsilon_{14})\right]$$
$$= E(u_1^4) \tag{5.7.15}$$
$$= 3\sigma_a^4$$

5.7.3 Sums of products of $\mathbf{Y}_{ij}$'s

This section includes sums of products of $\mathbf{Y}_{ij}$'s that will be used in evaluating expectations for $\hat{\theta}_1$ and $\hat{\theta}_2$.

$$B_i^2 = \left(\sum_j y_{ij}\right)^2 = \sum_{j_1}\sum_{j_2} y_{ij_1}y_{ij_2}$$
$$= \sum_{\substack{j_1 \\ \{j_1=j_2\}}} y_{ij_1}^2 + \sum_{j_1\neq j_2} y_{ij_1}y_{ij_2} \tag{5.7.16}$$

$$B_i^2\sum_j y_{ij}^2 = \left(\sum_j y_{ij}\right)^2 \sum_j y_{ij}^2$$
$$= \sum_{j_1}\sum_{j_2}\sum_{j_3} y_{ij_1}y_{ij_2}y_{ij_3}^2 \tag{5.7.17}$$
$$= \sum_{\substack{j_1 \\ \{j_1=j_2=j_3\}}} y_{ij_1}^4 + \sum_{\substack{j_1\neq j_3 \\ \{j_1=j_2\}}} y_{ij_1}^2 y_{ij_3}^2 + 2\sum_{\substack{j_1\neq j_3 \\ \{j_1=j_3;j_2=j_3\}}} y_{ij_1}y_{ij_3}^3 + \sum_{j_1\neq j_2\neq j_3} y_{ij_1}y_{ij_2}y_{ij_3}^2$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
B_{i_1}^2 \sum_{j_2} y_{i_2 j_2}^2 \;{}_{\{i_1 \neq i_2\}} &= \left(\sum_{j_1} y_{i_1 j_1} y_{i_1 j_2} \right) \sum_{j_3} y_{i_2 j_3}^2 \\
&= \sum_{j_1} \sum_{j_2} \sum_{j_3} y_{i_1 j_1} y_{i_1 j_2} y_{i_2 j_3}^2 \\
&= \sum_{\substack{j_1 \\ \{j_1 = j_2\}}} \sum_{j_3} y_{i_1 j_1}^2 y_{i_2 j_3}^2 + \sum_{j_1 \neq j_2} \sum_{j_3} y_{i_1 j_1} y_{i_1 j_2} y_{i_2 j_3}^2
\end{aligned}
\tag{5.7.18}
$$

$$
\begin{aligned}
B_i^4 &= \left(\sum_j y_{ij} \right)^4 \\
&= \sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_4} y_{ij_1} y_{ij_2} y_{ij_3} y_{ij_4} \\
&= \sum_{\substack{j_1 \\ \{j_1 = j_2 = j_3 = j_4\}}} y_{ij_1}^4 + 4 \sum_{\substack{j_1 \neq j_2 \\ \left\{ \begin{smallmatrix} j_1 = j_2 = j_3;\, j_1 = j_2 = j_4; \\ j_1 = j_3 = j_4;\, j_2 = j_3 = j_4 \end{smallmatrix} \right\}}} y_{ij_1} y_{ij_2}^3 + 3 \sum_{\substack{j_1 \neq j_2 \\ \{j_1 = j_2; j_1 = j_3; j_1 = j_4\}}} y_{ij_1}^2 y_{ij_2}^2 \\
&\quad + 6 \sum_{\substack{j_1 \neq j_2 \neq j_3 \\ \left\{ \begin{smallmatrix} j_1 = j_2;\, j_1 = j_3;\, j_1 = j_4 \\ j_2 = j_3;\, j_2 = j_4;\, j_3 = j_4 \end{smallmatrix} \right\}}} y_{ij_1} y_{ij_2} y_{ij_3}^2 + \sum_{j_1 \neq j_2 \neq j_3 \neq j_4} y_{ij_1} y_{ij_2} y_{ij_3} y_{ij_4}
\end{aligned}
\tag{5.7.19}
$$

$$
\begin{aligned}
B_{i_1}^2 B_{i_2}^2 \;{}_{\{i_1 \neq i_2\}} &= \left(\sum_{j_1} y_{i_1 j_1} \right)^2 \left(\sum_{j_2} y_{i_2 j_2} \right)^2 \\
&= \sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_4} y_{i_1 j_1} y_{i_1 j_2} y_{i_2 j_3} y_{i_2 j_4} \\
&= \sum_{\substack{j_1 \\ \{j_1 = j_2, j_3 = j_4\}}} \sum_{j_3} y_{i_1 j_1}^2 y_{i_2 j_3}^2 + 2 \sum_{\substack{j_1 \neq j_2 \\ \{j_1 = j_2, j_3 \neq j_4; j_1 \neq j_2; j_3 = j_4\}}} \sum_{j_3} y_{i_1 j_1} y_{i_1 j_2} y_{i_2 j_3}^2 \\
&\quad + \sum_{j_1 \neq j_2} \sum_{j_3 \neq j_4} y_{i_1 j_1} y_{i_1 j_2} y_{i_2 j_3} y_{i_2 j_4}
\end{aligned}
\tag{5.7.20}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5.7.4 Expectations of sums of products of $\mathbf{X}_{ij}$'s and $\mathbf{Y}_{ij}$'s

In this section expectations of sums of products of $\mathbf{X}_{ij}$'s and $\mathbf{Y}_{ij}$'s are derived in the following Section 5.8 to evaluate expectations for $\hat{\theta}_1$ and $\hat{\theta}_2$.

$$\begin{aligned} E(A_i|\mathbf{Y}) &= E\left(\sum_j z_{ij}\middle|\mathbf{Y}\right) \\ &= \sum_j y_{ij} E(\mathbf{X}_{ij}) \\ &= \frac{m}{n} B_i. \end{aligned} \tag{5.7.21}$$

$$\begin{aligned} E(A_i^2|\mathbf{Y}) &= E\left[\left(\sum_j z_{ij}\right)^2\middle|\mathbf{Y}\right] \\ &= E\left(\sum_{j_1}\sum_{j_2} z_{ij_1} z_{ij_2}\middle|\mathbf{Y}\right) \\ &= E\left(\sum_{j_1} z_{ij_1}^2 + \sum_{j_1\neq j_2} z_{ij_1} z_{ij_2}\middle|\mathbf{Y}\right) \\ &= \sum_{j_1} y_{ij_1}^2 E(\mathbf{X}_{ij_1}^2) + \sum_{j_1\neq j_2} y_{ij_1} y_{ij_2} E(\mathbf{X}_{ij_1}\mathbf{X}_{ij_2}) \\ &= \frac{m}{n}\sum_{j_1} y_{ij_1}^2 + \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\sum_{j_1\neq j_2} y_{ij_1} y_{ij_2} \\ &= \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\sum_{j_1} y_{ij_1}^2 + \sum_{j_1\neq j_2} y_{ij_1} y_{ij_2}\right) + \left(\frac{m}{n}\right)\left(\frac{n-m}{n-1}\right)\sum_{j_1} y_{ij_1}^2 \\ &= \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)B_i^2 + \left(\frac{m}{n}\right)\left(\frac{n-m}{n-1}\right)\sum_{j_1} y_{ij_1}^2. \end{aligned} \tag{5.7.22}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
E(A_{i_1}A_{i_2}|\mathbf{Y})_{\ \{i_1\neq i_2\}} &= E\left(\sum_{j_1}\sum_{j_2} z_{ij_2}z_{ij_2}\middle|\mathbf{Y}\right) \\
&= \sum_{j_1}\sum_{j_2} y_{ij_2}y_{ij_2}E(\mathbf{X}_{ij_1})E(\mathbf{X}_{ij_2}), \quad \text{applying 5.7.1,} \\
&= \left(\frac{m}{n}\right)^2 \sum_{j_1}\sum_{j_2} y_{i_1j_1}y_{i_2j_2}.
\end{aligned}
\tag{5.7.23}
$$

$$
\begin{aligned}
E(A_i^3|\mathbf{Y}) &= E\left(\sum_{j_1}\sum_{j_2}\sum_{j_3} z_{ij_1}z_{ij_2}z_{ij_3}\middle|\mathbf{Y}\right) \\
&= E\left(\sum_{\substack{j_1 \\ \{j_1=j_2=j_3\}}} z_{ij_1}^3 + 3\sum_{\substack{j_1\neq j_2 \\ \left\{\substack{j_1=j_2;\, j_1=j_3; \\ j_2=j_3}\right\}}} z_{ij_1}z_{ij_2}^2 + \sum_{j_1\neq j_2\neq j_3} z_{ij_1}z_{ij_2}z_{ij_3}\middle|\mathbf{Y}\right) \\
&= \sum_{j_1} y_{ij_1}^3 E(\mathbf{X}_{ij_1}^3) + 3\sum_{j_1\neq j_2} y_{ij_1}y_{ij_2}^2 E(\mathbf{X}_{ij_1}\mathbf{X}_{ij_2}^2) \\
&\quad + \sum_{j_1\neq j_2\neq j_3} y_{ij_1}y_{ij_2}y_{ij_3}E(\mathbf{X}_{ij_1}\mathbf{X}_{ij_2}\mathbf{X}_{ij_3}), \quad \text{applying 5.7.3,} \\
&= \frac{m}{n}\sum_{j_1} y_{ij_1}^3 + 3\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\sum_{j_1\neq j_2} y_{ij_1}y_{ij_2}^2 \\
&\quad + \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\frac{m-2}{n-2}\right)\sum_{j_1\neq j_2\neq j_3} y_{ij_1}y_{ij_2}y_{ij_3}.
\end{aligned}
\tag{5.7.24}
$$

$$
\begin{aligned}
E(A_{i_1}A_{i_2}^2|\mathbf{Y})_{\ \{i_1\neq i_2\}} &= E\left[\sum_{j_1} z_{i_1j_1}\left(\sum_{j_2} z_{i_2j_2}\right)^2\middle|\mathbf{Y}\right] \\
&= E\left(\sum_{j_1}\sum_{j_2}\sum_{j_3} z_{i_1j_1}z_{i_2j_2}z_{i_2j_3}\middle|\mathbf{Y}\right) \\
&= E\left(\sum_{j_1}\sum_{\substack{j_2 \\ \{j_2=j_3\}}} z_{i_1j_1}z_{i_2j_2}^2 + \sum_{j_1}\sum_{j_2\neq j_3} z_{i_1j_1}z_{i_2j_2}z_{i_2j_3}\middle|\mathbf{Y}\right)
\end{aligned}
$$

(continues on next page)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$E(A_{i_1}A_{i_2}^2|\mathbf{Y})_{\{i_1\neq i_2\}} = \sum_{j_1}\sum_{j_2} y_{i_1j_1}y_{i_2j_2}^2 E(\mathbf{X}_{i_1j_1})E(\mathbf{X}_{i_2j_2}^2)$$

$$+\sum_{j_1}\sum_{j_2\neq j_3} y_{i_1j_1}y_{i_2j_2}y_{i_2j_3}E(\mathbf{X}_{i_1j_1})(\mathbf{X}_{i_2j_2}\mathbf{X}_{i_2j_3})$$

applying 5.7.1 and 5.7.2

$$= \left(\frac{m}{n}\right)^2 \sum_{j_1}\sum_{j_2} y_{i_1j_1}y_{i_2j_2}^2$$

$$+\left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)\sum_{j_1}\sum_{j_2\neq j_3} y_{i_1j_1}y_{i_2j_2}^2. \tag{5.7.25}$$

$$E(A_i^4|\mathbf{Y}) = E\left[\left(\sum_j z_{ij}\middle|\mathbf{Y}\right)^4\right]$$

$$= E\left(\sum_{j_1}\sum_{j_2}\sum_{j_3}\sum_{j_4} z_{ij_1}z_{ij_2}z_{ij_3}z_{ij_4}\middle|\mathbf{Y}\right)$$

$$= E\Bigg(\sum_{\substack{j_1 \\ \{j_1=j_2=j_3=j_4\}}} z_{ij_1}^4 \; + \; 4\sum_{\substack{j_1\neq j_2 \\ \left\{\begin{array}{l} j_1=j_2=j_3;\,j_1=j_2=j_4; \\ j_1=j_3=j_4;\,j_2=j_3=j_4\end{array}\right\}}} z_{ij_1}z_{ij_2}^3 \; + \; 3\sum_{\substack{j_1\neq j_2 \\ \{j_1=j_2;j_1=j_3;j_1=j_4\}}} z_{ij_1}^2 z_{ij_2}^2 \; +$$

$$6\sum_{\substack{j_1\neq j_2\neq j_3 \\ \left\{\begin{array}{l} j_1=j_2;\,j_1=j_3;\,j_1=j_4 \\ j_2=j_3;\,j_2=j_4;\,j_3=j_4\end{array}\right\}}} z_{ij_1}z_{ij_2}z_{ij_3}^2 \; + \sum_{j_1\neq j_2\neq j_3\neq j_4} z_{ij_1}z_{ij_2}z_{ij_3}z_{ij_4}\Bigg|\mathbf{Y}\Bigg)$$

$$= \sum_{j_1} y_{ij_1}^4 E(\mathbf{X}_{ij_1}^4) + 4\sum_{j_1\neq j_2} y_{ij_1}y_{ij_2}^3 E(\mathbf{X}_{ij_1}\mathbf{X}_{ij_2}^3)$$

$$+3\sum_{j_1\neq j_2} y_{ij_1}^2 y_{ij_2}^2 E(\mathbf{X}_{ij_1}^2\mathbf{X}_{ij_2}^2)$$

$$+6\sum_{j_1\neq j_2\neq j_3} y_{ij_1}y_{ij_2}y_{ij_3}^2 E(\mathbf{X}_{ij_1}\mathbf{X}_{ij_2}\mathbf{X}_{ij_3}^2)$$

$$+\sum_{j_1\neq j_2\neq j_3\neq j_4} y_{ij_1}y_{ij_2}y_{ij_3}y_{ij_4}E(\mathbf{X}_{ij_1}\mathbf{X}_{ij_2}\mathbf{X}_{ij_3}\mathbf{X}_{ij_4}),$$

(continues on next page)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

applying 5.7.4

$$E(A_i^4|\mathbf{Y}) = \frac{m}{n}\sum_{j_1=1} y_{ij_1}^4 + 4\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\sum_{j_1\neq j_2} y_{ij_1}y_{ij_2}^3$$
$$+3\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\sum_{j_1\neq j_2} y_{ij_1}^2 y_{ij_2}^2$$
$$+6\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\frac{m-2}{n-2}\right)\sum_{j_1\neq j_2\neq j_3} y_{ij_1}y_{ij_2}y_{ij_3}^2$$
$$+\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\frac{m-2}{n-2}\right)\left(\frac{m-3}{n-3}\right)\sum_{j_1\neq j_2\neq j_3\neq j_4} y_{ij_1}y_{ij_2}y_{ij_3}y_{ij_4}.$$

(5.7.26)

$$E(A_{i_1}^2 A_{i_2}^2|\mathbf{Y})_{\{i_1\neq i_2\}} = E\left(\sum_{j_1}\sum_{j_2}\sum_{j_3}\sum_{j_4} z_{i_1j_1}z_{i_1j_2}z_{i_2j_3}z_{i_2j_4}\middle|\mathbf{Y}\right)$$
$$= E\left(\underset{\{j_1=j_2,j_3=j_4\}}{\sum_{j_1}\sum_{j_3}} z_{i_1j_1}^2 z_{i_2j_3}^2 + 2\underset{\{j_1\neq j_2,j_3=j_4;j_1=j_2,j_3\neq j_4\}}{\sum_{j_1\neq j_2}\sum_{j_3}} z_{i_1j_1}z_{i_1j_2}z_{i_2j_3}^2\right.$$
$$\left.+\sum_{j_1\neq j_2}\sum_{j_3\neq j_4} z_{i_1j_1}z_{i_1j_2}z_{i_2j_3}z_{i_2j_4}\middle|\mathbf{Y}\right)$$
$$= \sum_{j_1}\sum_{j_3} y_{i_1j_1}^2 y_{i_2j_3}^2 E(\mathbf{X}_{i_1j_1}^2)E(\mathbf{X}_{i_2j_3}^2)$$
$$+2\sum_{j_1\neq j_2}\sum_{j_3} y_{i_1j_1}y_{i_1j_2}y_{i_2j_3}^2 E(\mathbf{X}_{i_1j_1}\mathbf{X}_{i_1j_2})E(\mathbf{X}_{i_2j_3}^2)$$
$$+\sum_{j_1\neq j_2}\sum_{j_3\neq j_4} y_{i_1j_1}y_{i_1j_2}y_{i_2j_3}y_{i_2j_4}E(\mathbf{X}_{i_1j_1}\mathbf{X}_{i_1j_2})E(\mathbf{X}_{i_2j_3}\mathbf{X}_{i_2j_4}),$$

(continues on next page)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

applying 5.7.2,

$$
\begin{aligned}
E(A_{i_1}^2 A_{i_2}^2|\mathbf{Y})_{\{i_1\neq i_2\}} &= \left(\frac{m}{n}\right)^2 \sum_{j_1}\sum_{j_2} y_{i_1j_1}^2 y_{i_2j_2}^2 \\
&+2\left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)\sum_{j_1}\sum_{j_2\neq j_3} y_{i_1j_1}y_{i_1j_2}y_{i_2j_3}^2 \qquad (5.7.27)\\
&+\left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)^2\sum_{j_1\neq j_2}\sum_{j_3\neq j_4} y_{i_1j_1}y_{i_1j_2}y_{i_2j_3}y_{i_2j_4}.
\end{aligned}
$$

5.8 Moments for Evaluating Conditional and Unconditional variances and Covariance for $\hat{\theta}_1$ and $\hat{\theta}_2$

In this section the results from Section 5.7 are used to derive the first and second order marginal and conditional expectations needed for obtaining the within and between variances of $\hat{\theta}_1$ and $\hat{\theta}_2$.

5.8.1 Moments for $\hat{\theta}_1$

$$
\begin{aligned}
E(\hat{\theta}_1) &= E\left[E\left(\sum_i A_iB_i/k\right)\middle|\mathbf{Y}\right] \\
&= E\left[\frac{1}{k}\sum_i B_iE(A_i|\mathbf{Y})\right],
\end{aligned}
$$

applying 5.7.21,

$$
\begin{aligned}
&= E\left(\frac{1}{k}\sum_i B_i\frac{m}{n}B_i\right) \\
&= \left(\frac{1}{k}\right)\left(\frac{m}{n}\right)\sum_i E\left(\sum_j y_{ij}\right)^2 \\
&= \left(\frac{1}{k}\right)\left(\frac{m}{n}\right)\sum_i E\left(\sum_{j_1} y_{ij_1}^2 + \sum_{j_1\neq j_2} y_{ij_1}y_{ij_2}\right)
\end{aligned}
$$

(continues on next page)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

applying 5.7.6 and 5.7.7,

$$E(\hat{\theta}_1) = \frac{1}{k}\frac{m}{n}k(n\mu_2 + n(n-1)\mu_{11}) \\ = m(\mu_2 + (n-1)\mu_{11}). \tag{5.8.1}$$

$$\begin{aligned} E(\hat{\theta}_1^2) &= E\left\{E\left[\frac{1}{k^2}\left(\sum_i A_iB_i\right)\left(\sum_i A_iB_i\right)\middle|\mathbf{Y}\right]\right\} \\ &= E\left[E\left(\frac{1}{k^2}\sum_{i_1}\sum_{i_2} A_{i_1}B_{i_1}A_{i_2}B_{i_2}\middle|\mathbf{Y}\right)\right] \\ &= \frac{1}{k^2}E\left[E\left(\sum_{i_1=i_2} A_{i_1}^2B_{i_1}^2 + \sum_{i_1\neq i_2} A_{i_1}B_{i_1}A_{i_2}B_{i_2}\middle|\mathbf{Y}\right)\right] \\ &= \frac{1}{k^2}E\left[\sum_{i_1=i_2} B_{i_1}^2E(A_{i_1}^2|\mathbf{Y}) + \sum_{i_1\neq i_2} B_{i_1}B_{i_2}E(A_{i_1}A_{i_2}|\mathbf{Y})\right], \end{aligned}$$

applying 5.7.22 and 5.7.23,

$$\begin{aligned} &= \frac{1}{k^2}E\left\{\sum_{i_1=i_2} B_{i_1}^2\left[\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)B_{i_1}^2 + \left(\frac{m}{n}\right)\left(\frac{n-m}{n-1}\right)\sum_{j_3} y_{ij_3}^2\right]\right. \\ &\qquad \left. + \sum_{i_1\neq i_2} B_{i_1}B_{i_2}\left[\left(\frac{m}{n}\right)^2\sum_{j_1}\sum_{j_2} y_{ij_1}y_{ij_2}\right]\right\} \\ &= \frac{1}{k^2}E\left[\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\sum_{i_1} B_{i_1}^4 + \left(\frac{m}{n}\right)\left(\frac{n-m}{n-1}\right)\sum_{i_3}\sum_{j_3} B_{i_1}^2y_{i_1j_3}^2\right. \\ &\qquad \left. + \left(\frac{m}{n}\right)^2\sum_{i_1\neq i_2} B_{i_1}^2B_{i_2}^2\right]. \\ &= \frac{1}{k^2}\left\{\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)kn[\mu_4 + 4(n-1)\mu_{31} + 3(n-1)\mu_{22}\right. \\ &\qquad + 6(n-1)(n-2)\mu_{211} + (n-1)(n-2)(n-3)\mu_{1111}] \\ &\qquad + \left(\frac{m}{n}\right)\left(\frac{n-m}{n-1}\right)kn[\mu_4 + (n-1)\mu_{22} + 2(n-1)\mu_{31} \\ &\qquad\qquad + (n-1)(n-2)\mu_{211}] \\ &\qquad \left. + \left(\frac{m}{n}\right)^2k(k-1)n^2[\mu_2^2 + 2(n-1)\mu_{11}\mu_2 + (n-1)^2\mu_{11}^2]\right\}. \end{aligned} \tag{5.8.2}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
E\{[E(\hat{\theta}_1|\mathbf{Y})]^2\} &= E\left[\frac{1}{k^2}\left(\frac{m}{n}\right)^2\left(\sum_i B_i^2\right)^2\right] \\
&= \frac{1}{k^2}\left(\frac{m}{n}\right)^2 E\left(\sum_{i_1}\sum_{i_2} B_{i_1}^2 B_{i_2}^2\right) \\
&= \frac{1}{k^2}\left(\frac{m}{n}\right)^2 E\left[\sum_{i_1}\sum_{i_2}\left(\sum_{j_1}\sum_{j_2}\sum_{j_3}\sum_{j_4} y_{i_1j_1}y_{i_1j_2}y_{i_1j_3}y_{i_1j_4}\right)\right] \\
&= \frac{1}{k^2}\left(\frac{m}{n}\right)^2 E\left[\sum_{i_1}\left(\sum_{j_1} y_{i_1j_1}^4 + 4\sum_{j_1\neq j_2} y_{i_1j_1}y_{i_1j_2}^3 + 3\sum_{j_1\neq j_2} y_{i_1j_1}^2 y_{i_1j_2}^2\right.\right. \\
&\quad + 6\sum_{j_1\neq j_2\neq j_3} y_{i_1j_1}y_{i_1j_2}y_{i_1j_3}^2 \quad + \sum_{j_1\neq j_2\neq j_3\neq j_4} y_{i_1j_1}y_{i_1j_2}y_{i_1j_3}y_{i_1j_4}\Bigg) \\
&\quad + \sum_{i_1\neq i_2}\left(\sum_{j_1}\sum_{j_3} y_{i_1j_1}^2 y_{i_2j_3}^2 + 2\sum_{j_1\neq j_2}\sum_{j_3} y_{i_1j_1}y_{i_1j_2}y_{i_2j_3}^2\right. \\
&\quad \left.\left. + \sum_{j_1\neq j_2}\sum_{j_3\neq j_4} y_{i_1j_1}y_{i_1j_2}y_{i_2j_3}y_{i_2j_4}\right)\right] \\
&= \frac{1}{k^2}\left(\frac{m}{n}\right)^2\Bigg[k\Big(n\mu_4 + 4n(n-1)\mu_{13} + 3n(n-1)\mu_{22} \\
&\qquad + 6n(n-1)(n-2)\mu_{211} \\
&\qquad + n(n-1)(n-2)(n-3)\mu_{1111}\Big) \\
&\qquad + k(k-1)\left(n^2\mu_2^2 + 2n^2(n-1)\mu_{211} + n^2(n-1)^2\mu_{11}^2\right)\Bigg] \\
&= \frac{m^2}{kn}\Bigg[\Big(\mu_4 + 4(n-1)\mu_{31} + 3(n-1)\mu_{22} + 6(n-1)(n-2)\mu_{211} \\
&\qquad + (n-1)(n-2)(n-3)\mu_{1111}\Big) \\
&\qquad + n(k-1)\left(\mu_2^2 + 2(n-1)\mu_{211} + (n-1)^2\mu_{11}^2\right)\Bigg].
\end{aligned}
\tag{5.8.3}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5.8.2 Moments for $\hat{\theta}_2$

$$\begin{aligned}
E(\hat{\theta}_2) &= E\left[E\left(\sum_i A_i^2/k \middle| \mathbf{Y}\right)\right] \\
&= \frac{1}{k}E\left[\sum_i E(A_i^2|\mathbf{Y})\right], \text{ applying 5.7.22} \\
&= \frac{1}{k}E\left\{\sum_i\left[\frac{m}{n}\sum_{j_1} y_{ij_1}^2 + \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\sum_{j_1\neq j_2} y_{ij_1}y_{ij_2}\right]\right\} \\
&= \frac{1}{k}E\left\{\sum_i\left[\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)B_i^2 + \left(\frac{m}{n}\right)\left(\frac{n-m}{n-1}\right)\sum_{j_1} y_{ij_1}^2\right]\right\}. \qquad (5.8.4) \\
&= \frac{1}{k}E\left\{\sum_i\left[\frac{m}{n}\sum_{j_1} y_{ij_1}^2 + \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\sum_{j_1\neq j_2} y_{ij_1}y_{ij_2}\right]\right\} \\
&= \frac{1}{k}\frac{m}{n}kn\left(\mu_2 + \frac{m-1}{n-1}(n-1)\mu_{11}\right) \\
&= m\left(\mu_2 + (m-1)\mu_{11}\right).
\end{aligned}$$

$$\begin{aligned}
E(\hat{\theta}_2^2) &= E\left\{E\left[\left(\sum_{i=1}^k A_i^2/k\right)^2 \middle| \mathbf{Y}\right]\right\} \\
&= \frac{1}{k^2}E\left\{E\left[\left(\sum_{i=1}^k A_i^2\right)^2 \middle| \mathbf{Y}\right]\right\} \\
&= \frac{1}{k^2}E\left[E\left(\sum_{i_1}\sum_{i_2} A_{i_1}^2 A_{i_2}^2 \middle| \mathbf{Y}\right)\right] \\
&= \frac{1}{k^2}E\left[E\left(\sum_{i_1} A_{i_1}^4 + \sum_{i_1\neq i_2} A_{i_1}^2 A_{i_2}^2 \middle| \mathbf{Y}\right)\right] \\
&= \frac{1}{k^2}E\left\{\left(\sum_{i_1} E(A_{i_1}^4|\mathbf{Y}) + \sum_{i_1\neq i_2} E(A_{i_1}^2 A_{i_2}^2|\mathbf{Y})\right)\right\}, \text{ applying 5.7.26 and 5.7.27,}
\end{aligned}$$

(continues on next page)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
E(\hat{\theta}_2^2) = \frac{1}{k^2} E\Bigg\{ & \sum_{i_1} \Bigg[\frac{m}{n} \sum_{j_1} y_{1j_1}^4 + 4\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right) \sum_{j_1 \neq j_2} y_{ij_1} y_{ij_2}^3 \\
& + 3\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right) \sum_{j_1 \neq j_2} y_{ij_1}^2 y_{ij_2}^2 \\
& + 6\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\frac{m-2}{n-2}\right) \sum_{j_1 \neq j_2 \neq j_3} y_{ij_1} y_{ij_2} y_{ij_3}^2 \\
& + \left(\frac{m}{n}\right)\left(\frac{m-1}{m-1}\right)\left(\frac{m-2}{n-2}\right)\left(\frac{m-3}{n-3}\right) \sum_{j_1 \neq j_2 \neq j_3 \neq j_4} y_{ij_1} y_{ij_2} y_{ij_3} y_{ij_4} \Bigg] \\
& + \sum_{i_1 \neq i_2} \Bigg[\left(\frac{m}{n}\right)^2 \sum_{j_1} \sum_{j_3} y_{i_1 j_1}^2 y_{i_2 j_3}^2 \\
& \quad + 2\left(\frac{m}{n}\right)^2 \left(\frac{m-1}{n-1}\right) \sum_{i_1 \neq i_2} \sum_{j_3} y_{i_1 j_1} y_{i_1 j_2} y_{i_2 j_3}^2 \\
& \quad + \left(\frac{m}{n}\right)^2 \left(\frac{m-1}{n-1}\right)^2 \sum_{j_1 \neq j_2} \sum_{j_3 \neq j_4} y_{i_1 j_1} y_{i_1 j_2} y_{i_2 j_3} y_{i_2 j_4} \Bigg] \Bigg\} \\
= \frac{1}{k^2} \Bigg\{ & kn \Bigg[\frac{m}{n}\mu_4 + 4\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)(n-1)\mu_{31} + 3\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)(n-1)\mu_{22} \\
& + 6\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\frac{m-2}{n-2}\right)(n-1)(n-2)\mu_{211} \\
& + \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\frac{m-2}{n-2}\right)\left(\frac{n-3}{n-3}\right)(n-1)(n-2)(n-3)\mu_{1111} \Bigg] \\
& + k(k-1)n^2 \Bigg[\left(\frac{m}{n}\right)^2 \mu_2^2 + 2\left(\frac{m}{n}\right)^2 \left(\frac{m-1}{n-1}\right)(n-1)\mu_{11}\mu_2 \\
& \quad + \left(\frac{m}{n}\right)^2 \left(\frac{m-1}{n-1}\right)^2 (n-1)^2 \mu_{11}^2 \Bigg] \Bigg\} \\
= \frac{1}{k^2} \Bigg\{ & kn \Bigg[\frac{m}{n}\mu_4 + 4\left(\frac{m}{n}\right)(m-1)\mu_{31} + 3\left(\frac{m}{n}\right)(m-1)\mu_{22} \\
& + 6\left(\frac{m}{n}\right)(m-1)(m-2)\mu_{211} + \left(\frac{m}{n}\right)(m-1)(m-2)(m-3)\mu_{1111} \Bigg] \\
& + k(k-1)n^2 \left(\frac{m}{n}\right)^2 \mu_2^2 + 2\left(\frac{m}{n}\right)^2 (m-1)\mu_{11}\mu_2 + \left(\frac{m}{n}\right)^2 (m-1)^2 \mu_{11}^2 \Bigg\}
\end{aligned}
$$

(continues on next page)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
E(\hat{\theta}_2^2) = \frac{n}{k}\left(\frac{m}{n}\right)\Bigg\{ & \Big[\mu_4 + 4(m-1)\mu_{31} + 3(m-1)\mu_{22} + 6(m-1)(m-2)\mu_{211} \\
& + (m-1)(m-2)(m-3)\mu_{1111}\Big] \\
& + (k-1)n\frac{m}{n}\Big[\mu_2^2 + 2(m-1)\mu_{11}\mu_2 + (m-1)^2\mu_{11}^2\Big]\Bigg\} \\
= \frac{m}{k}\Bigg\{ & \Big[\mu_4 + 4(m-1)\mu_{31} + 3(m-1)\mu_{22} + 6(m-1)(m-2)\mu_{211} \\
& + (m-1)(m-2)(m-3)\mu_{1111}\Big] \\
& + (k-1)m\Big[\mu_2^2 + 2(m-1)\mu_{11}\mu_2 + (m-1)^2\mu_{11}^2\Big]\Bigg\}.
\end{aligned}
\tag{5.8.5}
$$

$$
\begin{aligned}
E[E(\hat{\theta}_2|\mathbf{Y})^2] &= E\Bigg\{\frac{1}{k^2}\Bigg[\sum_i\Bigg(\frac{m}{n}\sum_{j_1} y_{ij_1}^2 + \frac{m}{n}\frac{m-1}{n-1}\sum_{j_1\neq j_2} y_{ij_1}y_{ij_2}\Bigg)\Bigg]^2\Bigg\} \\
&= \frac{1}{k^2}E\Bigg\{\Bigg[\sum_i\Bigg(\frac{m}{n}\frac{m-1}{n-1}B_i^2 + \frac{m}{n}\frac{n-m}{m-1}\sum_j y_{ij}^2\Bigg)\Bigg]^2\Bigg\} \\
&= \frac{1}{k^2}E\Bigg[\sum_{i1}\sum_{i2}\Bigg(\frac{m}{n}\frac{m-1}{n-1}B_{i_1}^2 + \frac{m}{n}\frac{n-m}{n-1}\sum_{j3} y_{i_1j_3}^2\Bigg) \\
&\qquad\qquad \Bigg(\frac{m}{n}\frac{m-1}{n-1}B_{i_2}^2 + \frac{m}{n}\frac{n-m}{n-1}\sum_{j6} y_{i_2j_6}^2\Bigg)\Bigg] \\
&= \frac{1}{k^2}E\Bigg\{\sum_{i_1}\Bigg[\left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)^2 B_{i_1}^4 \\
&\qquad + 2\left(\frac{m}{n}\right)^2\frac{(m-1)(n-m)}{(n-1)^2}B_{i_1}^2\sum_{j_3} y_{i_1j_3}^2 \\
&\qquad + \left(\frac{m}{n}\right)^2\left(\frac{n-m}{n-1}\right)^2\Bigg(\sum_{j_3} y_{i_1j_3}^4 + \sum_{j_3\neq j_6} y_{i_1j_3}y_{i_2j_6}\Bigg)\Bigg] \\
&\quad + \sum_{i_1\neq i_2}\Bigg[\left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)^2 B_{i_1}^2B_{i_2}^2 \\
&\qquad + 2\left(\frac{m}{n}\right)^2\frac{(m-1)(n-m)}{(n-1)^2}B_{i_1}^2\sum_{j_3} y_{i_2j_3}^2 \\
&\qquad + \left(\frac{m}{n}\right)^2\left(\frac{n-m}{n-1}\right)^2\sum_{j_3}\sum_{j_6} y_{i_1j_3}^2 y_{i_2j_6}^2\Bigg]\Bigg\},
\end{aligned}
$$

(continues on next page)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

applying 5.7.19, 5.7.17,5.7.20 and 5.7.18,

$$
\begin{aligned}
E[E(\hat{\theta}_2|\mathbf{Y})^2] = \frac{1}{k^2}\Bigg\{ & \left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)^2 kn\Bigg([mu_4 + 4(n-1)\mu_{31} + 3(n-1)\mu_{22} \\
& \qquad + 6(n-1)(n-2)\mu_{211} \\
& \qquad + (n-1)(n-2)(n-3)\mu_{111}\Big] \\
& + 2\left(\frac{m}{n}\right)^2 \frac{(m-1)(n-m)}{(n-1)^2} kn\Big[\mu_4 + (n-1)\mu_{22} + 2(n-1)\mu_{31} \\
& \qquad + (n-1)(n-2)\mu_{211}\Big] \\
& + \left(\frac{m}{n}\right)^2\left(\frac{n-m}{n-1}\right)^2 kn\Big[\mu_4 + (n-1)\mu_{22}\Big] \\
& + k(k-1)n^2 \\
& \Bigg[\left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)^2\Big(\mu_2^2 + 2(n-1)\mu_{11}\mu_2 + (m-1)^2\mu_{11}^2\Big) \\
& \quad + 2\left(\frac{m}{n}\right)^2 \frac{(m-1)(n-m)}{(n-1)^2}\Big(\mu_2^2 + (n-1)\mu_{11}\mu_2\Big) \\
& \quad + \left(\frac{m}{n}\right)^2\left(\frac{n-m}{n-1}\right)^2 \mu_2^2\Bigg]\Bigg\}.
\end{aligned}
\tag{5.8.6}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5.8.3 Cross-Product Moments

$$E(\hat{\theta}_1\hat{\theta}_2) = E\left[E\left(\sum_{i=1}^{k} A_iB_i/k \sum_{i=1}^{k} A_i^2/k \middle| \mathbf{Y}\right)\right]$$

$$= \frac{1}{k^2}E\left[E\left(\sum_{i_1}\sum_{i_2} A_{i_1}B_{i_1}A_{i_2}^2 \middle| \mathbf{Y}\right)\right]$$

$$= \frac{1}{k^2}E\left[E\left(\sum_{i_1} A_{i_1}^3 B_{i_1} + \sum_{i_1\neq i_2} A_{i_1}B_{i_1}A_{i_2}^2 \middle| \mathbf{Y}\right)\right]$$

$$= \frac{1}{k^2}E\left[\left(\sum_{i_1} B_{i_1}E(A_{i_1}^3|\mathbf{Y}) + \sum_{i_1\neq i_2} B_{i_1}E(A_{i_1}A_{i_2}^2|\mathbf{Y})\right)\right],$$

applying 5.7.24 and 5.7.25,

$$= \frac{1}{k^2}E\Bigg[\sum_{i_1} B_{i_1}\Bigg(\frac{m}{n}\sum_{j_2} y_{i_1j_2}^3 + 3\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\sum_{j_1\neq j_2} y_{i_1j_1}y_{i_1j_2}^2$$

$$+ \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\frac{m-2}{n-2}\right)\sum_{j_1\neq j_2\neq j_3} y_{i_1j_1}y_{i_1j_2}y_{i_1j_3}\Bigg)$$

$$+ \sum_{i_1\neq i_2} B_{i_1}\Bigg(\left(\frac{m}{n}\right)^2\sum_{j_2}\sum_{j_3} y_{i_1j_2}y_{i_2j_3}^2$$

$$+ \left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)\sum_{j_2}\sum_{j_3\neq j_4} y_{i_1j_2}y_{i_2j_3}y_{i_2j_4}\Bigg)\Bigg]$$

(continues on next page)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
E(\hat{\theta}_1\hat{\theta}_2) = \frac{1}{k^2}E\Bigg\{\sum_{i_1}\Bigg[\frac{m}{n}\Bigg(\sum_{j_1} y_{i_1j_1}^4 + \sum_{j_1\neq j_2} y_{i_1j_1}y_{i_1j_2}^3\Bigg) \\
+ 3\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\Bigg(\sum_{j_2\neq j_3} y_{i_1j_2}^2y_{i_1j_3}^2 + \sum_{j_2\neq j_3} y_{i_1j_2}y_{i_1j_3}^3 \\
+ \sum_{j_1\neq j_2\neq j_3} y_{i_1j_1}y_{i_1j_2}y_{i_1j_3}^2\Bigg) \\
+ \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\frac{m-2}{n-2}\right)\Bigg(3\sum_{j_1\neq j_2\neq j_3} y_{i_1j_1}y_{i_1j_2}y_{i_1j_3}^2 \\
+ \sum_{j_1\neq j_2\neq j_3\neq j_4} y_{i_1j_1}y_{i_1j_2}y_{i_1j_3}y_{i_1j_4}\Bigg)\Bigg] \\
+ \sum_{i_1\neq i_2}\Bigg[\left(\frac{m}{n}\right)^2\Bigg(\sum_{j_2}\sum_{j_3} y_{i_1j_2}^2y_{i_2j_3}^2 + \sum_{j_1\neq j_2}\sum_{j_3} y_{i_1j_1}y_{i_1j_2}y_{i_2j_3}^2\Bigg) \\
+ \left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)\Bigg(\sum_{j_2}\sum_{j_3\neq j_4} y_{i_1j_2}^2y_{i_2j_3}y_{i_2j_4} \\
+ \sum_{j_1\neq j_2}\sum_{j_3\neq j_4} y_{i_1j_1}y_{i_1j_2}y_{i_2j_3}y_{i_2j_4}\Bigg)\Bigg]\Bigg\} \\
= \frac{1}{k^2}\Bigg\{kn\Bigg[\frac{m}{n}\Big(\mu_4 + (n-1)\mu_{31}\Big) \\
+ 3\left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\Big((n-1)\mu_{22} + (n-1)\mu_{31} + (n-1)(n-2)\mu_{211}\Big) \\
+ \left(\frac{m}{n}\right)\left(\frac{m-1}{n-1}\right)\left(\frac{m-2}{n-2}\right)\Big(3(n-1)(n-2)\mu_{211} \\
+ (n-1)(n-2)(n-3)\mu_{1111}\Big)\Bigg] \\
+ k(k-1)n^2\Bigg[\left(\frac{m}{n}\right)^2\Big(\mu_2^2 + (n-1)\mu_{11}\mu_2\Big) \\
+ \left(\frac{m}{n}\right)^2\left(\frac{m-1}{n-1}\right)\Big((n-1)\mu_2\mu_{11} + (n-1)^2\mu_{11}^2\Big)\Bigg]\Bigg\} \\
= \frac{m}{k}\Bigg\{\Big(\mu_4 + (n-1)\mu_{31}\Big) + 3(m-1)\Big(\mu_{22} + \mu_{31} + (n-2)\mu_{211}\Big) \\
+ (m-1)(m-2)\Big(3\mu_{211} + (n-3)\mu_{1111}\Big) \\
+ m(k-1)\Big[\mu^2 + (n-1)\mu_{11}\mu_2 + (m-1)\Big(\mu_2\mu_{11} + (n-1)\mu_{11}^2\Big)\Big]\Bigg\}
\end{aligned}
\tag{5.8.7}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
E[E(\hat{\theta}_1|\mathbf{Y})E(\hat{\theta}_2|\mathbf{Y})] &= E\Bigg\{\frac{1}{k}\left(\frac{m}{n}\right)\sum_{i_1} B_{i_1}^2 \frac{1}{k}\sum_{i_2}\Bigg[\frac{m}{n}\left(\frac{m-1}{n-1}\right)B_{i_2}^2 \\
&\qquad + \frac{m}{n}\left(\frac{n-m}{n-1}\right)\sum_{j_3} y_{i_2 j_3}^2\Bigg]\Bigg\} \\
&= \frac{1}{k^2}\left(\frac{m}{n}\right)^2 E\Bigg[\sum_{i_1}\sum_{i_2} B_{i_1}^2\left(\frac{m-1}{n-1}B_{i_2}^2 + \frac{n-m}{n-1}\sum_{j_3} y_{i_2 j_3}^2\right)\Bigg] \\
&= \frac{1}{k^2}\left(\frac{m}{n}\right)^2 E\Bigg[\sum_{i_1}\left(\frac{m-1}{n-1}B_{i_1}^4 + \frac{n-m}{n-1}B_{i_1}^2\sum_{j_3} y_{i_1 j_3}^2\right) \\
&\qquad + \sum_{i_1\neq i_2}\left(\frac{m-1}{n-1}B_{i_1}^2 B_{i_2}^2 + \frac{n-m}{n-1}B_{i_1}^2\sum_{j_3} y_{i_2 j_3}^2\right)\Bigg],
\end{aligned}
$$

applying 5.7.19.,5.7.17,5.7.20 and 5.7.18,

$$
\begin{aligned}
&= \frac{1}{k^2}\left(\frac{m}{n}\right)^2 \Bigg\{kn\left(\frac{m-1}{n-1}\right)\Big(\mu_4 + 4(n-1)\mu_{31} + 3(n-1)\mu_{22} \\
&\qquad + 6(n-1)(n-2)\mu_{211} \\
&\qquad + (n-1)(n-2)(n-3)\mu_{1111}\Big) \\
&\quad + kn\left(\frac{n-m}{n-1}\right)\Big(\mu_4 + (n-1)\mu_{22} + 2(n-1)\mu_{31} \\
&\qquad + (n-1)(n-2)\mu_{211}\Big) \\
&\quad + k(k-1)n^2\Bigg[\frac{m-1}{n-1}\Big(\mu_2^2 + 2(n-1)\mu_{11}\mu_2 \\
&\qquad + (n-1)^2\mu_{11}^2\Big) \\
&\qquad + \frac{n-m}{n-1}\Big(\mu_2^2 + (n-1)\mu_{11}\mu_2\Big)\Bigg]\Bigg\}
\end{aligned}
\tag{5.8.8}
$$

5.9 Evaluation of Expressions for Large Sample Approximations for $\hat{\Sigma}_{\hat{\Theta}}$, $\hat{\beta}$ and $\hat{\theta}$ and their Respective Components of Variance

In this section expressions for large sample approximations for the moments of $\hat{\theta}_1$ and $\hat{\theta}_2$ from Section 5.8 are evaluated and presented as functions of the data in Section

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5.9.1. These expressions are then used first to evaluate the mean and components of variance of $\hat{\beta}$ in Section 5.9.2, and then finally the components of variance of $\hat{\theta}$ in Section 5.9.3.

5.9.1 Expressions for Variances and Covariances of $\hat{\theta}_1$ and $\hat{\theta}_2$ and Their Between and Within Components

Using the results for first and second order moments for $\hat{\theta}_1$ and $\hat{\theta}_2$ from Sections 5.8, the expressions for the overall,between and within variances and covariance for $\hat{\theta}_1$ and $\hat{\theta}_2$ from the equations in Sections 5.6 were evaluated and simplified in Maple 6 as given below.

From the first and second order central moments for $\hat{\theta}_1$, 5.8.1 and 5.8.2, respectively,

$$var(\hat{\theta}_1) = \frac{m}{k}(n\sigma_a^2 + \sigma_e^2)(n\sigma_e^2 + m\sigma_e^2 + 2mn\sigma_a^2). \tag{5.9.1}$$

Applying 5.8.2 and 5.8.3 to 5.6.3,

$$var_{[w]}(\hat{\theta}_1) = \frac{m}{k}(n - m)(n\sigma_a^2 + \sigma_e^2)\sigma_e^2. \tag{5.9.2}$$

Applying 5.8.1 and 5.8.3 to 5.6.4,

$$var_{[b]}(\hat{\theta}_1) = \frac{2m^2}{k(n\sigma_a^2 + \sigma_e^2)^2}. \tag{5.9.3}$$

From the first and second order central moments for $\hat{\theta}_2$, 5.8.4 and 5.8.5, respectively,

$$var(\hat{\theta}_2) = \frac{2m^2}{k}(\sigma_e^2 + m\sigma_a^2)^2. \tag{5.9.4}$$

Applying 5.8.5 and 5.8.6 to 5.6.3,

$$var_{[w]}(\hat{\theta}_2) = \frac{2(n - m)m^2\sigma_e^2(2mn\sigma_a^2 + n\sigma_e^2 - 2\sigma_e^2 - 2m\sigma_a^2 + m\sigma_e^2)}{kn(n - 1)}. \tag{5.9.5}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Applying 5.8.4 and 5.8.6 to 5.6.4,

$$var_{[b]}(\hat{\theta}_2) = \frac{2m^2(m^2n^2\sigma_a^4 - m^2\sigma_a^4 n + 2m^2 n\sigma_a^2\sigma_e^2 + n\sigma_e^4 + m^2\sigma_e^4 - 2m\sigma_e^4 - 2m^2\sigma_a^2\sigma_e^2)}{kn(n-1)}. \qquad (5.9.6)$$

Using the first order central moments of $\hat{\theta}_1$, $\hat{\theta}_2$ and their cross product expectation, 5.8.1, 5.8.4 and 5.8.7, respectively,

$$cov(\hat{\theta}_1, \hat{\theta}_2) = \frac{2m^2}{k}(\sigma_e^2 + n\sigma_a^2)(\sigma_e^2 + m\sigma_a^2). \qquad (5.9.7)$$

Applying 5.8.7 and 5.8.8 to 5.6.5,

$$cov_{[w]}(\hat{\theta}_1, \hat{\theta}_2) = \frac{2m^2}{kn}(n-m)(\sigma_e^2 + n\sigma_a^2)\sigma_e^2. \qquad (5.9.8)$$

Applying 5.8.8, 5.8.1 and 5.8.4 to 5.6.6,

$$cov_{[b]}(\hat{\theta}_1, \hat{\theta}_2) = \frac{2m^3}{kn}(\sigma_e^2 + n\sigma_a^2)^2. \qquad (5.9.9)$$

5.9.2 Expressions for the Variance of $\hat{\beta}$ and its Between and Within Components

From Section 5.8 and 5.9 we have the moments and components of variance and covariance for $\hat{\theta}_1$ and $\hat{\theta}_2$ required to apply the formulas for the Taylor series approximation for the mean and variance of $\hat{\beta}$ in Section 5.4 and the components of the variance of $\hat{\beta}$ from Section 5.5., the expressions for the mean of $\hat{\beta}$ and overall, within and between variances for $\hat{\beta}$, as simplified in Maple 6, are as follows.

Substituting 5.8.1, 5.8.4, 5.9.4 and 5.9.7 into 5.4.1,

$$E(\hat{\beta}) \doteq \frac{(\sigma_e^2 + \sigma_a^2 n)}{(\sigma_e^2 + m\sigma_a^2)} = \beta. \qquad (5.9.10)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Substituting 5.8.1, 5.8.4, 5.9.1, 5.9.4 and 5.9.7 into 5.4.2,

$$\begin{aligned} var(\hat{\beta}) &\doteq \frac{(\sigma_e^2+\sigma_a^2 n)(n-m)\sigma_e^2}{mk(\sigma_e^2+\sigma_a^2 m)} \\ &= \frac{(n-m)\beta\sigma_e^2}{mk}. \end{aligned} \tag{5.9.11}$$

Substituting 5.8.1, 5.8.4, 5.9.2, 5.9.5 and 5.9.8 into 5.4.2,

$$\begin{aligned} var_{[w]}(\hat{\beta}) \doteq \sigma_e^2(\sigma_e^2+\sigma_a^2 n)(n-m)(m^2n^2\sigma_a^4+n^2\sigma_e^4-m^2\sigma_a^4 n \\ +2m^2n\sigma_a^2\sigma_e^2-2mn\sigma_a^2\sigma_e^2-n\sigma_e^4-2nm\sigma_e^4 \\ +2m^2\sigma_e^4)/\left(m(\sigma_e^2+m\sigma_a^2)^4 k(n-1)n\right). \end{aligned} \tag{5.9.12}$$

Substituting 5.8.1, 5.8.4, 5.9.3, 5.9.6 and 5.9.9 into 5.5.8,

$$var_{[b]}(\hat{\beta}) = \frac{2(n-m)^2(\sigma_e^2+\sigma_a^2 n)^2\sigma_e^4}{kn(n-1)(\sigma_e^2+\sigma_a^2 m)^4}. \tag{5.9.13}$$

The first order partial derivative of $\hat{\theta}$ with respect to $\hat{\beta}$ is

$$\frac{\partial\hat{\theta}}{\partial\hat{\beta}} = \frac{(\sigma_e^2+m\sigma_a^2)^2}{(n-m)(\sigma_a^2+\sigma_e^2)^2}. \tag{5.9.14}$$

5.9.3 Expressions for the Variance of $\hat{\theta}$ and its Between and Within Components

From Section5.9 we have the moments and variances and covariance for $\hat{\theta}_1$ and $\hat{\theta}_2$ required to obtain the Taylor series approximation for the variance of $\hat{\theta}$ and it's components as a function of $\hat{\beta}$ in Section 5.5. The expressions for the overall, within and between variances for $\hat{\theta}$, as simplified in Maple 6, are given below.

Substituting 5.9.11 and 5.9.14 into 5.5.9,

$$var(\hat{\theta}) \doteq \frac{(\sigma_e^2+m\sigma_a^2)^2(\sigma_e^2+\sigma_a^2 n)\sigma_e^2}{km(n-m)(\sigma_a^2+\sigma_e^2)^4}. \tag{5.9.15}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Substituting 5.9.12 and 5.9.14 into 5.5.10,

$$var_{[w]}(\hat{\theta}) \doteq (\sigma_e^2 + \sigma_a^2 n)(m^2\sigma_a^4 n^2 - m^2\sigma_a^4 n - 2\sigma_a^2 nm\sigma_e^2 + 2\sigma_a^2 m^2\sigma_e^2 n + \sigma_e^4 n^2 - \sigma_e^4 n - 2\sigma_e^4 mn + 2m^2\sigma_e^4)\sigma_e^2 / \left[(\sigma_a^2 + \sigma_e^2)^4 (n-m)n(n-1)mk\right]. \quad (5.9.16)$$

Substituting 5.9.13 and 5.9.14 into 5.5.11,

$$var_{[b]}(\hat{\theta}) = \frac{2(\sigma_e^2 + \sigma_a^2 n)^2 \sigma_e^4}{kn(n-1)(\sigma_a^2 + \sigma_e^2)^4}. \quad (5.9.17)$$

It is interesting to note here that the approximation for $var_{[b]}(\hat{\theta})$ depends on both variance components and n, but is completely independent of m, whereas $var_{[w]}(\hat{\theta})$ depends on both variance components, and both m and n. The section that follows will be concerned with the validity of the expressions derived for the within and between variances for $\hat{\theta}$.

5.10 Simulation Checks for Between and Within Variance Approximations

Simulated data sets were used to check the results for the between and within variance expressions for $\hat{\theta}$ from the previous section. 1000 data sets were simulated for $\theta = .0588, .2, .3509, .4019, .5, .6491, .8, .9412$, with $n = 8$ and $k = 1000$. For each data set, MR estimates were computed for each of 4 sub-samples for $m = 1, 2, \ldots, 7$.

SAS was used to compute estimates for the between and within variance estimates for $\hat{\theta}$ by analyzing one-way ANOVA's, treating the independent data sets as levels of a single random factor and sub-samples as the levels for reps. Expected mean square for the random component was used to estimate the between variance, and the MSE used to estimate the within variances.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 5.1 plots $var_{[w]}(\hat{\theta})$ against sub-sample size, m, for each simulated value of θ. The theoretical within variances plotted using "T" as a symbol, and within values estimated from the data simulation plotted as an "S". The closeness of the simulated and theoretical values for $var_{[w]}(\hat{\theta})$ indicate that there is a high degree of agreement between two, giving support to validity of the theoretical expression derived for $var_{[w]}(\hat{\theta})$.

Figure 5.2 plots $var_{[b]}(\hat{\theta})$ against sub-sample size,m, for each simulated value of θ. At each value of θ a single point in the graph represents the value of $var_{[b]}(\hat{\theta})$ a single sub-sample, corresponding to one of the values of m. The smooth curve represents the value of the theoretical $var_{[b]}(\hat{\theta})$ derived in the last section. This graph indicates that the calculated values of $var_{[b]}(\hat{\theta})$ for the simulations conform reasonably well to the function for the theoretical $var_{[b]}(\hat{\theta})$. There is, however, a great deal of variability among the simulated values, particularly so in the mid-region of the parameter space.

To understand the nature of the variability within the simulated values for $var_{[b]}(\hat{\theta})$ for fixed m values, $var_{[b]}(\hat{\theta})$ is plotted against m for each value of θ in figure 5.3. The graphs in this figure seem to indicate a moderate degree of variability among the simulated $var_{[b]}(\hat{\theta})$ values at each value of m. Furthermore, there appears to be no functional relationship between the simulated $var_{[b]}(\hat{\theta})$ and m, which is consistent with the theoretical expression of $var_{[b]}(\hat{\theta})$ derived, which is independent of m.

Collectively, figures 5.1, 5.2 and 5.3 give good support to the validity to the expressions for the within and between variances for $\hat{\theta}$ derived.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

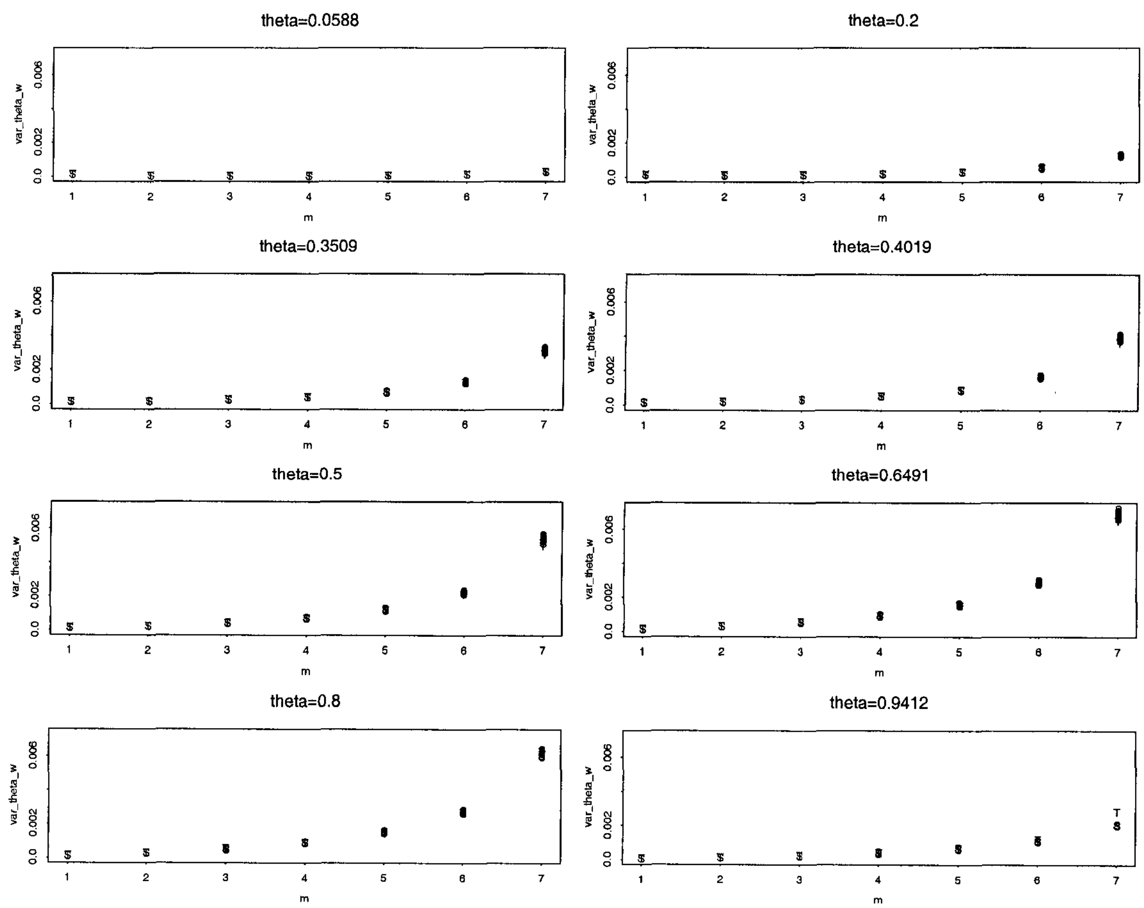

Figure 5.1: Variance of $\hat{\theta}_W$ plotted against sub-sample size (m) for several values of θ.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

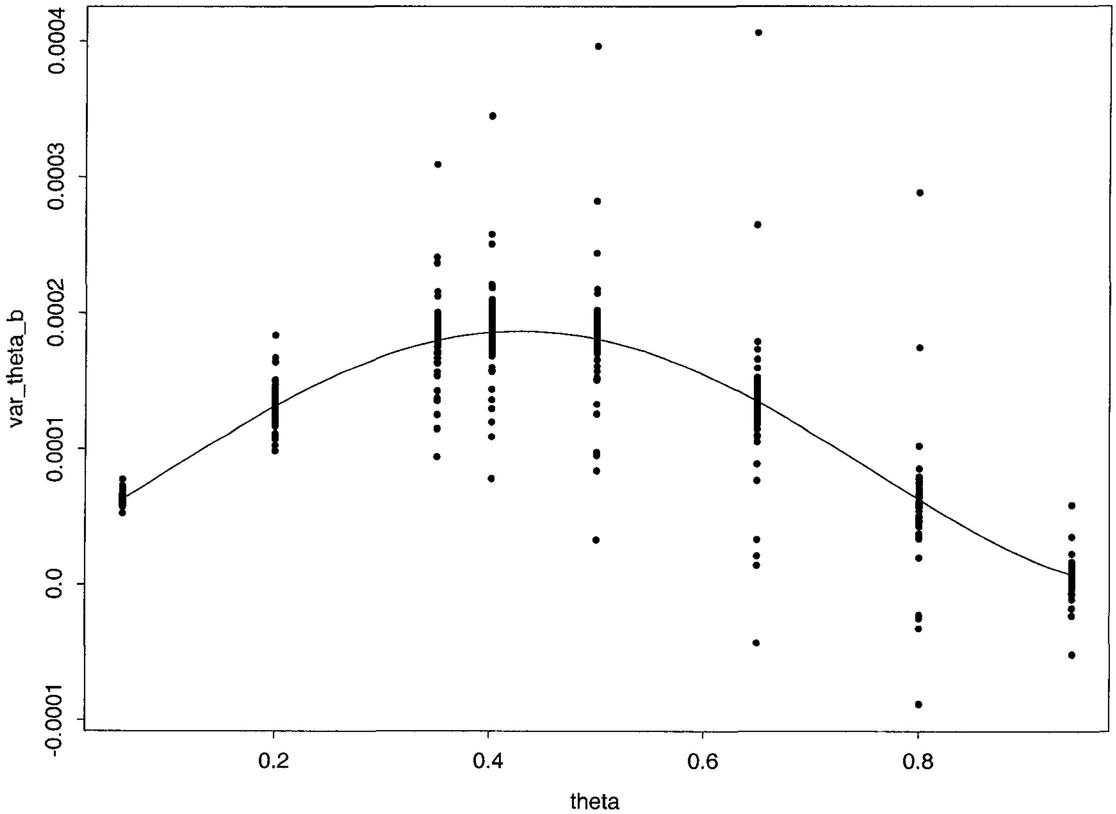

Figure 5.2: Variance of $\hat{\theta}_B$ plotted against θ for multiple values of m.

5.11 Asymptotic Relative Efficiency of MLE to MR for Single Sub-Sample

From Donner and Koval (1980), an expression for the asymptotic variance of the MLE of $\hat{\theta}$ is given by

$$var(\hat{\theta}_{MLE}) = \frac{2\lambda^2(n+\lambda)^2}{nk(1+\lambda)^4(n-1)}. \tag{5.11.1}$$

From the previous section, the asymptotic between variance for the MR estimator of $\hat{\theta}$ is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
var_{[b]}(\hat{\theta}_{MR}) &= \frac{2(\sigma_e^2+\sigma_a^2 n)^2\sigma_e^4}{(\sigma_a^2+\sigma_e^2)^4 n(n-1)k} \\
&= \frac{2(\lambda+n)^2\sigma_a^4\sigma_e^4}{nk(1+\lambda)^4\sigma_a^8(n-1)} \\
&= \frac{2(n+\lambda)^2\lambda^2}{nk(1+\lambda)^4(n-1)} \\
&= var(\hat{\theta}_{MLE})
\end{aligned}
$$

Therefore the ARE of $\hat{\theta}_{MR}$ (pooled estimate from multiple experiments) to the MLE is 1.

5.12 Proposed Bootstrap Confidence Intervals for Method-R, Mallinckrodt (1997)

Bootstrap confidence intervals for MR estimates were proposed by Mallinckrodt et al. (1997). Mallinckrodt states that the bootstrap variance for the MR estimator of θ is given by $var_{[b]}(\hat{\theta}_{MR})$. The procedure relies on an assumption that the bootstrap variance can by approximated without constructing the bootstrap distribution for the MR estimator. Specifically, it is assumed that for 50% sub-sampling, $var_{[w]}(\hat{\theta}_{MR}) \approx var_{[b]}(\hat{\theta}_{MR})$. Mallinckrodt et al. (1997) simulated 500 data sets for heritability values of $\theta = .1, .2, .3, .4, .5$ for each of two types of pedigree structure (random and from real data). For each data set, sub-samples of size 10, 20 and 50 were used to compute compute confidence intervals for θ at levels 80,80,95 and 99%. For the confidence intervals, T-intervals were computed for normalizing transforms (square root, arcsine and Box-Cox) of $\hat{\theta}_{MR}$. While reported overall coverage levels appear to be acceptable, one of several questions that linger is whether that $var_{[w]}(\hat{\theta}_{MR})$ adequately estimates $var_{[b]}(\hat{\theta}_{MR})$ to be called a bootstrap confidence interval. To justify estimating $var_{[b]}(\hat{\theta}_{MR})$ with $var_{[w]}(\hat{\theta}_{MR})$ under 50% sub-sampling, computed the ratio of the standard deviations (within/between) (reported), and also

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

tested for differences between the variances using ANOVA F-tests (not reported). Reported evidence on the empirical justification on approximating $var_{[b]}(\hat{\theta}_{MR})$ with $var_{[w]}(\hat{\theta}_{MR})$ is still ambiguous.

To further investigate whether $var_{[b]}(\hat{\theta}_{MR})$ can be estimated by $var_{[w]}(\hat{\theta}_{MR})$, we look once again to the one-way model for additional evidence. 1000 data sets were simulated for the one-way random effects model with fixed group size $n = 4$, number of groups $k = 4, 8, 16, 20, 36, 50, 76, 100, 500, 1000, 2000$ and heritability values $\theta = .0588, .2, .3509, .4019, .5, .6491, .8, .9412$. In each case, $\hat{\theta}^{(r,s)}$ values were computed for the $R = 1000$ data sets, for $S = 4$ sub-samples. The expected mean squares for the one-way ANOVA model 1.2.1 were used to estimate $var_{[w]}(\hat{\theta}_{MR})$ and $var_{[b]}(\hat{\theta}_{MR})$. Figure 5.4 plots the ratio of between variability to within variability for $\hat{\theta}_{MR}$, $R_{bw} = \frac{var_{[b]}(\hat{\theta}_{MR})}{var_{[w]}(\hat{\theta}_{MR})}$ against k. For $k \leq 2000$, R_{bw} appears to be close to 1 only for $\theta = .3509$, and does so slowly as k increases. Indeed, over most of the parameter space, R_{bw} can be quite far from 1. This would suggest that for the one-way random effects model, $var_{[w]}(\hat{\theta}_{MR})$ is a poor estimator for $var_{[b]}(\hat{\theta}_{MR})$. In light of these results, it would seem that in general, $var_{[w]}(\hat{\theta}_{MR})$ is not an adequate estimator of $var_{[b]}(\hat{\theta}_{MR})$, the bootstrap variance. And as such, confidence intervals computed using $var_{[w]}(\hat{\theta}_{MR})$ should not be referred to as bootstrap intervals.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

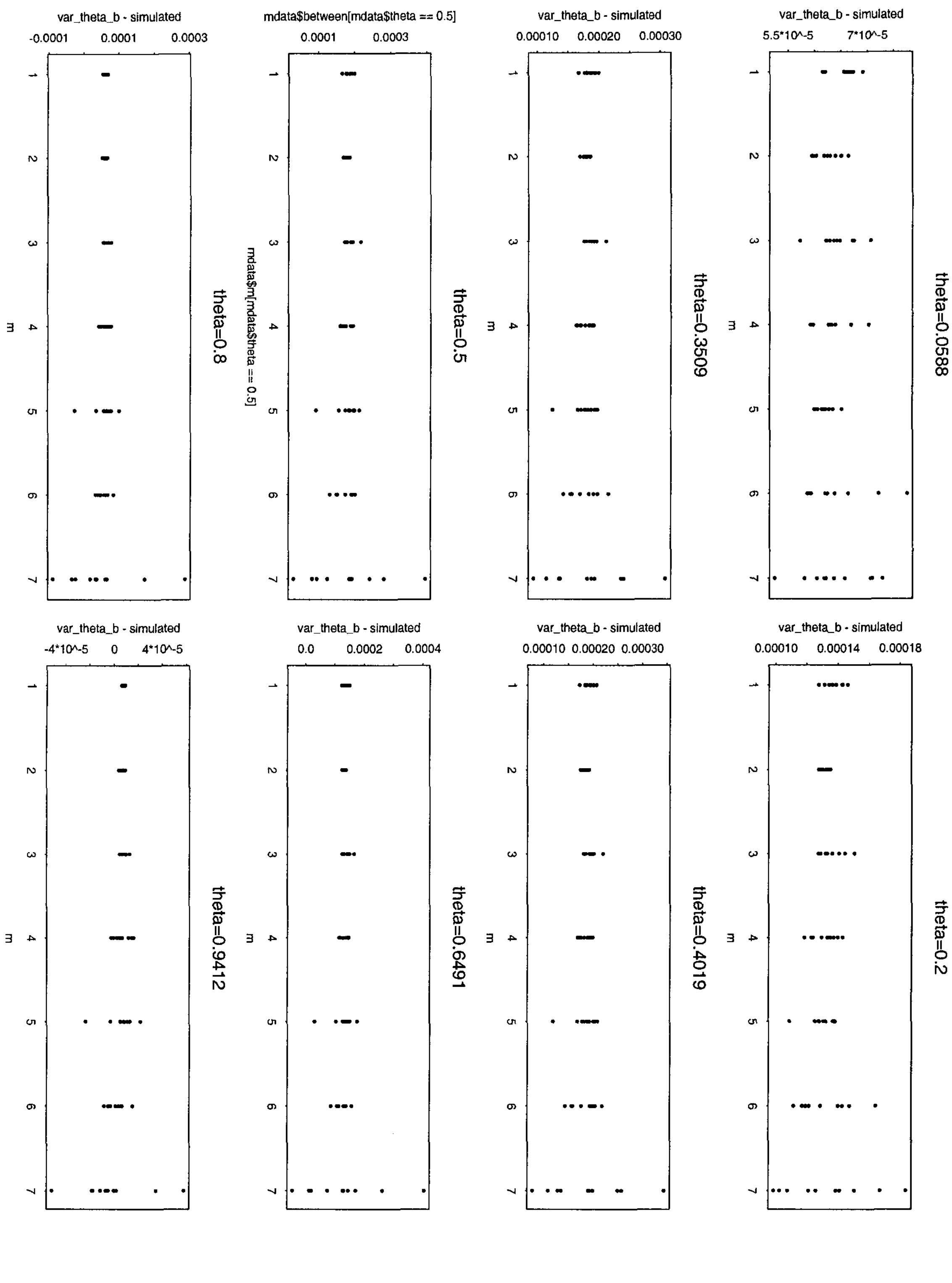

Figure 5.3: Variance of $\hat{\theta}_B$ plotted against sub-sample size m for several values of θ.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

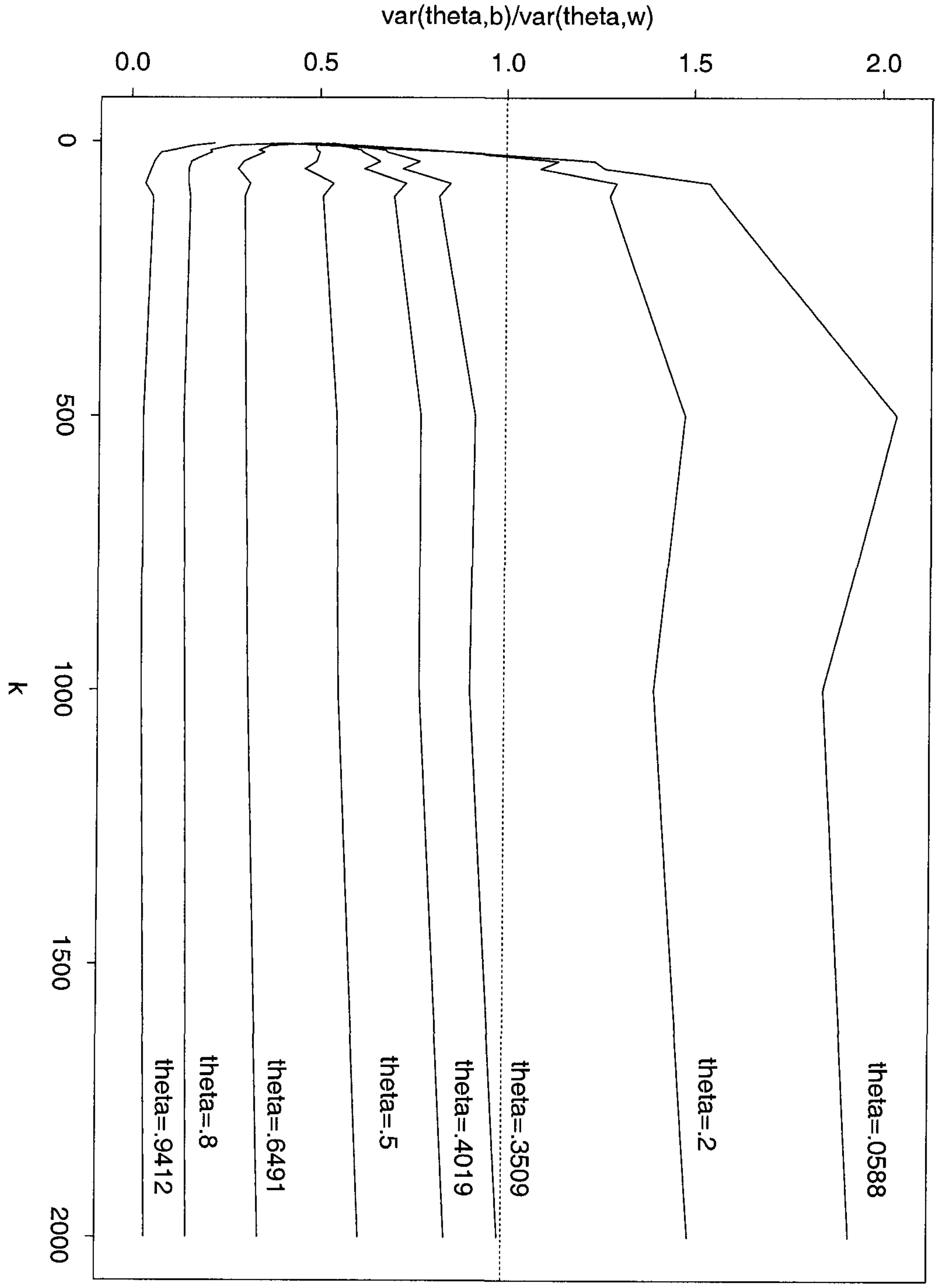

Figure 5.4: Ratio of between variance to within variance for $\hat{\theta}$ plotted against group size (k), for values of θ.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 6

BIVARIATE METHOD R

6.1 Introduction

In Chapter 2 the MR estimator was defined for the univariate case. In Chapter 3, properties of the MR estimator were examined in some detail for the simple balanced one-way random effects model. In this chapter, the arguments in Chapters 2 and 3 are extended in the study of the MR estimator for two traits. In particular, for the bivariate case, the characterization of the MR estimation through equating covariances between new and old predictions to covariances of "old" predictions, is shown to be equivalent to performing a multivariate regression of "old" predictions on "new" predictions. In the case of the balanced one-way bivariate MANOVA model, the multiple trait MR estimator provides a conditional MLE, given the sub-sample means, for the ratio of residual to genetic covariance matrices, Λ, where Λ is the multivariate analog of λ in the univariate case. An algorithm for this multivariate regression approach to MR is provided and implications of its use is discussed.

6.2 Bivariate Mixed Linear Model

The bivariate mixed linear model can be described as follows. Let $\mathbf{y}_i$ be the response vector of dimension $n \times 1$ for trait i $(i = 1, 2)$, with

$$\mathbf{y}_i = \mathbf{X}_i\gamma_i + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i. \tag{6.2.1}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For trait i $\{i = 1, 2\}$, γ_i is the $p \times 1$ fixed effects parameter vector with $n \times p$ design matrix $\mathbf{X}_i$; $\mathbf{u}_i \sim (\mathbf{0}, \mathbf{D}_{ii})$ is the $q \times 1$ random effects parameter vector. $\mathbf{Z}_i$ is the $n \times q$ design matrix corresponding to the random effects parameter vector, and $cov(\mathbf{u}_1, \mathbf{u}_2) = \mathbf{D_{12}}$; $\mathbf{e}_i \sim (\mathbf{0}, \mathbf{R}_{ii})$ is the $n \times 1$ random vector of residuals, and $cov(\mathbf{e}_1, \mathbf{e}_2) = \mathbf{R_{12}}$, and $\mathbf{e}_i$'s are independent of $\mathbf{u}_i$'s.

If we let

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \ \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \ \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \ \mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \text{ and}$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \end{bmatrix},$$

the model can be written as

$$\mathbf{y} = \mathbf{X}\gamma + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where $var(\mathbf{u}) = \mathbf{D} = \begin{bmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{12} & \mathbf{D}_{22} \end{bmatrix}$ and $var(\mathbf{e}) = \mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{12} & \mathbf{R}_{22} \end{bmatrix}$. Then

$$var(\mathbf{u}) = \mathbf{D} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}.$$

One special case of this model that we will use in this chapter is the two trait animal model, for which $\mathbf{u}_i \sim (\mathbf{0}, \mathbf{A}\sigma_i^2)$, where $\mathbf{A}$ is the $q \times q$ genetic relationship matrix; $cov(\mathbf{u}_1, \mathbf{u}_2) = \mathbf{A}\sigma_{12}$; $\mathbf{e}_i \sim (\mathbf{0}, \mathbf{I}\sigma_{e_i}^2)$ and $cov(\mathbf{e}_1, \mathbf{e}_2) = \mathbf{I}\sigma_{e_{12}}$.

6.3 Previous Work on Multiple Trait Method R

Method R estimation for multiple traits was first proposed by Reverter (1994c). For the two trait animal model, the MR estimates for trait heritabilities were obtained by equating the covariance between EBLUPS for whole and partial data to the corresponding covariance for EBLUPS for partial data. Let $\mathbf{u}_i$ and $\mathbf{u}_j$ be the vectors

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of the additive genetic random effects for traits i and j, respectively. Let $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{u}}_j$ be the whole sample BLUPS for traits i and j, respectively, and let $\tilde{\mathbf{u}}_i$ and $\tilde{\mathbf{u}}_j$ be the sub-sample BLUPS for traits i and j, respectively. Utilizing the result (see Section 6.4) that

$$cov(\tilde{\mathbf{u}}_i, \hat{\mathbf{u}}_j) = cov(\tilde{\mathbf{u}}_i, \tilde{\mathbf{u}}_j), \quad \text{for } i, j = 1, 2.$$

Reverter (1994c) defines his multiple trait MR estimators of the covariance parameters, as the values for the covariance parameters that satisfy the condition that $R_{ij} = 1$ for $i, j = 1, 2$, where R_{ij} is defined as

$$R_{ij} = \frac{\tilde{\mathbf{u}}_i' \mathbf{A}^{-1} \hat{\mathbf{u}}_j}{\tilde{\mathbf{u}}_i' \mathbf{A}^{-1} \tilde{\mathbf{u}}_j}, \quad \text{for } i, j = 1, 2. \tag{6.3.1}$$

When a covariance parameter of interest is overestimated, the corresponding regression coefficient is expected to be less than 1. If it is underestimated, the regression coefficient is expected to be greater than 1. As in Reverter (1994b), binary and linear iteration strategies were proposed in Reverter (1994c) for the multiple trait MR procedure. A multiplicative iterative algorithm for updating covariances was also presented for traits i and j. If $\mathbf{G}^n$ is the estimate of the genetic covariance matrix at the n^{th} iteration, then $\mathbf{G}^{n+1}$ is given by the equation below.

$$\mathbf{G}^{n+1} = \mathbf{G}^n * \begin{bmatrix} R_{ii}^n & R_{ij}^n \\ R_{ji}^n & R_{jj}^n \end{bmatrix}, \tag{6.3.2}$$

where the above product is a Hadamard product. The regression coefficients R_{ij} and R_{ji}, are in general, not equal, so the average of the two was used to guarantee symmetry. Convergence for this version of multiple trait MR was found to be quite slow, requiring many rounds for convergence. Misztal (1997) used the secant method to help reduce the number of rounds to convergence. Druet et.al.(2000), used a modified version of (6.3.1), suggested by Reverter(2000), that updates $\mathbf{G}$ by splitting the components of the regressions as follows.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\mathbf{G}^{n+1} = \begin{bmatrix} \tilde{\mathbf{u}}_i'\mathbf{A}^{-1}\hat{\mathbf{u}}_i & \tilde{\mathbf{u}}_j'\mathbf{A}^{-1}\hat{\mathbf{u}}_i \\ \tilde{\mathbf{u}}_i'\mathbf{A}^{-1}\hat{\mathbf{u}}_j & \tilde{\mathbf{u}}_j'\mathbf{A}^{-1}\hat{\mathbf{u}}_j \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{u}}_i'\mathbf{A}^{-1}\tilde{\mathbf{u}}_i & \tilde{\mathbf{u}}_j'\mathbf{A}^{-1}\tilde{\mathbf{u}}_i \\ \tilde{\mathbf{u}}_i'\mathbf{A}^{-1}\tilde{\mathbf{u}}_j & \tilde{\mathbf{u}}_j'\mathbf{A}^{-1}\tilde{\mathbf{u}}_j \end{bmatrix}^{-1} \mathbf{G}^n. \quad (6.3.3)$$

Druet(2000) reports that the the use of the above formula gives a $\mathbf{G}$ matrix that is asymmetric during iteration, but approaches symmetry on convergence. $\mathbf{G}$ can also be forced to be symmetric by averaging the off-diagonal components.

Publications applying multiple trait MR are still only a few in number. Aside from those cited above, most recently Druet (2000;2001 and 2002) has addressed the problem of covariance estimation across traits. Some analytical arguments for MR estimates for simple random effects models for the single and two trait cases.

6.4 Definition of Multiple Trait Method-R

For the two trait animal model application of equation 6.2.1, let $\hat{\mathbf{u}}_{2q\times1} = \begin{bmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix}$ be the *BLUP* for $\mathbf{u}_{2q\times1} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$ based on all the data, $\mathbf{y}_{2n\times1} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$, denoted by $BLUP(\mathbf{u}|\mathbf{y})$. Let $\tilde{\mathbf{u}}_{2q\times1} = \begin{bmatrix} \tilde{\mathbf{u}}_1 \\ \tilde{\mathbf{u}}_2 \end{bmatrix}$ be the *BLUP* for $\mathbf{u}$ based on a subset of $\mathbf{y}$. The MR algorithm relies on the result from the following theorem from Reverter (1994a). The proof provided is essentially that in Reverter (1994c), with minor corrections in notation.

Theorem 6.1 *For the bivariate mixed linear model, Let* $\hat{\mathbf{u}}$ *be the BLUP for* $\mathbf{u}$ *using all the data, and let* $\tilde{\mathbf{u}}$ *be the BLUP for* $\mathbf{u}$ *from a subset of the data. Then the following holds,*

$$cov(\hat{\mathbf{u}}, \tilde{\mathbf{u}}) = var(\tilde{\mathbf{u}}).$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Proof. Let $\mathbf{y}$ be the entire $2n \times 1$ observable random vector of responses and $\mathbf{u}$ the $2q \times 1$ unobservable vector of random effects. Then without loss in generality, the elements of $\mathbf{y}$ can be re-ordered and written as a partition, $\mathbf{y} = \begin{bmatrix} \mathbf{y}_{11} \\ \mathbf{y}_{12} \\ \mathbf{y}_{21} \\ \mathbf{y}_{22} \end{bmatrix}$, where the sub-vector $\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \end{bmatrix}$, for $i = 1, 2$; $\mathbf{y}_{i1}$ an $m \times 1$ random vector, and $\mathbf{y}_{i2}$ an $(n-m) \times 1$ random vector, for $i = 1, 2$. Let $\hat{\mathbf{u}}_i = BLUP(\mathbf{u}_i|\mathbf{y}_i)$ and let $\tilde{\mathbf{u}}_i = BLUP(\mathbf{u}_i|\mathbf{y}_{i1})$, for $i = 1, 2$. Let matrices $\mathbf{X}$ and $\mathbf{Z}$ represent the design matrices corresponding to fixed and random effects, respectively, for the re-ordered data vector. And let the matrices $\mathbf{D}$, $\mathbf{R}$ and $\mathbf{V}$ represent the covariance matrices corresponding to the re-ordered random vectors $\mathbf{u}$, $\mathbf{e}$ and $\mathbf{y}$, respectively. Then

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{u} \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{X}\gamma \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{V} & \mathbf{ZD} \\ \mathbf{DZ}' & \mathbf{D} \end{bmatrix} \right).$$

Let $\mathbf{C} = cov(\mathbf{y}, \mathbf{u}) = \mathbf{ZD}$. Then with respect to the partition on $\mathbf{y}$,

$$\mathbf{C} = \begin{bmatrix} cov(\mathbf{y}_{11}, \mathbf{u}) \\ cov(\mathbf{y}_{12}, \mathbf{u}) \\ cov(\mathbf{y}_{21}, \mathbf{u}) \\ cov(\mathbf{y}_{22}, \mathbf{u}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{11,1} & \mathbf{C}_{11,2} \\ \mathbf{C}_{12,1} & \mathbf{C}_{12,2} \\ \mathbf{C}_{21,1} & \mathbf{C}_{21,2} \\ \mathbf{C}_{22,1} & \mathbf{C}_{22,2} \end{bmatrix}.$$

Then the mean and variance of $\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{u} \end{bmatrix}$ is

$$\begin{bmatrix} \mathbf{y}_{11} \\ \mathbf{y}_{12} \\ \mathbf{y}_{21} \\ \mathbf{y}_{22} \\ \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{X}_{11}\gamma_1 \\ \mathbf{X}_{12}\gamma_1 \\ \mathbf{X}_{21}\gamma_2 \\ \mathbf{X}_{22}\gamma_2 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{V}_{11,11} & \mathbf{V}_{11,12} & \mathbf{V}_{11,21} & \mathbf{V}_{11,22} & \mathbf{C}_{11,1} & \mathbf{C}_{11,2} \\ \mathbf{V}_{12,11} & \mathbf{V}_{12,12} & \mathbf{V}_{12,21} & \mathbf{V}_{12,22} & \mathbf{C}_{12,1} & \mathbf{C}_{12,2} \\ \mathbf{V}_{21,11} & \mathbf{V}_{21,12} & \mathbf{V}_{21,21} & \mathbf{V}_{21,22} & \mathbf{C}_{21,1} & \mathbf{C}_{21,2} \\ \mathbf{V}_{22,11} & \mathbf{V}_{22,12} & \mathbf{V}_{22,21} & \mathbf{V}_{22,22} & \mathbf{C}_{22,1} & \mathbf{C}_{22,2} \\ \mathbf{C}'_{11,1} & \mathbf{C}'_{12,1} & \mathbf{C}'_{21,1} & \mathbf{C}'_{22,1} & \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{C}'_{11,2} & \mathbf{C}'_{12,2} & \mathbf{C}'_{21,2} & \mathbf{C}'_{22,2} & \mathbf{D}_{21} & \mathbf{D}_{22} \end{bmatrix} \right),$$

where $\mathbf{V}_{ij}$'s are the resulting partition on $\mathbf{V}$ for the re-ordered data vector, $\mathbf{y}$. From Schott (1997) theorem 4.3, there exists a triangular $N \times N$ matrix $\mathbf{T}$ having

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

non-negative diagonal elements, such that, $\mathbf{V} = \mathbf{T}\mathbf{T}'$. Further, since $\mathbf{V}$ is a positive definite matrix, $\mathbf{T}$ is unique with positive diagonal elements. Then $\mathbf{T}$ can be written with respect to the partition on $\mathbf{y}$ as

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_{11,11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{T}_{12,11} & \mathbf{T}_{22,12} & \mathbf{0} & \mathbf{0} \\ \mathbf{T}_{21,11} & \mathbf{T}_{21,12} & \mathbf{T}_{21,21} & \mathbf{0} \\ \mathbf{T}_{22,11} & \mathbf{T}_{22,12} & \mathbf{T}_{22,21} & \mathbf{T}_{22,22} \end{bmatrix}.$$

Let $\mathbf{z} = \mathbf{T}^{-1}\mathbf{y}$. Then the model for the transform $\mathbf{z}$ is

$$\mathbf{z} = \mathbf{T}^{-1}\mathbf{X}\gamma + \mathbf{T}^{-1}\mathbf{Z}\mathbf{u} + \mathbf{T}^{-1}\mathbf{e}.$$

The partition on $\mathbf{z}$ corresponding to that of $\mathbf{y}$ is

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{12} \\ \mathbf{z}_{12} \\ \mathbf{z}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11,11} & 0 & 0 & 0 \\ \mathbf{T}_{12,11} & \mathbf{T}_{22,12} & 0 & 0 \\ \mathbf{T}_{21,11} & \mathbf{T}_{21,12} & \mathbf{T}_{21,21} & 0 \\ \mathbf{T}_{22,11} & \mathbf{T}_{22,12} & \mathbf{T}_{22,21} & \mathbf{T}_{22,22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_{11} \\ \mathbf{y}_{12} \\ \mathbf{y}_{21} \\ \mathbf{y}_{22} \end{bmatrix}.$$

Let $\mathbf{W} = \mathbf{T}^{-1}\mathbf{X}$, with partition

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} \\ \mathbf{W}_{12} \\ \mathbf{W}_{21} \\ \mathbf{W}_{22} \end{bmatrix}, \tag{6.4.1}$$

and let $\mathbf{B} = cov(\mathbf{z}, \mathbf{u}) = \mathbf{T}^{-1}(\mathbf{I}_2 \otimes \mathbf{Z})\mathbf{D}$ with partition

$$\mathbf{B} = \begin{bmatrix} cov(\mathbf{z}_{11}, \mathbf{u}) \\ cov(\mathbf{z}_{12}, \mathbf{u}) \\ cov(\mathbf{z}_{21}, \mathbf{u}) \\ cov(\mathbf{z}_{22}, \mathbf{u}) \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{11,1} & \mathbf{B}_{11,2} \\ \mathbf{B}_{12,1} & \mathbf{B}_{12,2} \\ \mathbf{B}_{21,1} & \mathbf{B}_{21,2} \\ \mathbf{B}_{22,1} & \mathbf{B}_{22,2} \end{bmatrix}.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Then the mean and variance of $\begin{bmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{12} \\ \mathbf{z}_{21} \\ \mathbf{z}_{22} \\ \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$ is

$$\begin{bmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{12} \\ \mathbf{z}_{21} \\ \mathbf{z}_{22} \\ \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{W}_{11}\gamma \\ \mathbf{W}_{12}\gamma \\ \mathbf{W}_{21}\gamma \\ \mathbf{W}_{22}\gamma \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_{11,1} & \mathbf{B}_{11,2} \\ \mathbf{0} & \mathbf{I}_{n-m} & \mathbf{0} & \mathbf{0} & \mathbf{B}_{12,1} & \mathbf{B}_{12,2} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m & \mathbf{0} & \mathbf{B}_{21,1} & \mathbf{B}_{21,1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{n-m} & \mathbf{B}_{21,1} & \mathbf{B}_{21,2} \\ \mathbf{B}'_{11,1} & \mathbf{B}'_{12,1} & \mathbf{B}'_{21,1} & \mathbf{B}'_{22,1} & \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{B}'_{11,2} & \mathbf{B}'_{12,2} & \mathbf{B}'_{21,2} & \mathbf{B}'_{22,2} & \mathbf{D}_{21} & \mathbf{D}_{22} \end{bmatrix} \right).$$

Since $\hat{\mathbf{u}} = BLUP(\mathbf{u}|\mathbf{y}) = BLUP(\mathbf{u}|\mathbf{z})$, solving the MME for $\mathbf{u}$ with respect to the transformed data vector $\mathbf{z}$ gives

$$\begin{aligned} \hat{\mathbf{u}} &= \mathbf{D}(\mathbf{T}^{-1}\mathbf{Z})'(\mathbf{I}_N - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}')\mathbf{z} \\ &= \mathbf{D}(\mathbf{B}\mathbf{D}^{-1})'(\mathbf{I}_N - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}')\mathbf{z} \\ &= \mathbf{B}'(\mathbf{I}_N - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}')\mathbf{z} \\ &= \mathbf{B}'\mathbf{P}\mathbf{z}, \end{aligned} \tag{6.4.2}$$

where

$$\begin{aligned} \mathbf{P} &= \mathbf{I}_N - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}' \\ &= [\mathbf{P}_{ij,i'j'}], \text{ for } ij, i'j' = 11, 12, 21, 22, \end{aligned} \tag{6.4.3}$$

with

$$\mathbf{P}_{ij,i'j'} = \begin{cases} \mathbf{I}_m - \mathbf{W}_{ij,i'j'}(\mathbf{W}_{ij,i'j'}\mathbf{W}_{ij,i'j'})^{-}\mathbf{W}_{ij,i'j'}, & \text{if } i = i' = j = j' \\ \mathbf{I}_{N-m} - \mathbf{W}_{ij,i'j'}(\mathbf{W}_{ij,i'j'}\mathbf{W}_{ij,i'j'})^{-}\mathbf{W}_{ij,i'j'}, & \text{if } i = i', j = j' \\ -\mathbf{W}_{ij,i'j'}(\mathbf{W}_{ij,i'j'}\mathbf{W}_{ij,i'j'})^{-}\mathbf{W}_{ij,i'j'}, & \text{otherwise.} \end{cases} \tag{6.4.4}$$

Similarly, $\tilde{\mathbf{u}} = BLUP(\mathbf{u}|\mathbf{y}_1) = BLUP(\mathbf{u}|(\mathbf{z}_{11}, \mathbf{z}_{12})')$. Solving the MME for $\mathbf{u}$ with respect to $\begin{bmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{12} \end{bmatrix}$ gives

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\tilde{\mathbf{u}} = \mathbf{B'}_1\mathbf{P}_1 \begin{bmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{12} \end{bmatrix}, \tag{6.4.5}$$

where $\mathbf{B'}_1 = \begin{bmatrix} \mathbf{B}_{11,1} & \mathbf{B}_{11,2} \\ \mathbf{B}_{21,1} & \mathbf{B}_{21,2} \end{bmatrix}$ and $\mathbf{P}_1 = \begin{bmatrix} \mathbf{P}_{11,11} & \mathbf{P}_{11,12} \\ \mathbf{P}_{21,11} & \mathbf{P}_{21,21} \end{bmatrix}$.

Therefore

$$\begin{aligned} var(\tilde{\mathbf{u}}) &= \mathbf{B'}_1\mathbf{P}_1 Var\left(\begin{bmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{12} \end{bmatrix}\right)\mathbf{P}_1\mathbf{B}_1 \\ &= \mathbf{B'}_1\mathbf{P}_1\mathbf{I}_2 m\mathbf{P}_1\mathbf{B}_1 \\ &= \mathbf{B'}_1\mathbf{P}_1\mathbf{P}_1\mathbf{B}_1 \\ &= \mathbf{B'}_1 \begin{bmatrix} \mathbf{P}_{11,11} & 0 \\ 0 & \mathbf{P}_{21,21} \end{bmatrix} \mathbf{B}_1, \end{aligned} \tag{6.4.6}$$

since $\mathbf{P}_1$ is symmetric and idempotent, and

$$\begin{aligned} cov(\tilde{\mathbf{u}}, \hat{\mathbf{u}}) &= \mathbf{B}'_1\mathbf{P}_1 cov(\mathbf{z}_1, \mathbf{z})\mathbf{P}\mathbf{B} \\ &= \mathbf{B}'_1\mathbf{P}_1 \begin{pmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m & \mathbf{0} \end{pmatrix} \mathbf{P}\mathbf{B} \\ &= \mathbf{B}'_1 \begin{pmatrix} \mathbf{P}_{11,11} & \mathbf{0} & \mathbf{P}_{11,12} & \mathbf{0} \\ \mathbf{P}_{21,11} & \mathbf{0} & \mathbf{P}_{21,21} & \mathbf{0} \end{pmatrix} \mathbf{P}\mathbf{B} \\ &= \mathbf{B}'_1 \begin{pmatrix} \mathbf{P}_{11,11} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{21,21} \end{pmatrix} \mathbf{B}_1 \\ &= var(\tilde{\mathbf{u}}). \end{aligned} \tag{6.4.7}$$

□

6.5 Derivation of Equations for Bivariate Method-R

In this section we will use the result of Theorem 6.1 to derive the conditions used to define bivariate MR.

From equation 6.4.6 we have the following.

$$cov(\tilde{\mathbf{u}}_i, \hat{\mathbf{u}}_j) = cov(\tilde{\mathbf{u}}_i, \tilde{\mathbf{u}}_j), \text{ for } i = 1, 2;\ j = 1, 2.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Therefore

$$E(\tilde{\mathbf{u}}_i\hat{\mathbf{u}}_j') - E(\mathbf{u}_i)E(\mathbf{u}_j)' = E(\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_j') - E(\mathbf{u}_i)E(\mathbf{u}_j)',$$

$$\rightarrow E(\tilde{\mathbf{u}}_i\hat{\mathbf{u}}_j') = E(\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_j'), \qquad (6.5.1)$$

since $E(\tilde{\mathbf{u}}_i) = E(\hat{\mathbf{u}}_i) = \mathbf{0}$, for $i = 1, 2$. From LHS of equation 6.5.1,

$$\begin{aligned} tr(E(\tilde{\mathbf{u}}_i\hat{\mathbf{u}}_j')) &= \sum_{l=1}^{q} E(\tilde{u}_{il}\hat{u}_{jl}) \\ &= E(\sum_{l=1}^{q} \tilde{u}_{il}\hat{u}_{jl}) \\ &= E(tr(\tilde{\mathbf{u}}_i\hat{\mathbf{u}}_j')) \\ &= E(tr(\hat{\mathbf{u}}_j'\tilde{\mathbf{u}}_i)), \text{ from theorem 1.8.1 in Graybill (1976),} \\ &= E(\hat{\mathbf{u}}_j'\tilde{\mathbf{u}}_i) \\ &= E(\tilde{\mathbf{u}}_i'\hat{\mathbf{u}}_j), \text{ for } i = 1, 2;\ j = 1, 2. \end{aligned}$$

Similarly, for the RHS, we have

$$tr(E(\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_j')) = E(\tilde{\mathbf{u}}_i'\tilde{\mathbf{u}}_j), \text{ for } i = 1, 2;\ j = 1, 2.$$

Thus we have that

$$E(\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j) = E(\tilde{\mathbf{u}}_i'\tilde{\mathbf{u}}_j), \text{ for } i = 1, 2;\ j = 1, 2. \qquad (6.5.2)$$

Also, for any symmetric $q \times q$ matrix of weights, $\mathbf{W}$,

$$\begin{aligned} E(\hat{\mathbf{u}}_i'\mathbf{W}\hat{\mathbf{u}}_j) &= E\left(\sum_{r,s=1}^{q} \hat{u}_{ir}w_{rs}\tilde{u}_{js}\right) \\ &= E\left(\sum_{r=1}^{q} \hat{u}_{ir}w_{rr}\tilde{u}_{jr}\right) + E\left(\sum_{r\neq s}^{q} \hat{u}_{ir}w_{rs}\tilde{u}_{js}\right) \\ &= \sum_{r=1}^{q} w_{rr}E\left(\hat{u}_{ir}\tilde{u}_{jr}\right) + \sum_{r\neq s}^{q} w_{rs}E\left(\hat{u}_{ir}\tilde{u}_{js}\right) \\ &= \sum_{r=1}^{q} w_{rr}E\left(\tilde{u}_i\tilde{u}_i\right) + \sum_{i\neq j}^{q} d_{ij}E\left(\tilde{u}_i\tilde{u}_j\right) \\ &= E(\tilde{\mathbf{u}}_i'\mathbf{W}\tilde{\mathbf{u}}_j), \text{ for } i = 1, 2;\ j = 1, 2. \end{aligned} \qquad (6.5.3)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For the bivariate animal model, $\mathbf{W} = \mathbf{A}^{-1}$. Setting $\hat{\mathbf{u}}_i'\mathbf{A}^{-1}\hat{\mathbf{u}}_j$ and $\tilde{\mathbf{u}}_i'\mathbf{A}^{-1}\tilde{\mathbf{u}}_j$ to their respective expected values, and computing the ratio of LHS to RHS, we have the resulting statistic,

$$R_{ij} = \frac{\hat{\mathbf{u}}_i'\mathbf{A}^{-1}\hat{\mathbf{u}}_j}{\tilde{\mathbf{u}}_i'\mathbf{A}^{-1}\tilde{\mathbf{u}}_j} = 1, \text{ for } i = 1, 2; \; j = 1, 2. \tag{6.5.4}$$

For the bivariate animal model, the MR estimates of the covariance parameters are the values of the covariance parameters that satisfy the system of equations specified in (6.5.3).

6.5.1 Regression Characterization of Bivariate Method R

In this section we will look at the bivariate regression function of whole sample BLUP's given sub-sample BLUP's and it's relationship to bivariate MR.

Let random matrix of BLUPS for the two traits based on all the data be $\hat{\mathbf{U}}_{q\times 2} = [\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2]$, and let the random matrix of BLUPS for the two traits based on a subset of the data be $\tilde{\mathbf{U}}_{q\times 2} = [\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2]$. The bivariate regression model for $\hat{\mathbf{U}}$ on $\tilde{\mathbf{U}}$ is

$$\hat{\mathbf{U}} = \tilde{\mathbf{U}}\boldsymbol{\eta}_{\hat{U}|\tilde{U}} + \boldsymbol{\xi}_{\hat{U}|\tilde{U}}. \tag{6.5.5}$$

$\boldsymbol{\eta}_{\hat{U}|\tilde{U}}$ is a 2×2 matrix of regression coefficient parameters, with $\boldsymbol{\eta}_{\hat{U}|\tilde{U}} = [\boldsymbol{\eta}_1, \boldsymbol{\eta}_2]$, where $\boldsymbol{\eta}_i$ $(i = 1, 2)$ are both 2×1 vectors, and $\boldsymbol{\xi}_{\hat{U}|\tilde{U}}$ is a $q \times 2$ random matrix of regression residuals, with $\boldsymbol{\xi}_{\hat{U}|\tilde{U}} = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2]$, where $\boldsymbol{\xi}_i$ are both $q \times 1$ vectors $(i = 1, 2)$.

The weighted least squares (WLS) estimator of $\boldsymbol{\eta}_{\hat{U}|\tilde{U}}$, $\boldsymbol{\eta}_{\hat{U}|\tilde{U}}$, when using $\mathbf{A}^{-1}$ as the weighting matrix is

$$\hat{\boldsymbol{\eta}}_{\hat{U}|\tilde{U}} = (\tilde{\mathbf{U}}'\mathbf{A}^{-1}\tilde{\mathbf{U}})^{-1}\tilde{\mathbf{U}}'\mathbf{A}^{-1}\hat{\mathbf{U}}. \tag{6.5.6}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Applying the result from Rao (1973) Section 8a.2, the conditional mean of $\hat{\mathbf{u}}_1$ given $\tilde{\mathbf{u}}$ is

$$\begin{aligned} E(\hat{\mathbf{u}}_1 \mid \tilde{\mathbf{u}}) &= E(\hat{\mathbf{u}}_1) + cov(\hat{\mathbf{u}}_1, \tilde{\mathbf{u}})\,(var(\tilde{\mathbf{u}}))^{-1}\,(\tilde{\mathbf{u}} - E(\tilde{\mathbf{u}})) \\ &= \mathbf{u}_1 + cov(\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}})\,(var(\tilde{\mathbf{u}}))^{-1}\,(\tilde{\mathbf{u}} - \mathbf{u}) \\ &= \mathbf{u}_1 + (\mathbf{I}_q \mid \mathbf{0}_{q\times q})(\tilde{\mathbf{u}} - \mathbf{u}) \\ &= \mathbf{u}_1 + (\tilde{\mathbf{u}}_1 - \mathbf{u}_1) \\ &= \tilde{\mathbf{u}}_1. \end{aligned}$$

Similarly, it can be shown that $E(\hat{\mathbf{u}}_2 \mid \tilde{\mathbf{u}}) = \tilde{\mathbf{u}}_2$. Combining the conditional expectations given $\tilde{\mathbf{u}}$, we have $E(\hat{\mathbf{U}} \mid \tilde{\mathbf{U}}) = \tilde{\mathbf{U}}$. Which implies that

$$\begin{aligned} E(\hat{\eta}_{\hat{U}|\tilde{U}}) &= E\left[(\tilde{\mathbf{U}}'\mathbf{A}^{-1}\tilde{\mathbf{U}})^{-1}\tilde{\mathbf{U}}'\mathbf{A}^{-1}\hat{\mathbf{U}}\right] \\ &= (\tilde{\mathbf{U}}'\mathbf{A}^{-1}\tilde{\mathbf{U}})^{-1}\tilde{\mathbf{U}}'\mathbf{A}^{-1}E(\hat{\mathbf{U}} \mid \tilde{\mathbf{U}}) \\ &= (\tilde{\mathbf{U}}'\mathbf{A}^{-1}\tilde{\mathbf{U}})^{-1}\tilde{\mathbf{U}}'\mathbf{A}^{-1}\tilde{\mathbf{U}} \\ &= \mathbf{I}_2. \end{aligned} \tag{6.5.7}$$

Equation 6.5.6 provides an analogous means of defining bivariate MR estimates of the covariance parameters, as the values of the covariance parameters for which $\hat{\eta}_{\hat{U}|\tilde{U}}$ equals its expected value, $\mathbf{I}_2$.

6.6 Equivalence of Method R definitions of Equating Covariances to Multivariate Regression

In this section we will show that the approach to variance component estimation through the conditions of Reverter (1994c), is equivalent to that of MR estimation as defined through weighted multivariate regression.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For the balanced bivariate animal model, Reverter's conditions (from Section 6.3), that $R_{ij} = 1$ for $i, j = 1, 2$, gives the set of equations below.

$$\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1 = \tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1 \tag{6.6.1}$$

$$\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_2 = \tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2 \tag{6.6.2}$$

$$\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1 = \tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1 \tag{6.6.3}$$

$$\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2 = \tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2 \tag{6.6.4}$$

From the multivariate regression of Section 6.5, of $\hat{\mathbf{u}}$ on $\tilde{\mathbf{u}}$, setting the weighted least squares estimator of $\eta_{\hat{U}|\tilde{U}}$ to its expected value gives the system of equations below.

$$(\tilde{\mathbf{U}}'\mathbf{A}^{-1}\tilde{\mathbf{U}})^{-1}\tilde{\mathbf{U}}'\mathbf{A}^{-1}\hat{\mathbf{U}} = \mathbf{I}_2 \tag{6.6.5}$$

The LHS of the above equation 6.6.5

$$\begin{aligned}
&[(\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2)'\mathbf{A}^{-1}(\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2)]^{-1}[(\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2)'\mathbf{A}^{-1}(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2)] \\
&= \begin{bmatrix} \tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1 & \tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2 \\ \tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1 & \tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1 & \tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_2 \\ \tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1 & \tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2 \end{bmatrix} \\
&= \Delta^{-1} \begin{bmatrix} \tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2 & -\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1 \\ -\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2 & \tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1 & \tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_2 \\ \tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1 & \tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2 \end{bmatrix},
\end{aligned} \tag{6.6.6}$$

where $\Delta = (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2) - (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1))$.

Equating the off diagonal elements of the RHS and LHS of 6.6.5, we have the following.

From the upper right off-diagonal element of 6.6.6:

$$(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2)(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_2) - (\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2) = 0,$$

which implies that

$$(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2)(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_2) - (\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2). \tag{6.6.7}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From the lower left off-diagonal element of 6.6.6:

$$(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1) - (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1) = 0,$$

which implies that

$$(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1) = (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1). \tag{6.6.8}$$

Let

$$R_{11} = \frac{\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1}{\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1}, R_{12} = \frac{\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_2}{\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2}, R_{21} = \frac{\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1}{\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1} \text{ and } R_{22} = \frac{\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2}{\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2}.$$

From equation 6.6.7 we have $R_{22} = R_{12}$, and from equation 6.6.8 we have $R_{11} = R_{21}$.

Equating the diagonal elements of the matrix in 6.6.6 to 1, we have the following.

$$\begin{aligned}&(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1) - (\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1)\\ &= (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2) - (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)\end{aligned} \tag{6.6.9}$$

$$\begin{aligned}&(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2) - (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_2)\\ &= (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2) - (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)\end{aligned} \tag{6.6.10}$$

From 6.6.9 we have,

$$\begin{aligned}&(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)[(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1) - (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)] = (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)[(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1) - (\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)]\\ &\rightarrow (\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(R_{11} - 1) = (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)^2(R_{21} - 1),\end{aligned}$$

since $R_{11} = R_{21}$, we must have that

$$\begin{aligned}\{(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1) - (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)^2\}(R_{11} - 1) &= 0\\ \rightarrow \Delta(R_{11} - 1) &= 0\end{aligned} \tag{6.6.11}$$

Since $\tilde{\mathbf{U}}'\mathbf{A}^{-1}\tilde{\mathbf{U}}$ is positive definite, $\Delta > 0$. Therefore $R_{11} = R_{21} = 1$, which implies equations 6.6.1 and 6.6.3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Similarly, from 6.6.10,

$$(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)[(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2)-(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)] = (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)[(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_2)-(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)]$$
$$\rightarrow (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2)(R_{22}-1) = (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)^2(R_{12}-1), \tag{6.6.12}$$

since $R_{22}=R_{12}$, we must have that

$$\{(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2)-(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)^2\}(R_{22}-1)=0 \tag{6.6.13}$$
$$\rightarrow \Delta(R_{22}-1)=0 \tag{6.6.14}$$

Once again, since $\tilde{\mathbf{U}}'\mathbf{A}^{-1}\tilde{\mathbf{U}}$ is positive definite, $\Delta > 0$. Therefore $R_{22}=R_{12}=1$, which implies equations 6.6.2 and 6.6.4.

Hence, we have that the system of equations from the multivariate regression MR criterion implies the system of equations 6.6.1-6.6.4 of Reverter (1994c).

Conversely, we have the following.

From 6.6.1 and 6.6.3, we have $R_{11}=R_{21}$, which

$$\rightarrow \frac{\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1}{\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1} = \frac{\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1}{\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1}$$
$$\rightarrow (\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1) = (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_1)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_1) \rightarrow 6.6.8.$$

Also, from 6.6.2 and 6.6.4, we have $R_{22}=R_{12}$, which

$$\rightarrow \frac{\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2}{\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2} = \frac{\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_2}{\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2}$$
$$\rightarrow (\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\tilde{\mathbf{u}}_2)(\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2) = (\tilde{\mathbf{u}}_2'\mathbf{A}^{-1}\hat{\mathbf{u}}_2)(\tilde{\mathbf{u}}_1'\mathbf{A}^{-1}\hat{\mathbf{u}}_2) \rightarrow 6.6.7.$$

Since $\Delta \neq 0$, from 6.6.1 and 6.6.3 we have the following:

$$\Delta(R_{11}-1)=0 \rightarrow 6.6.11 \rightarrow 6.6.9.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From 6.6.2 and 6.6.4,

$$\Delta(R_{22} - 1) = 0 \rightarrow 6.6.12 \rightarrow 6.6.10.$$

Thus, Reverter's system of equations implies the system of equations for the multivariate regression MR criterion. Therefore, the two approaches are equivalent.

6.7 Bivariate MR for Balanced One-Way MANOVA Model

6.7.1 Introduction and Outline

In this section we will look at the bivariate balanced random effects MANOVA model. As in Chapter 3, the simplicity of this model allows for a convenient means of investigating properties of estimators derived from the bivariate MR procedure. The main result of this section is that the bivariate MR procedure gives a conditional MLE given the whole and sub-sample means, for Λ, the ratio of the residual covariance to the genetic covariance. Below is a brief outline of what is presented in this section.

1. Present the balanced bivariate one-way MANOVA model, corresponding MME and expressions for whole and sub-sample BLUPS. Define insufficient statistics from BLUPS. [Section 6.7.2]

2. Derive expressions for mean and variance of the joint distribution of the whole and sub-sample insufficient statistics.[Section 6.7.3]

3. Derive expressions for the conditional mean and variance whole sample insufficient statistics given the sub-sample insufficient statistics, and present the resulting regression model.[Section 6.7.4]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4. Present the multivariate regression function for whole sample insufficient statistics given the sub-sample insufficient statistics (analogous to multivariate regression in Section 6.5. The least squares estimator of the matrix of regression coefficients is shown to be the conditional MLE based on the insufficient statistics. [Section 6.7.5]

5. Express the BLUP's in terms of the insufficient statistics, then derive expressions for the joint mean and variance for whole and sub-sample BLUP's.[Section 6.7.6]

6. Derive expressions for the conditional mean and variance for whole sample BLUP's given the sub-sample BLUP's, and present the resulting regression model.[Section 6.7.7]

7. Revisit the multivariate regression model for whole sample BLUP's on sub-sample BLUP's. The least squares estimator of the matrix of regression coefficients is shown to be the conditional MLE given the whole and sub-sample means.[Section 6.7.8]

8. The least squares estimator for the matrix of regression coefficients of the BLUP's is defined in terms of the least squares estimator for the matrix of regression coefficients of the insufficient statistics. A closed form expression for the bivariate MR estimator for Λ is derived. [Section 6.7.9]

6.7.2 Bivariate Balanced One-Way MANOVA Model

The balanced bivariate MANOVA model can be described as follows. Let $\mathbf{y}_i$ be the response vector of dimension $kn \times 1$ for trait i ($i = 1, 2$), with

$$\mathbf{y}_i = \mu_i \mathbf{1}_{kn} + \mathbf{Z}_i \mathbf{u}_i + \mathbf{e}_i, \tag{6.7.1}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where for trait $i = 1, 2$, μ_i is a scalar fixed effects parameter, $\mathbf{u}_i$ is the $k \times 1$ random effects parameter vector with design matrix $\mathbf{Z}_i = \mathbf{I}_k \otimes \mathbf{1}_n$, and $\mathbf{e}_i$ is the $kn \times 1$ random vector of residuals, independent of $\mathbf{u}_i$. Then the MANOVA model is

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{kn} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{kn} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} \tag{6.7.2}$$

If we let $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$, $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$ and $\mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$, the model can be written as

$$\mathbf{y} = (\mathbf{I}_2 \otimes \mathbf{1}_{kn})\boldsymbol{\mu} + (\mathbf{I}_2 \otimes \mathbf{Z})\mathbf{u} + \mathbf{e}, \tag{6.7.3}$$

where $var(\mathbf{u}) = \Sigma_{\mathbf{u}} \otimes \mathbf{I}_k$, $var(\mathbf{e}) = \Sigma_{\mathbf{e}} \otimes \mathbf{I}_{kn}$ and the covariance matrices $\Sigma_{\mathbf{u}}$ and $\Sigma_{\mathbf{e}}$ are given by $\Sigma_{\mathbf{u}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$ and $\Sigma_{\mathbf{e}} = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} \\ \sigma_{e_{12}} & \sigma_{e_2}^2 \end{bmatrix}$.

The covariance matrix for $\mathbf{y}$ is then

$$\begin{aligned} var(\mathbf{y}) &= (\mathbf{I}_2 \otimes \mathbf{Z})(\Sigma_{\mathbf{u}} \otimes \mathbf{I}_k)(\mathbf{I}_2 \otimes \mathbf{Z}') + (\Sigma_{\mathbf{e}} \otimes \mathbf{I}_{kn}) \\ &= (\Sigma_{\mathbf{u}} \otimes \mathbf{I}_k \otimes \mathbf{J}_n) + (\Sigma_e \otimes \mathbf{I}_{kn}). \end{aligned}$$

where $\mathbf{J}_n$ is a $n \times n$ matrix with 1's as its elements.

The MME for the balanced bivariate MANOVA model is given by

$$\begin{bmatrix} nk\Sigma_e^{-1} & n(\Sigma_e^{-1} \otimes \mathbf{1}'_k) \\ n(\Sigma_e^{-1} \otimes \mathbf{1}_k) & \Sigma_e^{-1}(n\mathbf{I}_{2k} + \Lambda \otimes \mathbf{I}_k) \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} (\Sigma_e^{-1} \otimes \mathbf{1}'_{kn})\mathbf{y} \\ (\Sigma_e^{-1} \otimes \mathbf{I}_k \otimes \mathbf{1}'_n)\mathbf{y} \end{bmatrix}, \tag{6.7.4}$$

where $\Lambda = \Sigma_e \Sigma_u^{-1}$.

Directly from the MME (equation 6.7.4),

$$(n\mathbf{I}_2 \otimes \mathbf{1}_k)\boldsymbol{\mu} + [n(\mathbf{I_2} + \Lambda) \otimes \mathbf{I_k}]\mathbf{u} = \mathbf{y}_{i\cdot}. \tag{6.7.5}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Therefore

$$[(n\mathbf{I}_2 + \Lambda) \otimes \mathbf{I}_k]\mathbf{u} = \mathbf{y}_{i.} - (n\mathbf{I}_2 \otimes \mathbf{1}_k)\boldsymbol{\mu} \tag{6.7.6}$$

$$\longrightarrow \mathbf{u} = [(n\mathbf{I}_2 + \Lambda) \otimes \mathbf{I}_k]^{-1}[\mathbf{y}_{i.} - (n\mathbf{I}_2 \otimes \mathbf{1}_k)\boldsymbol{\mu}], \tag{6.7.7}$$

$$\longrightarrow \hat{\mathbf{u}} = [(n\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k][\mathbf{y}_{i.} - (n\mathbf{I}_2 \otimes \mathbf{1}_k)\hat{\boldsymbol{\mu}}]. \tag{6.7.8}$$

Now

$$\begin{aligned}(n\mathbf{I}_2 \otimes \mathbf{1}_k)\hat{\boldsymbol{\mu}} &= (n\mathbf{I}_2 \otimes \mathbf{1}_k)(n\mathbf{I}_2 \otimes \mathbf{1}'_{nk})\mathbf{y}/(nk) \\ &= (n\mathbf{I}_2 \otimes \mathbf{J}_k \otimes \mathbf{1}'_n)\mathbf{y}/(nk),\end{aligned} \tag{6.7.9}$$

which implies

$$\begin{aligned}\mathbf{y}_{i.} - (n\mathbf{I}_2 \otimes \mathbf{1}_k)\hat{\boldsymbol{\mu}} &= \left[(\mathbf{I}_2 \otimes \mathbf{I}_k \otimes \mathbf{1}'_n) - \left(\mathbf{I}_2 \otimes \frac{1}{k}\mathbf{J}_k \otimes \mathbf{1}'_n\right)\right]\mathbf{y} \\ &= \left[\mathbf{I}_2 \otimes \left(\mathbf{I}_k - \frac{1}{k}\mathbf{J}_k\right) \otimes \mathbf{1}'_n\right]\mathbf{y}.\end{aligned} \tag{6.7.10}$$

As in Chapter 3, let $\mathbf{P} = \mathbf{I}_k - \frac{1}{k}\mathbf{J}_k$ and let $\mathbf{B} = \mathbf{P} \otimes \mathbf{1}'_n$, then

$$\begin{aligned}\hat{\mathbf{u}} &= \left[(n\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k\right](\mathbf{I}_2 \otimes \mathbf{B})\mathbf{y} \\ &= \left[(n\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{B}\right]\mathbf{y}.\end{aligned} \tag{6.7.11}$$

For sub-samples of size m from n observations in each group, let the corresponding BLUP be $\tilde{\mathbf{u}}$. Define $\mathbf{G} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ to be the "selection" matrix, as described in Chapter 3. Then similar arguments as above, lead to the expression for the sub-sample BLUP,

$$\tilde{\mathbf{u}} = \left[(m\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{B}\mathbf{G}\right]\mathbf{y}. \tag{6.7.12}$$

In Chapter 3, insufficient statistics were defined from the BLUP's that, free of dependence of the unknown covariance parameters. In the univariate case, this

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

approach led to MR solutions free of the unknown covariance parameters and the sampling distribution of the MR estimator for a case of a single sub-sample. We will now define analogous insufficient statistics for the bivariate case

Define $\hat{\mathbf{v}}_{2k\times 1} = (\mathbf{I}_2 \otimes \mathbf{B})\mathbf{y} = \begin{bmatrix} \mathbf{B}\mathbf{y}_1 \\ \mathbf{B}\mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{v}}_1 \\ \hat{\mathbf{v}}_2 \end{bmatrix}$, and define $\tilde{\mathbf{v}}_{2k\times 1} = (\mathbf{I}_3 \otimes \mathbf{BG})\mathbf{y} = \begin{bmatrix} \mathbf{BG}\mathbf{y}_1 \\ \mathbf{BG}\mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{v}}_1 \\ \tilde{\mathbf{v}}_2 \end{bmatrix}$.

6.7.3 Joint Distribution of $\hat{\mathbf{v}}$ and $\tilde{\mathbf{v}}$

Since $E(\mathbf{B}\mathbf{y}_i) = E(\mathbf{BG}\mathbf{y}_i) = \mathbf{0}$, the joint distribution of $(\hat{\mathbf{v}}', \tilde{\mathbf{v}}')'$ is

$$\begin{pmatrix} \hat{\mathbf{v}} \\ \tilde{\mathbf{v}} \end{pmatrix} \sim \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}; \begin{pmatrix} (\mathbf{I}_2 \otimes \mathbf{B})\mathbf{V}(\mathbf{I}_2 \otimes \mathbf{B}') & (\mathbf{I}_2 \otimes \mathbf{B})\mathbf{V}(\mathbf{I}_2 \otimes \mathbf{BG}') \\ (\mathbf{I}_2 \otimes \mathbf{BG}')\mathbf{V}(\mathbf{I}_2 \otimes \mathbf{B}') & (\mathbf{I}_2 \otimes \mathbf{BG})\mathbf{V}(\mathbf{I}_2 \otimes \mathbf{BG}') \end{pmatrix} \right],$$

where $\mathbf{V} = var(\mathbf{y}) = \mathbf{Z}(\Sigma_u \otimes \mathbf{I}_k)\mathbf{Z}' + \Sigma_e \otimes \mathbf{I}_{nk}$; $\mathbf{Z} = \mathbf{I}_2 \otimes \mathbf{I}_k \otimes \mathbf{1}_n$. Therefore

$$\begin{aligned} \mathbf{V} &= (\mathbf{I}_2 \otimes \mathbf{I}_k \otimes \mathbf{1}_n)(\Sigma_u \otimes \mathbf{I}_k)(\mathbf{I}_2 \otimes \mathbf{I}_k \otimes \mathbf{1}_n') + \Sigma_e \otimes \mathbf{I}_{nk} \\ &= (\Sigma_u \otimes \mathbf{I}_k \otimes \mathbf{J}_n) + (\Sigma_e \otimes \mathbf{I}_{nk}). \end{aligned} \tag{6.7.13}$$

Computing the elements of the variance-covariance matrix of $(\hat{\mathbf{v}}', \tilde{\mathbf{v}}')'$,

$$\begin{aligned} var(\hat{\mathbf{v}}) &= (\mathbf{I}_2 \otimes \mathbf{B})\mathbf{V}(\mathbf{I}_2 \otimes \mathbf{B}') \\ &= (\mathbf{I}_2 \otimes \mathbf{P} \otimes \mathbf{1}_n')[(\Sigma_u \otimes \mathbf{I}_k \otimes \mathbf{J}_n) + (\Sigma_e \otimes \mathbf{I}_{nk})](\mathbf{I}_2 \otimes \mathbf{P} \otimes \mathbf{1}_n) \\ &= n^2(\Sigma_u \otimes \mathbf{P}) + n(\Sigma_e \otimes \mathbf{P}) \\ &= n(n\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}. \end{aligned} \tag{6.7.14}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
cov(\hat{\mathbf{v}}, \tilde{\mathbf{v}}) &= (\mathbf{I}_2 \otimes \mathbf{B})\mathbf{V}(\mathbf{I}_2 \otimes (\mathbf{BG})') \\
&= (\mathbf{I}_2 \otimes \mathbf{P} \otimes \mathbf{1}_n')[(\Sigma_u \otimes \mathbf{I}_k \otimes \mathbf{J}_n) + (\Sigma_e \otimes \mathbf{I}_{nk})](\mathbf{I}_2 \otimes \mathbf{P} \otimes \mathbf{1}_n^*) \\
&= nm\Sigma_u \otimes \mathbf{P} + m\Sigma_e \otimes \mathbf{P} \\
&= m(n\Sigma_u \otimes \mathbf{P} + \Sigma_e \otimes \mathbf{P}) \\
&= m(n\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}, \qquad (6.7.15)
\end{aligned}
$$

where $\mathbf{1}_n^* = (\mathbf{1}_m'|\mathbf{0}_{n-m}')'$, i.e. $\mathbf{1}_n^*$ is an n-dimensional vector with m 1's and $(n-m)$ 0's.

$$
\begin{aligned}
var(\tilde{\mathbf{v}}) &= [\mathbf{I}_2 \otimes (\mathbf{BG})]\mathbf{V}[\mathbf{I}_2 \otimes (\mathbf{BG})'] \\
&= (\mathbf{I}_2 \otimes \mathbf{P} \otimes \mathbf{1}_n^{*'})[(\Sigma_u \otimes \mathbf{I}_k \otimes \mathbf{J}_n) + (\Sigma_e \otimes \mathbf{I}_{nk})](\mathbf{I}_2 \otimes \mathbf{P} \otimes \mathbf{1}_n^*) \\
&= m^2(\Sigma_u \otimes \mathbf{P}) + m(\Sigma_e \otimes \mathbf{P}) \\
&= m(m\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}. \qquad (6.7.16)
\end{aligned}
$$

Thus $var[(\hat{\mathbf{v}}', \tilde{\mathbf{v}}')']$ can be written as

$$
var\begin{pmatrix} \hat{\mathbf{v}} \\ \tilde{\mathbf{v}} \end{pmatrix} = \begin{bmatrix} n(n\Lambda^{-1} + \mathbf{I}_2) & m(n\Lambda^{-1} + \mathbf{I}_2) \\ m(n\Lambda^{-1} + \mathbf{I}_2) & m(m\Lambda^{-1} + \mathbf{I}_2) \end{bmatrix} (\Sigma_e \otimes \mathbf{P}). \qquad (6.7.17)
$$

6.7.4 Conditional distribution of $\hat{\mathbf{v}}$ given $\tilde{\mathbf{v}}$

As in Section 3.3.2, applying the result from Rao (1973) Section 8a.2, the conditional mean and variance of $\hat{\mathbf{v}}$ given $\tilde{\mathbf{v}}$ are given by

$$
\begin{aligned}
E(\hat{\mathbf{v}} \mid \tilde{\mathbf{v}}) &= E(\hat{\mathbf{v}}) + cov(\hat{\mathbf{v}}, \tilde{\mathbf{v}})\,(var(\tilde{\mathbf{v}}))^-\,(\tilde{\mathbf{v}} - E(\tilde{\mathbf{v}})) \\
&= \mathbf{0} + [m(n\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}][m(m\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}]^-(\tilde{\mathbf{v}} - \mathbf{0}) \\
&= [(n\Lambda^{-1} + \mathbf{I}_2)(m\Lambda^{-1} + \mathbf{I}_2)^{-1} \otimes \mathbf{I}_k]\tilde{\mathbf{v}} \\
&= (\boldsymbol{\beta} \otimes \mathbf{I}_k)\tilde{\mathbf{v}},
\end{aligned} \qquad (6.7.18)
$$

where $\boldsymbol{\beta}_{2\times 2} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} = [(n\Lambda^{-1} + \mathbf{I}_2)(m\Lambda^{-1} + \mathbf{I}_2)^{-1}]$, and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
var(\hat{\mathbf{v}} \mid \tilde{\mathbf{v}}) &= var(\hat{\mathbf{v}}) - cov(\hat{\mathbf{v}}, \tilde{\mathbf{v}})\,(var(\tilde{\mathbf{v}}))^{-}\,cov(\tilde{\mathbf{v}}, \hat{\mathbf{v}}) \\
&= [n(n\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}] - [m(n\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}] \\
&\qquad [m(m\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}]^{-}[m(n\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}] \\
&= [n(n\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}] - [m(n\Lambda^{-1} + \mathbf{I}_2)(m\Lambda^{-1} + \mathbf{I}_2)^{-1} \\
&\qquad (n\Lambda^{-1} + \mathbf{I}_2)\Sigma_e \otimes \mathbf{P}] \\
&= (n\Lambda^{-1} + \mathbf{I}_2)(m\Lambda^{-1} + \mathbf{I}_2)^{-1}[n(m\Lambda^{-1} + \mathbf{I}_2) - m(n\Lambda^{-1} + \mathbf{I}_2)]\Sigma_e \otimes \mathbf{P} \\
&= (n\Lambda^{-1} + \mathbf{I}_2)(m\Lambda^{-1} + \mathbf{I}_2)^{-1}(n\mathbf{I}_2 - m\mathbf{I}_2)\Sigma_e \otimes \mathbf{P} \\
&= (n - m)(n\Lambda^{-1} + \mathbf{I}_2)(m\Lambda^{-1} + \mathbf{I}_2)^{-1}\Sigma_e \otimes \mathbf{P} \\
&= (n - m)\boldsymbol{\beta}\Sigma_e \otimes \mathbf{P} \\
&= \Sigma_v^* \otimes \mathbf{P},
\end{aligned} \tag{6.7.19}
$$

where $\Sigma_v^* = (n - m)\boldsymbol{\beta}\Sigma_e$.

Note: The matrix $\boldsymbol{\beta}\Sigma_e$ is symmetric and non-negative definite, thus Σ_v^* is symmetric and non-negative definite.

The model corresponding to the regression function in equation 6.7.18 is

$$\hat{\mathbf{v}} = (\boldsymbol{\beta} \otimes \mathbf{I}_k)\tilde{\mathbf{v}} + \boldsymbol{\varepsilon}_{\hat{v}|\tilde{v}}, \tag{6.7.20}$$

where $\boldsymbol{\varepsilon}_{\hat{v}|\tilde{v}_{2k\times 1}} \sim (\mathbf{0}, \Sigma_v^* \otimes \mathbf{P})$.

6.7.5 The Bivariate Regression Function for $\hat{\mathbf{V}}|\tilde{\mathbf{V}}$

Let $\hat{\mathbf{V}}_{k\times 2} = [\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2]$. and let $\tilde{\mathbf{V}}_{k\times 2} = [\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2]$. The bivariate regression model for $\hat{\mathbf{V}}$ on $\tilde{\mathbf{V}}$ is

$$\hat{\mathbf{V}} = \tilde{\mathbf{V}}\boldsymbol{\beta}_{\hat{V}|\tilde{V}} + \boldsymbol{\varepsilon}_{\hat{V}|\tilde{V}}, \tag{6.7.21}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where $\boldsymbol{\beta}_{\hat{V}|\tilde{V}}$ is a 2×2 matrix and $\boldsymbol{\varepsilon}_{\hat{V}|\tilde{V}}$ is a $k \times 2$ matrix. The bivariate regression model in (6.7.21) is essentially a model that collects the univariate multiple regressions of $\hat{\mathbf{v}}_1$ on the $\tilde{\mathbf{v}}_\mathbf{i}$'s and $\hat{\mathbf{v}}_2$ on the $\tilde{\mathbf{v}}_\mathbf{i}$'s ($i = 1, 2$). From the regression function in (6.7.18), the multiple regression models are

$$\hat{\mathbf{v}}_i = \tilde{\mathbf{V}}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \; i = 1, 2, \tag{6.7.22}$$

where $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})'$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22})'$. From (6.7.22), $\boldsymbol{\beta}_{\hat{V}|\tilde{V}} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and $\boldsymbol{\varepsilon}_{\hat{V}|\tilde{V}} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2)$. Combining the models for $\hat{\mathbf{v}}_1$ and $\hat{\mathbf{v}}_2$ given $\tilde{\mathbf{V}}$ in (6.7.22) as a single multiple linear regression model,

$$\begin{aligned} \hat{\mathbf{v}} &= \begin{pmatrix} \tilde{\mathbf{V}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{V}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{pmatrix} \\ &= (\mathbf{I}_2 \otimes \tilde{\mathbf{V}}) \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{pmatrix}. \end{aligned} \tag{6.7.23}$$

From (6.7.23) the weighted least squares (WLS) estimator of $\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$ is

$$\begin{aligned} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix}_{WLS} &= [(\mathbf{I}_2 \otimes \tilde{\mathbf{V}})'(\Sigma_v^* \otimes \mathbf{P})(\mathbf{I}_2 \otimes \tilde{\mathbf{V}})]^-(\mathbf{I}_2 \otimes \tilde{\mathbf{V}})'(\Sigma_v^* \otimes \mathbf{P})^- \begin{pmatrix} \hat{\mathbf{v}}_1 \\ \hat{\mathbf{v}}_2 \end{pmatrix} \\ &= (\Sigma_v^{*-1} \otimes \tilde{\mathbf{V}}'\mathbf{P}\tilde{\mathbf{V}})^-(\Sigma_v^{*-1} \otimes \tilde{\mathbf{V}}'\mathbf{P}) \begin{pmatrix} \hat{\mathbf{v}}_1 \\ \hat{\mathbf{v}}_2 \end{pmatrix} \\ &= (\mathbf{I}_2 \otimes (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}') \begin{pmatrix} \hat{\mathbf{v}}_1 \\ \hat{\mathbf{v}}_2 \end{pmatrix} \\ &= \begin{bmatrix} (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}' & \mathbf{0} \\ \mathbf{0} & (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}' \end{bmatrix} \begin{pmatrix} \hat{\mathbf{v}}_1 \\ \hat{\mathbf{v}}_2 \end{pmatrix}. \end{aligned} \tag{6.7.24}$$

This implies that the WLS estimator for $\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}$ is

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}_{WLS}} &= [(\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}'\hat{\mathbf{v}}_1, (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}'\hat{\mathbf{v}}_2] \\ &= (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}'\hat{\mathbf{V}} = \hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}_{OLS}}, \end{aligned} \tag{6.7.25}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where $\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}_{OLS}}$ is the ordinary least squares estimator (OLS) for $\boldsymbol{\beta}_{\hat{V}|\tilde{V}}$. For the remaining of this chapter we will use $\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}$ to denote the OLS estimator for $\boldsymbol{\beta}_{\hat{V}|\tilde{V}}$.

As in the univariate balanced on-way random effects model, $var(\hat{\mathbf{v}}|\tilde{\mathbf{v}})$ is singular, and thus, the likelihood does not exit. From Section 3.3.2, $\mathbf{P} = \mathbf{CDC}'$, where $\mathbf{D}$ is the diagonal matrix of eigenvalues for $\mathbf{P}$, i.e. $\mathbf{D}_{k\times k} = Diag(\mathbf{1}_{k-1}, 0)$. Let $\mathbf{C}^*_{(k-1)\times k}$ matrix resulting from deleting the last row of $\mathbf{C}$. Consider the linear transform $(\mathbf{I}_2 \otimes \mathbf{C}^*)$ on $\hat{\mathbf{v}}$. The conditional mean and variance for $(\mathbf{I}_2 \otimes \mathbf{C}^*)\hat{\mathbf{v}}$ are given below.

$$
\begin{aligned}
E[(\mathbf{I}_2 \otimes \mathbf{C}^*)\hat{\mathbf{v}}] &= (\mathbf{I}_2 \otimes \mathbf{C}^*)(\boldsymbol{\beta} \otimes \mathbf{I}_k)\tilde{\mathbf{v}} \\
&= (\boldsymbol{\beta} \otimes \mathbf{C}^*)\hat{\mathbf{v}}.
\end{aligned} \tag{6.7.26}
$$

$$
\begin{aligned}
var[(\mathbf{I}_2 \otimes \mathbf{C}^*)\hat{\mathbf{v}}] &= (\mathbf{I}_2 \otimes \mathbf{C}^*)(\Sigma_v^* \otimes \mathbf{P})(\mathbf{I}_2 \otimes \mathbf{C}^*) \\
&= \Sigma_v^* \otimes \mathbf{C}^*\mathbf{P}\mathbf{C}^{*'} \\
&= \Sigma_v^* \otimes \mathbf{C}^*\mathbf{C}'\mathbf{D}\mathbf{C}\mathbf{C}^{*'} \\
&= \Sigma_v^* \otimes \mathbf{I}_{k-1}.
\end{aligned} \tag{6.7.27}
$$

The multiple regression model generated by the transform $(\mathbf{I}_2 \otimes \mathbf{C}^*)$ on multiple regression function for $\hat{\mathbf{v}}$ from (6.7.23) is

$$
\begin{pmatrix} \mathbf{C}^*\hat{\mathbf{v}}_1 \\ \mathbf{C}^*\hat{\mathbf{v}}_2 \end{pmatrix} = [\mathbf{I}_2 \otimes (\mathbf{C}^*\tilde{\mathbf{v}}_1, \mathbf{C}^*\tilde{\mathbf{v}}_1)] \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{C}^*\boldsymbol{\varepsilon}_1 \\ \mathbf{C}^*\boldsymbol{\varepsilon}_2 \end{pmatrix}. \tag{6.7.28}
$$

The corresponding least squares estimator for $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ is given by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$
\begin{aligned}
\begin{pmatrix} \hat{\boldsymbol{\beta}}_1^* \\ \hat{\boldsymbol{\beta}}_2^* \end{pmatrix} &= \left\{[\mathbf{I}_2 \otimes (\mathbf{C}^*\tilde{\mathbf{v}}_1, \mathbf{C}^*\tilde{\mathbf{v}}_1)]' (\Sigma_v^* \otimes \mathbf{I}_{k-1}) [\mathbf{I}_2 \otimes (\mathbf{C}^*\tilde{\mathbf{v}}_1, \mathbf{C}^*\tilde{\mathbf{v}}_1)]\right\}^- \\
&\quad [\mathbf{I}_2 \otimes (\mathbf{C}^*\tilde{\mathbf{v}}_1, \mathbf{C}^*\tilde{\mathbf{v}}_1)]' (\Sigma_v^* \otimes \mathbf{I}_{k-1}) \begin{pmatrix} \mathbf{C}^*\hat{\mathbf{v}}_1 \\ \mathbf{C}^*\hat{\mathbf{v}}_2 \end{pmatrix} \\
&= \left[\Sigma_v^* \otimes \begin{pmatrix} \tilde{\mathbf{v}}_1'\mathbf{C}^{*'} \\ \tilde{\mathbf{v}}_2'\mathbf{C}^{*'} \end{pmatrix} \mathbf{I}_{k-1} (\mathbf{C}^*\tilde{\mathbf{v}}_1, \mathbf{C}^*\tilde{\mathbf{v}}_1)\right]^- \left[\Sigma_v^* \otimes \begin{pmatrix} \tilde{\mathbf{v}}_1'\mathbf{C}^{*'} \\ \tilde{\mathbf{v}}_2'\mathbf{C}^{*'} \end{pmatrix}\right] \begin{pmatrix} \mathbf{C}^*\hat{\mathbf{v}}_1 \\ \mathbf{C}^*\hat{\mathbf{v}}_2 \end{pmatrix} \\
&= \left[\Sigma_v^* \otimes \begin{pmatrix} \tilde{\mathbf{v}}_1'\tilde{\mathbf{v}}_1 & \tilde{\mathbf{v}}_1'\tilde{\mathbf{v}}_2 \\ \tilde{\mathbf{v}}_2'\tilde{\mathbf{v}}_1 & \tilde{\mathbf{v}}_2'\tilde{\mathbf{v}}_2 \end{pmatrix}\right]^- \left[\Sigma_v^* \otimes \begin{pmatrix} \tilde{\mathbf{v}}_1'\mathbf{C}^{*'} \\ \tilde{\mathbf{v}}_2'\mathbf{C}^{*'} \end{pmatrix}\right] \begin{pmatrix} \mathbf{C}^*\hat{\mathbf{v}}_1 \\ \mathbf{C}^*\hat{\mathbf{v}}_2 \end{pmatrix} \\
&= \left[\mathbf{I}_2 \otimes (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1} \begin{pmatrix} \tilde{\mathbf{v}}_1'\mathbf{C}^{*'} \\ \tilde{\mathbf{v}}_2'\mathbf{C}^{*'} \end{pmatrix}\right] \begin{pmatrix} \mathbf{C}^*\hat{\mathbf{v}}_1 \\ \mathbf{C}^*\hat{\mathbf{v}}_2 \end{pmatrix} \\
&= \begin{bmatrix} (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1} \begin{pmatrix} \tilde{\mathbf{v}}_1'\mathbf{C}^{*'} \\ \tilde{\mathbf{v}}_2'\mathbf{C}^{*'} \end{pmatrix} \mathbf{C}^*\hat{\mathbf{v}}_1 \\ (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1} \begin{pmatrix} \tilde{\mathbf{v}}_1'\mathbf{C}^{*'} \\ \tilde{\mathbf{v}}_2'\mathbf{C}^{*'} \end{pmatrix} \mathbf{C}^*\hat{\mathbf{v}}_2 \end{bmatrix} \\
&= \begin{bmatrix} (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}'\hat{\mathbf{v}}_1 \\ (\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}'\hat{\mathbf{v}}_2 \end{bmatrix} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix}_{OLS}.
\end{aligned}
\tag{6.7.29}
$$

note: Recall that $\mathbf{C}^{*'}\mathbf{C}^* = \mathbf{P}$, $\mathbf{P}\hat{\mathbf{v}}_i = \hat{\mathbf{v}}_i$ *and* $\mathbf{P}\tilde{\mathbf{v}}_i = \tilde{\mathbf{v}}_i$.

Thus, for the transformed bivariate regression model corresponding to (6.7.28)

$$
(\mathbf{C}^*\hat{\mathbf{v}}_1, \mathbf{C}^*\hat{\mathbf{v}}_2) = (\mathbf{C}^*\tilde{\mathbf{v}}_1, \mathbf{C}^*\tilde{\mathbf{v}}_2)\boldsymbol{\beta}_{\hat{V}|\tilde{V}} + (\mathbf{C}^*\boldsymbol{\varepsilon}_1, \mathbf{C}^*\boldsymbol{\varepsilon}_2), \tag{6.7.30}
$$

provided $(\mathbf{C}^*\tilde{\mathbf{v}}_1, \mathbf{C}^*\tilde{\mathbf{v}}_2)$ is full rank, it follows from Result 7.10 in Johnson and Wichern (1992), that if $[(\mathbf{C}^*\boldsymbol{\varepsilon}_1)', (\mathbf{C}^*\boldsymbol{\varepsilon}_2)']'$ is multivariate normal, $\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}$ is the MLE for $\boldsymbol{\beta}_{\hat{V}|\tilde{V}}$ *based on the whole and sub-sample insufficient statistics*, and $\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}$ has a normal distribution with $E(\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}) = \boldsymbol{\beta}_{\hat{V}|\tilde{V}}$ and $cov(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_k) = \sigma_{ik}^*[(\mathbf{C}^*\tilde{\mathbf{v}}_1, \mathbf{C}^*\tilde{\mathbf{v}}_2)'(\mathbf{C}^*\tilde{\mathbf{v}}_1, \mathbf{C}^*\tilde{\mathbf{v}}_2)]^{-1} = \sigma_{ik}^*(\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1}$ (from 6.7.29), where σ_{ik}^*, is the ik^{th} element of the covariance matrix Σ_v^*. Also, $\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}$ is independent of the MLE of Σ_v^* given by

$$
\begin{aligned}
\hat{\Sigma}_{\mathbf{v}}^* &= \frac{(\mathbf{C}^*\hat{\boldsymbol{\varepsilon}}_1, \mathbf{C}^*\hat{\boldsymbol{\varepsilon}}_2)'(\mathbf{C}^*\hat{\boldsymbol{\varepsilon}}_1, \mathbf{C}^*\hat{\boldsymbol{\varepsilon}}_2)}{k-1} \\
&= \frac{\hat{\boldsymbol{\varepsilon}}_{\hat{V}|\tilde{V}}'\hat{\boldsymbol{\varepsilon}}_{\hat{V}|\tilde{V}}}{k-1}
\end{aligned}
\tag{6.7.31}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where $\hat{\boldsymbol{\varepsilon}}_{\hat{V}|\tilde{V}} = \hat{\mathbf{V}} - \tilde{\mathbf{V}}\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}$.

Also, $\hat{\mathbf{v}}$ can be written as

$$\begin{aligned}\hat{\mathbf{v}} &= \mathbf{I}_2 \otimes \mathbf{B}\mathbf{y} \\ &= (\mathbf{I}_2 \otimes \mathbf{P} \otimes \mathbf{1}_n')\mathbf{y} \\ &= (\mathbf{I}_2 \otimes k\mathbf{P})\bar{\mathbf{y}}.\end{aligned}$$

Therefore $(\mathbf{I}_2 \otimes \mathbf{C}^*)\hat{\mathbf{v}} = (\mathbf{I}_2 \otimes k\mathbf{C}^*\mathbf{P})\bar{\mathbf{y}}$. The matrix $(\mathbf{I}_2 \otimes k\mathbf{C}^*\mathbf{P})$ has full row rank equal to the dimension of the error space of the design matrix for the fixed effects, $(\mathbf{I}_2 \otimes \mathbf{1}_{kn})$. Thus, from Section 1.3.3, it follows that $\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}$ and $\hat{\Sigma}_v^*$ are also the *REML estimators for* $\boldsymbol{\beta}_{\hat{V}|\tilde{V}}$ *and* Σ_v^* *based on the whole and sub-sample means.*

It should be noted that from equation 6.7.19, an estimator for Σ_e as a function of $\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}$ is

$$\begin{aligned}\hat{\Sigma}_e &= (n-m)^{-1}\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}^{-1}\hat{\Sigma}_{\mathbf{v}}^* \\ &= \hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}^{-1}\frac{(\hat{\mathbf{V}} - \tilde{\mathbf{V}}\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}})'(\hat{\mathbf{V}} - \tilde{\mathbf{V}}\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}})}{(k-1)(n-m)}.\end{aligned} \tag{6.7.32}$$

The result above is attractive for being a function of likelihood based estimators for $\boldsymbol{\beta}_{\hat{V}|\tilde{V}}$ and Σ_v^*. It's main drawback is that the result may not be symmetric or positive definite. Imposing constraints such as those proposed in Section 6.3 may provide usable estimates for covariance, but is not guaranteed to maintain optimality properties.

6.7.6 Joint Distribution of $\hat{\mathbf{u}}$ and $\tilde{\mathbf{u}}$

From the definition of the insufficient statistics, we are able to express the BLUP's, $\hat{\mathbf{u}}$ and $\tilde{\mathbf{u}}$ as linear transforms of the insufficient statistics $\hat{\mathbf{v}}$ and $\tilde{\mathbf{v}}$. In this section, we will use expressions for $\hat{\mathbf{u}}$ and $\tilde{\mathbf{u}}$ in terms of $\hat{\mathbf{v}}$ and $\tilde{\mathbf{v}}$ to derive the joint mean and variance for $\hat{\mathbf{u}}$ and $\tilde{\mathbf{u}}$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

From Section 6.7.2 we can write the BLUP's, $\hat{\mathbf{u}}$ and $\tilde{\mathbf{u}}$ in terms of $\hat{\mathbf{v}}$ and $\tilde{\mathbf{v}}$ as follows.

$$\hat{\mathbf{u}} = [(n\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k]\hat{\mathbf{v}}. \tag{6.7.33}$$

$$\tilde{\mathbf{u}} = [(m\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k]\tilde{\mathbf{v}}. \tag{6.7.34}$$

Since $E(\hat{\mathbf{v}}) = \mathbf{0}$ and $E(\tilde{\mathbf{v}}) = \mathbf{0}$, and are linear functions of the whole and sub-sample BLUP's, respectively, it is easy to see that $E(\hat{\mathbf{u}}) = \mathbf{0}$ and $E(\tilde{\mathbf{u}}) = \mathbf{0}$. From equation 6.7.17, we can compute the elements of $var[(\hat{\mathbf{u}}', \tilde{\mathbf{u}}')']$.

$$\begin{aligned} var(\hat{\mathbf{u}}) &= [(n\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k] var(\hat{\mathbf{v}}) [(n\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k]' \\ &= [(n\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k]\{[n(n\Lambda^{-1} + \mathbf{I}_2)](\Sigma_e \otimes \mathbf{P})\}[(n\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{I}_k] \\ &= n\Lambda^{-1}\Sigma_e(n\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P}. \end{aligned} \tag{6.7.35}$$

$$\begin{aligned} cov(\hat{\mathbf{u}}, \tilde{\mathbf{u}}) &= [(n\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k] cov(\hat{\mathbf{v}}, \tilde{\mathbf{v}}) [(m\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k]' \\ &= [(n\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k]\{[m(n\Lambda^{-1} + \mathbf{I}_2)](\Sigma_e \otimes \mathbf{P})\}[(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{I}_k] \\ &= m\Lambda^{-1}\Sigma_e(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P}. \end{aligned} \tag{6.7.36}$$

$$\begin{aligned} var(\tilde{\mathbf{u}}) &= [(m\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k] var(\tilde{\mathbf{v}}) [(m\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k]' \\ &= [(m\mathbf{I}_2 + \Lambda)^{-1} \otimes \mathbf{I}_k]\{[m(m\Lambda^{-1} + \mathbf{I}_2)](\Sigma_e \otimes \mathbf{P})\}[(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{I}_k] \\ &= m\Lambda^{-1}\Sigma_e(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P}. \end{aligned} \tag{6.7.37}$$

Therefore $var[(\hat{\mathbf{u}}', \tilde{\mathbf{u}}')']$ can be written as

$$var\begin{pmatrix} \hat{\mathbf{u}} \\ \tilde{\mathbf{u}} \end{pmatrix} = \begin{bmatrix} n\Lambda^{-1}\Sigma_e(n\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P} & m\Lambda^{-1}\Sigma_e(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P} \\ m\Lambda^{-1}\Sigma_e(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P} & m\Lambda^{-1}\Sigma_e(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P} \end{bmatrix}, \tag{6.7.38}$$

which is consistent with Theorem 6.1, as $cov(\hat{\mathbf{u}}, \tilde{\mathbf{u}}) = var(\tilde{\mathbf{u}})$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

6.7.7 Conditional Mean and Variance of $\hat{\mathbf{u}}$ given $\tilde{\mathbf{u}}$

Following the arguments in Section 6.7.4, we now present expressions for the condition mean and variance of $\hat{\mathbf{u}}$ given $\tilde{\mathbf{u}}$

$$\begin{aligned} E(\hat{\mathbf{u}}|\tilde{\mathbf{u}}) &= E(\hat{\mathbf{u}}) + cov(\hat{\mathbf{u}}, \tilde{\mathbf{u}})\,(var(\tilde{\mathbf{u}}))^{-}\,(\tilde{\mathbf{u}} - E(\tilde{\mathbf{u}})) \\ &= \mathbf{0} + \mathbf{I}_{2k}(\tilde{\mathbf{u}} - \mathbf{0}) = \tilde{\mathbf{u}}. \end{aligned} \tag{6.7.39}$$

and

$$\begin{aligned} var(\hat{\mathbf{u}}|\tilde{\mathbf{u}}) &= var(\hat{\mathbf{u}}) - cov(\hat{\mathbf{u}}, \tilde{\mathbf{u}})\,(var(\tilde{\mathbf{u}}))^{-}\,cov(\tilde{\mathbf{u}}, \hat{\mathbf{u}}) \\ &= [n\Lambda^{-1}\Sigma_e(n\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P}] - [m\Lambda^{-1}\Sigma_e(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P}] \\ &\qquad [m\Lambda^{-1}\Sigma_e(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P}]^{-}[m\Lambda^{-1}\Sigma_e(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P}], \\ &= [n\Lambda^{-1}\Sigma_e(n\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P}] - [m\Lambda^{-1}\Sigma_e(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P}], \\ &= [n\Lambda^{-1}(m\mathbf{I}_2 + \Lambda)' - m\Lambda^{-1}(m\mathbf{I}_2 + \Lambda)'](n\mathbf{I}_2 + \Lambda)^{-1'}(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P}. \end{aligned} \tag{6.7.40}$$

Now

$$\begin{aligned} [n\Lambda^{-1}(m\mathbf{I}_2 + \Lambda)' - m\Lambda^{-1}(m\mathbf{I}_2 + \Lambda)'] &= nm\Lambda^{-1} + n\Lambda^{-1}\Lambda' - mn\Lambda^{-1} - m\Lambda^{-1}\Lambda' \\ &= (n - m)\Lambda^{-1}\Lambda'. \end{aligned} \tag{6.7.41}$$

Substituting equation 6.7.41 into 6.7.40,

$$\begin{aligned} var(\hat{\mathbf{u}}|\tilde{\mathbf{u}}) &= (n - m)\Lambda^{-1}\Lambda'\Sigma_e(n\mathbf{I}_2 + \Lambda)^{-1'}(m\mathbf{I}_2 + \Lambda)^{-1'} \otimes \mathbf{P} \\ &= \Sigma_u^* \otimes \mathbf{P}, \end{aligned} \tag{6.7.42}$$

where $\Sigma_u^* = (n - m)\Lambda^{-1}\Lambda'\Sigma_e(n\mathbf{I}_2 + \Lambda)^{-1'}(m\mathbf{I}_2 + \Lambda)^{-1'}$.

The model corresponding to the regression function in equation 6.7.39 is

$$\hat{\mathbf{u}} = \tilde{\mathbf{u}} + \boldsymbol{\xi}_{\hat{u}|\tilde{u}},$$

where $\boldsymbol{\xi}_{\hat{u}|\tilde{u}_{2k\times 1}} \sim (\mathbf{0}, \Sigma_u^* \otimes \mathbf{P})$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

6.7.8 Multivariate Regression model for BLUP's

In Section 6.5 we defined $\hat{\mathbf{U}}_{k\times 2} = [\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2]$ and $\tilde{\mathbf{U}}_{k\times 2} = [\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2]$. The bivariate regression model for $\hat{\mathbf{U}}$ on $\tilde{\mathbf{U}}$ is

$$\hat{\mathbf{U}} = \tilde{\mathbf{U}}\boldsymbol{\eta}_{\hat{U}|\tilde{U}} + \boldsymbol{\xi}_{\hat{U}|\tilde{U}}.$$

$\boldsymbol{\eta}_{\hat{U}|\tilde{U}}$ is a 2×2 matrix with $\boldsymbol{\eta}_{\hat{U}|\tilde{U}} = [\boldsymbol{\eta}_1, \boldsymbol{\eta}_2]$, where $\boldsymbol{\eta}_i$ $(i = 1, 2)$, are both 2×1 vectors, and $\boldsymbol{\xi}_{\hat{U}|\tilde{U}}$ is a $k\times 2$ matrix with $\boldsymbol{\xi}_{\hat{V}|\tilde{V}} = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2]$, where $\boldsymbol{\xi}_i$ $(i = 1, 2)$ are both $k\times 1$ vectors.

As in equation 6.7.23, the combined multiple regressions of $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ on $\tilde{\mathbf{U}}$ we have

$$\begin{aligned}\hat{\mathbf{u}} &= \begin{pmatrix} \tilde{\mathbf{U}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{U}} \end{pmatrix}\begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix} \\ &= (\mathbf{I}_2 \otimes \tilde{\mathbf{U}})\begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}. \end{aligned} \tag{6.7.43}$$

Repeating the arguments used in equations 6.7.24 and 6.7.25, it is easily seen that

$$\hat{\boldsymbol{\eta}}_{\hat{U}|\tilde{U}_{WLS}} = \hat{\boldsymbol{\eta}}_{\hat{U}|\tilde{U}_{OLS}} \tag{6.7.44}$$

Recall, that in Section 6.5, it was shown that $E\left(\hat{\boldsymbol{\eta}}_{\hat{U}|\tilde{U}}\right) = \mathbf{I}_2$. Next we will show that by setting $\hat{\boldsymbol{\eta}}_{\hat{U}|\tilde{U}}$ to it's expected value, we can obtain an expression for the MR estimator of Λ.

6.7.9 MR estimators for Λ

We can express $\hat{\mathbf{U}}$ in terms of the insufficient statistics $\hat{\mathbf{V}}$ as follows.

$$\begin{aligned}\hat{\mathbf{U}} &= (\hat{\mathbf{u}}, \tilde{\mathbf{u}}) \\ &= \left[\frac{1}{|n\mathbf{I}_2 + \Lambda|}\left((n+\lambda_{22})\hat{\mathbf{v}}_1 - \lambda_{12}\hat{\mathbf{v}}_2\right), \frac{1}{|n\mathbf{I}_2 + \Lambda|}\left((n+\lambda_{11})\hat{\mathbf{v}}_2 - \lambda_{21}\hat{\mathbf{v}}_1\right)\right] \\ &= (\hat{\mathbf{v}}, \tilde{\mathbf{v}})\frac{1}{|n\mathbf{I}_2 + \Lambda|}\begin{pmatrix} n+\lambda_{22} & -\lambda_{21} \\ -\lambda_{12} & n+\lambda_{11} \end{pmatrix} \\ &= \hat{\mathbf{V}}(n\mathbf{I}_2 + \Lambda)^{-1'}. \end{aligned} \tag{6.7.45}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Similarly, $\tilde{\mathbf{U}} = \tilde{\mathbf{V}}(m\mathbf{I}_2 + \Lambda)^{-1'}$. Substituting for $\hat{\mathbf{U}}$ and $\tilde{\mathbf{U}}$ in $\hat{\boldsymbol{\eta}}_{\hat{U}|\tilde{U}}$.

$$\begin{aligned}
\hat{\boldsymbol{\eta}}_{\hat{U}|\tilde{U}} &= (\tilde{\mathbf{U}}'\tilde{\mathbf{U}})^{-1}\tilde{\mathbf{U}}'\hat{\mathbf{U}} \\
&= [(m\mathbf{I}_2 + \Lambda)^{-1}\tilde{\mathbf{V}}'\tilde{\mathbf{V}}(m\mathbf{I}_2 + \Lambda)^{-1'}]^{-1}[(m\mathbf{I}_2 + \Lambda)^{-1}\tilde{\mathbf{V}}'\hat{\mathbf{V}}(n\mathbf{I}_2 + \Lambda)^{-1'}] \\
&= (m\mathbf{I}_2 + \Lambda)'(\tilde{\mathbf{V}}'\tilde{\mathbf{V}})^{-1}(\tilde{\mathbf{V}}'\hat{\mathbf{V}})(n\mathbf{I}_2 + \Lambda)^{-1'} \\
&= (m\mathbf{I}_2 + \Lambda)'\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}(n\mathbf{I}_2 + \Lambda)^{-1'}. \qquad (6.7.46)
\end{aligned}$$

Setting $\hat{\boldsymbol{\eta}}_{\hat{U}|\tilde{U}}$ equal to its expected value, we have,

$$\mathbf{I}_2 = (m\mathbf{I}_2 + \Lambda)'\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}(n\mathbf{I}_2 + \Lambda)^{-1'}. \qquad (6.7.47)$$

Now we solve for Λ in terms of $\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}$. From equation 6.7.44,

$$\begin{aligned}
(n\mathbf{I}_2 + \Lambda)' &= (m\mathbf{I}_2 + \Lambda)'\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}} \\
\rightarrow (n\mathbf{I}_2 - m\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}) &= \Lambda'(\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}} - \mathbf{I}_2) \\
\rightarrow \Lambda' &= (n\mathbf{I}_2 - m\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}})(\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}} - \mathbf{I}_2)^{-1}.
\end{aligned}$$

Thus, the bivariate MR estimator for Λ is

$$\hat{\Lambda} = (\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}} - \mathbf{I}_2)^{-1'}(n\mathbf{I}_2 - m\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}})'. \qquad (6.7.48)$$

By the invariance property of MLE's, it follows that if the residuals are multivariate normal, the bivariate MR estimator of Λ is a *conditional MLE given the whole and sub-sample insufficient statistics*, and *conditional REML estimator given the whole and sub-sample group means.*

As was the case for Σ_e in Section 6.7.5 we can write an expression for an estimator for Σ_u from equations 6.7.32 and 6.7.48.

$$\begin{aligned}
\hat{\Sigma}_u &= \hat{\Lambda}^{-1}\hat{\Sigma}_e \\
&= (n\mathbf{I}_2 - m\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}})^{-1'}(\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}} - \mathbf{I}_2)'\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}}^{-1}\frac{(\hat{\mathbf{V}} - \tilde{\mathbf{V}}\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}})'(\hat{\mathbf{V}} - \tilde{\mathbf{V}}\hat{\boldsymbol{\beta}}_{\hat{V}|\tilde{V}})}{(k-1)(n-m)}. \qquad (6.7.49)
\end{aligned}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This result also may not be symmetric or positive definite, and as mentioned in Section 6.7.5, imposing constraints will likely result in a loss in optimality.

6.8 Correlation Structure for the Balanced One-Way MANOVA Model

In this section we will take is a closer look at of the correlation structure for the bivariate balanced one-way MANOVA model. In doing this, we are able to identify trait intraclass correlations as elements of the the correlation matrix of the data vector $\mathbf{y}$.

For the balanced one-way MANOVA model, let the subscripts for the random vector $\mathbf{y}$ for trait, group, observation be $i = 1, 2;\ j = 1, \ldots, k;\ l = 1, \ldots, n$, respectively. The covariance structure for $\mathbf{y}$ is as follows.

$$
\begin{aligned}
&cov(y_{ijl}, y_{ijl}) = var(y_{ijl}) = \sigma_i^2 + \sigma_{e_i}^2 = \sigma_{(i)}^2 \\
&cov(y_{ijl}, y_{ijl'}) \underset{l \neq l'}{=} \sigma_i^2 \\
&cov(y_{ijl}, y_{ij'l'}) \underset{j \neq j', l \neq l'}{=} 0 \\
&cov(y_{ijl}, y_{i'jl}) \underset{i \neq i'}{=} \sigma_{ii'} + \sigma_{e_{ii'}} = \sigma_{(ii')} \\
&cov(y_{ijl}, y_{i'jl'}) \underset{i \neq i', l \neq l'}{=} \sigma_{ii'} \\
&cov(y_{ijl}, y_{i'j'l'}) \underset{i \neq i', j \neq j', l \neq l'}{=} 0
\end{aligned}
$$

The resulting correlation structure is as follows.

$$
\begin{aligned}
&corr(y_{ijl}, y_{ijl}) = 1 \\
&cov(y_{ijl}, y_{ijl'}) \underset{l \neq l'}{=} \frac{\sigma_i^2}{\sigma_{(i)}^2} = \rho_i \\
&corr(y_{ijl}, y_{ij'l'}) \underset{j \neq j', l \neq l'}{=} 0 \\
&corr(y_{ijl}, y_{i'jl}) \underset{i \neq i'}{=} \frac{\sigma_{(ii')}}{\sigma_{(i)}\sigma_{(i')}} = \rho_{(ii')} \\
&corr(y_{ijl}, y_{i'jl'}) \underset{i \neq i', l \neq l'}{=} \frac{\sigma_{ii'}}{\sigma(i)\sigma(i')} = \rho_{ii'} \\
&corr(y_{ijl}, y_{i'j'l'}) \underset{i \neq i', j \neq j', l \neq l'}{=} 0
\end{aligned}
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The ρ_i represents the intraclass correlation coefficient for trait i. $\rho_{(ii')}$ will be referred to as the "total" class inter-trait correlation coefficient for traits i and i', and $\rho_{ii'}$ is the class inter-trait correlation coefficient for traits i and i'.

In terms of the elements of covariance matrices, Σ_u and Σ_e, the correlations can be expressed in matrix form as follows.

$$\begin{pmatrix} \rho_1 & \rho_{12} \\ \rho_{12} & \rho_2 \end{pmatrix} = (D_u + D_e)^{-1/2}\Sigma_u(D_u + D_e)^{-1/2} \tag{6.8.1}$$

and

$$\begin{pmatrix} 1-\rho_1 & \rho_{(12)} - \rho_{12} \\ \rho_{(12)} - \rho_{12} & 1-\rho_2 \end{pmatrix} = (D_u + D_e)^{-1/2}\Sigma_e(D_u + D_e)^{-1/2}, \tag{6.8.2}$$

where $\mathbf{D}_u$ and $\mathbf{D}_e$ are the diagonal matrices corresponding to the covariance matrices, Σ_u and Σ_e, respectively.

In terms of the correlations, the covariance parameters can be expressed as follows for $i = 1, 2$ and $i' = 1, 2$.

$$\sigma_i^2 = \rho_i \sigma_{(i)}^2$$

$$\sigma_{e_i}^2 = (1 - \rho_i)\sigma_{(i)}^2$$

$$\sigma_{ii'} = \rho_{ii'}\sigma_{(i)}\sigma_{(i')}$$

$$\sigma_{(ii')} = \rho_{(ii')}\sigma_{(i)}\sigma_{(i')}$$

$$\sigma_{e_{ii'}} = \sigma_{(ii')} - \sigma_{ii'} = (\rho_{(ii')} - \rho_{ii'})\sigma_{(i)}\sigma_{(i')}$$

Thus, the covariance matrices Σ_e and Σ_u, can be written as follows:

$$\Sigma_u = \begin{bmatrix} (1-\rho_1)\sigma_{(1)}^2 & (\rho_{(12)} - \rho_{12})\sigma_{(1)}\sigma_{(2)} \\ (\rho_{(12)} - \rho_{12})\sigma_{(1)}\sigma_{(2)} & (1-\rho_2)\sigma_{(2)}^2 \end{bmatrix}$$

$$\Sigma_e = \begin{bmatrix} \rho_1\sigma_{(1)}^2 & \rho_{12}\sigma_{(1)}\sigma_{(2)} \\ \rho_{12}\sigma_{(1)}\sigma_{(2)} & \rho_2\sigma_{(2)}^2 \end{bmatrix}$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Thus, the matrix $\Lambda = \Sigma_e \Sigma_u^{-1}$ can be written in terms of elements of the correlation matrix as follows:

$$\Lambda = \frac{1}{\rho_1\rho_2 - \rho_{12}^2} \begin{bmatrix} (1-\rho_1)\rho_2 - (\rho_{(12)} - \rho_{12})\rho_{12} & (\rho_1\rho_{(12)} - \rho_{12})\frac{\sigma_{(1)}}{\sigma_{(2)}} \\ (\rho_2\rho_{(12)} - \rho_{12})\frac{\sigma_{(2)}}{\sigma_{(1)}} & (1-\rho_2)\rho_1 - (\rho_{(12)} - \rho_{12})\rho_{12} \end{bmatrix}$$
$$= \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix},$$

which is not symmetric in general.

6.9 Outline for Bivariate Method R via Multi-trait Regression

The steps for estimation for Method R using multivariate regression of whole sample BLUE's on sub-sample BLUE's are as follows.

1. Choose a starting values for Σ_u and fix $\Sigma_e = \mathbf{I}_2$. Compute $\mathbf{A}^{-1}$.

2. Compute the insufficient statistics $\hat{\mathbf{V}}^{(j)}$ and $\tilde{\mathbf{V}}^{(j)}$ for the j^{th} iteration.

3. Compute the matrix of regression coefficients $\hat{\beta}_{\hat{V}|\tilde{V}} = (\tilde{\mathbf{V}}'\mathbf{A}^{-1}\tilde{\mathbf{V}})^{-1}(\tilde{\mathbf{V}}'\mathbf{A}^{-1}\hat{\mathbf{V}})$ for the j^{th} iteration, where $\hat{\mathbf{V}} = \hat{\mathbf{V}}^{(j)}$ and $\tilde{\mathbf{V}} = \tilde{\mathbf{V}}^{(j)}$.

4. Compute $\hat{\Lambda}$ using equation 6.7.47.

5. Estimate Σ_e for the j^{th} iteration using one of the following.

(i) Using the MME residual sums of squares. In which case the estimator for $\hat{\Sigma}_e$ should be positive definite.

(ii) Use equation 6.7.32 to estimate Σ_e. It follows from the invariance property of MLE's that this estimator is optimal provided the result is symmetric and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

positive definite. If it is not symmetric and positive definite, (6.7.32) does not belong to the parameter space for $\hat{\Sigma}_e$, and is thus not a MLE. In such a case the elements of (6.7.32) can be adjusted by methods such as those proposed in Section 6.3, to yield an estimator that is non-optimal, but is symmetric and positive definite.

6. Estimate $\hat{\Sigma}_u$ for the j^{th} iteration using one of the following.

(i) Use the current estimate of Λ and the estimate of MME residual sums of squares. In which case the estimator for $\hat{\Sigma}_e$ should be positive definite.

(ii) Use equation 6.7.49 to estimate Σ_u. It follows from the invariance property of MLE's that this estimator is optimal provided the result is symmetric and positive definite. If it is not symmetric and positive definite, (6.7.49) does not belong to the parameter space for $\hat{\Sigma}_u$, and is thus not a MLE. In such a case the elements of (6.7.49) can be adjusted by methods such as those proposed in Section 6.3, to yield an estimator that is non-optimal, but is symmetric and positive definite.

7. Estimate the elements of the correlation matrix using equations 6.8.1 and 6.8.2.

8. Iterate steps 2 through 7 using the updated estimates of $\hat{\Sigma}_u$ and $\hat{\Sigma}_e$, continue until

$max(|\hat{\rho}_1{}^{(j+1)} - \hat{\rho}_1{}^{(j)}|, |\hat{\rho}_2{}^{(j+1)} - \hat{\rho}_2{}^{(j)}|, |\hat{\rho_{12}}^{(j+1)} - \hat{\rho_{12}}^{(j)}|, |\hat{\rho_{(12)}}^{(j+1)} - \hat{\rho_{(12)}}^{(j)}|) < \varepsilon$,

where ε is the desired level of precision.

6.10 Remarks

The regression characterization of multi-trait MR, as in the univariate case, provides some insight to the properties of the MR estimator in the multivariate setting. As

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

seen for the balanced bivariate one-way MANOVA model, the MR estimates of the matrix of regression coefficients and Λ, provide a conditional MLE's based on the insufficient statistics and conditional REML estimators given the whole and sub-sample means.

In Chapter 3, we saw that MR ceased to be optimal in the unbalanced case. This result is no doubt also true for the bivariate unbalanced one-way MANOVA model. The algorithm for the bivariate regression approach to MR in Section 6.9 may be of use for future small to moderate sample sized simulations, similar to those presented in Chapter 4.

The issue of out-of-bounds estimates received some attention for the univariate case, in the discussion of the one-way balanced ANOVA model in Chapter 3. It is reasonable to assume that out-of-bounds estimates will be a problem for small to moderate sample problems. As the size of the sample increases, the incidence of estimates outside the parameter space should decrease. Indeed, for practical purposes, use multi-trait MR would be advised only for problems in which size of the data set becomes a significant computational issue for more traditional methods of multivariate estimation. These are precisely the situations in which we expect the out-of-bounds problem to be of minimal concern.

In summary, the results for the bivariate one-way random effects model extend quite nicely from the those observed for the univariate one-way random effects models. More work is needed to understand the theoretical properties of the MR procedure for more complex models. Logical extensions of this work would include expanding the one-way model to a 2-factor mixed effects model, balanced and unbalanced, to study the impact of estimating fixed effects; incoorperating different methods of sampling (random vs. stratified); the study of nested models, mixed effects and random effects, balanced and unbalanced, and so on.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Bibliography

C. R. Henderson, O. Kempthorne, S. S. and vonKrosgk, C. (1959). The estimation of environmental genetic traits from records subject to culling. *Biometrics*, 15:192.

Druet, T. (2002). *Estimation of additive and dominance genetic variances with Method R.* PhD thesis, Faculte Uinversitaire des Sciences Agronomiques d e Gembloux, Gembloux, Belgium.

Druet, T., Misztal, I., Duangjinda, M., Reverter, A., and Gengler, N. (2000). Covariance estimation with Method R. *Journal of Dairy Science*, 83 (Suppl. 1):56 (Abstr.).

Druet, T., Misztal, I., Duangjinda, M., Reverter, A., and Gengler, N. (2001). Estimation of genetic covariances with Method R. *Journal of Animal Science*, 79:605–615.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall/CRC, Boca Raton, Florida.

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Trans. R. Soc.*, 52:399–433.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. Lond., Ser. A*, 222:309–368.

Fisher, R. A. (1925a). *Statistical Methods of Research Workers.* Oliver & Boyd, Edinburg, London, first edition.

Fisher, R. A. (1925b). Theory of statistical estimation. *Proc. Camb. Phil. Soc. A*, 22:700–725.

Ganguli, M. (1941). A note on nested sampling. *Sankhya*, 5:449–452.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Gauss, K. (1809). *Theoria Motus Corporum Celestrium in Sectionibus Conics Solem Ambientum.* Perthes and Besser, Hamburg.

George, A., Liu, J., and Ng, E. (1980). *Userguide for SPARSPAK: Waterloo sparse linear equations package.* CS-78-30, Department of Computer Science, University of Waterloo, Ontario, Canada.

Gilmour, A. R., Thompson, R., and Cullis, B. R. (1995). Average information reml: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, 51:1440–1450.

Golden, B. L., Brinks, J. S., and Bourdon, R. M. (1991). A performance programmed method for computing inbreeding coefficients from large data sets for use in mixed model analysis. *J. Anim Sci.*, 69:3564–3573.

Graser, H. U., Smith, S. P., and Tier, B. (1987). A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *Journal of Animal Science*, 64:1362–1370.

Graybill, F. A. (1954). On quadratic estimates of variance components. *Annals of Mathematical Statistics*, 25:367–372.

Graybill, F. A. (1969). *Introduction to Matrices with Applications in Statistics.* Wadsworth Publishing Co., Inc., Belmont, California.

Graybill, F. A. (1976). *Theory and Application of the Linear Model.* Duxbury, North Scituate, Massachusetts.

Graybill, F. A. and Hultquist, R. A. (1961). Theorems concerning eisenhart's model ii. *Annals of Mathematical Statistics*, 32:261–269.

Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54:93–108.

Harville, D. A. (1969). Quadratic unbiased estimation of variance components for the on-way classification. *Biometrika*, 56:313–326. Corrigenda **57**, 226.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Harville, D. A. (1977). Maximum-likelihood approaches to variance component estimation and to related problems. *Journal of American Statistical Association*, 72:320–340.

Hemmerle, W. J. and Hartley, H. O. (1973). Computing maximum-likelihood estimates for the mixed a.o.v. model using the w transformation. *Technometrics*, 15:819–831.

Henderson, C. R. (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics*, 21:309–310.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9:226–252.

Henderson, C. R. (1973). Maximum likelihood estimation of variance components. *unpublished manuscript.*

Henderson, C. R. (1976). A simple method for computing the inverse of a numerator matrix used in prediction of breeding values. *Biometrics*, 32:69–84.

Henderson, C. R. (1984). *Application of Linear Models in Animal Breeding.* University of Guelph Press, Ontario, Canada.

Herbach, L. H. (1959). Properties of model ii type analysis of variance tests, a: optimum nature of the f-test for model ii in the balanced case. *Annals of Mathematical Statistics*, 30:939–959.

Hill, B. M. (1965). Inference about variance components in the one-way model. *J. Amer. Stat. Assoc.*, 60:806–825.

Hill, B. M. (1967). Correlated errors in the random model. *J. Amer. Stat. Assoc.*, 62:1387–1400.

Hocking, R. R. and Kutner, M. H. (1975). Some analytical and numerical comparisons of estimators for the a.o.v. model. *Biometrics*, 31:19–28.

Jenrich, R. J. and Sampson, P. F. (1976). Newton-rhapson and related algorithms for maximum likelihood variance component estimation. *Technometrics*, 18:11–17.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Johnson, D. L. and Thompson, R. (1995). Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and a quasi newton procedure. *Journal of Dairy Science*, 78:449–456.

Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis.* Prentice Hall, Inc., New Jersey, third edition.

Kempthorne, O. (1977). Status of quantitative genetic theory. In et.al., E. P., editor, *Proceedings International Conference Quantitative Genetics*, pages 719–760, Ames, Iowa. Iowa State University Press.

Khuri, A. I. and Sahai, H. (1985). Variance components analysis: a selective literature survey. *Int. Stat. Rev.*, 53:279–300.

LaMotte, L. R. (1970). A class of estimators of variance components. Technical report no. 10, Department of Statistics, University of Kentucky, Lexington, Kentucky.

LaMotte, L. R. (1971). Locally best quadratic estimators of variance components. Technical report no. 22, Department of Statistics, University of Kentucky, Lexington, Kentucky.

LaMotte, L. R. (1973). Invariance quadratic estimators in the random, one-way anova model. *Biometrics*, 32:793–804.

LaMotte, L. R. (1973a). On the non-negative quadratic unbiased estimation variance components. *Journal of the American Statistical Association*, 68:728–730.

LaMotte, L. R. (1973b). Quadratic estimation of variance components. *Biometrics*, 29:311–330.

Legendre, A. (1806). *Nouvelles Méthodes pour la Determination des Orbites des Comètes; avec un Supplément Contenant Divers Perfectionnements de ces Méthodes et leur Application aux deux Comètes de 1805.* Coucier, Paris.

Lehmann, E. L. (1983). *Theory of Point Estimation.* Springer-Verlag, New York.

Lehmann, E. L. (1999). *Elements of Large Sample Theory.* Springer-Verlag, New York.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Li, S. H. and Kotz, J. H. (1978). Components of variance estimation for the split-plot design. *Journal of the American Statistical Association*, 73:147–152.

Lindstrom, M. J. and Bates, D. M. (1988). Newton-rhaphson and em algorithms for mixed-linear effects with repeated measures data. *Journal of the American Statistical Association*, 83:1014–1022.

Mallinckrodt, C. H., Golden, B. L., and Reverter, A. (1997). Approximate confidence intervals for heritability from Method R estimates. *Journal of Animal Science*, 75:2041–2046.

Miller, J. J. (1973). Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance. Technical report no. 12, Department of Statistics, Stanford University, Stanford, California.

Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of analysis of variance. *Ann. Stat.*, 5:746–762.

Misztal, I. (1997). Estimation of variance components with large-scale dominance models. *Journal of Dairy Science*, 80:965–974.

Misztal, I. (2000). *Computational Techniques in Animal Breeding.* University of Georgia, Athens, Georgia. Class Notes.

Neyman, J., Iwaszkiewicz, K., and Kolodziejczyk, S. T. (1997). Statistical problems in agricultural experimentation. *J. R. Stat. Soc. Suppl.*, 2:107–154.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58:545–554.

Plackett, R. (1972). Studies in the history of probability and statistics. xxix. the discovery of least squares. *Biometrika*, 59:239–251.

Press, W. H., Flannery, S. A., Teukolsky, S. A., and T.Vetterling, W. (1989). *Numerical recipes.* Cambridge University Press, New York.

Quaas, R. L. (1976). Computing the diagonal elements of the inverse of a large numerator relationship matrix. *Biometrics*, 32:949–953.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Quaas, R. L. (1989). Transformed mixed model equations: A recursive algorithm to eliminate $\mathbf{a}^{-1}$. *J. Dairy Sci.*, 72:1937–1941.

Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *J. Amer. Stat. Assoc.*, 65:161–172.

Rao, C. R. (1971a). Estimation of variance and covariance components - minque theory. *J. Multivariate Analysis*, 1:257–275.

Rao, C. R. (1971b). Minimum variance quadratic unbiased estimation of variance components. *J. Multivariate Analysis*, 1:445–456.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. *J. Amer. Stat. Assoc.*, 67:112–115.

Rao, C. R. (1973). *Linear statistical inference and it's applications.* John Wiley & Sons, New York, second edition.

Rao, C. R. and Kleffe, J. (1988). *Estimation of Variance Components and Applications.* North-Holland, Amsterdam.

Reverter, A. (1994). *Method R: A procedure for the estimation of variance and covariance components.* PhD thesis, Colorado State University, Fort Collins, Colorado.

Reverter, A. (1998). Empirical evidence of the optimality of Method R estimates. In *Proc. 6th World Cong. Genet. Appl. Livest. Prod.*, volume 25, pages 533–536, Armidale, Australia.

Reverter, A., Golden, B. L., Bourdon, R. M., and Brinks, J. S. (1994a). Technical note: Detection of bias in genetic predictions. *Journal of Animal Sciences*, 72:2247–2253.

Reverter, A., Golden, B. L., Bourdon, R. M., and Brinks, J. S. (1994b). Method R variance components procedure: Application on the simple breeding value model. *Journal of Animal Sciences*, 72:2247–2253.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reverter, A., Kaiser, C. J., and Mallinckrodt, C. H. (1998). A bootstrap approach to confidence regions for genetic parameters from Method R estimates. *Journal of Animal Sciences*, 76:2263–2271.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bull.*, 2:110–114.

Schaeffer, L. R. (1998). *Variance Components Estimation Course Notes.* Genteics and Animal Breeding Unit, University of New England, Armidale NSW, Australia.

Scheffé, H. (1956). Alternative models for the analysis of variance. *Annals of Math. Stat.*, 27:251–271.

Scheffé, H. (1959). *The Analysis of Variance.* John Wiley & Sons, New York.

Schenkel, F. S. and Scaeffer, L. R. (2000). Effects of nonrandom parental selection on estimation of variance components. *J. Anim. Breed. Genet.*, 117:225–239.

Schott, J. R. (1997). *Matrix Analysis for Statistics.* John Wiley & Sons, New York.

Searle, S. R. (1970). Large sample variances of maximum likelihood estimators of variance components using unbalanced data. *Biometrics*, 26:505–524.

Searle, S. R. (1979). Notes on variance component estimation. a detailed account of maximum likelihood and kindred methodology. Technical report bu-673-m, Biometrics Unit, Cornell University, Ithaca, New York.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance components.* John Wiley & Sons, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

Smith, S. P. and U.Graser, H. (1986). Estimating variance components in a class of mixed models by restricted maximum likelihood. *Journal of Dairy Science*, 69:1156.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Swallow, W. H. and F.Monahan, J. (1984). Monte-carlo comparison of anova, minque,reml and ml estimators of variance components. *Technometrics*, 26:47–57.

Tippett, L. H. C. (1931). *The Methods of Statistics.* Williams and Norgate, London, first edition.

Townsend, E. C. (1968). *Unbiased estimators of variance components in simple unbalanced designs.* PhD thesis, Cornell University, Ithaca, New York.

Townsend, E. C. and Searle, S. (1971). Best quadratic unbiased estimation of variance components from unbalanced data in the one-way classification. *Biometrics*, 27:643–657.

W.A. Thompson, J. (1962). The problem of negative estimates of variance components. *Ann. Math. Stat.*, 33:273–289.

Wald, A. (1940). A note on the analysis of variance with unequal class frequencies. *Ann. Math. Stat*, 11:96–100.

Wald, A. (1941). On the analysis of variance in case of multiple classifications with unequal class frequencies. *Ann. Math. Stat*, 12:346–350.

Winsor, C. P. and L.Clarke, G. (1940). Statistical study of variation in the catch of plankton nets. *Sears Foundation J. Marine Res.*, 3:1–34.

Yates, F. and Zacopanay, I. (1935). The estimation of the efficiency of sampling with special reference to sampling for yield in cereal experiments. *J. Agric. Sci.*, 25:545–577.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.