Winter 1997

Analysis of Repeated Measures Data Under Circular Covariance

Andrew Montgomery Hartley
Old Dominion University

ANALYSIS OF

REPEATED MEASURES DATA

UNDER CIRCULAR COVARIANCE

by

Andrew Montgomery Hartley
B.S. 1986, Allegheny College
M.S. 1994, Old Dominion University

A Thesis submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

APPLIED AND COMPUTATIONAL MATHEMATICS

OLD DOMINION UNIVERSITY
December 1997

Approved by:

Dayanand N. Naik (Director)

Richardean Benjamin-Coleman (Member)

Larry D. Lee (Member)

Edward P. Markowski (Member)

John P. Morgan (Member)

ABSTRACT

ANALYSIS OF
REPEATED MEASURES DATA
UNDER CIRCULAR COVARIANCE.

Andrew Montgomery Hartley
Old Dominion University, 1997
Director: Dr. Dayanand N. Naik

Circular covariance is important in modelling phenomena in epidemiological, communications and numerous physical contexts. We introduce and develop a variety of methods which make it a more versatile tool. First, we present two classes of estimators for use in the presence of missing observations. Using simulations, we show that the mean squared errors of the estimators of one of these classes are smaller than those of the Maximum Likelihood (ML) estimators under certain conditions. Next, we propose and discuss a parsimonious, autoregressive type of circular covariance structure which involves only two parameters. We specify ML and other types of estimators of these parameters, and present techniques for selection between various covariance structures related to circular covariance. Finally, we consider estimation assuming that observations on different individuals are correlated in various ways. This model is generalized for use when varying numbers of observations are taken on individuals. In all these contexts, we combine the measurements on individuals with covariates of varying dimensions, and consider estimation of the correlation between the observations and the covariates.

I wish to dedicate this thesis to my wife, children and parents,
who have encouraged me in all my endeavors.

ACKNOWLEDGMENTS

Above and beyond all, I want to express my gratitude to my Savior and Lord, Jesus Christ, for the purpose, direction and sustenance I have gained through knowing Him. Any one of innumerable obstacles could have easily made this research frustrating, unfulfilling, costly, difficult or impossible, but He made my way straight. At the close of this chapter of my life, I am immensely richer, in memories, musically professionally, mathematically and even financially.

Without an advisor as devoted as mine, Dr. Dayanand Naik, this endeavor scarcely would have begun. Dr. Naik continuously and selflessly suggested research topics, directed me to important literature and other resources, read my work thoroughly, and made probing comments which compelled me to investigate phenomena and ideas deeply enough to make a contribution to my field. Most importantly, his attitude of inquisitiveness has enlivened in me a fascination for and delight in research. For these contributions and much more, I will be ever in his debt. His associate, Dr. Ravinda Khattree also provided helpful discussion.

To the faculty of the Mathematics and Statistics Department, I owe sincere thanks. Both before and during the writing of this thesis, they went beyond the call in answering questions and clarifying mathematical issues. The opportunities to which they directed and commended me were invaluable in my professional development.

My parents have supported unfailingly every task to which my wife and I have set our hands. Their wisdom of experience has set us on the right path, time and again. The profound joy they express in our activities and accomplishments has spurred us on to new challenges.

Our children, Abigail and Samuel, are a constant inspiration. Their concern for my wife, myself and each other, and their eagerness to please, remind me to press on in my work with diligence and dedication. My wife, Karen made countless personal and career sacrifices to allow the completion of this thesis. My Pastor, Dr. Hennie Becker and his wife, Eleanor lifted me to God in prayer at all the right times (always!).

To all these and many others, I owe a great deal and give warm and sincere thanks.

TABLE OF CONTENTS

LIST OF TABLES

1 PRELIMINARIES AND INTRODUCTION

[1] In many biological and physical situations in which measurements on individuals are taken at relatively uniform intervals in a circular fashion, it happens that each measurement is equally correlated with the two measurements next to it, equally correlated with the two measurements next removed from it, and so forth. For instance, the disease incidence rate in each (relatively homogeneous) geographical sector around the business district of a city may correlate equally with those of neighboring sectors. Alternatively, during an outbreak of a disease, the incidence rate in any sector around the initial etiological agent may correlate with those in neighboring sectors. As a third example, in oil exploration, an explosive charge is frequently placed and detonated at the earth's surface at the center of a circle of microphones. The microphones subsequently record the echoes they receive from the lower strata of soil and rock. Characteristics of the echoes received by each microphone may be equally correlated with those of the two microphones next to it, equally correlated with those of the two next removed from it, and so on. In any of these three examples, the degree of correlation between any two points around the circle is a function of the number of points between them, moving along the shorter distance around the circumference. This type of covariance is known as circular covariance.

If the observations x_{ij}, $j = 1, 2, ..., a$ within individual i, $i = 1, 2, ..., n$ are correlated according to a circular covariance structure, and these are assembled in a vector $\mathbf{x}_i$, then we may write $cov\,(\mathbf{x}_i) = \mathbf{C}$, where $\mathbf{C}$ has elements

$$(\mathbf{C})_{l,l'} = cov(x_{il}, x_{il'}) = \begin{cases} \sigma_{|l-l'|}, & |l-l'| \le a/2, \\ \sigma_{a-|l-l'|}, & |l-l'| > a/2. \end{cases} \tag{1}$$

The pattern of the middle terms of $\mathbf{C}$ depends on whether a is odd or even, as is shown below

[1] The journal model for this thesis is *Communications In Statistics.*

for $a = 4$ and $a = 5$; if a is odd, then $\sigma_{(a-1)/2}$ appears twice in each row of **C**.

$$\begin{bmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \sigma_1 \\ \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 \\ \sigma_1 & \sigma_2 & \sigma_1 & \sigma_0 \end{bmatrix}, \quad \begin{bmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \sigma_2 & \sigma_1 \\ \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 \\ \sigma_1 & \sigma_2 & \sigma_2 & \sigma_1 & \sigma_0 \end{bmatrix}.$$

Applications of circular covariance matrices abound in public health contexts. In a study of the distribution of typhus fever in cities in the Southeastern United States, for instance, Maxcy (1926) concluded that the focal point of the spread of this disease in most cities was the heart of the business district, where the greatest number of contacts occur between infectious and susceptible persons. Maxcy's findings confirmed the validity of the model proposed by Hamer (1906), which considered the potential for the spread in any community of any one of a broad class of infectious diseases to be the product of the number of contacts between susceptible and infectious individuals, and a "transmission constant" β. Thus, it is plausible to suggest (at least in most cities) that the incidence rates of any one of these diseases in the sectors around the city center correlate more or less equally with the rate of the city center, and that (as we have suggested at the outset of this section) the rates in the sectors are correlated with each other in a circular fashion.

Modelling the circular covariance structure of the disease rates in these sectors may offer significant advantages in studying the effectiveness of interventions performed to improve public health. For instance, a popular method of conducting a "community trial" (Lilienfeld and Stolley, 1994, p.181) is to select two or more cities which are as similar as possible in pertinent baseline measurements and other, qualitative characteristics, and then intervene in one or more of them. The intervention usually consists of disease control measures such as education or immunization programs, or renovation efforts in public housing. Differences between the cities in the distribution of a disease or of another public health measure are ultimately examined to assess the effectiveness of the intervention. However, the rates of transmission of many contagions, such as influenza, are affected to a large extent by external factors (weather, disease exposure, etc.), and therefore an important source of variation in the differences between the incidence rates of the cities is the effect of these factors that bear upon the entire cities to different extents after the baseline measurements are taken. Furthermore, as noted by Anderson (1982) and described in detail for measles in New York City by York and London (1973), many of these disease rates exhibit strong (often negative) autoregressive patterns from year to year, and disease rates of entire cities commonly are at different points in their autoregressive cycles. A method which may reduce the effect of the variation between cities due to these factors, is to allocate control and treatment locales among sectors of the same city. If the (likely circular) covariance parameters (and, ideally, the means) of the incidence rates of these sectors have been estimated with reasonable precision, then simple likelihood ratio tests, multivariate linear regression and other statistical tools may be implemented to quantify and test for the effects of the intervention in the treatment sectors.

Countless examples of the usefulness of circular covariance matrices appear in mechanical engineering contexts, as well. For instance, even a slight out-of-balance condition in certain rotating mechanical components can have a negative impact on the lifespans of both the component itself and its parent equipment. Currently, the most sensitive means of quickly detecting such a condition, as described by O'Connor (1993a), refracts laser beams, assessing the asymmetry of the profile of the component through the pattern of misalignments measured by the beams. However, the tests must often be conducted outside of a vacuum, and doing so introduces interferences due to wind currents, airborne particulates, and extraneous magnetic charges. Thus, it sometimes becomes necessary to account for randomness in the recording of the component's profile. In an adequately balanced component, if one were to extract measurements from a number of evenly spaced points along the period of one rotation, the patterns of misalignment ("false positives") produced at these points by these purely environmental perturbations could be expected to follow a circular covariance structure with a uniform mean for the measurements around the circle.

There is also a need to quickly detect out-of-balance conditions among the blades of the rotors in steam turbines, in distributions of weights, temperatures and vibrations, in the presence of a myriad of varying environmental, load and velocity factors (O'Connor, 1993b). Under normal conditions, any of these types of measurements on the blades could be expected to follow a circular covariance pattern.

Lastly, whenever it is necessary to investigate the turbidity (or another characteristic) of a fluid in an enclosed, circular space (such as the interior of a pipe), and measurements are taken at relatively uniform intervals around the circumference of the space, these measurements could be modelled using circular covariance matrices.

Olkin and Press (1969) completed the first work combining understanding of the spectral decomposition of **C** with a statistical model. They examined a model involving normally distributed data with circular covariance wherein the first row of **C** is, applying (1), with $m = int\,(a/2)$,

$$\begin{aligned}(\sigma_0, \sigma_1, \sigma_2, ..., \sigma_{m-1}, \sigma_m, \sigma_{m-1}, ..., \sigma_1),\ a &= 2m, \qquad (2)\\ (\sigma_0, \sigma_1, \sigma_2, ..., \sigma_{m-1}, \sigma_m, \sigma_m, \sigma_{m-1}, ..., \sigma_1),\ a &= 2m+1.\end{aligned}$$

They assume $(\sigma_0, \sigma_1, \sigma_2, ..., \sigma_m)$ are unrelated parameters, subject only to the restriction **C**>0, that is, that **C** is positive definite. As illustrated in (1), the subsequent rows are all circulants of this first row. Olkin and Press showed the first $m+1$ eigenvalues of **C** to be $\delta_j = (\sigma_0, \sigma_1, \sigma_2, ..., \sigma_m)\, B_j$, where B_j is the j^{th}column of **B**, $j = 1, 2, ..., m+1$, and **B** is the nonsingular $(m+1) \times (m+1)$ matrix having elements[2]

[2] Olkin and Press write that the subscript on the α_j in b_{jl} indicates the column in which it appears. That this is an error becomes evident when one calculates the eigenvalues $\delta_i, i = 1, 2, ..., a$ of **C** using a lemma in Basilevsky (1983, p.223). When the first row of **C** is specified by (2), the lemma gives

$$\delta_i = \sum_{l=1}^{m+1} \sigma_{l-1} \cos\left[\frac{2\pi i(l-1)}{a}\right] + \sum_{l=m+2}^{a} \sigma_{a+1-l} \cos\left[\frac{2\pi i(l-1)}{a}\right] = \sum_{l=1}^{m+1} \alpha_l \sigma_{l-1} \cos\left[\frac{2\pi i(l-1)}{a}\right].$$

Hence the subscript on each $\alpha_j, j = 1, 2, ..., m+1$ must correspond to the row, not the column, of **B** in which α_j appears. Also, using the fact that $\cos\theta = \cos(2\pi c - \theta)$, for any integer c and angle θ, their expression

$$\cos\left[\frac{2\pi(j-1)(a-l+1)}{a}\right]$$

in the definition of b_{jl} simplifies slightly to the expression we give here.

$$b_{jl} = \alpha_j \cos\left[\frac{2\pi(j-1)(l-1)}{a}\right]. \tag{3}$$

Here $\alpha_1 = 1, \alpha_2 = \alpha_3 = \ldots = \alpha_m = 2$ and $\alpha_{m+1} = \begin{cases} 1, & a = 2m, \\ 2, & a = 2m+1. \end{cases}$ Note the first row (but not the first column) of $\mathbf{B}$ is $\mathbf{1}'$, a $(1 \times (m+1))$ vector of unities. The last $a - m - 1$ eigenvalues of $\mathbf{C}$ are generated using $\delta_j = \delta_{a-j+2}, j = m+2, m+3, ..., a$. The multiplicity of δ_j is $\alpha_j, j = 1, 2, ..., m+1$, and the positive definiteness of $\mathbf{C}$ implies $\delta_j > 0$. That the elements of $\mathbf{B}$ do not depend on $(\sigma_0, \sigma_1, \sigma_2, ..., \sigma_m)$ is both remarkable and convenient; it implies that estimators of $\delta' = (\delta_1, \delta_2, ..., \delta_{m+1})$ can easily be mapped to estimators of $(\sigma_0, \sigma_1, \sigma_2, ..., \sigma_m)$. Olkin and Press (1969) took advantage of these properties of $\mathbf{C}$ to derive the Maximum Likelihood Estimator (MLE) of the δ_j and thus of the σ_j. The developments in all the sections of this thesis make use of this MLE. Olkin and Press also derived Likelihood Ratio Tests for selecting between spherical, circular and general (unrestricted) covariance.

A more recent paper treating circular covariance matrices in the context of statistical models was that of Khattree and Naik (1994a), which extended the results of Olkin and Press by pairing a covariate (such as a "parent's" score) p_i having variance σ_{pp} with each $\mathbf{x}_i$ (scores on "siblings" (sibs)) having circular covariance. In the framework of multivariate normality, Khattree and Naik calculated the ML estimate of the interclass correlation coefficient $\rho_{ps} = cov(p_i, x_{ij}) / \sqrt{\sigma_{pp}\sigma_0}$, and identified its distribution under various sets of assumptions about the mean structure and about σ_{pp}.

Because such "parent-sibling" terminology is quite common in discussions of repeated measures models, we adopt it throughout most of this thesis. It is not envisioned, however, that many particular characteristics about siblings will have circular covariance patterns, unless the sibs are born or hatched at roughly evenly spaced points in time throughout the year, and the characteristics depend on the season in which the sibs are born or hatched. Thus, this "parent-sibling" language is purely a conceptual aid, not intended to suggest a particular application.

This thesis extends the ideas of Olkin and Press (1969) and Khattree and Naik (1994a) in a number of ways. Section 2 considers parameter estimation given non-normality or when elements may be missing from the $\mathbf{x}_i$. Two estimation procedures are developed; the first enjoys the benefit of unbiasedness, even when the (possibly incomplete) data arise from a non-normal distribution, and the second (drawing important ideas from the Expectation and Maximization (EM) Algorithm) is superior due to the smaller mean squared errors of its estimators, given data arising from normal or t-distributions, when na is small or when a small proportion of the observations are missing.

In Section 3, we extend Olkin and Press' work by introducing a highly parsimonious, "autoregressive" case of circular covariance in which the parameters $\boldsymbol{\sigma}' = (\sigma_0, \sigma_1, ..., \sigma_m)$ are all functions of just two parameters (σ^2, ρ). After describing the model and its advantages, we derive the ML estimate of (σ^2, ρ), as well as an alternative, relatively nonparametric estimate of ρ whose performance is superior to that of the MLE for small sample sizes. Then, we describe two methods of selecting between spherical covariance, "autoregressive" circular covariance and the general, relatively unrestricted circular covariance which Olkin and Press discuss. We also present estimation methods for use when the numbers of sibs vary between families, using the MLE when the sample size is large, and the relatively nonparametric, alternative estimator of ρ when the sample size is small or the data are non-normal.

In Section 4, we estimate the parameters when observations in different families may be correlated. That is to say, assume families may be grouped into "cousinships," and any two sibs in different families in the same cousinship are stochastically related through one or more "compound symmetry" covariance parameters. Estimators are developed under these conditions and various assumptions about numbers of families in each cousinship, numbers of sibs in each family and the inter-family covariance structure.

All sections extend the ideas of Khattree and Naik by adding a parent's score to each vector $\mathbf{x}_i$ (and, in some cases, a grandparent's score to each cousinship of families) for each case that has been cited above as an extension of Olkin and Press' work. Estimation procedures have been outlined for every one of these cases.

2 ESTIMATION, DATA MISSING COMPLETELY AT RANDOM

2.1 Introduction. While the methods of Olkin and Press (1969) can be used to find the MLE of the circular covariance parameters given a complete, roughly multivariate normal dataset, they cannot be immediately implemented when data are missing, or when the assumption of normality is severely violated. In this section we develop two distinct methods for estimation when data are incomplete or non-normal. The first method estimates all the covariance parameters without bias and the second, more efficient (for small datasets, assuming normality) method applies certain aspects (while employing commonsense substitutes of other aspects) of the Expectation and Maximization (EM) Algorithm (Little and Rubin, 1987). With simulations, mean squared errors of the estimators produced by these methods are compared assuming various multivariate normal distributions and several of the t distributions. We also develop parameter estimation when the observations (the scores on "siblings") following circular covariance are combined with a covariate ("parent's" measurement) on each family.

In a case related to the present, missing data case, families are available with different numbers of sibs in each family, and the covariance between any two sibs is a function of the number of sibs between them, moving along the shorter arc between them. More formally, this is to assume that two or more distinct family sizes $a_1 < a_2 < ... < a_c$ are observed, and in any family having a_k sibs,

$$cov(x_{ij}, x_{ij'}) = \begin{cases} \sigma_{|j-j'|}, |j-j'| \le a_k/2, \\ \sigma_{a_k-|j-j'|}, |j-j'| > a_k/2, \end{cases}$$

x_{ij} being the score of the j^{th} sib in the i^{th} family of the group. One may refer to this setup as "unbalanced," to distinguish it from the "missing data" case of this section, in which all families have the same number of sibs but some of the sibs' scores are unavailable. The "unbalanced" case requires different estimators than the ones introduced in this present section; Section 4 of this thesis discusses a version of the "unbalanced" case, deriving an overall estimator of each σ_l by grouping families according to the a_k, estimating all the σ_l possible within each group, and combining the resulting within group estimators.

2.2 Unbiased Estimation of Circular Covariance Parameters, No Missing Data. To shorten calculations in later subsections, we first consider the case in which no data are missing. Assume the data consist of n families' measurements $\mathbf{x}_i$, $i = 1, 2, ..., n$, each of which is a $(a \times 1)$-variate (possibly non-normal) vector having mean $\mu\mathbf{1}$ and covariance $\mathbf{C}$, where the elements of $\mathbf{C}$, following Section 1, are

$$(\mathbf{C})_{jk} = \begin{cases} \sigma_{|j-k|}, |j-k| \leq a/2, \\ \sigma_{a-|j-k|}, |j-k| > a/2. \end{cases}$$

Many situations, including some listed in Section 1, will necessitate different mean structures from that given by $\mu\mathbf{1}$. In these cases, the methods of the present section will not apply.

Let $m = int\,(a/2)$. Note that whereas $\sigma_l = \mathbf{E}\left[(x_{ij} - \mu)(x_{i,j-l} - \mu)\right]$ if $j + l \leq a$, it may be shown that

$$\mathbf{E}\left[(x_{ij} - \bar{x}_i)(x_{i,j-l} - \bar{x}_i)\right] = \sigma_l - a^{-1} \sum \sigma_j.$$

Here, $\bar{x}_i$ is the sample mean of the i^{th} family, and the sum is over the first row[3] of $\mathbf{C}$, which is $(\sigma_0, \sigma_1, ..., \sigma_{m-1}, \sigma_m, \sigma_{m-1}, ..., \sigma_1)$ if $a = 2m$ and $(\sigma_0, \sigma_1, ..., \sigma_{m-1}, \sigma_m, \sigma_m, \sigma_{m-1}, ..., \sigma_1)$ if $a = 2m + 1$. Hence it seems that, whether or not data are missing, any estimator of σ_l composed of sums of the form $(x_{ij} - \bar{x}_i)(x_{i,j-l} - \bar{x}_i)$ is biased unless $\sum \sigma_j = 0$.

Estimators involving sums of terms of the form $\left(x_{ij} - \bar{x}_{(i)}\right)\left(x_{i,j-l} - \bar{x}_{(i)}\right)$, $j + l \leq a$, where $\bar{x}_{(i)} = \left[(n-1)\,a\right]^{-1} \sum_{g \neq i} \sum_{j=1}^{a} x_{gj}$ is the sample mean of all measurements excluding those of the i^{th} family, are biased as well, despite the independence of x_{ij} and $\bar{x}_{(i)}$, and that of $x_{i,j-l}$ and $\bar{x}_{(i)}$. Specifically,

$$\mathbf{E}\left[\left(x_{ij} - \bar{x}_{(i)}\right)\left(x_{i,j-l} - \bar{x}_{(i)}\right)\right] = \sigma_l + \frac{\sum \sigma_j}{(n-1)a}.$$

However, terms of the forms $x_{ij}(x_{i,j-l} - \bar{x}_{(i)}), j + l \leq a$, and $x_{ij}(x_{i,j-a-l} - \bar{x}_{(i)}), j + a - l \leq a$, are unbiased for σ_l:

$$\mathbf{E}\left[x_{ij}(x_{i,j-l} - \bar{x}_{(i)})\right] = \mathbf{E}\left[x_{ij}(x_{i,j-a-l} - \bar{x}_{(i)})\right] = \sigma_l + \mu^2 - \mu^2 = \sigma_l.$$

[3] Equivalently, any row; the rows of $\mathbf{C}$ are all circulants of the first row, moving the elements of the first row to the right and wrapping the last element back to the left.

In what follows, we estimate σ_l by taking an average $\tilde{\sigma}_l$ of terms of this form. This estimator is generalized in the next subsection to the case in which data may be missing.

For $l = 0, 1, ..., m$ let $\mathbf{A}_l$=$\{(i, j, k) : j \leq k, cov(x_{ij}, x_{ik}) = \sigma_l\}$. Also let c_l be the cardinality of $\mathbf{A}_l$. Then, upon proposing

$$\tilde{\sigma}_l = c_l^{-1} \sum_{\mathbf{A}_l} x_{ij}(x_{ik} - \bar{x}_{(i)})$$

as an estimator of σ_l, we have $\mathbf{E}(\tilde{\sigma}_l) = \sigma_l$.

Let us describe the elements of $\mathbf{A}_l$, and find an expression for c_l. When a is even and $l = a/2$, upon distributing the x_{ij} among the terms x_{ik} and $\bar{x}_{(i)}$ of the second factor of each term in $\tilde{\sigma}_l$, the first set of terms of $c_l\tilde{\sigma}_l$ is

$$\sum_{\mathbf{A}_l} x_{ij}x_{ik} = \sum_{i=1}^{n} \sum_{j=1}^{l} x_{ij}x_{i,j-l} \tag{4}$$

and $c_l = na/2$. Next, for any a and any $l < a/2$, the first set of terms of $c_l\tilde{\sigma}_l$ is

$$\sum_{\mathbf{A}_l} x_{ij}x_{ik} = \sum_{i=1}^{n} \left[\sum_{j=1}^{a-l} x_{ij}x_{i,j-l} + \sum_{j=1}^{l} x_{ij}x_{i,j-a-l} \right] \tag{5}$$

and $c_l = na$. For instance, take $a = 6$ and $l = 3 = a/2$. Fixing i, the only pairs (x_{ij}, x_{ik}) such that $cov(x_{ij}, x_{ik}) = \sigma_l$ are

$$(x_{i1}, x_{i4}), (x_{i2}, x_{i5}) \text{ and } (x_{i3}, x_{i6}),$$

so that $c_l = 3n$. However, if $l = 2 < a/2$, the number of pairs (x_{ij}, x_{ik}) such that $cov(x_{ij}, x_{ik}) = \sigma_l$ doubles. These pairs are

$$(x_{i1}, x_{i3}), (x_{i2}, x_{i4}), (x_{i3}, x_{i5}), (x_{i4}, x_{i6}), (x_{i1}, x_{i5}) \text{ and } (x_{i2}, x_{i6}),$$

keeping the second subscript of the second coordinate in each pair greater than that of the first, to conform with the specifications of the elements of $\mathbf{A}_l$. Here, $c_l = 6n = an$. In any case,

$$c_l = \begin{cases} an, l < a/2, \\ an/2, l = a/2. \end{cases}$$

Expressing each $\tilde{\sigma}_l$ as a quadratic form will aid in the calculation of its variance. We first express each $\sum_{j=1}^{a-l} x_{ij}x_{i,j-l}$ as a quadratic form. For any $l = 0, 1, 2, ..., a$, let $\mathbf{S}_l$ be obtained by shifting all the elements of an $a \times a$ identity matrix upwards by l positions. Note that $\mathbf{S}_l$ has

$a-l$ nonzero elements and rank $a-l$, that $\mathbf{S}_0$ is an identity matrix and $\mathbf{S}_a$ is a zero matrix, and that premultiplying $\mathbf{x}_i$ by $\mathbf{S}_l$ shifts the elements of $\mathbf{x}_i$ upwards l positions, leaving zeros in its last l positions. Therefore,

$$\begin{aligned} \mathbf{x}_i'\mathbf{S}_l\mathbf{x}_i &= x_{i1}x_{i,1-l} + x_{i2}x_{i,2-l} + \ldots + x_{i,a-l}x_{ia} = \sum_{j=1}^{a-l} x_{ij}x_{i,j-l} \text{ and} \\ \mathbf{x}_i'\mathbf{S}_{a-l}\mathbf{x}_i &= x_{i1}x_{i,1-a-l} + x_{i2}x_{i,2-a-l} + \ldots + x_{il}x_{ia} = \sum_{j=1}^{l} x_{ij}x_{i,j-a-l}. \end{aligned}$$

So, in the i^{th} family (noting again that for $l=0$, we have $\mathbf{S}_{a-l}=\mathbf{0}$),

$$\sum_{(\mathbf{A}_l,\mathrm{fix}(i))} x_{ij}x_{ik} = \mathbf{x}_i'\left(\mathbf{S}_l + I_{\{l<a/2\}}\mathbf{S}_{a-l}\right)\mathbf{x}_i, \tag{6}$$

where, for any event A, I_A is the indicator function of A. Setting $X' = (\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_n')$ yields

$$\sum_{\mathbf{A}_l} x_{ij}x_{ik} = X'\left[I_n \otimes (\mathbf{S}_l + I_{\{l<a/2\}}\mathbf{S}_{a-l})\right]X.$$

The second sum of terms of $\hat{\sigma}_l$ is $\sum_{\mathbf{A}_l} x_{ij}x_{ik}^0\bar{x}_{(i)}$, inserting x_{ik}^0 so as to retain the dependence of the summation[4] on k. To write this sum as a quadratic form, as in (6), considering two cases $l=a/2$ and $l<a/2$, noting how the sum in (4) and (5) were expressed, we have

$$\sum_{\mathbf{A}_l} x_{ij}x_{ik}^0\bar{x}_{(i)} = \sum_{i=1}^{n}\bar{x}_{(i)}\left[\sum_{j=1}^{l} x_{ij} + I_{\{l<a/2\}}\sum_{j=1}^{a-l} x_{ij}\right].$$

Now, whether or not $l<a/2$, we can write

$$\bar{x}_{(i)} = s^{-1}\mathbf{1}_s'\mathbf{D}_{(i)}X = s^{-1}X'\mathbf{D}_{(i)}'\mathbf{1}_s,$$

where $s=(n-1)a$ and $\mathbf{D}_{(i)}$ is a modified $na \times na$ identity matrix, obtained by deleting the rows of identity which correspond to the measurements in X on the i^{th} family ($\mathbf{D}_{(i)}$ has dimensions $(n-1)a \times na$, and is of full row rank). Then for this i,

$$\begin{aligned} \bar{x}_{(i)}\sum_{j=1}^{l} x_{ij} &= s^{-1}X'\mathbf{D}_{(i)}'\mathbf{1}_s\mathbf{1}_a'\mathbf{S}_l'\mathbf{x}_i \text{ and} \\ \bar{x}_{(i)}\sum_{j=1}^{a-l} x_{ij} &= s^{-1}X'\mathbf{D}_{(i)}'\mathbf{1}_a'\mathbf{S}_{a-l}'\mathbf{x}_i \end{aligned}$$

[4] For the case $x_{ik}=0$, define $0^0=0$.

so that the second set of terms in $\tilde{\sigma}_l$ expressed as quadratic form is

$$\sum_{A_l} x_{ij} x_{ik}^0 \bar{x}_{(i)}$$
$$= s^{-1} X' \left[D'_{(1)}, D'_{(2)}, \ldots, D'_{(n)} \right] \left\{ I_n \otimes \left[1_s 1'_a \left(S'_l + I_{\{l<a/2\}} S'_{a-l} \right) \right] \right\} X.$$

Assembling the two sets of terms in $\tilde{\sigma}_l$ gives

$$c_l \tilde{\sigma}_l = X' \left[I_n \otimes \left(S_l + I_{\{l<a/2\}} S_{a-l} \right) \right] X - \tag{7}$$
$$s^{-1} X' \left[D'_{(1)}, D'_{(2)}, \ldots, D'_{(n)} \right] \left\{ I_n \otimes \left[1_s 1'_a \left(S'_l + I_{\{l<a/2\}} S'_{a-l} \right) \right] \right\} X.$$

Simple analytic expressions for the variances of these $\tilde{\sigma}_l$ do not seem to be available, as will be discussed below in Subsection 2.5. We will refer to the $\tilde{\sigma}_l$ and, later, the corresponding estimator $\tilde{\sigma}_{psl}$ of σ_{ps}, as "leave one out" (LOO) estimators, due to their dependence on the "leave one out" sample means $\bar{x}_{(i)}$.

2.3 Unbiased Estimation of Circular Covariance Parameters, Missing Data. It is logical, when data are missing, to first assume normality and attempt ML estimation of the σ_l. However, maximization of the likelihood seems problematic. As noted in Section 1, Olkin and Press (1969) accomplished ML estimation in the full data case by finding the MLE of the eigenvalues δ_j of C (the eigenvectors Γ_j of C being functions of trigonometric functions only, unrelated to the σ_l). The covariance matrix C_i^* of the observed part of the i^{th} family is obtained by deleting from C the rows and columns corresponding to the unavailable measurements in this family. One might endeavor to adapt the methods of Olkin and Press, therefore, by specifying the relation between the (δ_j, Γ_j) and the spectral decomposition $\left(\delta_{ij}^*, \Gamma_{ij}^*\right)$ of C_i^*.

However, no relation between the spectral decomposition of a matrix and that of one of its submatrices has been found in the literature, except in special, trivial cases. In fact, Hadi (1988), working in the context of linear regression, showed that in the general case no closed form equations can relate the eigenvalues of a matrix to those of one of its submatrices obtained in this way.

Developing such an adaptation of Olkin and Press' work may be worthwhile if a relation, closed form or otherwise, can be established between the δ^{*}_{ij} and the σ_l and if the Γ^{*}_{ij}, like the Γ_j, can be expressed without reference to the σ_l. We leave this adaptation (and ML estimation given missing data in general) to a future paper, and continue by generalizing the LOO estimators introduced in the last subsection to the missing data case.

When data are missing, elements of $\mathbf{A}_l$ are removed and c_l decreases commensurately. Define the random variables z_{ij} and $y_{ij}, i = 1, 2, ..., n$ and $j = 1, 2, ..., int(3a/2)$, as[5]

$$z_{ij} = \begin{cases} x_{ij}, \ x_{ij} \text{ available}, \\ 0, \text{ otherwise}, \end{cases} \quad \text{and } y_{ij} = \begin{cases} 1, \ x_{ij} \text{ available}, \\ 0, \text{ otherwise}, \end{cases}$$

with the stipulation that $z_{ij} = y_{ij} = 0$ if $j > a$. Also define

$$\begin{aligned} \mathfrak{B}_l &= \{(i,j,k) : j \leq k, cov(z_{ij}, z_{ik}) = \sigma_l, y_{ij}y_{ik} = 1\}, \mathbf{z}_i' = (z_{i1}, z_{i2}, ..., z_{ia}), \\ \text{and } Z' &= (\mathbf{z}_1', \mathbf{z}_2', ..., \mathbf{z}_n'). \end{aligned}$$

Let a_i be the number of available measurements in the i^{th} family, and $a_{(i)} = a_{\bullet} - a_i$ be the number of available measurements in the entire dataset excluding those in this family. Here, in contrast to the last subsection, we estimate σ_l by

$$\tilde{\sigma}_l = c_l^{-1} \sum_{\mathfrak{B}_l} z_{ij}(z_{ik} - \bar{z}_{(i)}),$$

where $\bar{z}_{(i)} = a_{(i)}^{-1} \mathbf{1}_s' \mathbf{D}_{(i)} Z$ and $c_l = card(\mathfrak{B}_l)$.

The first sum of terms $\sum_{\mathfrak{B}_l} z_{ij} z_{ik}$ is expressed as a quadratic form in exactly the same manner in which $\sum_{\mathbf{A}_l} x_{ij} x_{ik}$ was expressed when no data were missing, replacing all x_{ij} there by z_{ij}. This is true because if x_{ij} or x_{ik} is not available, then $z_{ij} z_{ik} = 0$ contributes nothing to $\tilde{\sigma}_l$. This sum is therefore

$$\sum_{\mathfrak{B}_l} z_{ij} z_{ik} = Z' \left[\mathbf{I}_n \otimes (\mathbf{S}_l + I_{\{l<a/2\}} \mathbf{S}_{a-l}) \right] Z.$$

The second sum of terms in $c_l \tilde{\sigma}_l$ is $\sum_{\mathfrak{B}_l} z_{ij} y_{ik} \bar{z}_{(i)}$, inserting y_{ik} to retain the dependence of the summation in $\tilde{\sigma}_l$ on k. Let the k^{th} element of the $a \times 1$ column vector $\mathbf{e}_{li}$ be $y_{i,k-l}, l = 0, 1, 2, ..., a$ and $k = 1, 2, ..., a$ so that, recalling $y_{ik} = 0$ if $k > a$, the last l elements of $\mathbf{e}_{li}$ each are zero. Also

[5] Defining z_{ij} and y_{ij} for $j = a + 1, a + 2, ..., int(3a/2)$ is necessary to properly define the vectors $\mathbf{e}_{li}$ below.

let $e' = (e'_{01}, e'_{02}, \dots, e'_{0n})$. We may then encompass both of the cases $l = a/2$ and $l < a/2$ as in (4) and (5), by writing

$$\sum_{\mathfrak{B}_l} z_{ij} y_{ik} \tilde{z}_{(i)}$$

$$= Z' \left[\frac{D'_{(1)}}{a_{(1)}}, \frac{D'_{(2)}}{a_{(2)}}, \dots, \frac{D'_{(n)}}{a_{(n)}} \right] block \left[\begin{array}{c} (e'_{l1} + I_{\{l<a/2\}} e'_{a-l,1}), (e'_{l2} + I_{\{l<a/2\}} e'_{a-l,2}), \\ \dots, (e'_{ln} + I_{\{l<a/2\}} e'_{a-l,n}) \end{array} \right] Z,$$

where the *block* function joins matrices diagonally.

It remains to find c_l for the case involving missing data. Recall that e_{li} has elements

$$(e_{li})_{k1} = y_{i,k-l} = \begin{cases} 1, & x_{i,k-l} \text{ available,} \\ 0, & \text{otherwise.} \end{cases}$$

Therefore (noting, for the case $l = 0$, that $e_{a-0,i}$ is a zero vector),

$$c_l = (e'_{01}, e'_{02}, \dots, e'_{0n}) \begin{bmatrix} e_{l1} + I_{\{l<a/2\}} e_{a-l,1} \\ e_{l2} + I_{\{l<a/2\}} e_{a-l,2} \\ \vdots \\ e_{ln} + I_{\{l<a/2\}} e_{a-l,n} \end{bmatrix}.$$

We can now express $\tilde{\sigma}_l$ for the case involving missing data as a quadratic form:

$$\tilde{\sigma}_l = \left\{ (e'_{01}, e'_{02}, \dots, e'_{0n}) \left(\begin{bmatrix} e_{l1} \\ e_{l2} \\ \vdots \\ e_{ln} \end{bmatrix} + I_{\{l<a/2\}} \begin{bmatrix} e_{a-l,1} \\ e_{a-l,2} \\ \vdots \\ e_{a-l,n} \end{bmatrix} \right) \right\}^{-1} \times$$

$$Z' \left\{ \begin{array}{c} I_n \otimes (S_l + I_{\{l<a/2\}} S_{a-l}) \\ -block \left[\begin{array}{c} (e_{l1} + I_{\{l<a/2\}} e_{a-l,1}) 1'_s, (e_{l2} + I_{\{l<a/2\}} e_{a-l,2}) 1'_s, \\ \dots, (e_{ln} + I_{\{l<a/2\}} e_{a-l,n}) 1'_s \end{array} \right] \begin{bmatrix} a_{(1)}^{-1} D_{(1)} \\ a_{(2)}^{-1} D_{(2)} \\ \vdots \\ a_{(n)}^{-1} D_{(n)} \end{bmatrix} \end{array} \right\} Z,$$

which reduces to (7) if no data are missing. Recalling $e' = (e'_{01}, e'_{02}, \dots, e'_{0n})$ simplifies the expression corresponding to c_l slightly; then, since $e_{li} = S_l e_{0i}$, we have

$$[e'_{l1}, e'_{l2}, \dots, e'_{ln}]' = (I_n \otimes S_l)\, e.$$

Using this and a similar expression for $\mathbf{e}_{a-l,i}$ yields $c_l = \mathbf{e}'\left[\mathbf{I}_n \otimes (\mathbf{S}_l + I_{\{l<a/2\}}\mathbf{S}_{a-l})\right]\mathbf{e}$.

To simplify the matrix producing $c_l\tilde{\sigma}_l$, we recall that $\mathbf{D}_{(i)}$ is a modification of a $na \times na$ identity matrix, removing the rows of identity corresponding to the i^{th} family, so that the elements of $a_{(i)}^{-1}\mathbf{D}_{(i)}$ and $\mathbf{e}_{li}$ are

$$\begin{aligned}\left(a_{(i)}^{-1}\mathbf{D}_{(i)}\right)_{jk} &= \begin{cases} a_{(i)}^{-1}, j = k - l - aI_{\{ia<k\}}, \\ 0, \text{otherwise}, \end{cases} \\ \text{and } (\mathbf{e}_{0i})_{ql} &= y_{iq},\end{aligned}$$

$j = 1, 2, ..., s; k = 1, 2, ..., na; q = 1, 2, ..., a$ and $i = 1, 2, ..., n$. Thus

$$\begin{aligned} a_{(i)}^{-1}\mathbf{e}_{li}\mathbf{1}_s'\mathbf{D}_{(i)} &= a_{(i)}^{-1}\mathbf{e}_{li}\mathbf{1}_s'\begin{bmatrix} \mathbf{I}_{(i-1)a} & \mathbf{0}_{(i-1)a\times a} & \mathbf{0}_{(i-1)a\times(n-i)a} \\ \mathbf{0}_{(n-i)a\times(i-1)a} & \mathbf{0}_{(n-i)a\times a} & \mathbf{I}_{(n-i)a} \end{bmatrix} \\ &= a_{(i)}^{-1}\left[\mathbf{e}_{li}, \mathbf{e}_{li}, ..., \mathbf{e}_{li}, \mathbf{0}_a, \mathbf{0}_a, ..., \mathbf{0}_a, \mathbf{e}_{li}, \mathbf{e}_{li}, ..., \mathbf{e}_{li}\right] \end{aligned}$$

with vectors of zeroes appearing in the $(i-1)a+1, (i-1)a+2, ..., ia$ columns, or

$$\begin{aligned} a_{(i)}^{-1}\mathbf{e}_{li}\mathbf{1}_s'\mathbf{D}_{(i)} &= a_{(i)}^{-1}\left[\mathbf{e}_{li}\mathbf{1}_{(i-1)a}', \mathbf{0}_{a\times a}, \mathbf{e}_{li}\mathbf{1}_{(n-i)a}'\right] \\ &= a_{(i)}^{-1}\mathbf{e}_{li}\left[\mathbf{1}_{(i-1)a}', \mathbf{0}_{1\times a}, \mathbf{1}_{(n-i)a}'\right]. \end{aligned}$$

Extending this expansion, in the matrix producing $c_l\tilde{\sigma}_l$, we therefore have

$$a_{(i)}^{-1}\left(\mathbf{e}_{li} + I_{\{l<a/2\}}\mathbf{e}_{a-l,i}\right)\mathbf{1}_s'\mathbf{D}_{(i)} = a_{(i)}^{-1}\left(\mathbf{e}_{li} + I_{\{l<a/2\}}\mathbf{e}_{a-l,i}\right)\left[\mathbf{1}_{(i-1)a}', \mathbf{0}_{1\times a}, \mathbf{1}_{(n-i)a}'\right]$$

Concatenating the $a_{(i)}^{-1}\left(\mathbf{e}_{li} + I_{\{l<a/2\}}\mathbf{e}_{a-l,i}\right)\mathbf{1}_s'\mathbf{D}_{(i)}, i = 1, 2, ..., n$ vertically gives

$$\begin{aligned} &block\begin{pmatrix} a_{(1)}^{-1}\left(\mathbf{e}_{l1} + I_{\{l<a/2\}}\mathbf{e}_{a-l,1}\right), a_{(2)}^{-1}\left(\mathbf{e}_{l2} + I_{\{l<a/2\}}\mathbf{e}_{a-l,2}\right), \\ ..., a_{(n)}^{-1}\left(\mathbf{e}_{ln} + I_{\{l<a/2\}}\mathbf{e}_{a-1,n}\right) \end{pmatrix} \times \\ &\left[(\mathbf{J}_n - \mathbf{I}_n) \otimes \mathbf{1}_a'\right], \end{aligned}$$

where $\mathbf{1}_{0a}'$ has one row and no columns. Hence, $\tilde{\sigma}_l$ reduces to

$$\tilde{\sigma}_l = \left\{\mathbf{e}'\left[\mathbf{I}_n \otimes (\mathbf{S}_l + I_{\{l<a/2\}}\mathbf{S}_{a-l})\right]\mathbf{e}\right\}^{-1} \times \tag{8}$$

$$Z'\left\{\begin{array}{c}\mathbf{I}_n\otimes\left(\mathbf{S}_l+I_{\{l<a/2\}}\mathbf{S}_{a-l}\right)\\ -block\left[\begin{array}{c}a_{(1)}^{-1}\left(\mathbf{e}_{l1}+I_{\{l<a/2\}}\mathbf{e}_{a-l,1}\right),a_{(2)}^{-1}\left(\mathbf{e}_{l2}+I_{\{l<a/2\}}\mathbf{e}_{a-l,2}\right),\\ \dots,a_{(n)}^{-1}\left(\mathbf{e}_{lN}+I_{\{l<a/2\}}\mathbf{e}_{a-l,n}\right)\end{array}\right]\times\\ \left[(\mathbf{J}_n-\mathbf{I}_n)\otimes\mathbf{1}_a'\right]\end{array}\right\}Z.$$

which greatly simplifies calculating $var(\tilde{\sigma}_l)$. Note also that $\tilde{\sigma}_0>0$ with probability one for continuous distributions, i.e.(noting $\mathbf{S}_a=\mathbf{0}$ and $\mathbf{e}_a=\mathbf{0}$)

$$\left\{\mathbf{I}_{na}-block\left[a_{(1)}^{-1}\mathbf{e}_{01},a_{(2)}^{-1}\mathbf{e}_{02},\dots,a_{(n)}^{-1}\mathbf{e}_{0n}\right]\left[(\mathbf{J}_n-\mathbf{I}_n)\otimes\mathbf{1}_a'\right]\right\}$$

is nonnegative definite, with $\tilde{\sigma}_0=0$ if and only if all the nonzero components of Z are identical.

2.4 Unbiased Estimation of Interclass Covariance, Missing Data. With each $\mathbf{x}_i$, a parent's measurement p_i may be taken, having mean μ_p, such that $\sigma_{ps}=cov(p_i,x_{ij})$, $j=1,2,\dots,a$. In any family with no missing data, the covariance structure is therefore

$$\mathbf{C}_{aug}=\begin{bmatrix}\sigma_{pp} & \sigma_{ps}\mathbf{1}_a'\\ \sigma_{ps}\mathbf{1}_a & \mathbf{C}\end{bmatrix}.$$

Assuming that some of the x_{ij} but none of the p_i may be missing, we seek to estimate σ_{ps}. Let $\mu_s=\mathbf{E}(x_{ij})$, $i=1,2,\dots,n$ and $j=1,2,\dots,a$. For these parameters, p, s, ps and pp are not indices, but only subscripts. Noting $\mathbf{E}(p_ix_{ij})=\sigma_{ps}+\mu_p\mu_s$, the independence of p_i and $\bar{x}_{(i)}$ implies $\mathbf{E}\left[p_i(x_{ij}-\bar{x}_{(i)})\right]=\sigma_{ps}$. Therefore, we might develop

$$\tilde{\sigma}_{ps1}=c_{ps}^{-1}\sum_{\mathfrak{Z}_{ps}}p_i(z_{ij}-\bar{z}_{(i)})$$

as an unbiased estimator of σ_{ps}, where

$$\mathfrak{Z}_{ps}=\{(i,j):y_{ij}=1,i=1,2,\dots,n\text{ and }j=1,2,\dots,int(3a/2)\}$$

and $card(\mathfrak{Z}_{ps})=a_\bullet=\sum_i a_i$ is the number of available measurements equicorrelated with the p_i. When all data are present, $a_\bullet=na$. Estimator $\tilde{\sigma}_{ps1}$ may be considered to be the interclass analog to $\tilde{\sigma}_l$. However, if we define $\bar{z}=z_{\bullet\bullet}/a_\bullet=\sum_{ij}z_{ij}/a_\bullet$, then

$$\tilde{\sigma}_{ps}=\frac{\sum_{\mathfrak{Z}_{ps}}p_i(z_{ij}-\bar{z})}{a_\bullet-\sum_i a_i^2/a_\bullet}$$

is another unbiased estimator of σ_{ps}, and (as will be discussed in the next section) $var(\tilde{\sigma}_{ps}) \leq var(\tilde{\sigma}_{ps1})$, with strict inequality (remarkably) if and only if data is missing. When no data is missing,

$$\tilde{\sigma}_{ps} = \left(\frac{n}{n-1}\right)\hat{\sigma}_{ps},$$

where $\hat{\sigma}_{ps}$ is the (biased) MLE (assuming normality) of σ_{ps} derived by Khattree and Naik (1994a). If each of the pairs (p_i, x_{ij}) are bivariate normal, the MLE of the variance σ_{pp} of p_i is simply $n^{-1}\sum(p_i - \bar{p})^2$, as the x_{ij}, $j = 1, 2, ..., a$ provide no information about this parameter.

To express $\tilde{\sigma}_{ps1}$ and $\tilde{\sigma}_{ps}$ as quadratic forms, let $\mathbf{Z}_i' = (p_i, \mathbf{z}_i')$, $i = 1, 2, ..., n$ and $\mathbf{Z}' = (\mathbf{Z}_1', \mathbf{Z}_2', ..., \mathbf{Z}_n')$ (recall Z is not the same as $\mathbf{Z}$, as it does not include the parents' scores). If $\mathbf{w}$ is obtained by concatenating a zero at the top of a $a \times 1$ vector of unities and $\mathbf{a}' = (a_1, a_2, ..., a_n)$, then

$$\tilde{\sigma}_{ps1} = a_\bullet^{-1}\mathbf{Z}'\left\{\left(\mathbf{I}_n - (\mathbf{J}_n - \mathbf{I}_n)\begin{bmatrix} \frac{a_1}{a_{(1)}} & 0 & \cdots & 0 \\ 0 & \frac{a_2}{a_{(2)}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{a_n}{a_{(n)}} \end{bmatrix}\right) \otimes \left[\mathbf{w}, \mathbf{0}_{(a-1)\times a}\right]\right\}\mathbf{Z} \text{ and}$$

$$\tilde{\sigma}_{ps} = \left(a_\bullet - a_\bullet^{-1}\sum_i a_i\right)^{-1}\mathbf{Z}'\left\{\left(\mathbf{I}_n - \frac{\mathbf{1}_n\mathbf{a}'}{a_\bullet}\right) \otimes \left[\mathbf{w}, \mathbf{0}_{(a-1)\times a}\right]\right\}\mathbf{Z}, \tag{9}$$

mentioning $\tilde{\sigma}_{ps1}$ only because it is the interclass analog to the LOO estimators $\tilde{\sigma}_l$.

We have now proposed unbiased estimators of $\sigma_l, l = 0, 1, 2, ..., m$, and of σ_{ps} when p_i is available for each of the families. The initial definitions of these estimators are far more practicable than the corresponding quadratic form expressions when actually calculating the estimators, due to the large matrices involved in the quadratic forms. However, the latter make possible the calculation of their variances, to which we turn in the next subsection.

2.5 Variances of the Unbiased Estimators . Let the nonzero elements of the diagonal $n(a+1)\times n(a+1)$ matrix $\mathbf{D}_{aug}$ be unities corresponding to the observed (rather than imputed, zero) elements of $\mathbf{Z}$, so that $\mathbf{D}_{aug}(\mathbf{I}_n \otimes \mathbf{C}_{aug})\mathbf{D}_{aug}$, and $\mathbf{D}_{aug}\left(\mathbf{I}_n \odot [\mu_p, \mu_s\mathbf{1}']'\right) = \mathbf{D}_{aug}\boldsymbol{\mu}$ are the covariance and mean of the (possibly) incomplete dataset $\mathbf{Z}$ including the parents' measurements

p_t, conditional on $\mathbf{D}_{aug}\mathbf{1}_{n(a-1)}$ (note $\mathbf{e} = \mathbf{D}_{aug}\mathbf{1}_{n(a-1)}$). If $\mathbf{F}_{ps}$ is the symmetric matrix such that $\tilde{\sigma}_{ps} = \mathbf{Z}'\mathbf{F}_{ps}\mathbf{Z}$, from linear models theory, the cumulants of $(\tilde{\sigma}_{ps}|\mathbf{e})$ are

$$\kappa_{ps,r} = 2^{r-1}(r-1)!\left\{\begin{array}{c} tr\left\{\left[\mathbf{F}_{ps}\mathbf{D}_{aug}\left(\mathbf{I}_n \otimes \mathbf{C}_{aug}\right)\mathbf{D}_{aug}\right]^r\right\} \\ +r\boldsymbol{\mu}'\mathbf{D}_{aug}\mathbf{F}_{ps}\left[\mathbf{D}_{aug}\left(\mathbf{I}_n \otimes \mathbf{C}_{aug}\right)\mathbf{D}_{aug}\mathbf{F}_{ps}\right]^{r-1}\mathbf{D}_{aug}\boldsymbol{\mu} \end{array}\right\}, r \in \mathbb{N}.$$

Likewise, let the nonzero elements of the diagonal $na \times na$ matrix $\mathbf{D}$ be diagonal unities corresponding to the observed elements of Z. Making $\mathbf{F}_l$ the symmetric matrix so that $\tilde{\sigma}_l = Z'\mathbf{F}_l Z$, $l = 0, 1, 2, \ldots, m$, the cumulants of $(\tilde{\sigma}_l|\mathbf{e})$ are

$$\kappa_{l,r} = 2^{r-1}(r-1)!\left\{\begin{array}{c} tr\left\{\left[\mathbf{F}_l\mathbf{D}\left(\mathbf{I}_n \otimes \mathbf{C}\right)\mathbf{D}\right]^r\right\} \\ +r\mu_s^2\mathbf{e}'\mathbf{F}_l\left[\mathbf{D}\left(\mathbf{I}_n \otimes \mathbf{C}\right)\mathbf{D}\mathbf{F}_l\right]^{r-1}\mathbf{e} \end{array}\right\},$$

noting $\mathbf{E}(Z|\mathbf{e}) = \mu_s\mathbf{e}$ and $cov(Z|\mathbf{e}) = \mathbf{D}(\mathbf{I}_n \otimes \mathbf{C})\mathbf{D}$. In particular,

$$\kappa_{ps,1} = \mathbf{E}(\tilde{\sigma}_{ps}) = \mathbf{E}(\tilde{\sigma}_{ps}|\mathbf{e}) = \sigma_{ps} \text{ and } \kappa_{l,1} = \mathbf{E}(\tilde{\sigma}_l) = \mathbf{E}(\tilde{\sigma}_l|\mathbf{e}) = \sigma_l,$$

as was shown in the last subsection. Also

$$\kappa_{ps,2} = \mathbf{Var}(\tilde{\sigma}_{ps}|\mathbf{e}) = 2\left\{\begin{array}{c} tr\left\{\left[\mathbf{F}_{ps}\mathbf{D}_{aug}\left(\mathbf{I}_n \otimes \mathbf{C}_{aug}\right)\mathbf{D}_{aug}\right]^2\right\} \\ +2\boldsymbol{\mu}'\mathbf{D}_{aug}\mathbf{F}_{ps}\mathbf{D}_{aug}\left(\mathbf{I}_n \otimes \mathbf{C}_{aug}\right)\mathbf{D}_{aug}\mathbf{F}_{ps}\mathbf{D}_{aug}\boldsymbol{\mu} \end{array}\right\},$$

and

$$\kappa_{l,2} = \mathbf{Var}(\tilde{\sigma}_l|\mathbf{e}) = 2\left\{\begin{array}{c} tr\left\{\left[\mathbf{F}_l\mathbf{D}\left(\mathbf{I}_n \otimes \mathbf{C}\right)\mathbf{D}\right]^2\right\} \\ +2\mu_s^2\mathbf{e}'\mathbf{F}_l\mathbf{D}\left(\mathbf{I}_n \otimes \mathbf{C}\right)\mathbf{D}\mathbf{F}_l\mathbf{e} \end{array}\right\}. \tag{10}$$

Even when no data are missing, simplification of each $\kappa_{l,2}$ does not seem possible, and missing data make the expressions even less tractable, due to the series of unities and zeros $\mathbf{e}_{li}$ appearing in the $\mathbf{F}_l$, $cov(\mathbf{Z}|\mathbf{e}) = \mathbf{D}_{aug}(\mathbf{I}_n \otimes \mathbf{C}_{aug})\mathbf{D}_{aug}$ and $cov(Z|\mathbf{e}) = \mathbf{D}(\mathbf{I}_n \otimes \mathbf{C})\mathbf{D}$. To illustrate, the symmetric $\mathbf{F}_l$ may be expressed as $\frac{\mathbf{A}_l - \mathbf{A}_l'}{2}$, where $\mathbf{A}_l$ is the matrix in (8) producing $\tilde{\sigma}_l$. The matrix $2c_l\mathbf{F}_l$ is then

$$\begin{array}{c} \mathbf{I}_n \otimes \left(\mathbf{S}_l + \mathbf{S}_{-l} + I_{\{l<a/2\}}\left(\mathbf{S}_{a-l} + \mathbf{S}_{l-a}\right)\right) \\ -block\left[\begin{array}{c} a_{(1)}^{-1}\left(\mathbf{e}_{l1} + I_{\{l<a/2\}}\mathbf{e}_{a-l,1}\right), a_{(2)}^{-1}\left(\mathbf{e}_{l2} + I_{\{l<a/2\}}\mathbf{e}_{a-l,2}\right), \\ \ldots, a_{(n)}^{-1}\left(\mathbf{e}_{ln} + I_{\{l<a/2\}}\mathbf{e}_{a-l,n}\right) \end{array}\right]\left[(\mathbf{J}_n - \mathbf{I}_n) \otimes \mathbf{1}_a'\right] \\ -\left[(\mathbf{J}_n - \mathbf{I}_n) \otimes \mathbf{1}_a\right] block\left[\begin{array}{c} a_{(1)}^{-1}\left(\mathbf{e}_{l1}' + I_{\{l<a/2\}}\mathbf{e}_{a-l,1}'\right), a_{(2)}^{-1}\left(\mathbf{e}_{l2}' + I_{\{l<a/2\}}\mathbf{e}_{a-l,2}'\right), \\ \ldots, a_{(n)}^{-1}\left(\mathbf{e}_{ln}' + I_{\{l<a/2\}}\mathbf{e}_{a-l,n}'\right) \end{array}\right], \end{array}$$

generalizing the definition of $\mathbf{S}_l$ so that if $-a \leq l < 0$, the unities of $\mathbf{I}_a$ are shifted down l positions, and $\mathbf{S}_l = \mathbf{S}'_{-l}$ for all $l \in \mathbb{N}$.

Hence it is seen that no means is evident of simplifying $c_l\mathbf{F}_l$ in the most general case (for arbitrary e, a, n and $\mathbf{C}$) to form a manageable expression for $\mathbf{F}_l$ which can be analytically pre- and post-multiplied by $\mathbf{D}\left(\mathbf{I}_n \bigotimes \mathbf{C}\right)\mathbf{D}$ (or $\mathbf{D}_{aug}\left(\mathbf{I}_n \otimes \mathbf{C}_{aug}\right)\mathbf{D}_{aug}$) and then squared, cubed, etc. The expressions for $c_l\mathbf{F}_l$ are complicated even if no data is missing, in which case all $y_{ij} = 1, a_i = a$ and $\mathbf{e}_{li} = \mathbf{S}_l\mathbf{1}$. It may be shown that

$$var\left(\tilde{\sigma}_{ps1}\right) = var\left(\tilde{\sigma}_{ps}\right) = \frac{\sigma_{pp}\sum\sigma_j + a\sigma_{ps}^2}{a\left(n-1\right)}$$

when no data is missing. When data is missing, though, convenient expressions for $var\left(\tilde{\sigma}_{ps1}|\mathbf{e}\right)$ and $var\left(\tilde{\sigma}_{ps}|\mathbf{e}\right)$ do not exist due to the difficulty of writing the $\mathbf{e}'_{0i}\mathbf{C}\mathbf{e}_{0i}$ without the use of matrices. Only the variances of $\left(\tilde{\sigma}_{ps1}|\mathbf{e}\right), \left(\tilde{\sigma}_{ps}|\mathbf{e}\right)$ and $\left(\tilde{\sigma}_l|\mathbf{e}\right)$ may be calculated, by machine, for whatever $\mathbf{C}_{aug}$, a and n are postulated and for whatever particular e is observed.

In the sections that follow, we will wish to compare the LOO estimators with the estimators we will form by extending the principles of Expectation and Maximization. It is preferable to compare the unconditional (on e) variances of these estimators, under reasonable assumptions about the missing data mechanism, rather than their variances conditional on, say, a small number of selected missing data patterns. Since

$$\begin{aligned} var\left(\tilde{\sigma}_l\right) &= var\left(\mathbf{E}\left(\tilde{\sigma}_l|\mathbf{e}\right)\right) + \mathbf{E}\left(var\left(\tilde{\sigma}_l|\mathbf{e}\right)\right) \\ &= 0 + \mathbf{E}\left(var\left(\tilde{\sigma}_l|\mathbf{e}\right)\right), \end{aligned}$$

estimation of $\mathbf{E}\left(var\left(\tilde{\sigma}_l|\mathbf{e}\right)\right)$ constitutes that of $var\left(\tilde{\sigma}_l\right)$. Estimating $\mathbf{E}\left(var\left(\tilde{\sigma}_l|\mathbf{e}\right)\right)$ may be accomplished by calculating $var\left(\tilde{\sigma}_l|\mathbf{e}\right)$ for a large number of different e, chosen at random from the 2^{na} possible missing data patterns, with plausible distributional assumptions for e. Analogous statements can be made about the estimation of the unconditional $var\left(\tilde{\sigma}_{ps}\right)$ and $var\left(\tilde{\sigma}_{ps1}\right)$. Because simulations seemed to show that $var\left(\tilde{\sigma}_{ps}|\mathbf{e}\right) < var\left(\tilde{\sigma}_{ps1}|\mathbf{e}\right)$ uniformly in the missing data case, at this point we discontinue mention of $\tilde{\sigma}_{ps1}$.

One reasonable assumption is that the e_{ij} are iid Bernoulli $\left(1 - p_m\right)$ random variables, so that $\mathbf{P}(e_{ij} = 0) = p_m$ is the probability that x_{ij} is missing. Since p_m does not depend on x_{ij}, the process represents a Missing Completely at Random (MCAR) mechanism, as discussed by Little and Rubin (1987). For each of several combinations of n, a, p_m and setting the covariance parameters to $(\sigma_0, \sigma_1, \sigma_2, \sigma_3) = (4, 2.5, 2, 1.5)$, one hundred missing data patterns were chosen at random (with replacement, except that any pattern for which any $a_{(i)} = 0$ was discarded), and the means and standard errors of the resulting one hundred variances of each of $(\tilde{\sigma}_{ps}|\mathbf{e})$ and $(\tilde{\sigma}_l|\mathbf{e})$ were calculated. These means of variances estimate the unconditional (on $\mathbf{e}$) variances of the $\tilde{\sigma}_{ps}$ and $\tilde{\sigma}_l$, since we have randomized over all possible values of $\mathbf{e}$. Table I displays the means and standard errors of $var\,(\tilde{\sigma}_l)$ only; those of $var\,(\tilde{\sigma}_{ps})$ are very similar.

Table I: Estimated Means and Standard Errors of Conditional Variances of LOO Estimators.

p_m	n, a		$l = 0$	$l = 1$	$l = 2$	$l = 3$
.1	10,4	$mean(var(\bar{\sigma}_l))$	1.8682	1.8878	1.9888	
.1		$StdErr(var(\bar{\sigma}_l))$	.0465	.1363	.1077	
.1	10,7	$mean(var(\bar{\sigma}_l))$	1.3404	1.3379	1.2902	1.3675
.1		$StdErr(var(\bar{\sigma}_l))$	.0342	.0814	.0704	.0782
.1	50,4	$mean(var(\bar{\sigma}_l))$	.3508	.3597	.3814	
.1		$StdErr(var(\bar{\sigma}_l))$	.0034	.0093	.0111	
.1	50,7	$mean(var(\bar{\sigma}_l))$	.2495	.2492	.2386	.2525
.1		$StdErr(var(\bar{\sigma}_l))$	.0023	.0046	.0047	.0053
.5	10,4	$mean(var(\bar{\sigma}_l))$	2.6045	3.8330	4.7459	
.5		$StdErr(var(\bar{\sigma}_l))$	.4903	1.1884	1.4121	
.5	10,7	$mean(var(\bar{\sigma}_l))$	1.7378	2.6382	2.4413	2.2241
.5		$StdErr(var(\bar{\sigma}_l))$	.1159	.6063	.5521	.3405
.5	50,4	$mean(var(\bar{\sigma}_l))$	.4987	.8326	.8894	
.5		$StdErr(var(\bar{\sigma}_l))$	.0274	.1448	.1070	
.5	50,7	$mean(var(\bar{\sigma}_l))$	.3347	.4996	.4543	.4807
.5		$StdErr(var(\bar{\sigma}_l))$	.0129	.0467	.0393	.0445

It is seen from the tabulated values that for most of the combinations of C_{aug}, a and n, the sample coefficients of variance (the quotient of sample mean and sample variance) of the conditional sample cumulants of $\bar{\sigma}_{ps}$ and each $\bar{\sigma}_l$ are quite large, especially as n increases and when p_m is small. Hence we may assume that, for the vast majority of missing data patterns **e**, the conditional (on **e**) cumulants do not vary too much from the unconditional cumulants. This implies that, as will be desired, the above estimates of the unconditional variances may be used in their places without introducing a significant source of error.

Of interest is a comparison of the performance of $\tilde{\sigma}_{ps}$ and each $\tilde{\sigma}_l$ with that of the corresponding MLEs $\hat{\sigma}_{ps}$ and $\hat{\sigma}_l$ derived by Khattree and Naik (1994a) and Olkin and Press (1969), respectively, in the case of normal, complete data (no missing observations). Khattree and Naik described the modifications to Olkin and Press' MLE of $\boldsymbol{\sigma}$ given the assumption $\mathbf{E}(\mathbf{x}_i) = \mu\mathbf{1}$. It may be shown that for each $l = 0, 1, ..., m$, $\hat{\sigma}_l$, like $\tilde{\sigma}_l$, is unbiased for σ_l, and together the $\hat{\sigma}_l$ have covariance

$$cov \begin{bmatrix} \hat{\sigma}_0 \\ \hat{\sigma}_1 \\ \vdots \\ \hat{\sigma}_m \end{bmatrix} = (\mathbf{B}^{-1})' \begin{bmatrix} \frac{2\delta_1^2}{n} & 0 & 0 & \cdots & & 0 \\ 0 & \frac{\delta_2^2}{n} & 0 & \cdots & & 0 \\ 0 & 0 & \frac{\delta_3^2}{n} & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \\ & & & & \frac{\delta_m^2}{n} & 0 \\ 0 & 0 & \cdots & & 0 & \frac{2\delta_{m+1}^2}{\alpha_{m+1} n} \end{bmatrix} \mathbf{B}^{-1},$$

so that comparing the mean squared errors of the $\hat{\sigma}_l$ with those of the $\tilde{\sigma}_l$ is equivalent to comparing their variances. Also, it may be shown that (in the full data case with normally distributed data)

$$\mathbf{E}(\hat{\sigma}_{ps}) = \frac{(n-1)}{n}\sigma_{ps} \text{ and } var(\hat{\sigma}_{ps}) = \frac{(n-1)}{n^2 a}\left(\sigma_{pp}\sum\sigma_j + a\sigma_{ps}^2\right),$$

so that

$$\text{MSE}(\hat{\sigma}_{ps}) = var(\hat{\sigma}_{ps}) + [bias(\hat{\sigma}_{ps})]^2 = \frac{(n-1)\sigma_{pp}\sum\sigma_j + an\sigma_{ps}^2}{n^2 a}$$

and hence $\text{MSE}(\hat{\sigma}_{ps}) - \text{MSE}(\tilde{\sigma}_{ps}) = \frac{(1-2n)\sigma_{pp}\sum\sigma_j - na\sigma_{ps}^2}{n^2 a(n-1)}$ is negative, provided that $\sum\sigma_j > \frac{na\sigma_{ps}^2}{\sigma_{pp}(1-2n)}$. Alternatively,

$$\frac{\text{MSE}(\hat{\sigma}_{ps})}{\text{MSE}(\tilde{\sigma}_{ps})} = \frac{(n-1)}{n}\frac{\left(\frac{n-1}{n}\sigma_{pp}\sum\sigma_j + a\sigma_{ps}^2\right)}{\left(\sigma_{pp}\sum\sigma_j + a\sigma_{ps}^2\right)}$$

can be viewed as a kind of relative efficiency of these estimators and is, for most reasonable parameter values, less than unity. The advantages of $\tilde{\sigma}_{ps}$ over $\hat{\sigma}_{ps}$, therefore, are its unbiasedness (even without the assumption of normality) and its immediate definition when data is missing.

In Table II are compared the variances (which are mean squared errors, due to unbiasedness) of the $\tilde{\sigma}_l$ with those of the $\hat{\sigma}_l$ for various combinations of n and a with no missing data.

for the normal population having the particular combination of circular covariance parameters $(\sigma_{pp}, \sigma_{ps}, \sigma_0, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7) = (4,2,4,2.5,2,1.5,1.3,1.15,1)$ (though some of these parameters may not appear in the density, depending on a). The MLE $\hat{\sigma}_l$ are guaranteed to achieve asymptotic efficiency, whereas the information contained in the data about σ_l ignored by the $\tilde{\sigma}_l$ (due to the "leave-one-out" sample means $\bar{z}_{(i)}$ they implement) is bound to increase the variances of the latter above efficiency for all sample sizes. However, for moderate na it is seen in the tabulated values of $var(\tilde{\sigma}_l)/var(\hat{\sigma}_l)$ that the majority of the $\tilde{\sigma}_l$ have lower variances than do the $\hat{\sigma}_l$ and that, where $\hat{\sigma}_l$ has lower variance than $\tilde{\sigma}_l$, the difference in variances is slight. The variances of the $\hat{\sigma}_l$ are seen to decrease below those of the $\tilde{\sigma}_l$ as na becomes large. The theoretical variances of the $\tilde{\sigma}_l$ were derived from (10), substituting a vector of unities for e (corresponding to a no missing data pattern).

Table II: Comparison of Mean Squared Errors of LOO Estimators with MSEs of ML Estimators given No Missing Data.

n, a	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$
2,7	.9460	.9828	1.0128	1.0027			
2,10	.9267	.9581	.9950	1.0102	.9967	.9753	
2,13	.9114	.9391	.9733	1.0000	1.0062	.9957	.9736
5,7	.9783	.9931	1.0051	1.0010			
5,10	.9707	.9833	.9980	1.0041	.9986	.9901	
5,13	.9646	.9756	.9893	1.0008	1.0025	.9982	.9894
10,7	1.0110	1.0035	1.0026	1.0006			
10,10	1.0156	1.0085	1.0010	1.0020	1.0007	1.0049	
10,13	1.018	1.0123	1.0053	1.0052	1.0012	1.0008	1.0053
15,7	1.0072	1.0023	1.0016	1.0004			

Exact variances were not calculated for $\tilde{\sigma}_l$ for $na > 130$, because of the enormous matrices involved in these quadratic forms.

Given the low MSEs of the $\tilde{\sigma}_l$ relative to those of the $\hat{\sigma}_l$ for moderate sample sizes, we might hope that an estimator $\tilde{\rho}_{ps} = \tilde{\sigma}_{ps}/\sqrt{\tilde{\sigma}_{pp}\tilde{\sigma}_0}$ could be constructed of the interclass correlation $\rho_{ps} = \sigma_{ps}/\sqrt{\sigma_{pp}\sigma_0}$, utilizing some suitable estimator $\tilde{\sigma}_{pp}$ of σ_{pp}, which would perform as well as or better than the MLE $\hat{\rho}_{ps} = \hat{\sigma}_{ps}/\sqrt{\hat{\sigma}_{pp}\hat{\sigma}_0}$ of ρ_{ps} for small or moderate na. The uniform minimum variance unbiased estimator (UMVUE)

$$\tilde{\sigma}_{pp} = \frac{\sum_i (p_i - \bar{p})^2}{n-1}$$

would probably be expected to serve as the best "suitable estimator" of σ_{pp}, given the unbiasedness of $\tilde{\sigma}_{ps}$ and $\tilde{\sigma}_0$.

In fact, however, for normal data, the MSEs of $\tilde{\rho}_{ps}$ (constructed in this way) and $\hat{\rho}_{ps}$ seem almost identical across a wide range of (n, a), with a very slight advantage evident in the use of $\hat{\rho}_{ps}$. It appears that the biases inherent in the $\hat{\sigma}_{pp}, \hat{\sigma}_{ps}$ and $\hat{\sigma}_0$ in large part mitigate each other, combining to produce a relatively efficient MLE having small bias, even for small na. Simulations involving t-distributions did not seem to change these comparisons significantly. It seems that the advantages of $\tilde{\sigma}_{ps}$ and $\tilde{\sigma}_l$ are their unbiasedness, relatively small MSEs (in the case of the $\tilde{\sigma}_l$) for moderate na, usefulness when normality cannot be assumed (since the behaviors of the ML estimates, and the "EM" estimates described in the following subsection, have not been identified), and (not least) their immediate extensions to any missing data pattern **e** for which the missing data mechanism can be ignored. In the estimation of interclass correlation, $\tilde{\sigma}_{ps}$ and $\tilde{\sigma}_l$ do not appear to offer an advantage over the maximum likelihood approach except when data is missing, or when normality cannot be assumed.

<u>2.6 Development of EM Algorithm Estimators and Evaluation of their Performance.</u> The methods of this subsection draw on important ideas of the popular Expectation and Maximization (EM) Algorithm; nonetheless, important differences exist between EM and what we outline. The estimates given in this subsection therefore bear the name "EM" in quotation marks.

Following the formulation of EM given by Little and Rubin (1987), the E step, if it were possible, in the case of circular covariance with missing data and uniform sib mean, would begin by assuming a particular multivariate distribution for the hypothetical, complete (no missing data) dataset, and choosing appropriate sufficient statistics for the covariance parameters to be estimated in the resulting likelihood. The expectations of these statistics would be calculated, conditional on both the observed data and the current approximations of the parameters. The M step would follow by maximizing the likelihood with respect to the parameters, using the estimated sufficient statistics in the place of the actual values (had they been observed). Iterations then would continue between the E and M steps.

The obvious choices for the sufficient statistics to be estimated would be the sums of squares of the canonical variables $\sum_i y_{ij}^2$, $j = 1, 2, ..., a$ (these are not the y_{ij} introduced in the last subsection) which Olkin and Press (1969) implement to maximize the likelihood, assuming normality throughout. Unfortunately, methods for estimating these sums are not obvious. We respond to this difficulty by estimating instead the sums $\sum_i \sum_j x_{ij} = x_{\bullet\bullet}$, $\sum_i \sum_j p_i x_{ij}$ and $\sum_i x_{ij} x_{il}$, $j, l = 1, 2, ..., a$. While these sums do not seem to allow ML estimation, they are sufficient statistics for μ_s, σ_{ps} and the σ_l in the likelihood of the hypothetical complete dataset, and can be estimated when data is missing.

When possible in any multivariate situation with missing data, estimating $x_{\bullet\bullet}$ is accomplished using the ML regression of each missing data point x_{ij} on the observed part of the i^{th} family. At the t^{th} iteration, we denote the current estimator of x_{ij} by $x_{ij}^{(t)}$, which is simply x_{ij} if x_{ij} is available.

If the missing data followed a monotone pattern, the procedure for estimating the missing data points x_{ij} for a given j and all i, would be to calculate the augmented sample covariance matrix (Little and Rubin, 1987) of $(x_{i1}, x_{i2}, ..., x_{ij})$ using the data from the families i on which all the components of $(x_{i1}, x_{i2}, ..., x_{ij})$ are observed. The sweep and reverse sweep operators would then conveniently provide estimates of the regression of the x_{ij} on $(x_{i1}, x_{i2}, ..., x_{ij-1})$, $i = 1, 2, ..., n$. However, since a nonmonotone missing data pattern is a far more realistic development of the model now at hand, we seek alternate methods of imputing values to the missing data.

Another way of estimating the missing data $\mathbf{x}_{i1}$ of the i^{th} family using the observed sibs' scores $\mathbf{x}_{i2}$ from the same family and the current estimates of the parameters assumes that $\mathbf{E}\left(\mathbf{x}_{i1}|\mathbf{x}_{i2}, \mu^{(t)}, \mathbf{C}^{(t)}\right)$ is linear in $\mathbf{x}_{i2}$, so that

$$\mathbf{E}\left(\mathbf{x}_{i1}|\mathbf{x}_{i2}, \mathbf{C}^{(t)}\right) = \mu^{(t)}\mathbf{1} + \mathbf{C}_{12,i}^{(t)}\left(\mathbf{C}_{22,i}^{(t)}\right)^{-1}\left(\mathbf{x}_{i2} - \mu^{(t)}\mathbf{1}\right), \tag{11}$$

where $\mathbf{C}_{22,i}^{(t)}$ and $\mathbf{C}_{12,i}^{(t)}$ are the appropriate submatrices of $\mathbf{C}^{(t)}$. When a parent's score is available from each family, (11) could easily be modified appropriately. However, it is felt that calculating $\left(\mathbf{C}_{22,i}^{(t)}\right)^{-1}$ for each iteration, for each family, would consume excessive computer resources, especially when the dimension a of each circle (family) is large (increasing the potential dimensions of the $\mathbf{C}_{22,i}^{(t)}$).

A faster method also supposes that if x_{ij} is missing and x_{ik} is available, then $\mathbf{E}(x_{ij}|x_{ik})$ is linear in x_{ik}, so that the conditional mean and variance of x_{ij} given x_{ik} are $\mu_s + \frac{\sigma_l}{\sigma_0}(x_{ik} - \mu_s)$ and (assuming normality) $\sigma_0\left(1 - \left(\frac{\sigma_l}{\sigma_0}\right)^2\right) = \sigma_0 - \frac{\sigma_l^2}{\sigma_0}$, where

$$l = \begin{cases} |j-k|, & |j-k| \le a/2, \\ a - |j-k|, & \text{otherwise.} \end{cases}$$

Similarly, assuming normality, the conditional mean and variance of x_{ij} given p_i are $\mu_s + \frac{\sigma_{ps}}{\sigma_{pp}}(p_i - \mu_p)$ and (assuming normality) $\sigma_0 - \frac{\sigma_{ps}^2}{\sigma_{pp}}$. Letting the current estimates of the σ_l at the t^{th} iteration be $\sigma_l^{(t)}, l = 0, 1, ..., m$ and that of σ_{ps} be $\sigma_{ps}^{(t)}$, the following statistics have approximately the same expectations as the mean of the posterior distribution of x_{ij} given the available data:

$$\bar{x}_{ij,k}^{(t)} = \bar{z} + \frac{\sigma_l^{(t)}}{\sigma_0^{(t)}}(z_{ik} - \bar{z}); \; x_{ik} \text{ available, } k \neq j, \text{ and}$$

$$\bar{x}_{ij,p}^{(t)} = \bar{z} + \frac{\sigma_{ps}^{(t)}}{\sigma_0^{(t)}}(p_i - \bar{p}),$$

where $\bar{p}$ is the sample mean of all the parents' scores (all of which are available, by assumption) and $\bar{z}$ is the sample mean of all available x_{ij}. In fact, for $q = 1, 2, ..., a$, $\bar{x}_{ij,q}^{(t)}$, if available, is the MLE of the posterior mean of x_{ij} given the respective predictive observation and the current parameter estimates. For each missing x_{ij}, a_i ($a_i + 1$ if parents' scores are available) linear regressions of x_{ij} on the observed part of the scores on the i^{th} family are available: at the t^{th} E step, combining these regressions can be expected to yield a satisfactorily efficient predictor $x_{ij}^{(t)}$ of x_{ij}, with an acceptably small bias. We combine them in each iteration using weights inversely proportional to the estimates of their sampling variances $\sigma_0^{(t)} - \frac{\left(\sigma_l^{(t)}\right)^2}{\sigma_0^{(t)}}$ (for $\tilde{x}_{ij,k}^{(t)}$) and $\sigma_0^{(t)} - \frac{(\sigma_{ps}^{(t)})^2}{\sigma_0^{(t)}}$ (for $\tilde{x}_{ij,p}^{(t)}$). That is to say, if x_{ij} is not available, then at the t^{th} iteration, we put

$$x_{ij}^{(t)} = \frac{\sum_{q=1}^{a} w_{ijq}\tilde{x}_{ij,q}^{(t)}\left(v_l^{(t)}\right)^{-1} + \tilde{x}_{ij,p}^{(t)}\left(v_{ps}^{(t)}\right)^{-1}}{\sum_{q=1}^{a} w_{ijq}\left(v_l^{(t)}\right)^{-1} + \left(v_{ps}^{(t)}\right)^{-1}}, \qquad (12)$$

omitting $\tilde{x}_{ij,p}^{(t)}$ (and its associated weight, in the denominator of $x_{ij}^{(t)}$) if no parents' measurements are taken or if σ_{ps} is known to be zero. The notation in (12) uses the following additional variables:

$$w_{ijq} = \begin{cases} 1, \ \tilde{x}_{ij,q} \text{ available,} \\ 0, \text{ otherwise} \end{cases}$$

is an indicator; $v_l^{(t)} = \sigma_0^{(t)} - \frac{\left(\sigma_l^{(t)}\right)^2}{\sigma_0^{(t)}}$, if $w_{ijq} = 1$, is the current estimate of the conditional variance of x_{ij} given x_{iq} for $q = 1, 2, ..., a$;

$$l = l(j,q) = \begin{cases} |j-q|\,, |j-q| \leq a/2, \\ a - |j-q|\,, \text{ otherwise;} \end{cases}$$

$v_{ps}^{(t)} = \sigma_0^{(t)} - \frac{(\sigma_{ps}^{(t)})^2}{\sigma_0^{(t)}}$ is the current estimate of the conditional variance of x_{ij} given $\tilde{x}_{ij,p}^{(t)}$; and the index (t) on all estimates indicates the current iteration. Because (12) includes information about x_{ij} given by its parent's measurement p_i, the "EM" estimators, unlike the $\tilde{\sigma}_l$, can be expected to improve in efficiency in the presence of the p_i (and, more generally, with increases in $corr(p_i, x_{ij})$) since the $\tilde{\sigma}_l$ do not use information from the p_i at all.

The expectations of $\tilde{x}_{ij,k}^{(t)}$ and $\tilde{x}_{ij,p}^{(t)}$ are only approximately equal to the conditional expectation of the missing x_{ij} given the available data because the first moments of $\frac{\sigma_l^{(t)}}{\sigma_0^{(t)}}$ and $\frac{\sigma_{ps}^{(t)}}{\sigma_0^{(t)}}$ are not known to be $\frac{\sigma_l}{\sigma_0}$ and $\frac{\sigma_{ps}}{\sigma_{pp}}$, respectively, and the covariances of these random fractions with $z_{i,j-l}, z_{i,j-l}, z_{i,j-a-l}, z_{i,j-a+l}, \bar{z}, p_i$ and $\bar{p}$, are not zero. No theory exists to calculate the expectations of these fractions, because the fractions derive from "EM" iterations. The highly appealing Laplace Approximations developed by Lieberman (1994) could be used to estimate the errors of the ratios of quadratic forms $\frac{\tilde{\sigma}_l}{\tilde{\sigma}_0}$ and $\frac{\tilde{\sigma}_{ps}}{\tilde{\sigma}_{pp}}$. However, the biases of the $\frac{\sigma_l^{(t)}}{\sigma_0^{(t)}}$ and $\frac{\sigma_{ps}^{(t)}}{\sigma_0^{(t)}}$ used in estimating the posterior means of the missing x_{ij} cannot be quantified; justification for their use in the (likelihood-based) E step rests on their status as the best substitutes for ML estimates.

The sum of the $\tilde{x}_{ij}^{(t)}$ estimates $x_{\bullet\bullet}$ in the likelihood of the complete data. The second part of the modified E step is to substitute suitable values in this likelihood for the sums of squares and cross products $\sum_i x_{ij}x_{il}, j, l = 1, 2, ..., a$. We calculate

$$\mathbf{E}\left(\sum_i x_{ij}x_{ip}|\mathbf{Z}, \mathbf{C}_{aug}^{(t)}\right) = \sum_i \left(x_{ij}^{(t)}x_{ip}^{(t)} + c_{jpi}^{(t)}\right)$$

where $\mathbf{C}_{aug}^{(t)}$ is the estimate of $\mathbf{C}_{aug}$ at the t^{th} iteration and

$$c_{jpi}^{(t)} = \begin{cases} cov(x_{ij}, x_{ip}|\mathbf{Z}, \mathbf{C}_{aug}^{(t)}), \; y_{ij} + y_{ip} = 0, \\ 0, \; y_{ij} + y_{ip} > 0, \end{cases} = \begin{cases} \sigma_l^{(t)}, \; y_{ij} + y_{ip} = 0, \\ 0, \; y_{ij} + y_{ip} > 0. \end{cases}$$

The lag l is given by $l = |j - p|$ if $|j - p| \leq a/2$ and $l = a - |j - p|$ otherwise in these expressions, as usual.

Having found suitable substitutes for the sufficient statistics we have chosen, we estimate the parameters $\boldsymbol{\sigma}$ directly, without reference to $\boldsymbol{\delta}$, since the ML estimation procedures of Olkin and Press, and of Khattree and Naik, cannot be applied here because they use different sufficient statistics than we have estimated to calculate the ML estimates of the eigenvalues of $\mathbf{C}$. Because $\sigma_0 = var(x_{i1}) = var(x_{i2}) = ... = var(x_{ia})$, $\sigma_1 = cov(x_{i1}, x_{i2}) = cov(x_{i2}, x_{i3}) = ... = cov(x_{i,a-1}, x_{ia}) = cov(x_{ia}, x_{i1})$, etc., we put

$$\sigma_l^{(t)} = \begin{cases} a^{-1}\left[\begin{array}{c} \sum_{j=1}^{a-l}\left[n^{-1}\sum_{i=1}^{n}(x_{ij}^{(t)}x_{i,j-l}^{(t)} + c_{j,j-l,i}^{(t)}) - \left(\mu_s^{(t)}\right)^2\right] \\ +\sum_{j=1}^{l}\left[n^{-1}\sum_{i=1}^{n}(x_{ij}^{(t)}x_{i,j-a-l}^{(t)} + c_{j,j-a-l,i}^{(t)}) - \left(\mu_s^{(t)}\right)^2\right] \end{array}\right], & l < a/2, \\ \frac{2}{a}\left[\sum_{j=1}^{l}\left[n^{-1}\sum_{i=1}^{n}(x_{ij}^{(t)}x_{i,j-l}^{(t)} + c_{j,j-l,i}^{(t)}) - \left(\mu_s^{(t)}\right)^2\right]\right], & l = a/2, \end{cases}$$

$$= \begin{cases} \frac{1}{Na}\left[\begin{array}{c} \sum_{j=1}^{a-l}\sum_{i=1}^{n}\left(x_{ij}^{(t)}x_{i,j-l}^{(t)} + c_{j,j-l,i}^{(t)}\right) \\ +\sum_{j=1}^{l}\sum_{i=1}^{n}\left(x_{ij}^{(t)}x_{i,j-a-l}^{(t)} + c_{j,j-a-l,i}^{(t)}\right) \end{array}\right] - \left(\mu_s^{(t)}\right)^2, & l < a/2, \\ \frac{2}{na}\sum_{j=1}^{l}\sum_{i=1}^{n}(x_{ij}^{(t)}x_{i,j-l}^{(t)} + c_{j,j-l,i}^{(t)}) - \left(\mu_s^{(t)}\right)^2. & l = a/2, \end{cases}$$

as the estimate of σ_l and

$$\sigma_{ps}^{(t)} = (na)^{-1} \sum_{i=1}^{n} p_i \sum_{j=1}^{a} x_{ij}^{(t)} - \hat{\mu}_p \mu_s^{(t)}$$

as that of σ_{ps}, all at iteration t. While other divisors within $\sigma_l^{(t)}$ and $\sigma_{ps}^{(t)}$ could be devised which attempt to correct for the biases which result from the losses in the degrees of freedom in the observed part of the data, the estimates we employ use the divisor na, as they are the substitutes for ML estimates in our adaptation of EM. No adjusting constants analogous to the $c_{jpi}^{(t)}$ in the $\sigma_l^{(t)}$ are needed in $\sigma_{ps}^{(t)}$ since none of the p_i are missing. Here $na\mu_s^{(t)} = \sum_{i,j} x_{ij}^{(t)}$ is the current estimate of $na\mu_s$, and is the proxy for the sufficient statistic $x_{\bullet\bullet}$ in the likelihood of the hypothetical complete data. Note that $(\hat{\mu}_p, \hat{\sigma}_{pp})$, the ML estimate of μ_p and the variance σ_{pp} of the parents' measurements, is available without the use of either E or M steps since none of the parents' measurements are missing. Once the $\sigma_l^{(t)}$ and $\sigma_{ps}^{(t)}$ are calculated, a new iteration may begin. Upon convergence, the final estimators are denoted by $\tilde{\sigma}_{pse}$ and $\tilde{\sigma}_{le}, l = 0, 1, ..., m$.

When no data are missing, all the $c_{jli}^{(t)}$ are zero and the sufficient statistics noted above are available from the data, requiring no estimation of their own. Even in this case, analytic expressions for the variances of the $\tilde{\sigma}_{le}$, like those of the LOO estimators, appear to be intractable. The expected values of the "EM" estimators, though, are easily found to be

$$\mathbf{E}(\tilde{\sigma}_{pse}) = \left(\frac{n-1}{n}\right)\sigma_{ps} \text{ and } \mathbf{E}(\tilde{\sigma}_{le}) = \sigma_l - \frac{\sum \sigma_j}{na},$$

for the case in which no observations are missing, the sum involving j being over any row of $\mathbf{C}$.

For each of the one hundred missing data patterns e sampled above in estimating the cumulants of $\tilde{\sigma}_{ps}$ and the $\tilde{\sigma}_l$ for each of the choices of combinations of $\mathbf{C}_{aug}$, p_m, n and a, an incomplete (having missing elements) dataset of $n + 1$'e observations, following the missing data pattern e, from a normal population having circular covariance matrix $\mathbf{C}_{aug}$ and zero mean was simulated. "EM" estimates were obtained for each of these datasets, stopping the iterations in each case when the sum of the absolute differences

$$\left|\sigma_{ps}^{(t)} - \sigma_{ps}^{(t-1)}\right| + \sum_{l=0}^{m} \left|\sigma_l^{(t)} - \sigma_l^{(t-1)}\right|$$

was less than .001, or when the number of iterations had reached one hundred. The sample mean errors and variances of the resulting one thousand estimates $\tilde{\sigma}_{pse}$ and $100(m+1)$ estimates $\tilde{\sigma}_{le}$ were calculated for each combination of C_{aug}, p_m, n and a. Table III lists these sample mean errors (sample biases).

Table III: Estimation of Biases of EM Estimators.

p_m	n, a	$\tilde{\sigma}_{pse}$	$\tilde{\sigma}_{0e}$	$\tilde{\sigma}_{1e}$	$\tilde{\sigma}_{2e}$	$\tilde{\sigma}_{3e}$
.1	5,4	.0065	-.4482	-.6944	-.6004	
.1	5,7	-.0829	-.3144	-.6072	-.5380	-.4353
.1	15,4	-.0601	-.1346	-.3468	-.1794	
.1	15,7	-.1313	-.0126	-.2965	-.2443	-.1739
.1	25,4	-.0534	-.0876	-.3703	-.2405	
.1	25,7	-.0363	-.0169	-.3067	-.2393	-.1704
.5	15,4	.0058	-.3184	-1.3307	-1.055	
.5	15,7	.0368	-.1570	-1.3471	-1.0804	-.7465
.5	25,4	.0531	-.0567	-1.2415	-.9859	
.5	25,7	.0733	.1597	-1.2657	-.9412	-.6502

Table IV compares the variances (mean squared errors) of the "EM" estimators with the estimated (sample mean) unconditional (on e) variances of the (unbiased) $\tilde{\sigma}_{ps}$ and $\tilde{\sigma}_l$. The tabled values are r =MSE($\tilde{\sigma}_{pse}$ or $\tilde{\sigma}_{le}$)/MSE($\tilde{\sigma}_{ps}$ or $\tilde{\sigma}_l$). The mean MSEs of the $\tilde{\sigma}_{ps}$ and $\tilde{\sigma}_l$, used in computing the values of r, can be considered to be close to the true variances of the $\tilde{\sigma}_{ps}$ and $\tilde{\sigma}_l$, because of the uniformly small standard errors of the variances randomizing over all possible **e**.

Table IV: Comparison of Mean Squared Errors of LOO Estimators with MSEs of EM Estimators.

p_m	n, a	ps	$l = 0$	$l = 1$	$l = 2$	$l = 3$
.1	5,4	.5535	.7942	.5789	.5776	
.1	5,7	.7812	.8422	.8165	.7674	.7378
.1	15,4	1.4291	.8134	.7574	.6572	
.1	15,7	.9411	.8420	.8547	.8415	.892
.1	25,4	1.0256	1.22	1.0944	.9646	
.1	25,7	.9984	1.1536	1.0351	1.0039	.9515
.5	15,4	.5938	.814	1.0871	.492	
.5	15,7	.7941	.9522	1.2654	.9372	.5486
.5	25,4	.7943	.803	1.5003	.6910	
.5	25,7	1.3432	1.295	1.8628	1.2624	.7269

Computer simulations showed that, for t-distributions of various degrees of freedom, the comparisons between the $(\tilde{\sigma}_{pse}, \tilde{\sigma}_{le})$and the $(\tilde{\sigma}_{ps}, \tilde{\sigma}_l)$ were substantially the same as those of Table IV. As a fixed point of reference for this table, in the normal, full data case with $(n, a) = (5, 4)$ and with the above parameter values, we have MSE$(\hat{\sigma}_{ps})$ /MSE$(\tilde{\sigma}_{ps}) = .6827$, which is fairly close to MSE$(\tilde{\sigma}_{pse})$ /MSE$(\tilde{\sigma}_{ps}) = .5535$ for $(n, a) = (5, 4)$ at $p_m = .1$ in Table IV.

Summarizing the comparisons of the $\tilde{\sigma}_{pse}$ and $\tilde{\sigma}_{le}$ with the $\tilde{\sigma}_{ps}$ and $\tilde{\sigma}_l$, the variances of the $\tilde{\sigma}_{le}$, but not the $\tilde{\sigma}_l$, can be expected to decrease with increases in $corr(p_i, x_{ij})$, because the $\tilde{\sigma}_{le}$ make use of the information about the missing data contained in p_i. In fact, if $|corr(p_i, x_{ij})|$ is close to 1. then the availability of p_i makes immaterial the potential missingness of the x_{ij}. Predictably, these improvements in efficiency are especially marked when the proportion of missing observations is large.

The $\tilde{\sigma}_{ps}$ and $\tilde{\sigma}_l$ have the additional slight advantage over $\tilde{\sigma}_{pse}$ and $\tilde{\sigma}_{le}$ of being calculated in a noniterative fashion. While calculating the "EM" estimates above, on two occasions convergence was not reached in the 100 iterations allowed. Furthermore, if the data are severely non-normal, the expectations of the "EM" estimators are not known and may be highly undesirable in some cases; however, computer simulations showed that the ratios of the MSEs of the two sets of estimators do not change significantly over a wide range (in terms of degrees of freedom) of t-distributions. On the other hand, $(\tilde{\sigma}_{ps}, \tilde{\sigma}_l)$ are unbiased for any distributional assumptions.

In the case of normally or t distributed data, the MSEs of the "EM" estimators are, on the whole, less than those of $(\tilde{\sigma}_{ps}, \tilde{\sigma}_l)$, especially when na or p_m is small. The advantage of unbiasedness in the latter apparently has been bought with a price in terms of efficiency.

If data is distributed normally or according to a t distribution, the MSE of $\tilde{\sigma}_{pse}$ seems to be smaller than that of $\tilde{\sigma}_{ps}$ for a wide range of parameter values, sample sizes and missing data patterns. Under these assumptions, $\tilde{\sigma}_{pse}$ will be preferred in applications unless unbiasedness is of much greater value than efficiency. Simulations in a future paper may show that the MSE of $\tilde{\sigma}_{ps}$ may be reduced when the divisor of $\tilde{\sigma}_{ps}$ is changed from $a_\bullet - \sum_i a_i^2/a_\bullet$ to a figure which more optimally balances bias with efficiency.

2.7 Analysis of Simulated Dataset. As an illustration, we apply the iterative ("EM") and non-iterative estimation procedures defined above to a dataset generated from a central multivariate t_4 distribution. Assume any family having no missing data values is composed of a parent and six siblings and has covariance

$$C_{aug} = \begin{bmatrix} 8 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 6 & 4.6 & 4.2 & 3.8 & 4.2 & 4.6 \\ 1 & 4.6 & 6 & 4.6 & 4.2 & 3.8 & 4.2 \\ 1 & 4.2 & 4.6 & 6 & 4.6 & 4.2 & 3.8 \\ 1 & 3.8 & 4.2 & 4.6 & 6 & 4.6 & 4.2 \\ 1 & 4.2 & 3.8 & 4.2 & 4.6 & 6 & 4.6 \\ 1 & 4.6 & 4.2 & 3.8 & 4.2 & 4.6 & 6 \end{bmatrix},$$

and let any sibling's score have probability .2 of being missing. The scores of ten such families were simulated, and certain siblings' observations were subsequently replaced with zeros according to the outcome of a stream of iid Bernoulli(.8) random variables y_{ij}; below, each row is a family's vector of observations, with the parent's score appearing in its first position:

$$\begin{bmatrix} -3.32 & 6.34 & 2.84 & .86 & 2.56 & 4.22 & 3.88 \\ -3.10 & -3.20 & -3.92 & -3.96 & -4.89 & -2.96 & -1.51 \\ 2.26 & .49 & .03 & .32 & 0 & 1.52 & -.64 \\ 2.31 & 2.20 & 1.18 & -.19 & .62 & 1.98 & -.72 \\ -2.24 & -2.96 & -1.09 & -1.06 & 0 & 0 & -2.84 \\ 2.10 & .16 & -.82 & -.16 & .15 & 1.64 & .59 \\ -1.11 & -.74 & 0 & -.19 & -1.43 & 0 & -1.60 \\ -1.55 & -5.15 & 0 & .05 & 0 & -3.20 & 0 \\ -.03 & .34 & -1.52 & -.91 & .66 & 1.45 & 0 \\ -1.83 & .15 & -.21 & -.97 & -1.99 & 0 & 0 \end{bmatrix}$$

The resultant unbiased estimators are (treating μ_p and μ_s as unknown)

$$(\tilde{\sigma}_{ps}, \tilde{\sigma}_0, \tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\sigma}_3) = (1.077, 5.6315, 4.7682, 4.4111, 4.0714).$$

"EM" iterations, on the other hand, produced the following estimators:

$$(\tilde{\sigma}_{pse}, \tilde{\sigma}_{0e}, \tilde{\sigma}_{1e}, \tilde{\sigma}_{2e}, \tilde{\sigma}_{3e}) = (.9549, 5.2885, 3.3421, 3.0135, 2.9438).$$

With the exception of $\tilde{\sigma}_{pse}$, all the "EM" estimators have negative errors (reflecting their negative biases) which are larger, in absolute value, than those of the LOO estimators. Nonetheless, because of the moderate sample size, t distribution, moderate p_m and availability of covariates (parents' scores), the $(\tilde{\sigma}_{pse}, \tilde{\sigma}_{le})$ can in the long run be expected to perform better than the $(\tilde{\sigma}_{ps}, \tilde{\sigma}_{l})$ in minimizing mean squared error, if more datasets are generated from this distribution.

3 AUTOREGRESSIVE CIRCULAR COVARIANCE

3.1 Introduction. In this section we introduce and develop a highly parsimonious case of circular covariance which, when its assumptions adequately characterize the data, has important advantages over general (unrestricted) circular covariance. All the covariance terms in this model depend on only two underlying parameters $\rho \in (-1, 1)$ and $\sigma^2 > 0$. It possesses the following structure within the i^{th} family:

$$Cov(x_{ij}, x_{il}) = \begin{cases} \sigma^2 \rho^{|j-l|}, & |j-l| \le a/2, \\ \sigma^2 \rho^{a-|j-l|}, & |j-l| > a/2. \end{cases} \qquad (13)$$

We may refer to this phenomenon as "autoregressive" circular covariance, as it is the circular analog to the covariance of the AR(1) time series models. As with all circular covariance structures, the middle terms differ depending on whether a is even or odd, as displayed below for $a = 4$ and $a = 5$:

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho & \rho^2 & \rho & 1 \end{bmatrix}, \quad \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^2 & \rho \\ \rho & 1 & \rho & \rho^2 & \rho^2 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho^2 & \rho & 1 & \rho \\ \rho & \rho^2 & \rho^2 & \rho & 1 \end{bmatrix}.$$

If the autoregresssive circular covariance assumption is a valid representation of the data in an application, the advantage of modelling only two parameters, instead of the many required given general circular covariance, can be significant. This advantage is especially great if n/a is small, indicating many parameters are to be estimated by observations on only a few families. In fact, wherever circular covariance is used to model data, we advocate that autoregressive circular covariance be considered as one possible structure.

As observed in Section 1, Olkin and Press (1969) examined a less restrictive model involving normally distributed data with circular covariance wherein the first row of $\mathbf{C}$ is

$$\begin{aligned} \sigma^2(1, \rho_1, \rho_2, \ldots, \rho_{m-1}, \rho_m, \rho_{m-1}, \ldots, \rho_1), \ a &= 2m, \text{ or} \qquad (14) \\ \sigma^2(1, \rho_1, \rho_2, \ldots, \rho_{m-1}, \rho_m, \rho_m, \rho_{m-1}, \ldots, \rho_1), \ a &= 2m+1. \end{aligned}$$

with $(\rho_1, \rho_2, ..., \rho_m)$ being independent parameters, subject only to the restriction C>0. In the case we are now considering, $\rho_j = \rho^j, j = 1, 2, ..., m$, so that

$$\delta_j = \sigma^2(1, \rho, \rho^2, ..., \rho^m)B_j, \; j = 1, 2, ..., m+1 \tag{15}$$

are the first $m+1$ eigenvalues of $\mathbf{C}$, and the last $a-m-1$ eigenvalues of $\mathbf{C}$ are (again) obtained using $\delta_j = \delta_{a-j+2}, \; j = m+2, m+3, ..., a$.

3.2 Maximum Likelihood Estimation: Equal Numbers of Measurements within Families. Let $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ia})' \sim N_a(\mu\mathbf{1}, \mathbf{C})$, $i = 1, 2, ..., n$ and $\mathbf{X}' = (\mathbf{x}'_1, \mathbf{x}'_2, ..., \mathbf{x}'_n)$ where $\mathbf{1}$, as in the previous section, is a vector of unities of the appropriate length, and $\mathbf{C}$ has elements defined by (13). Define $\mathbf{\Gamma}$ as the $a \times a$ symmetric orthogonal matrix having elements

$$\gamma_{jl} = a^{-1/2}\left\{\cos\left[\frac{2\pi(j-1)(l-1)}{a}\right] + \sin\left[\frac{2\pi(j-1)(l-1)}{a}\right]\right\}. \tag{16}$$

Then the rows (columns) of $\mathbf{\Gamma}$ are the eigenvectors of $\mathbf{C}$, so that $\mathbf{C} = \mathbf{\Gamma} diag(\delta_1, ..., \delta_a)\mathbf{\Gamma}$ (Basilevsky, 1983). Also, create the canonical variables

$$\begin{aligned} \mathbf{y}_i &= (y_{i1}, y_{i2}, ..., y_{ia})' = \mathbf{\Gamma}\mathbf{x}_i, \; i = 1, 2, ..., n; \text{ then} \\ \mathbf{y}_i &\sim N_a(\boldsymbol{\mu}, diag(\delta_1, ..., \delta_a)) = N_a(\boldsymbol{\mu}, \sigma^2 diag(d_1, ..., d_a)), \end{aligned}$$

where the d_j are functions of ρ only, not depending on σ^2, that is, $\delta_j = \sigma^2 d_j$, and (noting all rows of $\mathbf{\Gamma}$ but the first are orthogonal to $\mathbf{1}$) $\boldsymbol{\mu}' = (\sqrt{a}\mu, 0, 0, ..., 0)$. Here

$$diag(\delta_1, ..., \delta_a) = \begin{cases} diag\,(\delta_1, \delta_2, ..., \delta_{m-1}, \delta_m, \delta_{m-1}, ..., \delta_2)\,, & a = 2m, \\ diag\,(\delta_1, \delta_2, ..., \delta_{m+1}, \delta_{m+1}, \delta_m, \delta_{m-1}, ..., \delta_2)\,, & a = 2m+1. \end{cases}$$

Let $\mathbf{Y}' = (\mathbf{y}'_1, \mathbf{y}'_2, ..., \mathbf{y}'_n)$. Then $\mathbf{Y}$ is an orthogonal (one-to-one) transformation of $\mathbf{X}$, and consequently contains all the information in $\mathbf{X}$ about $(\mu, \boldsymbol{\delta}')$ and therefore about (σ^2, ρ). Furthermore, the MLE of these parameters expressed in $\mathbf{Y}$ is equivalent to that expressed in $\mathbf{X}$.

Olkin and Press (1969) obtained the MLE of $(\boldsymbol{\delta}', \sigma^2, \rho'_r)$ in terms of $\mathbf{Y}$ for their (less restrictive) model. We modify their approach first, as did Khattree and Naik (1994a), to accommodate our requirement that the mean vector be estimated from $\{\boldsymbol{\mu} = \mu\mathbf{1} : \mu \in \Re\}$: that is, while they estimate the mean of the j^{th} measurement in each family by its MLE $\hat{\mu}_j = \bar{x}_{\bullet j}$, we estimate the overall mean μ by its MLE $\hat{\mu} = \bar{x}_{\bullet\bullet}$. Additionally, while they calculate the matrix of sums of squares and cross products as S=(s_{jl}), with $s_{jl} = \sum_{i=1}^{n}(x_{ij} - \hat{\mu}_j)(x_{il} - \hat{\mu}_l)$, we let $s_{jl} = \sum_{i=1}^{n}(x_{ij} - \hat{\mu})(x_{il} - \hat{\mu})$. No further modifications to the argument of Olkin and Press are necessary to obtain $\hat{\boldsymbol{\delta}}$ in our situation. The MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{na}\sum_{i=1}^{n}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})'\left[diag(d_1, d_2, ..., d_a)\right]^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}), \tag{17}$$

solving the score equation in σ^2 while making use of the fact that $\sum_{j=1}^{a} d_j = \sigma^{-2}tr\,(\mathbf{C}) = a$. Until estimates of the d_j are available, from $\hat{\boldsymbol{\delta}}$ and using the fact that $\boldsymbol{\rho} = \left(1, \rho, \rho^2, ..., \rho^m\right)' = \boldsymbol{\sigma}^{-2}\left(\mathbf{B}^{-1}\right)'\boldsymbol{\delta}$, an initial estimate of σ^2 can be calculated as $\left(\hat{\sigma}^2\right)^{(1)} = \mathbf{b}^1\hat{\boldsymbol{\delta}}$, where $\mathbf{b}^j$ is the j^{th} row of $\left(\mathbf{B}^{-1}\right)'$.

To find the MLE of the autoregressive parameter ρ using the canonical variables $\mathbf{Y}$, let $\boldsymbol{\eta}_i = \mathbf{y}_i - \boldsymbol{\mu}$ have elements $\eta_{ij}, j = 1, 2, ..., a$ and note that the first derivative with respect to ρ of the loglikelihood is

$$\frac{\partial \ln p\,(\mathbf{y})}{\partial \rho} = \frac{1}{2}\sum_j \frac{\partial \delta_j}{\partial \rho}\left(\frac{\sum_i \eta_{ij}^2}{\delta_j^2} - \frac{n}{\delta_j}\right).$$

Solving $\frac{\partial \ln p(\mathbf{y})}{\partial \rho} = 0$ iteratively can be accomplished using either the Newton Raphson (NR) algorithm $\hat{\rho}^{(t-1)} = \hat{\rho}^{(t)} - \left(H_{\rho\rho}\right)^{-1}S(\rho)$ or the Fisher Scoring (FS) algorithm $\hat{\rho}^{(t-1)} = \hat{\rho}^{(t)} + \left(I(\rho)\right)^{-1}S(\rho)$ where $H_{\rho\rho}$ is the second derivative $\frac{\partial^2 \ln p(\mathbf{y})}{\partial \rho^2}$, $I(\rho) = -\mathbf{E}\left(H_{\rho\rho}\right)$ is the information about ρ, and $S(\rho) = \frac{\partial \ln p(\mathbf{y})}{\partial \rho}$ is the score of ρ. To implement both of these methods, we calculate

$$\frac{\partial^2 \ln p\,(\mathbf{y})}{\partial \rho^2} = \frac{1}{2}\sum_j\left[\frac{\partial^2 \delta_j}{\partial \rho^2}\left(\frac{\sum_i \eta_{ij}^2}{\delta_j^2} - \frac{n}{\delta_j}\right) + \left(\frac{\partial \delta_j}{\partial \rho}\right)^2\left(\frac{n}{\delta_j^2} - \frac{2\sum_i \eta_{ij}^2}{\delta_j^3}\right)\right],$$

which has expected value equal to the negative value of the information about ρ :

$$I(\rho) = -\mathbf{E}\left(\frac{\partial^2 l}{\partial \rho^2}\right) = \frac{n}{2}\sum_j\left[\left(\frac{\partial \delta_j}{\partial \rho}\right)^2 \delta_j^{-2}\right]. \tag{18}$$

Recalling that $\delta_j = \sigma^2(1, \rho, \rho^2, ..., \rho^m)B_j$, and that $\mathbf{B}$ does not depend on ρ,

$$\frac{\partial \delta_j}{\partial \rho} = \sigma^2(0, 1, 2\rho, 3\rho^2, ..., m\rho^{m-1})B_j \text{ and } \frac{\partial^2 \delta_j}{\partial \rho^2} = \sigma^2(0, 0, 2, 6\rho, ..., m\,(m-1)\,\rho^{m-2})B_j.$$

Substituting these first and second derivatives into $H_{\rho\rho}$, $I(\rho)$ and $S(\rho)$, using the current estimate of ρ in place of ρ, together complete the elements of the equations required for the NR and FS iterations. For either of these methods, a new estimator of σ^2 is calculated using (17) at each iteration.

Simulating one hundred datasets from normal distributions with autoregressive circular covariance, for each of the several combinations of n, a and the covariance parameters ρ and σ^2 shown in the comparison of $\hat{\rho}$ with the alternate estimator $\bar{\rho}$ of ρ below, it was found that NR and FS produced essentially the same results for $\hat{\rho}$, although FS usually converged (to the same convergence criterion) in fewer iterations. While multiple roots of $\frac{\partial \ln p(y)}{\partial(\rho,\sigma^2)} = 0$ may exist, both NR and FS iterations seemed to converge, in simulations, to the true values (ρ, σ^2).

3.3 Selection of Appropriate Covariance Structure. Consider the following (nested) hypotheses:

$$H_I \;:\; Cov(\mathbf{x}_i) = \sigma^2\mathbf{I} \text{ (sphericity)},$$

$$H_\rho \;:\; Cov(x_{ij}, x_{il}) = \begin{cases} \sigma^2\rho^{|j-l|}, & |j-l| \le a/2, \\ \sigma^2\rho^{a-|j-l|}, & |j-l| > a/2, \end{cases} \quad \text{(autoregressive circular covariance)},$$

$$H_c \;:\; Cov(x_{ij}, x_{il}) = \begin{cases} \sigma_{|j-l|}, & |j-l| \le a/2, \\ \sigma_{a-|j-l|}, & |j-l| > a/2. \end{cases} \quad \text{("unrestricted" circular covariance)}.$$

Although the σ_j, given H_c, must meet the restriction C>0, we may call H_c the hypothesis of unrestricted circular covariance, to distinguish it from H_ρ.

Olkin and Press (1969) give the likelihood ratio test (LRT) (Khattree and Naik, 1995) of H_I versus H_c. Our acquisition of the MLE $\hat{\rho}$ in the last subsection permits the use of the LRTs of H_I versus H_ρ and H_ρ versus H_c. Whether H_I, H_ρ or H_c describes the covariance structure, if δ_j are the eigenvalues of the appropriate covariance matrix, the loglikelihood is

$$-\frac{n}{2}\sum_j \ln\delta_j - \frac{1}{2}\sum_i\sum_j \frac{\eta_{ij}^2}{\delta_j}.$$

Under H_I, $\delta_j = \sigma^2$ for all j, so that the covariance structure depends on a single parameter, and $\hat{\sigma}^2 = \frac{1}{Na}(\mathbf{X}-\hat{\mu}\mathbf{1})'(\mathbf{X}-\hat{\mu}\mathbf{1})$ completes ML estimation of all covariance parameters, in which case the MLE of $\ln L$ simplifies to $-\frac{Na}{2}\left[\ln\left(\hat{\sigma}^2\right)+1\right]$. Assuming H_ρ, the $1\times(m+1)$-vector of distinct eigenvalues is given by $\boldsymbol{\delta}' = \sigma^2(1,\rho,\rho^2,...,\rho^m)\mathbf{B}$ and depends on two independent parameters σ^2 and ρ. In this case the MLE of $\ln L$ simplifies to $-\frac{n}{2}\left[\sum_j \ln\hat{\delta}_j + a\right]$. Next, H_c implies that $\delta' = (\sigma_0, \sigma_1, \sigma_2, ..., \sigma_m)\mathbf{B}$ depends on the $m+1$ independent parameters $\sigma_0, \sigma_1, \sigma_2, ..., \sigma_m$, and the MLE of $\ln L$ is again $-\frac{n}{2}\left[\sum_j \ln \hat{\delta}_j + a\right]$ with the sole difference from H_ρ that the $\hat{\delta}_j$ are derived differently.

Again letting B_j be the j^{th} column of B, the MLE of $-2\ln(L_\rho/L_c)$, which has an approximate chi-squared distribution with $(m+1)-(2) = m-1$ degrees of freedom when H_ρ is true, can be expressed as

$$-2\ln\lambda_{\rho c} = -2\left[\ln\hat{L}_\rho - \ln\hat{L}_c\right] = n\sum_{j=1}^{m-1}\alpha_j \ln\frac{\hat{\sigma}^2(1,\hat{\rho},\hat{\rho}^2,...,\hat{\rho}^m)B_j}{(\hat{\sigma}_0,\hat{\sigma}_1,\hat{\sigma}_2,...,\hat{\sigma}_m)B_j}. \quad (19)$$

The LRT statistic for H_I versus H_ρ is similar; each of these statistics can be interpreted as the log of the ratio of the sample generalized variances under the two hypotheses concerned.

For each of the two nestings H_I versus H_ρ (which actually tests $\rho = 0$) and H_ρ versus H_c, the sizes of the LRT under the combinations of a, n and ρ listed below were estimated by generating one thousand normal datasets from distributions under the null hypothesis, and calculating $-2\ln\lambda$ for each simulation. This test statistic fell in the critical region $\left\{-2\ln\lambda : -2\ln\lambda > \chi^2_{d,.05}\right\}$ (causing type I error; here d is the difference in the number of parameters estimated) on the proportions of occasions listed in Table V. In each case, the proportion estimates the size of the test. The variance σ^2 is everywhere fixed at 1:

Table V: Estimated Sizes (Rejection Proportions) of LRTs of Sphericity versus Autoregressive Circular Covariance, and Autoregressive versus Unrestricted Circular Covariance.

H_I versus H_ρ			H_ρ versus H_c			
n	a	.	n	a	$\rho = -.4$	$\rho = .4$
10	4	.064	10	4	.073	.062
10	7	.058	10	7	.056	.066
30	4	.057	30	4	.052	.064
30	7	.052	30	7	.048	.061

It seems the size of each test is a small amount larger, on average, than the desired 5%, as long as na is moderately large.

To examine the power of the LRT of H_I versus H_ρ under various nonzero values of ρ, one thousand normal datasets having autoregressive circular covariance were simulated for each of several different combinations of n, a and ρ while fixing $\sigma^2 = 1$. The proportions of rejections (estimating the power) were tabulated in each case in Table VI:

Table VI: Estimated Power (Rejection Proportion) of LRT of Sphericity versus Autogregressive Circular Covariance.

n	a	$\rho = -.4$	$\rho = -.2$	$\rho = .2$	$\rho = .4$
10	4	.784	.269	.171	.675
10	7	.923	.428	.335	.894
30	4	.997	.617	.582	.994
30	7	1.00	.843	.804	1.00

As is expected and desired, the LRT seems to have greatest power when $|\rho|$ is large.

To examine the power of the LRT of H_ρ versus H_c under various cases of H_c, it is necessary, when choosing a non-autoregressive structure, to develop some notion of the relation between the covariance parameters and the eigenvalues $\boldsymbol{\delta} = \mathbf{B}'\boldsymbol{\sigma}$ of the covariance matrix $\mathbf{C}$. Obviously, this relationship is important because there is a need to keep all $\delta_j > 0$. However, it is also important to have some idea of the somewhat surprising effects the choices of the parameters $\boldsymbol{\sigma}' = (\sigma_0, \sigma_1, ..., \sigma_m)$ have on the power of the LRT, as this power depends on the extent to which the eigenvalues differ from those of an autoregressive $\mathbf{C}$.

As shown by (19), the distribution of the LRT statistic depends on $\boldsymbol{\sigma}$ through $\boldsymbol{\delta}$; hence, the power depends on the distributions of the MLEs of $\boldsymbol{\delta}$ calculated assuming the null and alternative hypotheses. Because the $\hat{\delta}_j$ appear in the denominator in the log in $-2\ln\lambda$, the variance of this statistic can be expected to increase dramatically when one or more of the expected values of the $\hat{\delta}_j$ approach zero. This will certainly occur when any of the true δ_j approach zero, and it is plausible to suggest that, if the alternative hypothesis H_a is true, one or more of the $\mathrm{E}(\hat{\delta}_j)$, calculated under the null hypothesis H_0, will approach zero even when the true δ_j do not do so. Because $\ln(L_1/L_0) > 0$ when H_0 is nested within H_a, this increased variance implies an increased frequency of rejection of the LRT, and hence a higher power of this test.

As measured by the LRT, therefore, circular covariance structures are "different" in terms of the differences between their eigenvalues, not so much in terms of differences between their covariance parameters $\boldsymbol{\sigma}$. This fact is demonstrated empirically in the tables below, where it is seen that small deviations from autoregressive circular covariance derived by $\sigma^2(1, \rho \pm \varepsilon_1, \rho^2 \pm \varepsilon_2, \ldots)$, where the ε_i are small numbers, often result in high power of the LRT.

The importance of the distinction between the effects of deviations imputed to the σ_j and those of deviations imputed to the δ_j becomes evident upon consideration of some examples. For instance, the $\mathbf{C}$ associated with the choice $a = 4$ and $(\sigma_0, \sigma_1, \sigma_2) = (1, .75, .50)$ is singular. The $\mathbf{C}$ associated with $(\sigma_0, \sigma_1, \sigma_2) = (1, .75, .675)$ and $a = 4$ is nearly so, as its smallest eigenvalue is 0.7, only 6% of the largest eigenvalue. These two facts are surprising because it would seem that both sets of covariance parameters are quite "close", for example, to the $(\sigma_0, \sigma_1, \sigma_2) = (1, .75, .5625)$ which appears in autoregressive circular covariance with $(\sigma^2, \rho) = (1, .75)$. For $a = 7$, the choice $(\sigma_0, \sigma_1, \sigma_2, \sigma_3) = (1, .75, .675, .50)$ also produces an extremely small eigenvalue (when compared with the largest eigenvalue), and is thus shown to be very "different" from autoregressive circular covariance with $(\sigma^2, \rho) = (4, .75)$, which possesses covariance parameters $(\sigma_0, \sigma_1, \sigma_2, \sigma_3) = (1, .75, .5625, .4225)$. Furthermore, data arising from non-autoregressive circular covariance with these parameters are actually extremely different from that arising from autoregressive circular covariance, as is shown below by the high power of the LRT under these alternatives.

Consequent to the need for $\mathbf{C}$ to have $\delta_j > 0$ which are more or less evenly proportioned, we proceed as follows in constructing non-autoregressive circular covariance structures. To investigate the power of the LRT when the circular covariance parameters $\boldsymbol{\sigma}$ are highly non-autoregressive, we assume that the covariance between any two observations increases with the lag between them, and choose $(\sigma_0, \sigma_1, \sigma_2, \sigma_3) = (1, .25, .375, .50)$, which results in a circular covariance matrix with relatively evenly spaced eigenvalues $(\delta_1, \delta_2, \delta_3) = (1.875, .625, .875)$ if $a = 4$ and $(\delta_1, \delta_2, \delta_3, \delta_4) = (3.25, .2439, .8366, .7947)$ if $a = 7$.

To construct non-autoregressive circular covariance matrices which generate data more closely resembling that which arise from autoregressivity, in each case we start by choosing a ρ which would seem reasonable as an autoregressive circular covariance parameter, find the eigenvalues of the resultant covariance matrix, and adjust one or more of these eigenvalues slightly. Pre- and post-multiplying the eigenvalues, in diagonal form, by the eigenvectors contained in $\mathbf{\Gamma}$ produces a new, positive definite matrix if all the adjusted eigenvalues are positive. Assuring that the new matrix is not itself an autoregressive circular covariance matrix is then accomplished by examining the ratios between the elements of its first row.

Assuming autoregressivity and setting $(\sigma^2, \rho) = (1, .4)$ implies

$$\begin{aligned} \boldsymbol{\delta}' &= (1, .4, .16)\mathbf{B} = (1.96, .84, .36) \text{ when } a = 4, \text{ and} \\ \boldsymbol{\delta}' &= (1, .4, .16, .064)\mathbf{B} = (2.248, 1.3123, .6135, .4502) \text{ when } a = 7, \end{aligned}$$

in each case using the $\mathbf{B}$ matrix applicable to the dimension a. We adjust one or more of the smallest of these eigenvalues upwards for each a, first by a small amount and then by a larger amount. For $a = 4$, adjusting δ_3 to .5 implies $\boldsymbol{\sigma}' = (1.035, .365, .195)$, and adjusting it to .8 implies $\boldsymbol{\sigma}' = (1.11, .29, .27)$; neither of these $\boldsymbol{\sigma}$ satisfy the autoregressive assumption, although $|\sigma_l| > |\sigma_{l-1}|$ for each l and each adjustment. Dividing each of these $\boldsymbol{\sigma}'$ by its σ_0 normalizes them to $(1, .3527, .1884)$ and $(1, .2613, .2432)$. For $a = 7$,

$$\begin{aligned} \text{adjusting } \delta_4 \text{ to .6 implies } \boldsymbol{\sigma}' &= (1.0428, .3615, .1867, .0545), \\ \text{and adjusting } (\delta_3, \delta_4) \text{ to } (.9, .8) \text{ implies } \boldsymbol{\sigma}' &= (1.1818, .2918, .1485, .0928); \end{aligned}$$

both of these $\boldsymbol{\sigma}'$ approximate, though do not replicate, $\sigma^2(1, \rho, \rho^2, \rho^3)$ for appropriate (σ^2, ρ). Dividing each of these σ' by its σ_0 normalizes them to $(1, .3467, .179, .0523)$ and $(1, .2469, .1257, .0785)$.

One thousand datasets were simulated for each of several combinations of (n, a), all with normal, non-autoregressive circular covariance having the parameters σ listed below, which were derived using the adjustments described above. The LRT statistic $-2\ln\lambda_{pc} = -2\ln\frac{L_p}{L_c}$ fell into the critical region on the proportions of instances listed in Table VII, in each case estimating the power of the LRT.

Table VII: Estimated Power (Rejection Proportion) of LRT of Autoregressive versus Unrestricted Circular Covariance.

(n, a)	σ'	Rejection Proportion
(10,4)	(1,.75,.675)	.354
(10,7)	(1,.75,.675,.50)	.769
(30,4)	(1,.75,.675)	.675
(30,7)	(1,.75,.675,.50)	1.00
(10,4)	(1,.25,.375)	.257
(10,7)	(1,.25,.375,.50)	.981
(30,4)	(1,.25,.375)	.749
(30,7)	(1,.25,.375,.50)	1.00
(10,4)	(1,.2613,.2432)	.085
(10,7)	(1,.2469,.1257,.0785)	.066
(30,4)	(1,.2613,.2432)	.294
(30,7)	(1,.2469,.1257,.0785)	.098
(10,4)	(1,.3527,.1884)	.066
(10,7)	(1,.3467,.179,.0523)	.080
(30,4)	(1,.3527,.1884)	.077
(30,7)	(1,.3467,.179,.0523)	.152

As expected, the power of the test depends on the amount of perturbation away from autoregressivity.

Another means of testing for covariance structure is that provided by Akaike (1973a, 1973b). See Jones (1994) for a description of this method. Akaike's Information Criterion, AIC$=-2\ln L+2d$, where d is the number of estimated parameters, is calculated for each proposed model. The decision rule then chooses the model yielding the lowest AIC. We here use the criterion to choose between the three alternatives for covariance considered above: spherical, autoregressive circular, and unrestricted circular. It is known (Jones, 1994) that AIC is not a consistent estimator of the order of an autoregression. Therefore, an argument could be made in the present context in favor of the modified AIC due to Schwarz (1978), SC$=-2\ln L+d\ln n$ (where n is the total number of units) over the original AIC. However, since switching between decision rules based on $-2\ln L$ simply redistributes the probabilities of choosing among the possible structures, the election of one of these rules over another actually must depend on the cost of misclassification associated with each wrong decision. We present results using the original AIC, as the proportions of correct decisions using this criterion seem to be approximately equal for the three covariance structures.

Given $m = m(a)$, the AIC statistics for the three covariance structures under consideration are

$$-2\ln L + 2(m+1) \quad : \quad \text{Unrestricted Circular,}$$
$$-2\ln L + 2(2) \quad : \quad \text{Autoregressive Circular,}$$
$$-2\ln L + 2(1) \quad : \quad \text{Spherical,}$$

ignoring for the moment the number of estimated mean parameters.

First, one thousand datasets were simulated for each of several combinations of (n, a, ρ), and the proportions of correct choices based on AIC among the three alternatives were tabulated in Table VIII. Note that if $\rho = 0$, a correct decision is one in favor of spherical covariance.

Table VIII: Estimated Accuracy Rate (Proportion of Correct Decisions) of AIC under Sphericity and Autoregressive Circular Covariance.				
n	a	$\rho = -.4$	$\rho = 0$	$\rho = .4$
10	4	.743	.776	.671
10	7	.819	.802	.799
30	4	.834	.784	.841
30	7	.833	.805	.872

Next, one thousand datasets were simulated for each of several combinations of (n, a), all with normal, non-autoregressive circular covariance having parameters $\boldsymbol{\sigma}' = (1, .25, .375, .50)$, $(1, .2613, .2432)$ or $(1, .2469, .1257, .0785)$, as applicable. The proportions of correct decisions based on AIC among the three alternatives for covariance structure are listed in Table IX:

Table IX: Estimated Accuracy Rate (Proportion of Correct Decisions) of AIC under Unrestricted Circular Covariance.			
n	a	$\boldsymbol{\sigma}'$	Proportion
10	4	$(1, .25, .375)$	.380
10	7	$(1, .25, .375, .50)$	.976
30	4	$(1, .25, .375)$	.835
30	7	$(1, .25, .375, .50)$	1.00
10	4	$(1, .2613, .2432)$	.208
10	7	$(1, .2469, .1257, .0785)$	.125
30	4	$(1, .2613, .2432)$	.505
30	7	$(1, .2469, .1257, .0785)$	.234

Again, correct decisions are less frequent when the unrestricted circular covariance closely resembles an autoregressive circular covariance structure (the last four cases in the table). The low accuracy displayed by this criterion for $a = 4$, even when the unrestricted circular covariance is highly non-autoregressive, is perhaps due in large part to the fact that, with this dimension (as well as with $a = 5$), just one parameter estimate, $\hat{\sigma}_2$, is useful in differentiating between autoregressive and unrestricted circular covariance. In fact, if $a < 4$, all possible circular covariance structures are also at the same time both autoregressive circular and compound symmetry structures. The lower power of the LRT seen above in differentiating between autoregressive and unrestricted circular covariance for $a = 4$ can probably be explained in this way, as well.

3.4 An Alternative Estimator of the Autoregressive Parameter. By the properties of efficient likelihood estimation assuming any of the regular families of distributions, the MLE $\hat{\rho}$ is guaranteed to be asymptotically efficient and consistent, having variance in the limit $1/I(\rho)$. However, for small $Na\rho$, another estimator $\tilde{\rho}$ has very favorable qualities compared with $\hat{\rho}$, which we develop in what follows.

Given the MLE $\hat{\boldsymbol{\delta}}$, estimators of m powers of ρ are available as $\hat{\rho}^j = \hat{\sigma}^{-2}\mathbf{b}^{j-1}\hat{\boldsymbol{\delta}}, j = 1, 2, ..., m$, where $\mathbf{b}^j$ is the j^{th} row of $(\mathbf{B}^{-1})'$. Since for $j = 1, 2..., m$, $\mathbf{b}^{j-1}\boldsymbol{\delta}/\mathbf{b}^j\boldsymbol{\delta} = \rho$, we have m estimators of ρ as $\tilde{\rho}_j = \mathbf{b}^{j-1}\hat{\boldsymbol{\delta}}/\mathbf{b}^j\hat{\boldsymbol{\delta}}, j = 1, 2, ..., m$. The final estimate $\tilde{\rho}$ will be found as a linear combination of the $\tilde{\rho}_j$.

Because of the efficient likelihood estimation properties of $\hat{\boldsymbol{\delta}}$ (Lehmann, 1983), it is known that

$$\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \rightarrow N_{m-1}(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\delta})),$$

where

$$\mathbf{I}(\boldsymbol{\delta}) = \frac{n}{2} diag\left(\frac{\alpha_1}{\delta_1^2}, \frac{\alpha_2}{\delta_2^2}, ..., \frac{\alpha_{m-1}}{\delta_{m-1}^2}\right)$$

is the Fisher information matrix of $\boldsymbol{\delta}$.

Suppose

$$\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_{m-1})' = (\mathbf{b}^1\boldsymbol{\delta}, \mathbf{b}^2\boldsymbol{\delta}, ..., \mathbf{b}^{m-1}\boldsymbol{\delta})' = (\mathbf{B}^{-1})'\boldsymbol{\delta}$$

and

$$\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_{m-1}\right)' = (\mathbf{b}^1\hat{\boldsymbol{\delta}}, \mathbf{b}^2\hat{\boldsymbol{\delta}}, ..., \mathbf{b}^{m-1}\hat{\boldsymbol{\delta}})' = (\mathbf{B}^{-1})'\hat{\boldsymbol{\delta}};$$

then,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sqrt{n}\left((\mathbf{B}^{-1})'\hat{\boldsymbol{\delta}} - (\mathbf{B}^{-1})'\boldsymbol{\delta}\right) \longrightarrow N_{m-1}\left(0, (\mathbf{B}^{-1})'\mathbf{I}^{-1}(\boldsymbol{\delta})\mathbf{B}^{-1}\right).$$

If we denote

$$\begin{aligned}
g_i(\boldsymbol{\theta}) &= g_i = \frac{\theta_{i+1}}{\theta_i} = \frac{\mathbf{b}^{i+1}\boldsymbol{\delta}}{\mathbf{b}^i\boldsymbol{\delta}} = \rho, \\
g_i(\hat{\boldsymbol{\theta}}) &= \hat{g}_i = \frac{\hat{\theta}_{i+1}}{\hat{\theta}_i} = \frac{\mathbf{b}^{i+1}\hat{\boldsymbol{\delta}}}{\mathbf{b}^i\hat{\boldsymbol{\delta}}} = \hat{\rho}_i, \\
\mathbf{g}(\boldsymbol{\theta}) &= (g_1, g_2, ..., g_m)' = \mathbf{g} = \rho\mathbf{1}_m, \text{ and} \\
\mathbf{g}(\hat{\boldsymbol{\theta}}) &= \hat{\mathbf{g}} = (\hat{g}_1, \hat{g}_2, ..., \hat{g}_m)' = (\hat{\rho}_1, \hat{\rho}_2, ..., \hat{\rho}_m)',
\end{aligned}$$

then by the Delta Theorem (Lehmann, 1983) we have

$$\sqrt{n}(\hat{\mathbf{g}} - \mathbf{g}) = \sqrt{n}(\hat{\mathbf{g}} - \mathbf{1}\rho) \rightarrow N_m(0, \mathbf{D}_g(\mathbf{B}^{-1})'\mathbf{I}^{-1}(\boldsymbol{\delta})\mathbf{B}^{-1}\mathbf{D}_g'),$$

where $\mathbf{D}_g$ is the $m \times (m+1)$ derivative of $\mathbf{g}$ having elements

$$(\mathbf{D}_g)_{ij} = \begin{cases} \sigma^{-2}\rho^{1-i}, \; j = i+1, \\ -\sigma^{-2}\rho^{2-i}, \; j = i, \\ 0, \; i+1 \neq j \neq i. \end{cases}$$

Now it can easily be shown, using a version of the extended Cauchy-Schwartz inequality (Johnson and Wichern, 1992, p. 66), that the final estimate $\tilde{\rho}$ of ρ which is optimal in the sense of having minimum asymptotic variance among all consistent linear estimators based on the elements of $\hat{\mathbf{g}}$, is

$$\tilde{\rho} = \frac{\mathbf{1}'\left(\mathbf{D}_g(\mathbf{B}^{-1})'\mathbf{I}^{-1}(\boldsymbol{\delta})\mathbf{B}^{-1}\mathbf{D}_g'\right)^{-1}}{\mathbf{1}'\left(\mathbf{D}_g(\mathbf{B}^{-1})'\mathbf{I}^{-1}(\boldsymbol{\delta})\mathbf{B}^{-1}\mathbf{D}_g'\right)^{-1}\mathbf{1}}\hat{\mathbf{g}}.$$

$\tilde{\rho}$ has asymptotic variance $\left(\mathbf{1}'\left(\mathbf{D}_g(\mathbf{B}^{-1})'\mathbf{I}^{-1}(\boldsymbol{\delta})\mathbf{B}^{-1}\mathbf{D}_g'\right)^{-1}\mathbf{1}\right)^{-1}$ and the $(i,j)^{th}$ element of $\mathbf{D}_g(\mathbf{B}^{-1})'\mathbf{I}^{-1}(\boldsymbol{\delta})\mathbf{B}^{-1}\mathbf{D}_g'$ is, after much simplification,

$$\left(\mathbf{D}_g(\mathbf{B}^{-1})'\mathbf{I}^{-1}(\boldsymbol{\delta})\mathbf{B}^{-1}\mathbf{D}_g'\right)_{ij} = 2N^{-1}\sigma^{-4}\rho^{2-i-j}\sum_{l=1}^{m-1}(b^{i-1,l} - \rho b^{il})\frac{\delta_l^2}{\alpha_l}(b^{j-1,l} - \rho b^{jl}), \quad (20)$$

where b^{ij} is the (ij) element of $(\mathbf{B}^{-1})'$.

In practical situations, the MLE $\left(\hat{\sigma}^2, \hat{\delta}\right)$ is substituted for σ^2 and δ in the above expressions. Then, subsequent estimates $\tilde{\rho}^{(t)}, t = 1, 2, 3, ...$ are substituted into the expressions until convergence is reached.

One thousand datasets were simulated from each of the normally distributed populations with autoregressive circular covariance having the combinations of n, a and ρ listed in Table X, with σ^2 fixed at 1. The sample mean squared errors of $\tilde{\rho}$ are compared with those of the MLE $\hat{\rho}$, where $\hat{\rho}$ is approximated using FS algorithm iterations. It is seen that MSE($\tilde{\rho}$)<MSE($\hat{\rho}$) for small $Na\rho$. Furthermore, while $\tilde{\rho}$ is derived in an iterative fashion, as is $\hat{\rho}$, the iterations leading to $\tilde{\rho}$ always converged, whereas both the FS and NR algorithms were found not to converge on several instances. However, the performance of $\hat{\rho}$, when FS converges, quickly overtakes that of $\tilde{\rho}$ as $Na\rho$ increases, implying $\hat{\rho}$ is to be preferred for datasets of appreciable size.

Table X: Comparison of Mean Squared Error of Linear Combination Estimator with MSE of MLE.

ρ	n	a	MSE($\tilde{\rho}$)/MSE($\hat{\rho}$)
-.6	2	4	.4310
-.6	2	7	.5325
-.2	2	4	.7555
-.2	2	7	.8974
0	2	4	.8649
0	2	7	.9645
.2	2	4	1.6214
.2	2	7	1.3819

3.5 Estimation, Unequal Numbers of Measurements within Families. The immediately preceding subsections identified methods of estimating ρ when all families have the same number of measurements. When numbers of measurements vary between families, a modification of these methods is necessary. Although not dealing specifically with circular covariance models, Gleser (1992) and C.A.B. Smith (1957) have each approached the estimation of mean and covariance parameters in the case of unequal numbers of measurements on families (albeit with altogether different models). Gleser pooled families into g groups according to the numbers of measurements on each family, obtaining maximum likelihood estimators $\hat{\boldsymbol{\theta}}'(i) = (\hat{\theta}_1(i), \hat{\theta}_2(i), ..., \hat{\theta}_p(i))$ of the p parameters within the i^{th} group of families, $i = 1, 2, ..., g$, and then solving for the overall estimates $\hat{\boldsymbol{\theta}}' = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_p)$ in the equation

$$\sum_{i=1}^{g} \hat{\mathbf{I}}_i(\hat{\theta} - \hat{\boldsymbol{\theta}}(i)) = 0,$$

where $\hat{\mathbf{I}}_i(\boldsymbol{\theta})$ is the estimate of the information in any family in group i on $\boldsymbol{\theta}$. To apply these methods in our estimation of (ρ, σ^2), if an appreciable number of families are available in each group, we would combine the MLE $(\hat{\rho}(i), \hat{\sigma}^2(i))$ from group $i, i = 1, 2, ..., g$, which were specified in the last subsection.

On the other hand, C.A.B. Smith discussed estimation of the variance components in an unbalanced one way random effects general linear model. Smith in essence pooled blocks (families) into groups according to size, as did Gleser (1992). From the i^{th} group, $i = 1, 2, ..., g$ he obtained a preliminary estimate $v_B(i)$ of the variance V_B of the random effect. Smith then estimated the variance of each within-group estimator as

$$w_i^{-1} = \frac{v_A}{n_i} + v_B(i),$$

where n_i is number of observations on each block in the i^{th} group and v_A is the usual estimate of the error variance. The within-group estimators of V_B were then iteratively combined and improved, using weights inversely proportional to progressively improving estimates of their variances (proportional to the w_i). It can be shown that this method minimizes the variance of the overall estimator, and produces a linear combination of the within-group estimators analogous to the minimum variance estimator of ρ found in the last subsection using the extended Cauchy-Schwartz inequality as a linear combination of the $\tilde{\rho}_l$.

We adopt a modification of C.A.B. Smith's method here. As he combined the within-group $v_B(i)$, $i = 1, 2, ..., g$, we combine within-group estimators of (μ, σ^2, ρ). Group families according to the number of measurements on each family, letting I be the maximum number of measurements on any family. Also, for $i = 1, 2, ..., I$, let N_i be the number of families having i measurements. Let x_{ijk} be the k^{th} measurement on the j^{th} family having i measurements, $i : N_i > 0$; $j = 1, 2, ..., N_i$ and $k = 1, 2, ..., i$. Define

$$\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, ..., x_{iji})' \sim N_i(\mu \mathbf{1}, \mathbf{C}_i),$$

where the elements of $\mathbf{C}_i$ are

$$Cov(x_{ijk}, x_{ijk'}) = \begin{cases} \sigma^2 \rho^{|k-k'|}, & |k-k'| \leq i/2, \\ \sigma^2 \rho^{i-|k-k'|}, & |k-k'| > i/2. \end{cases} \tag{21}$$

$\mathbf{C}_i$ has eigenvalues $\delta_{ij}, j = 1, 2, ..., i$, which can be expressed in a manner similar to (15). Also, let $m_i = int\,(i/2)$ and $\mathbf{X}_i' = (\mathbf{x}_{i1}', \mathbf{x}_{i2}', ..., \mathbf{x}_{iN_i}')$, $i : N_i > 0$. The MLE

$$\hat{\mu} = \frac{\sum_{i:N_i>0} \delta_{i1}^{-1} x_i}{\sum_{i:N_i>0} i N_i \delta_{i1}^{-1}}$$

of μ is easily found using the loglikelihood involving $\mathbf{X}'$=$(\mathbf{X}_1', \mathbf{X}_2', ..., \mathbf{X}_I')$. Solving the score equation for μ is facilitated by expressing each $\mathbf{C}_i$ as $\mathbf{C}_i = \mathbf{\Gamma}_i diag\,(\delta_{i1}, \delta_{i2}, ..., \delta_{ii})\,\mathbf{\Gamma}_i$, where $\mathbf{\Gamma}_i$ has rows Γ_{ij} and is the obvious adaptation of (16) to the dimension i.

At this point, the within-group MLEs of ρ can easily be combined to produce an asymptotically efficient estimate of this parameter. Equation (18) specified the information about ρ when all families have the same size; adding the subscript i to denote the group (i.e., the dimension) for the present context, we have that

$$I_i(\rho) = \frac{N_i}{2} \sum_j \left[\left(\frac{\partial \delta_{ij}}{\partial \rho} \right)^2 \delta_{ij}^{-2} \right]$$

is the reciprocal of the asymptotic variance of the MLE of ρ within the group in which i measurements are taken on each family. Combining these MLEs using weights proportional to the $I_i(\rho)$ produces the overall asymptotically efficient estimate of ρ.

When all of the $\rho i N_i, i = 2, 3, ..., I$ are small, we can specify a better overall estimate $\tilde{\rho}$ of ρ than the overall combination of the within-group MLEs $\hat{\rho}(i)$, as the latter are less efficient than the within-group combination estimators $\tilde{\rho}(i)$. Discard all families having a single measurement each, as they provide no information about ρ. Group the other families according to the number of measurements available on each of them, as was done for the estimation of σ^2 and μ above. Within the group having i measurements on each family, and following the development of the last subsection, create the matrix $\mathbf{B}_i = (b_{ijl})$ using an adaptation of the $\mathbf{B}$ matrix described in Section 1, and calculate the MLEs $\hat{\boldsymbol{\delta}}_i$ of the vectors of eigenvalues $\boldsymbol{\delta}_i$. Also let $\tilde{\rho}_l(i)$ be the "combination" estimate of ρ obtained by dividing the l^{th} and $(l+1)^{th}$ linear transformations of $\hat{\boldsymbol{\delta}}_i$ induced by the rows of $(\mathbf{B}_i^{-1})'$, i.e.,

$$\tilde{\rho}_l(i) = \frac{\mathbf{b}_i^{l-1}\hat{\boldsymbol{\delta}}_i}{\mathbf{b}_i^{l}\hat{\boldsymbol{\delta}}_i},\ l = 1, 2, ..., m_i.$$

Define $M = \max_i\{m_i\} = m_I$ and $L_l = \min_i\{i : m_i \geq l\}$, $l = 1, 2, ..., M$. Because $\tilde{\rho}_l(i)$ is not available if $l \geq m_i + 1$ (since $(\mathbf{B}_i^{-1})'$ has only $m_i + 1$ rows), for a given l, L_l can be viewed as the smallest i so that $\tilde{\rho}_l(i)$ exists. For $l = 1, 2, ..., M$, let

$$\mathbf{r}_l' = (\tilde{\rho}_l(L_l), \tilde{\rho}_l(L_l + 1), ..., \tilde{\rho}_l(I))$$

be the vector combining across groups of families the estimates of ρ using the l^{th} and $(l+1)^{th}$ orthogonal transformations of $\hat{\boldsymbol{\delta}}_i$. For $i = L_l, L_l + 1, ..., I - 1$, the $\tilde{\rho}_l(i)$ is not present in $\mathbf{r}_l$ if no families in the sample have i measurements; let k_l be the resultant row dimension of $\mathbf{r}_l$. The strategy we follow is to (i) combine the $\tilde{\rho}_l(i)$, $i = 2, 3, ..., I$ as $\tilde{\rho}_l = \mathbf{u}_l'\mathbf{r}_l$ for some vectors $\mathbf{u}_l$ so that $\mathbf{u}_l'\mathbf{1} = 1$, $l = 1, 2, ..., M$, (ii) set $\mathbf{q}' = (\tilde{\rho}_1, \tilde{\rho}_2, ..., \tilde{\rho}_M)$ and then (iii) find the best combination $\tilde{\rho} = \mathbf{w}'\mathbf{q}$ of the $\tilde{\rho}_l$ for some vector $\mathbf{w}$ so that $\mathbf{w}'\mathbf{1} = 1$.

To find the elements of $\mathbf{u}_l$, we note that for a given l, the $\tilde{\rho}_l(i)$, $i = L_l, L_l + 1, ..., M$ are mutually independent, as they are obtained from different groups of families. Therefore, the covariance of $\mathbf{r}_l$ is $cov(\mathbf{r}_l) = diag(v_{L_l l}^{-1}, v_{L_l+1,l}^{-1}, ..., v_{Il}^{-1})$ where, from (20), it is known that

$$v_{il}^{-1} = var\left[\tilde{\rho}_l(i)\right] \doteq \frac{2\rho^{2-2l}}{N_i\sigma^4}\sum_{a=1}^{m_i - 1}\left(b_i^{l-1,a} - \rho b_i^{la}\right)^2 \frac{\delta_{ia}^2}{\alpha_{ia}}.$$

In $\tilde{\rho}_l$, we wish to assign weights to the elements of $\mathbf{r}_l$ inversely proportional to their estimated sampling variances. This is achieved by creating the $1 \times k_l$ vector

$$\mathbf{u}_l' = (u_{L_l,l}, u_{L_l-1,l}, ...u_{Il}) = \left(\sum v_{il}\right)^{-1} (v_{L_l,l}, v_{L_l-1,l}, ..., v_{I,l})$$

and letting $\tilde{\rho}_l = \mathbf{u}_l'\mathbf{r}_l$, the sum being over all $i : N_i > 0$ and $i \geq L_l$.

To find the elements of $\mathbf{w}' = (w_1, w_2, ..., w_M)$, we find the terms in the asymptotic covariance of $\mathbf{q}' = (\tilde{\rho}_1, \tilde{\rho}_2, ..., \tilde{\rho}_M)$,

$$(cov(\mathbf{q}))_{jl} = cov(\tilde{\rho}_j, \tilde{\rho}_l) = cov(\mathbf{u}_j'\mathbf{r}_j, \mathbf{u}_l'\mathbf{r}_l) = \mathbf{u}_j' \left[cov(\mathbf{r}_j, \mathbf{r}_l)\right] \mathbf{u}_l,$$

and yet again use the extended Cauchy-Schwartz inequality to minimize the asymptotic variance of $\bar{\rho}$. We consequently find the elements of $cov(\mathbf{r}_j, \mathbf{r}_l)$, which has dimension $k_j \times k_l$ (nonsquare for $m_j \neq m_l$, though possibly square even if $j \neq l$ since different j can map to the same m_j). Matrix $cov(\mathbf{r}_j, \mathbf{r}_l)$ has nonzero diagonal elements proceeding leftward and up from its last [i.e., its $(k_j \times k_l)^{th}$ element $cov\left[\tilde{\rho}_j(I), \tilde{\rho}_l(I)\right]$]. If, for instance, $j \leq l$, then $m_j \leq m_l$ and $k_j \geq k_l$, so that $cov(\mathbf{r}_j, \mathbf{r}_l)$ is at least as high as it is wide, and the first nonzero element of $cov(\mathbf{r}_j, \mathbf{r}_l)$ is $cov[\tilde{\rho}_j(L_l), \tilde{\rho}_l(L_l)]$ and appears in its $(k_j - k_l + 1, 1)$ position. In this case,

$$(cov(\mathbf{q}))_{jl} = \sum_{i \geq L_l : N_i > 0} u_{ij} cov[\tilde{\rho}_j(i), \tilde{\rho}_l(i)] u_{il}.$$

On the other hand, if $j \geq l$, then $cov(\mathbf{r}_j, \mathbf{r}_l)$ is at least as wide as it is high and

$$(cov(\mathbf{q}))_{jl} = \sum_{i \geq L_l : N_i > 0} u_{ij} cov[\tilde{\rho}_j(i), \tilde{\rho}_l(i)] u_{il}.$$

Let $L_{jl} = \max\{L_j, L_l\} = \min\{i : m_i \geq j \text{ and } m_i \geq l\}$; then regardless of whether j is as large as l, we can write the elements of $cov(\mathbf{q})$ as

$$(cov(\mathbf{q}))_{jl} = \sum_{i \geq L_{jl} : N_i > 0} u_{ij} cov[\tilde{\rho}_j(i), \tilde{\rho}_l(i)] u_{il}.$$

Here, for each $i, i = 2, 3, ..., I$, $cov[\tilde{\rho}_j(i), \tilde{\rho}_l(i)]$ is the $(j, l)^{th}$ element of $\mathbf{D}_{gi} \left(\mathbf{B}_i^{-1}\right)' \mathbf{I}_i^{-1}(\delta) \mathbf{B}_i^{-1} \mathbf{D}_{gi}'$, defining the matrices of this product by applying the results of the last subsection to the dimension i. Consequently, using (20), the nonzero elements of $cov(\mathbf{r}_j, \mathbf{r}_l)$ can be specified as

$$cov[\tilde{\rho}_j(i), \tilde{\rho}_l(i)] = \frac{2\rho^{2-i-j}}{N_i \sigma^4} \sum_{a=1}^{m_i - 1} (b_i^{j-1,a} - \rho b_i^{ja}) \frac{\delta_{ia}^2}{\alpha_{ia}} (b_i^{l-1,a} - \rho b_i^{la}), \; i = I, I-1, ..., L_{jl}.$$

where b_i^{ja} is the $(j,a)^{th}$ element of $(\mathbf{B}_i^{-1})'$. Our estimate of ρ then becomes

$$\bar{\rho} = \frac{\mathbf{1}'(cov(\mathbf{q}))^{-1}\mathbf{q}}{\mathbf{1}'(cov(\mathbf{q}))^{-1}\mathbf{1}},$$

having asymptotic expected value ρ and variance $\left(\mathbf{1}'(cov(\mathbf{q}))^{-1}\mathbf{1}\right)^{-1}$.

Now we proceed to find an overall estimate $\bar{\sigma}^2$ of σ^2. Even the families, if there are any, having only one measurement each will provide information about σ^2, although they provide none about ρ. So, for $i = 1, 2, ..., I$, let $\mathbf{b}_i^j$ be the j^{th} row of $(\mathbf{B}_i^{-1})'$, $j = 1, 2, ..., m_i + 1$. If $N_i > 0$, then $\hat{\boldsymbol{\delta}}_i' = (\hat{\delta}_{i1}, \hat{\delta}_{i2}, ..., \hat{\delta}_{im_i+1})$, the MLEs of the eigenvalues $\boldsymbol{\delta}_i' = (\delta_{i1}, \delta_{i2}, ..., \delta_{im_i-1})$ of $\mathbf{C}_i$ can be found within group i of N_i families (each having i measurements). This is accomplished following the methods of Olkin and Press (1969). Then an initial estimate of σ^2 from the families in the i^{th} group can be found following the methods outlined in Subsection (3.2) as $\hat{\sigma}^2(i)^{(1)} = \mathbf{b}_i^! \hat{\boldsymbol{\delta}}_i$. Subsequent iterations then proceed between the $\hat{\sigma}^2(i)^{(t)}$ and the $\hat{\rho}(i)^{(t)}$ in a manner similar to that described in that subsection, which will converge to the MLE $(\hat{\sigma}^2(i), \hat{\rho}(i))$ within group i. To combine the $\hat{\sigma}^2(i)$ into an overall estimator $\bar{\sigma}^2$, we first calculate the information $I_i(\sigma^2) = -\mathbf{E}\left[\frac{\partial^2}{\partial(\sigma^2)^2} \ln p_i(\mathbf{X}_i)\right]$ (inserting the subscript i everywhere to indicate the dimension of each family) that the i^{th} group of families contains about σ^2.

For the i so that $N_i > 0$, create the canonical variables $\mathbf{y}_{ij}' = (y_{ij1}, y_{ij2}, ..., y_{iji})$, where within the j^{th} family having i measurements, $y_{ijk} = \Gamma_{ik}' \mathbf{x}_{ij}, k = 1, 2, ...i$, Γ_{ik} being the k^{th} column of $\boldsymbol{\Gamma}_i$. Then $\mathbf{y}_{ij} = \boldsymbol{\Gamma}_i \mathbf{x}_{ij}, j = 1, 2, ..., N_i$. Each $\mathbf{y}_{ij}$ has mean $\boldsymbol{\mu}_i = \boldsymbol{\Gamma}_i \mathbf{1}\mu = (\sqrt{i}\mu, 0, 0, ..., 0)'$ and the likelihood of the parameters in $\mathbf{y}_{ij}$ is

$$p_i(\mathbf{y}_{ij}) = (2\pi)^{-i/2} |\mathbf{C}_i|^{-1/2} \exp\left\{-\frac{(\mathbf{y}_{ij} - \boldsymbol{\mu}_i)' diag(\delta_{i1}^{-1}, \delta_{i2}^{-1}, ..., \delta_{ii}^{-1})(\mathbf{y}_{ij} - \boldsymbol{\mu}_i)}{2}\right\}.$$

Here $\delta_{il} = \delta_{i,i-l-2}$, $l = 2, 3, ..., i$. If $\boldsymbol{\delta}_i' = (\delta_{i1}, \delta_{i2}, ..., \delta_{i,m_i-1})$ and $\boldsymbol{\rho}_i' = (1, \rho, \rho^2, ..., \rho^{m_i})$, the relation between $\boldsymbol{\delta}_i$ and $\boldsymbol{\rho}_i$ is given by

$$\boldsymbol{\delta}_i' = \sigma^2(\boldsymbol{\rho}_i' B_{i1}, \boldsymbol{\rho}_i' B_{i2}, ..., \boldsymbol{\rho}_i' B_{i,m_i-1}),$$

where B_{il} is the l^{th} column of $\mathbf{B}_i$. Defining $B_{il} = B_{i,i-l-2}$ for $l = m_i + 2, m_i + 3, ..., i$, gives

$$\ln p_i(\mathbf{y}_{ij}) = const - \frac{1}{2}\sum_{l=1}^{i} \ln(\sigma^2 B_{il} \boldsymbol{\rho}_i) - \frac{1}{2}\sum_{l=1}^{i} \frac{(y_{ijl} - \mu_{il})^2}{\sigma^2 \boldsymbol{\rho}_i' B_{il}}.$$

where $\mu_{il} = \begin{cases} \sqrt{i}\mu, l = 1 \\ 0, l = 2, 3, ..., i. \end{cases}$ Thus,

$$\Psi_{\sigma^2,i}(\mathbf{y}_{ij}) = \frac{\partial}{\partial\sigma^2} \ln p(\mathbf{y}_{ij}) = -\frac{1}{2}\sum_{l=1}^{i}\left(\frac{1}{\sigma^2} - \frac{(y_{ijl} - \mu_{il})^2}{\sigma^4 \rho_i' B_{il}}\right)$$

and so

$$\Psi'_{\sigma^2,i}(\mathbf{y}_{ij}) = \frac{1}{2}\sum_{l=1}^{i}\left(\frac{1}{\sigma^4} - 2\frac{(y_{ijl} - \mu_{il})^2}{\sigma^6 \rho_i' B_{il}}\right)$$

implies that the information about σ^2 conveyed by a family in the i^{th} group (i.e., a family having i measurements) is

$$I_i(\sigma^2) = -\frac{1}{2}\sum_{l=1}^{i}\left(\frac{1}{\sigma^4} - \frac{2}{\sigma^4}\right) = \frac{i}{2\sigma^4},$$

so that the information about σ^2 within the i^{th} group of N_i families is proportional to iN_i. Consequently, in the overall estimate $\tilde{\sigma}^2$ of σ^2, the MLE $\hat{\sigma}^2(i)$ of σ^2 within i^{th} group should receive a weight proportional to iN_i, so we let

$$\tilde{\sigma}^2 = \frac{\sum_{i:N_i>0} iN_i\hat{\sigma}^2(i)}{\sum_{k:N_k>0} kN_k}. \tag{22}$$

Because these $\hat{\sigma}^2(i)$ are mutually independent, their covariance matrix is diagonal. Consequently, this method of combining them is equivalent to using the extended Cauchy-Schwartz inequality to find their consistent minimum variance linear combination $(\hat{\sigma}^2(1), \hat{\sigma}^2(2), ..., \hat{\sigma}^2(I))\mathbf{v}$ for some vector $\mathbf{v}$ subject to the restriction that $\mathbf{v}'\mathbf{1} = 1$. Setting the derivative with respect to σ^2 of the entire loglikelihood (involving $\mathbf{X}$) equal to zero shows that the MLE $\hat{\sigma}^2$ of σ^2 is also a linear combination of the within-group MLEs $\hat{\sigma}^2(i)$. The MLE is asymptotically efficient for $\sum_{i:N_i>0} iN_i \to \infty$ (increasing without bound at least one family's number of measurements or one group's number of families, or the number of groups), but cannot be more asymptotically efficient than the consistent minimum variance linear combination $\tilde{\sigma}^2$; hence $\tilde{\sigma}^2$ is asymptotically efficient as well.

3.6 Estimation of Interclass Correlation, Unequal Numbers of Measurements within Families.

In applications in which the measurements on a family follow a circular covariance structure, these measurements can be equicorrelated with some other measurement. Citing once more our example involving measurements on disease incidence rates in sectors around a city center. these measurements may correlate uniformly with a measurement on another temporary or permanent characteristic of the city, such as the amount of rainfall in some preceding period of time. The main interest here is to estimate the correlation between the rainfall and each incidence rate.

Utilizing the "parent-sib" language common in discussions of repeated measures models, the case in which all the families have a parent and the same number of sibs, and the covariance parameters $(\sigma^2, \rho_1, \rho_2, ..., \rho_{m_I})$ are all functionally independent, has in effect been examined by Khattree and Naik (1994a). Assuming a multivariate normal distribution with a uniform mean for the sibs' scores, they specified the MLE for the mean and covariance parameters, as well as for the interclass correlation between each parent's score p_{ij} and any one of the sibs' scores. We seek to extend their arguments by allowing for different numbers of sibs within families, while restricting the (circular) covariance structure between sibs' scores by assuming that the elements of $\mathbf{C}_i$ are all σ^2 times powers of ρ, as described by (21).

With $(x_{ij1}, x_{ij2,...,} x_{iji})$ defined as in the last subsection and $\mathbf{x}_{ij} = (p_{ij}, x_{ij1}, x_{ij2}, ..., x_{iji})'$ as the family's entire set of measurements, assume

$$\mathbf{x}_{ij} \sim N_{i-1}\left(\begin{bmatrix} \mu_p \\ \mu_s \mathbf{1} \end{bmatrix}, \begin{bmatrix} \sigma_{pp} & \sigma_{ps}\mathbf{1}' \\ \sigma_{ps}\mathbf{1} & \mathbf{C}_i \end{bmatrix}\right),$$

$i : N_i > 0$ and $j = 1, 2, ..., N_i$, where $\mathbf{C}_i$ is as defined in (21), and let $\rho_{ps} = \sigma_{ps} / \left[\sigma_{pp}\sigma^2\right]^{1/2}$ be the interclass correlation.

The initial strategy here is to specify the MLE $(\hat{\sigma}_{pp}(i), \hat{\sigma}_{ps}(i), \hat{\sigma}^2(i))$ for $(\sigma_{pp}, \sigma_{ps}, \sigma^2)$ within each group of families. Subsequently, we calculate two consistent estimators of ρ_{ps} and compare their variances. In finding the first estimate $\dot{\rho}_{ps}$, we combine the $(\hat{\sigma}_{pp}(i), \hat{\sigma}_{ps}(i), \hat{\sigma}^2(i))$ across the groups to obtain overall estimators $\tilde{\sigma}_{pp}, \tilde{\sigma}_{ps}$, and $\tilde{\sigma}^2$ having minimum variances among all consistent linear combinations of the $(\hat{\sigma}_{pp}(i), \hat{\sigma}_{ps}(i), \hat{\sigma}^2(i))$. Then, another use of the Delta Theorem allows the calculation of the limiting distribution of

$$\dot{\rho}_{ps} = \tilde{\sigma}_{ps} / \left[\tilde{\sigma}_{pp}\tilde{\sigma}^2\right]^{1/2}.$$

The second estimator $\tilde{\rho}_{ps}$ is the minimum variance consistent linear combination

$$\tilde{\rho}_{ps} = \frac{\sum_{i:N_i>0} \hat{\rho}_{ps}(i) w_i}{\sum_{i:N_i>0} w_i}$$

where the w_i are appropriate weights and the $\hat{\rho}_{ps}(i)$ are within-group sample correlation coefficients (MLEs)

$$\hat{\rho}_{ps}(i) = \frac{\hat{\sigma}_{ps}(i)}{\sqrt{\hat{\sigma}_{pp}(i)\,\hat{\sigma}^2(i)}}.$$

Of the two estimates of ρ_{ps}, it seems reasonable to choose the more efficient one, assuming their small-sample biases are comparable and given (as will be shown) they are both consistent.

Khattree and Naik (1994a) in effect found the MLE $(\hat{\sigma}_{pp}(i), \hat{\sigma}_{ps}(i))$ within each group of families having the same number of sibs. Their estimator remains the MLE under our "autoregressive" assumption; this assumption changes merely the estimation of the circular covariance parameters, which involve σ^2 and the autoregressive parameter ρ. In the last subsection, we produced the minimum variance consistent linear combination $\tilde{\sigma}^2$ of the $\hat{\sigma}^2(i)$ when no parents' scores are available. However, the presence of the parents' scores increases the information about σ^2, as is shown below. Hence, our main task with respect to $\hat{\rho}_{ps}$ consists of finding the Fisher information, first in each group and then over all groups, of $\boldsymbol{\theta}' = (\theta_1, \theta_2, \theta_3) = (\sigma_{pp}, \sigma_{ps}, \sigma^2)$, specifying overall estimators of σ_{pp},σ_{ps} and σ^2. With respect to $\tilde{\rho}_{ps}$, in each group we must calculate the asymptotic variance of each $\hat{\rho}_{ps}(i)$. These derivations allow the calculation of the asymptotic variances of the two estimators of ρ_{ps}, which we compare.

Let $\hat{\boldsymbol{\theta}}'(i) = (\hat{\theta}_1(i), \hat{\theta}_2(i), \hat{\theta}_3(i)) = (\hat{\sigma}_{pp}(i), \hat{\sigma}_{ps}(i), \hat{\sigma}^2(i))$, $i : N_i > 0$ be the vectors of within-group MLEs of $\boldsymbol{\theta}$. Let the elements of the information matrix about $\boldsymbol{\theta}$ in any family in the i^{th} group, and in all the groups, be described by

$$I_i(k,l) = -\mathbf{E}\left(\frac{\partial^2 \ln p(\mathbf{x}_{ij})}{\partial\theta_l \partial\theta_k}\right) \text{ and } I(k,l) = \sum_{i:N_i>0} N_i I_i(k,l),$$

modifying slightly the commonly used notations so that $\mathbf{I}_i(\boldsymbol{\theta}) = (I_i(k,l))$ and $\mathbf{I}(\boldsymbol{\theta}) = (I(k,l))$. The matrix sum $\mathbf{I}(\boldsymbol{\theta})$ will be used in the approximation of the distribution of $\hat{\rho}_{ps}$, and the $\mathbf{I}_i(\boldsymbol{\theta})$ will be used in that of $\tilde{\rho}_{ps}$. Using the efficient likelihood properties of the $\hat{\theta}_k(i)$, we have $(I_i(k,k))^{1/2}\sqrt{N_i}(\hat{\theta}_k(i) - \theta_k) \to N(0,1)$ and $\mathbf{I}_i^{1/2}(\boldsymbol{\theta})\sqrt{N_i}(\hat{\boldsymbol{\theta}}(i) - \boldsymbol{\theta}) \to N(0, \mathbf{I})$ as $N_i \to \infty$. Let

$\mathbf{q}_k' = (\hat{\theta}_k(1), \hat{\theta}_k(2), ..., \hat{\theta}_k(I))$ be the vector of MLEs of θ_k from the groups of families having $N_i > 0$. The minimum variance consistent linear combination estimator of θ_k is $\bar{\theta}_k = \mathbf{l}_k'\mathbf{q}_k$, where

$$\mathbf{l}_k' \doteq \left[\sum_{i:N_i>0} N_i \hat{I}_i(k,k)\right]^{-1} \left(N_1\hat{I}_1(k,k), N_2\hat{I}_2(k,k), ..., N_I\hat{I}_I(k,k)\right).$$

Also let $\bar{\boldsymbol{\theta}}' = (\mathbf{l}_1'\mathbf{q}_1, \mathbf{l}_2'\mathbf{q}_2, \mathbf{l}_3'\mathbf{q}_3)$; we wish to estimate $cov\left(\bar{\boldsymbol{\theta}}\right) = \boldsymbol{\Omega}$ and then find the limiting distribution of $\hat{\rho}_{ps} = g(\bar{\boldsymbol{\theta}}) = \frac{\bar{\theta}_2}{\sqrt{\bar{\theta}_1\bar{\theta}_3}}$. For $k, l = 1, 2, 3$, if the N_i are large, we may consider the estimate $\hat{I}_i^{(k,l)}$ of $I_i^{(k,l)}$, the (k,l) element of $\mathbf{I}_i^{-1}(\boldsymbol{\theta})$, to be sufficiently close to $cov(\hat{\theta}_k(i), \hat{\theta}_l(i))$. Then,

$$\begin{aligned} var(\mathbf{l}_k'\mathbf{q}_k) &\doteq \sum_{i:N_i>0}\left[\left(\frac{N_i\hat{I}_i(k,k)}{\sum_{l:N_l>0} N_l\hat{I}_l(k,k)}\right)^2 \frac{\hat{I}_i^{(k,k)}}{N_i}\right] \\ &= \left(\sum_{l:N_l>0} N_l\hat{I}_l(k,k)\right)^{-2} \sum_{i:N_i>0} N_i\left(\hat{I}_i(k,k)\right)^2 \hat{I}_i^{(k,k)} \end{aligned}$$

and, more generally, for $j, k = 1, 2, 3$,

$$\begin{aligned} cov\left(\mathbf{l}_j'\mathbf{q}_j, \mathbf{l}_k'\mathbf{q}_k\right) &\doteq \mathbf{l}_j' cov(\mathbf{q}_j, \mathbf{q}_k)\mathbf{l}_k \\ &= \mathbf{l}_j' diag\left(\frac{\hat{I}_1^{(j,k)}}{N_1}, \frac{\hat{I}_2^{(j,k)}}{N_2}, ..., \frac{\hat{I}_I^{(j,k)}}{N_I}\right)\mathbf{l}_k \\ &\doteq \left[\left(\sum_{l:N_l>0} N_l\hat{I}_l(j,j)\right)\left(\sum_{l:N_l>0} N_l\hat{I}_l(k,k)\right)\right]^{-1} \times \\ &\quad \sum_{i:N_i>0}\left[N_i\hat{I}_i(j,j)\hat{I}_i^{(j,k)}\hat{I}_i(k,k)\right]. \end{aligned} \tag{23}$$

Because the elements of $\bar{\boldsymbol{\theta}}$ are consistent minimum variance linear combinations of (efficient) maximum likelihood estimators, and because of the additivity of both variances and information across independent groups of families, the elements of $\bar{\boldsymbol{\theta}}$ are efficient for estimating $\boldsymbol{\theta}$. Furthermore, as $\sum N_i \to \infty$, $\boldsymbol{\Omega}^{-1/2}\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \to N(0,\mathbf{I})$ and $\left(\mathbf{D}_g\boldsymbol{\Omega}\mathbf{D}_g'\right)^{-1/2}\left(g(\bar{\boldsymbol{\theta}}) - \rho_{ps}\right) = \left(\mathbf{D}_g\boldsymbol{\Omega}\mathbf{D}_g'\right)^{-1/2}\left(\hat{\rho}_{ps} - \rho_{ps}\right) \to N(0,1)$, where $\mathbf{D}_g$ is the (1×3) derivative of g with respect to $\boldsymbol{\theta}$. With respect to the approximation of the distribution of $\hat{\rho}_{ps}$, it remains to find $\hat{I}_i^{(j,k)}$, $N_i > 0$, $j, k = 1, 2, 3$, to which we next proceed.

To find each $I_i(j,j)$, we follow Khattree and Naik, for the i^{th} group of families creating the nonsingular $\mathbf{\Phi}_i = (i^{-1}\mathbf{1}.\Gamma_{i2}, \Gamma_{i3}, ..., \Gamma_{ii})$ using the

$$\Gamma_{ij} = \begin{array}{c} i^{-1/2}\left\{\cos\left(\frac{2\pi(j-1)0}{i}\right) + \sin\left(\frac{2\pi(j-1)0}{i}\right), \cos\left(\frac{2\pi(j-1)1}{i}\right) + \sin\left(\frac{2\pi(j-1)1}{i}\right), ..., \right. \\ \left. \cos\left(\frac{2\pi(j-1)(i-1)}{i}\right) + \sin\left(\frac{2\pi(j-1)(i-1)}{i}\right)\right\} \end{array}$$

defined in subsection (3.5). Then, we employ the canonical variables $\mathbf{y}_{ij} = (p_{ij}, y_{ij1}, y_{ij2}, ..., y_{iji})' = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{\Phi}_i' \end{bmatrix} \mathbf{x}_{ij}$. The covariance of $\mathbf{y}_{ij}$ is

$$cov(\mathbf{y}_{ij}) = \begin{bmatrix} \theta_1 & \theta_2 & 0 & 0 & \dots & 0 \\ \theta_2 & \lambda_i & 0 & 0 & \dots & 0 \\ 0 & 0 & \theta_3\kappa_{i2} & 0 & \dots & 0 \\ 0 & 0 & 0 & \theta_3\kappa_{i3} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \theta_3\kappa_{ii} \end{bmatrix}, \quad (24)$$

where $\lambda_i = var(y_{ij1}) = \theta_3\kappa_{i1}$. The κ_{ik} here are

$$\begin{aligned} \kappa_{i1} &= i^{-2}\mathbf{1}'\mathbf{C}_i\mathbf{1} = i^{-1}\left[1 + 2\sum_{l=1}^{m_i-1}\rho^l + \alpha_{im_i}\rho^{m_i}\right] \text{ and} \\ \kappa_{ik} &= 1 + 2\sum_{l=1}^{m_i-1}\rho^l \cos\frac{2\pi(k-1)l}{i} + \alpha_{im_i}\rho^{m_i}, k = 2, 3, ..., i. \end{aligned}$$

Also, defining $d_i = \theta_1\theta_3\kappa_{i1} - \theta_2^2$ in $[cov(\mathbf{y}_{ij})]^{-1}$, the loglikelihood involving each $\mathbf{y}_{ij}$ is

$$\begin{aligned} \ln p_i(\mathbf{y}_{ij}) &= const - \frac{1}{2}\ln\left[\left(\theta_1\theta_3\kappa_{i1} - \theta_2^2\right)\theta_3^{(i-1)}\prod_{l=2}^{i}\kappa_{il}\right] \\ &- \frac{1}{2}\left[\begin{array}{c} (p_{ij} - \mu_p)^2\frac{\theta_3\kappa_{i1}}{d_i} - 2(p_{ij} - \mu_p)(y_{ij1} - \mu_s)\frac{\theta_2}{d_i} + (y_{ij1} - \mu_s)^2\frac{\theta_1}{d_i} \\ + \sum_{l=2}^{i}\frac{y_{ijl}^2}{\kappa_{il}\theta_3} \end{array}\right]. \end{aligned}$$

The information about $\boldsymbol{\theta}$ in any family having i measurements is therefore

$$\begin{aligned} \mathbf{I}_i(\boldsymbol{\theta}) &= -\mathbf{E}\left[\frac{\partial^2 \ln p_i(\mathbf{y}_{ij})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = \left(\theta_1\lambda_i - \theta_2^2\right)^{-2} \times \\ & \begin{bmatrix} \lambda_i^2/2 & -\theta_2\lambda_i & \theta_2^2\kappa_{i1}/2 \\ \cdot & \theta_1\lambda_i + \theta_2^2 & -\theta_1\theta_2\kappa_{i1} \\ \cdot & \cdot & \left(\begin{array}{c} i\theta_1^2\lambda_i^2 + \\ (i-1)\theta_2^2\left(2\theta_1\lambda_i + \theta_2^2\right) \end{array}\right)/2\theta_3^2 \end{bmatrix}, \end{aligned}$$

using the expectations

$$\mathbf{E}\left[(p_{ij}-\mu_p)^2\right] = \theta_1,\ \mathbf{E}\left[(y_{ij1}-\mu_s)^2\right] = \kappa_{i1}\theta_3,$$
$$\mathbf{E}\left(y_{ijk}^2\right) = \kappa_{ik}\theta_3 \text{ and } \mathbf{E}\left[(p_{ij}-\mu_p)(y_{ij1}-\mu_s)\right] = \theta_2,$$

for $k = 2, 3, ..., i$. Combining this information across all families, the information in all the data about $\boldsymbol{\theta}$ is

$$\mathbf{I}(\boldsymbol{\theta}) = \sum_{i:N_i>0} N_i\left(\theta_1\lambda_i - \theta_2^2\right)^{-2} \times \begin{bmatrix} \lambda_i{}^2/2 & -\theta_2\lambda_i & \theta_2^2\kappa_{i1}/2 \\ \cdot & \theta_1\lambda_i+\theta_2^2 & -\theta_1\theta_2\kappa_{i1} \\ \cdot & \cdot & \left(\begin{array}{c} i\theta_1^2\lambda_i^2+ \\ (i-1)\,\theta_2^2\left(\theta_2^2-2\theta_1\lambda_i\right)\end{array}\right)/2\theta_3^2 \end{bmatrix}.$$

Within the i^{th} group of N_i families, $i : N_i > 0$, the methods of Khattree and Naik and (for θ_3) those outlined in Subsection 3.5 can be adopted to find the within-group MLE $\hat{\boldsymbol{\theta}}(i)$ $=\left(\hat{\theta}_1(i), \hat{\theta}_2(i), \hat{\theta}_3(i)\right)'$ of $\boldsymbol{\theta}$. Finally, minimizing the asymptotic variance of each $\bar{\theta}_j$ (and therefore of $trace\,(\boldsymbol{\Omega})$) is accomplished by setting

$$\mathbf{l}_1' = \left\{\sum_{i:N_i>0}\frac{N_i\lambda_i^2}{\left(\theta_1\lambda_i^2-\theta_2^2\right)^2}\right\}^{-1} \times \left[\frac{N_1\lambda_1^2}{\left(\theta_1\lambda_1-\theta_2^2\right)^2}, \frac{N_2^2\lambda_2^2}{\left(\theta_1\lambda_2-\theta_2^2\right)^2}, ..., \frac{N_I\lambda_I^2}{\left(\theta_1\lambda_I-\theta_2^2\right)^2}\right],$$

$$\mathbf{l}_2 = \left\{\sum_{i:N_i>0}\frac{N_i\left(\theta_1\lambda_i^2+\theta_2^2\right)}{\left(\theta_1\lambda_i^2-\theta_2^2\right)^2}\right\}^{-1} \times \left[\frac{N_1\left(\theta_1\lambda_1+\theta_2^2\right)}{\left(\theta_1\lambda_1-\theta_2^2\right)^2}, \frac{N_2\left(\theta_1\lambda_2+\theta_2^2\right)}{\left(\theta_1\lambda_2-\theta_2^2\right)^2}, ..., \frac{N_I\left(\theta_1\lambda_I+\theta_2^2\right)}{\left(\theta_1\lambda_I-\theta_2^2\right)^2}\right],$$

$$\mathbf{l}_3' = \left[\sum_{i:N_i>0}\frac{N_i\left(\begin{array}{c} i\theta_1^2\lambda_i^2+ \\ (i-1)\,\theta_2^2\left(\theta_2^2-2\theta_1\lambda_i\right)\end{array}\right)}{\left(\theta_1\lambda_i^2-\theta_2^2\right)^2}\right]^{-1} \times$$

$$\left(\frac{N_1\left(\begin{array}{c}1\theta_1^2\lambda_i^2+\\(1-1)\theta_2^2\left(\theta_2^2-2\theta_1\lambda_1\right)\end{array}\right)}{\left(\theta_1\lambda_1-\theta_2^2\right)^2},\frac{N_2\left(\begin{array}{c}2\theta_1^2\lambda_2^2+\\(2-1)\theta_2^2\left(\theta_2^2-2\theta_1\lambda_2\right)\end{array}\right)}{\left(\theta_1\lambda_2-\theta_2^2\right)^2},\ldots,\frac{N_I\left(\begin{array}{c}I\theta_1^2\lambda_i^2+\\(I-1)\theta_2^2\left(\theta_2^2-2\theta_1\lambda_I\right)\end{array}\right)}{\left(\theta_1\lambda_I-\theta_2^2\right)^2}\right),$$

and $\mathbf{q}_k' = (\tilde{\theta}_k(1),\tilde{\theta}_k(2),...,\tilde{\theta}_k(I)), k=1,2,3.$

Because the elements of the $\mathbf{l}_j$ involve $\boldsymbol{\theta}$, the estimations must be performed iteratively, at each step substituting into the $\mathbf{l}_j$ the current estimates of $\boldsymbol{\theta}$.

We now use the preceding arguments to derive the large-sample variances of $\dot{\rho}_{ps} = \tilde{\theta}_2/\sqrt{\tilde{\theta}_1\tilde{\theta}_3}$ and $\bar{\rho}_{ps} = \frac{\sum_{i:N_i>0}\tilde{\rho}_{ps}(i)\left(var(\tilde{\rho}_{ps}(i))\right)^{-1}}{\sum_{l:N_l>0}\left(var(\tilde{\rho}_{ps}(l))\right)^{-1}}$. With respect to $\dot{\rho}_{ps}$, Cramer's rule obtains

$$\mathbf{I}_i^{-1}(\boldsymbol{\theta}) = i^{-1}\begin{bmatrix}2\frac{i\theta_1^2\lambda_i^2-(i-1)\theta_2^4}{\lambda_i} & 2\frac{\pi_i\theta_2}{\lambda_i} & 2\frac{\theta_2^2}{\kappa_{i1}}\\ . & \pi_i+\theta_2^2 & 2\theta_2\theta_3\\ . & . & 2\theta_3^2\end{bmatrix}, \quad (25)$$

setting $\pi_i = i\theta_1\lambda_i - (i-1)\theta_2^2$. This inverse is easily estimated, substituting $\left(\tilde{\theta}_1,\tilde{\theta}_2,\tilde{\theta}_3\right)$ for $\boldsymbol{\theta}$ and $\tilde{\kappa}_{i1} = \kappa_{i1}(\tilde{\rho})$ (as derived in the last subsection) for κ_{i1} into the above. This result, with (23), produces the distinct elements of $\boldsymbol{\Omega} = cov\tilde{\boldsymbol{\theta}}$ as

$$var(\tilde{\theta}_1) \doteq I^{(1,1)} \doteq \frac{2}{\theta_3}\left(\sum_{l:N_l>0} N_l\frac{\kappa_{l1}^2}{d_l^2}\right)^{-2}\sum_{i:N_i>0} N_i\frac{\kappa_{i1}^3\left(i\theta_1^2\lambda_i^2-(i-1)\theta_2^4\right)}{d_i^4 i},$$

$$cov(\tilde{\theta}_1,\tilde{\theta}_2) \doteq I^{(1,2)} \doteq \frac{2\theta_2}{\theta_3}\left[\left(\sum_{l:N_l>0} N_l\frac{\kappa_{l1}^2}{d_l^2}\right)\left(\sum_{l:N_l>0} N_l\frac{\theta_1\lambda_l+\theta_2^2}{d_l^2}\right)\right]^{-1}\times \sum_{i:N_i>0} N_i\frac{\kappa_{i1}\pi_i\left(\theta_1\lambda_i+\theta_2^2\right)}{id_i^4},$$

$$cov(\tilde{\theta}_1,\tilde{\theta}_3) \doteq I^{(1,3)} \doteq 2\theta_2^2\left[\left(\sum_{l:N_l>0} N_l\frac{\kappa_{l1}^2}{d_l^2}\right)\left(\sum_{l:N_l>0} N_l\frac{\left(\begin{array}{c}l\theta_1^2\lambda_l^2+\\(l-1)\theta_2^2\left(\theta_2^2-2\theta_1\lambda_l\right)\end{array}\right)}{d_l^2}\right)\right]^{-1}\times \sum_{i:N_i>0} N_i\frac{\kappa_{i1}}{id_i^4}\left(i\theta_1^2\lambda_i^2+(i-1)\theta_2^2\left(\theta_2^2-2\theta_1\lambda_i\right)\right).$$

$$\begin{aligned}
var(\tilde{\theta}_2) &\doteq I^{(2,2)} \doteq \left(\sum_{l:N_l>0} N_l \frac{\theta_1\lambda_l+\theta_2^2}{d_l^2}\right)^{-2} \sum_{i:N_i>0} N_i \left(\frac{\theta_1\lambda_i+\theta_2^2}{d_i^2}\right)^2 \frac{\pi_i+\theta_2^2}{i}, \\
cov(\tilde{\theta}_2,\tilde{\theta}_3) &\doteq I^{(2,3)} \doteq 2\theta_2\theta_3 \left[\begin{array}{c} \left(\sum_{l:N_l>0} N_l \frac{\theta_1\lambda_l-\theta_2^2}{d_l^2}\right) \times \\ \left(\sum_{l:N_l>0} N_l \frac{l\theta_1^2\lambda_l^2+(l-1)\theta_2^2(\theta_2^2-2\theta_1\lambda_l)}{d_l^2}\right) \end{array}\right]^{-1} \times \\
&\sum_{i:N_i>0}\left[N_i \frac{(\theta_1\lambda_i+\theta_2^2)\left[i\theta_1^2\lambda_i^2+(i-1)\theta_2^2(\theta_2^2-2\theta_1\lambda_i)\right]}{i d_i^4}\right], \text{ and} \\
var(\tilde{\theta}_3) &\doteq I^{(3,3)} \doteq 2\theta_3^2 \left(\sum_{l:N_l>0} N_l \frac{\left(\begin{array}{c} l\theta_1^2\lambda_l^2+ \\ (l-1)\theta_2^2(\theta_2^2-2\theta_1\lambda_l) \end{array}\right)}{d_l^2}\right)^{-2} \times \\
&\sum_{i:N_i>0} N_i \frac{\left(\begin{array}{c} i\theta_1^2\lambda_i^2+ \\ (i-1)\theta_2^2(\theta_2^2-2\theta_1\lambda_i) \end{array}\right)^2}{i d_i^4}.
\end{aligned}$$

We remark that a large-sample confidence region for $\boldsymbol{\theta}'= (\theta_1, \theta_2, \theta_3)$, the components of ρ_{ps}, can be calculated on the basis of the consistency of each element of $\tilde{\boldsymbol{\theta}}$ and the expressions for the elements of $\boldsymbol{\Omega}$ specified in (23).

Another use of the Delta Theorem now gives the large-sample distribution of $\dot{\rho}_{ps}$ as

$$\begin{aligned}
\left(\mathbf{D}_g\boldsymbol{\Omega}\mathbf{D}_g'\right)^{-1/2}\left(\dot{\rho}_{ps}-\rho_{ps}\right) &\rightarrow \mathcal{N}(0,1), \text{ implying} \\
var\left(\dot{\rho}_{ps}\right) &\doteq \mathbf{D}_g\boldsymbol{\Omega}\mathbf{D}_g',
\end{aligned}$$

taking the elements of $\boldsymbol{\Omega}$ as those identified in (23) and $\mathbf{D}_g = \frac{1}{\sqrt{\theta_1\theta_3}}\left[\frac{-\theta_2}{2\theta_1}, 1, \frac{-\theta_2}{2\theta_3}\right]$ as the derivative of g with respect to $\boldsymbol{\theta}$.

Now we focus on the distribution of $\tilde{\rho}_{ps}$. Having let $\hat{\boldsymbol{\theta}}'(i) = (\hat{\theta}_1(i), \hat{\theta}_2(i), \hat{\theta}_3(i))$ using the methods of Khattree and Naik and those of Subsection 3.2, we start by finding the asymptotic distribution of $\tilde{\rho}_{ps}(i) = g(\hat{\boldsymbol{\theta}}(i)) = \frac{\hat{\theta}_2(i)}{\sqrt{\hat{\theta}_1(i)\hat{\theta}_3(i)}}$ as a within-group estimator of $g(\boldsymbol{\theta}) = \rho_{ps}$. The $g(\hat{\boldsymbol{\theta}}(i)), i : N_i > 0$ will be combined to produce an overall estimate $\tilde{\rho}_{ps}$. It is known from the Delta Theorem that

$$\left(\mathbf{D}_g\mathbf{I}_i^{-1}(\boldsymbol{\theta})\mathbf{D}_g'\right)^{-1/2}\sqrt{N_i}(g(\hat{\boldsymbol{\theta}}(i))-\rho_{ps}) \rightarrow \mathcal{N}_1(0,1).$$

From the expression for $\mathbf{I}_i^{-1}(\boldsymbol{\theta})$ in (25), we therefore have, after considerable simplification,

$$\begin{aligned} var(g(\hat{\boldsymbol{\theta}}(i))) &= var(\hat{\rho}_{ps}(i)) \doteq N_i^{-1}\mathbf{D}_g\mathbf{I}_i^{-1}(\boldsymbol{\theta})\mathbf{D}_g' \\ &= \frac{id_i^2\left[\theta_2^2\left(\theta_1\lambda_i+\theta_2^2-4\theta_1\right)+2\theta_1^2\lambda_i\right]+\theta_2^2\left(\theta_2^4-\theta_1\theta_2^2+d_i\theta_1\right)}{2N_i i\theta_1^3\theta_3\lambda_i}. \end{aligned}$$

Combining the $\hat{\rho}_{ps}(i) = g(\hat{\boldsymbol{\theta}}(i))$ across groups by assigning weights to them inversely proportional to their variances produces a consistent minimum variance linear combination estimator

$$\tilde{\rho}_{ps} = \frac{\sum_{i:N_i>0}\hat{\rho}_{ps}(i)\frac{N_i i\lambda_i}{id_i^2\left[\theta_2^2(\theta_1\lambda_i+\theta_2^2-4\theta_1)+2\theta_1^2\lambda_i\right]+\theta_2^2(\theta_2^4-\theta_1\theta_2^2+d_i\theta_1)}}{\sum_{l:N_l>0}\frac{N_l l\lambda_l}{ld_l^2\left[\theta_2^2(\theta_1\lambda_l+\theta_2^2-4\theta_1)+2\theta_1^2\lambda_l\right]+\theta_2^2(\theta_2^4-\theta_1\theta_2^2+d_l\theta_1)}}.$$

In practice, the consistent minimum variance linear combination estimators $(\tilde{\sigma}_{pp}, \tilde{\sigma}_{ps}^2, \tilde{\sigma}^2, \tilde{\rho})$ are used in place of the parameters in this expression for $\tilde{\rho}_{ps}$. To find the approximate expected value and asymptotic variance of $\tilde{\rho}_{ps}$, letting

$$\begin{aligned} w_i &= \left[var\left(\hat{\rho}_{ps}(i)\right)\right]^{-1} \\ &= \frac{2N_i i\theta_1^3\theta_3\lambda_i}{id_i^2\left[\theta_2^2\left(\theta_1\lambda_i+\theta_2^2-4\theta_1\right)+2\theta_1^2\lambda_i\right]+\theta_2^2\left(\theta_2^4-\theta_1\theta_2^2+d_i\theta_1\right)} \end{aligned}$$

gives

$$\begin{aligned} var\left(\tilde{\rho}_{ps}\right) &\doteq \frac{\sum_{i:N_i>0} w_i^2 w_i^{-1}}{\left(\sum_{i:N_i>0} w_i\right)^2} \\ &= \frac{1}{2\theta_1^3\theta_3}\left(\sum_{i:N_i>0}\frac{N_i i\lambda_i}{id_i^2\left[\theta_2^2\left(\theta_1\lambda_i+\theta_2^2-4\theta_1\right)+2\theta_1^2\lambda_i\right]+\theta_2^2\left(\theta_2^4-\theta_1\theta_2^2+d_i\theta_1\right)}\right)^{-1}. \end{aligned}$$

We now compare the large-scale variances of $\dot{\rho}_{ps}$ and $\tilde{\rho}_{ps}$ which, ignoring the biases of both estimators, amounts to specifying the relation, for large $\sum N_i$, of

$$r = \frac{var\left(\dot{\rho}_{ps}\right)}{var\left(\tilde{\rho}_{ps}\right)} = \left(\mathbf{D}_g\boldsymbol{\Omega}\mathbf{D}_g'\right)\left(\sum_{i:N_i>0} w_i\right)$$

to unity. In particular, $\dot{\rho}_{ps}$ is more (less) efficient than $\tilde{\rho}_{ps}$ whenever r is less than (exceeds) 1.

Computer calculations of the theoretical variances (using the above expressions with the parameters $(\boldsymbol{\theta},\rho)$ themselves, not their estimates) across a wide range of sample sizes and parameter values show that r depends chiefly on ρ_{ps} and the the evenness of the spacing of the eigenvalues of

$$cov\left(\mathbf{x}_{ij}\right) = \begin{bmatrix} \theta_1 & \theta_2\mathbf{1}' \\ \theta_2\mathbf{1} & \mathbf{C}_i \end{bmatrix}.$$

To be brief, we may say that generally, $r > 1$, indicating that $\tilde{\rho}_{ps}$ is more efficient than $\dot{\rho}_{ps}$ for the majority of cases. The exceptions occur when $\rho_{ps} = 0$ (in which case $r = 1$) or when $|\theta_2|$ is so large in relation to $(\theta_1, \theta_3, \rho)$ that too much of $trace\,(cov\,(\mathbf{x}_{ij}))$ belongs to the largest of the eigenvalues (in which case $r < 1$). The value of r depends hardly at all on the N_i or the number of groups.

Computer calculations showed $r < 1$ whenever $|\rho_{ps}|$ is close to 1; however, r may be less than 1 even for relatively small values of $|\rho_{ps}|$, provided that some of the eigenvalues of $cov\,(\mathbf{x}_{ij})$ are close to zero in relation to the largest eigenvalues. As an example, the case $(\theta', \rho) = (2, 1.2, 3, .5)$, $N_4 = 5$, $N_7 = 2$ and all other $N_i = 0$ possesses a relatively small $|\rho_{ps}|$ (i.e., .4898); yet $r = .4997 < 1$, because the smallest of the eight eigenvalues of $cov\,(\mathbf{x}_{ij})$ for $i = 7$ are quite small in relation to the first eigenvalue. In fact, the first eigenvalue is almost half of $trace\,(cov\,(\mathbf{x}_{ij}))$.

However, $\tilde{\rho}_{ps}$ is more efficient than $\dot{\rho}_{ps}$ in most cases; it usually is to be preferred unless $|\rho_{ps}|$ is known to be very close to unity.

4 COMPOUND SYMMETRY WITHIN COUSINSHIPS

4.1 Introduction. In certain biological and physical settings, a disturbance may emanate from a single source to various nodes, from each of which other disturbances are transmitted to a number of receivers in a circular fashion. For instance, a communications satellite might send signals to central earth stations, from each of which the signals are relayed to points (such as end users of the information conveyed from the satellite) on a circle around it. Assuming that the strengths (or another characteristic) of the signals received by the points around any single station follow a circular covariance structure, it is of interest to estimate the circular covariance parameters, the correlation between the signal strength at the points and at the stations, and the correlation between the signal strength at the stations and at the satellite. Such patterns arise also in accelerometers used in certain automobiles to deploy air bags; these devices may be arranged in groups around the car to record disturbances experienced by the car, each device consisting of a circle of sensors whose measurements are analyzed by a microprocessor.

To adapt the "parent-sib-family" terminology of the previous sections, we may assume that degrees of a quantitative feature tend to be passed from grandparent to parent, and subsequently from parent to child, and the covariance of the feature among the siblings is circular. We will seek to estimate the correlations of the feature between generations. In other words, assume sibships, each of which is multivariate normally distributed having circular covariance, can be grouped into "cousinships" (csp's) so that the measurements on the sibs in any two different families within the same csp are associated by a single "compound symmetry" parameter V. Also, extend the terminology so that with each cousinship may be associated a "grandparent." Then, under some sets of assumptions, we calculate the ML estimator of the intergenerational covariances, V and the circular covariance parameters $(\sigma_0, \sigma_1, ...\sigma_m)$. Under other, less restrictive assumptions, we present ANOVA-type estimators based on reductions of sums of squares which are easier than the MLE to develop and calculate.

4.2 Uniform Family and Cousinship Sizes with no Scores of Parents or Grandparents . In this subsection, assume that nc families, each with a sibs, are arranged in n csp's having c families each, and that the covariance between any two sibs in the same family is

$$(\mathbf{C})_{l,l'} = cov(x_{ijl}, x_{ijl'}) = \begin{cases} \sigma_{|l-l'|}, & |l-l'| \leq a/2, \\ \sigma_{a-|l-l'|}, & |l-l'| > a/2, \end{cases} \tag{26}$$

x_{ijl} being the score on the l^{th} sib in the j^{th} family in the i^{th} csp. Without loss of generality, assume in this and the subsequent subsections, where estimation is by maximum likelihood, that all the sibs' scores have mean zero[6]. Also let the covariance between any two sibs in different families in the same csp ("cousins") be

$$cov(x_{ijl}, x_{ij'l'}) = V,$$

$j \neq j'$. The covariance structure within csp i can thus be written succinctly as

$$cov(\mathbf{x}_i) = cov\left([x_{i11}, x_{i12}, ..., x_{ica}]'\right) = \mathbf{I}_c \otimes (\mathbf{C} - V\mathbf{J}_a) + V\mathbf{J}_{ac}$$

where $\mathbf{J}_k, k \in \mathbb{N}$ is a square matrix of unities of dimension k.

We mention briefly that the data can also be represented using a random effects general linear models framework. If G_i are random effects having zero mean and variance V, representing the effects of the grandparents, then the data may be written as

$$\begin{aligned} x_{ijk} &= \mu + G_i + e_{ijk}, \\ i &= 1, 2, ..., n;\ j = 1, 2, ..., c;\ k = 1, 2, ..., a. \end{aligned}$$

The (circular) specifications for the covariance structure of the errors e_{ijk} would, in this case, be that

$$\begin{aligned} cov\left(\mathbf{e}_{ij}\right) &= cov\left([e_{ij1}, e_{ij2}, ...e_{ija}]'\right) = \mathbf{C} - V\mathbf{J} \text{ and} \\ cov\left(\mathbf{e}_{ij}, \mathbf{e}_{i'j'}\right) &= \mathbf{0},\ i \neq i' \text{ or } j \neq j', \end{aligned}$$

[6] No generality is lost because, if estimation is by maximum likelihood, we may at any time adjust the data so that they are deviations from the ML estimates for the mean parameters.

making use of the $\mathbf{C}$ defined in (26). General linear models could be constructed, if desired, corresponding to the assumptions of the subsequent subsections of this section, as well. These models would be analogous to those employed by Khattree and Naik (1994b).

Reduction of the data of this subsection to the canonical variables $\mathbf{y}_i$, $i = 1, 2, ..., n$ by the one to one transformation $\mathbf{y}_i = (\mathbf{I}_c \otimes \mathbf{\Gamma})\, \mathbf{x}_i$ is convenient because $\mathbf{\Gamma}$ reduces $\mathbf{C}$ to diagonal form

$$\mathbf{\Lambda} = \begin{cases} diag(\delta_1, \delta_2, ..\delta_{m-1}, \delta_m, ...\delta_2), & a = 2m, \\ diag(\delta_1, \delta_2, ..\delta_{m-1}, \delta_{m+1}, \delta_m, ...\delta_2), & a = 2m+1. \end{cases}$$

Noting all the rows (columns) of $\mathbf{\Gamma}$ but the first are orthogonal to $\mathbf{1}$, the covariance of each $\mathbf{y}_i$ is

$$\begin{aligned} cov(\mathbf{y}_i) &= (\mathbf{I}_c \otimes \mathbf{\Gamma})\,(\mathbf{I}_c \otimes (\mathbf{C} - V\mathbf{J}_a) + V\mathbf{J}_{ac})\,(\mathbf{I}_c \otimes \mathbf{\Gamma}) \\ &= \mathbf{I} \otimes (\mathbf{\Lambda} - aV\mathbf{f}\mathbf{f}') + aV\,(\mathbf{1} \otimes \mathbf{f})\,(\mathbf{1}' \otimes \mathbf{f}') \end{aligned}$$

where $\mathbf{f}' = (1, 0, ..., 0)$ is $1 \times a$. The covariance of each family's set of canonical variables, excluding the first, is thus

$$cov\left([y_{ij2}, y_{ij3}, ..., y_{ija}]'\right) = \begin{cases} diag(\delta_2, \delta_3, ..\delta_{m-1}, \delta_m, ...\delta_2), & a = 2m, \\ diag(\delta_2, \delta_3, ..\delta_{m-1}, \delta_{m-1}, \delta_m, ...\delta_2), & a = 2m+1, \end{cases}$$

and the $(y_{ij2}, y_{ij3}, ..., y_{ija})$ are independent of each other and of all the y_{ij1}. Assuming normality, the MLE of $(\delta_2, \delta_3, ..., \delta_{m-1})$ can consequently be obtained using the methods of Olkin and Press (1969). On the other hand, the first canonical variables of the families of each csp have a compound symmetry structure:

$$cov\left([y_{i11}, y_{i21}, ..., y_{ic1}]'\right) = \theta_1 \mathbf{I}_c + \theta_2 \mathbf{J}_c$$

where $\theta_1 = \delta_1 - aV$ and $\theta_2 = aV$. Letting $p = \sum_i \left(\sum_{j=1}^c y_{ij1}\right)^2$ and $s = \sum_{i,j} y_{ij1}^2$, and using the well known ML estimator of compound symmetry covariance parameters given balanced data, unbiased estimators (which will be MLE assuming normality) of (θ_1, θ_2) are (is)

$$\left(\hat{\theta}_1, \hat{\theta}_2\right) = \left(\frac{sc - p}{nc(c-1)}, \frac{p - s}{nc(c-1)}\right), \tag{27}$$

implying unbiased estimators (MLE, assuming normality) of (V, δ_1) are (is)

$$\left(\hat{V}, \hat{\delta}_1\right) = \left(a^{-1}\hat{\theta}_2, \hat{\theta}_1 + a\hat{V}\right) = \left(\frac{p - s}{anc(c-1)}, \frac{s}{nc}\right). \tag{28}$$

The expression for $\hat{\delta}_1$ is especially noteworthy because it is the same as the MLE Olkin and Press identified. Thus it is shown, in this balanced case, that the possibility of correlation between cousins does not change ML estimation of δ. The same cannot be asserted for the unbalanced cases below.

It can be shown (see Lehmann, 1983, p. 197) that, assuming normality, the values in (27) represent the uniform minimum variance unbiased (UMVU) estimators of (θ_1, θ_2). Using a theorem in Lehmann (1983, p. 77), it is seen that (28) consequently gives the UMVU estimators of (V, δ_1).

If the mean structure is generalized so that the normally distributed sibs have a common, unknown mean, the UMVU estimators of (θ_1, θ_2) become

$$\left(\hat{\theta}_1, \hat{\theta}_2\right) = \left(\frac{cs - p}{nc(c-1)}, c^{-1}\left[\frac{p - n^{-1}w_{\bullet\bullet}^2}{c(n-1)} - \frac{cs - p}{nc(c-1)}\right]\right).$$

implying the UMVU estimators of (V, δ_1) are

$$\left(\hat{V}, \hat{\delta}_1\right) = \left(\frac{1}{ac}\left[\frac{p - n^{-1}w_{\bullet\bullet}^2}{c(n-1)} - \frac{cs - p}{nc(c-1)}\right], \frac{p - w_{\bullet\bullet}^2}{c^2 n(n-1)} + \frac{s}{nc}\right).$$

Most of the results of the subsequent subsections of this section can be viewed as generalizations of (28), with the addition in some cases of a parent's score to each family's vector of observations, and a grandparent's score to each csp's vector of observations, under various assumptions about the covariances between members of different families.

4.3 Unequal Family and Cousinship Sizes with no Scores of Parents or Grandparents . In this subsection, let the distinct family sizes be $a_1 < a_2 < \ldots < a_c$, and let csp i have b_{ij} families of size a_j, allowing the possibilitiy that some of the b_{ij} may be zero. We assume, placing the sibs in any family on a circle whose circumference is proportional to the family size, that the covariance between any two sibs is a function of the length of the shorter arc between them, and that this function does not change with family size. E.g., families with three and four sibs have covariance structures

$$\begin{bmatrix} \sigma_0 & \sigma_1 & \sigma_1 \\ . & \sigma_0 & \sigma_1 \\ symm. & . & \sigma_0 \end{bmatrix} \text{ and } \begin{bmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \sigma_1 \\ . & \sigma_0 & \sigma_1 & \sigma_2 \\ . & . & \sigma_0 & \sigma_1 \\ symm. & . & . & \sigma_0 \end{bmatrix},$$

respectively, with the parameters $\boldsymbol{\sigma}' = (\sigma_0, \sigma_1, \sigma_2)$ constant for both family sizes. Retain the assumption that any two cousins have covariance V.

Let x_{ijkl} be the l^{th} observation, and y_{ijkl} be the l^{th} canonical variable, in the k^{th} family having a_j sibs in csp i. The vector $\mathbf{y}_{ijk}$ of canonical variables from this family is obtained by premultiplying the vector $\mathbf{x}_{ijk}$ of sibs' scores by $\boldsymbol{\Gamma}_j$, an orthogonal matrix whose elements and dimensions are derived using an obvious generalization of the $\boldsymbol{\Gamma}$ matrix given in Subsection (3.2), depending on the family size a_j.

Under these assumptions and generalizing the results of the last subsection, setting $\mathbf{z}'_{ijk} = (y_{ijk2}, y_{ijk3}, \dots, y_{ijka_j})$ implies

$$cov(\mathbf{z}_{ijk}) = \begin{cases} diag(\delta_{j2}, \delta_{j3}, ..\delta_{j,m_j-1}, \delta_{jm_j}, ...\delta_{j2}), & a_j = 2m_j, \\ diag(\delta_{j2}, \delta_{j3}, ..\delta_{j,m_j-1}, \delta_{j,m_j-1}, \delta_{jm_j}, ...\delta_{j2}), & a_j = 2m_j + 1. \end{cases}$$

Here, $m_j = int(a_j/2)$ and $\delta_{jl}, l = 1, 2, 3, \dots, m_j + 1$ is the l^{th} distinct eigenvalue of the covariance matrix $\mathbf{C}_j$ pertaining to the observations in a family having a_j sibs. Following the development of the last subsection, the $\mathbf{z}_{ijk}$ are independent of each other and of

$$\begin{aligned} \mathbf{w}'_i &= (y_{i111}, y_{i121}, \dots, y_{i1b_{i1}1}, \dots, y_{ic11}, y_{ic21}, \dots y_{icb_{ic}1}) \\ &= (w_{i11}, w_{i12}, \dots, w_{i1b_{i1}}, \dots, w_{ic1}, w_{ic2}, \dots, w_{icb_{ic}}). \end{aligned}$$

To express the covariance structure more specifically,

$$\begin{aligned} cov(\mathbf{y}_i) &= block\left[\mathbf{I}_{b_{i1}} \otimes (\boldsymbol{\Lambda}_1 - a_1 V \mathbf{f}_1 \mathbf{f}'_1), \dots, \mathbf{I}_{b_{ic}} \otimes (\boldsymbol{\Lambda}_c - a_c V \mathbf{f}_c \mathbf{f}'_c)\right] \\ &\quad + V \begin{bmatrix} \sqrt{a_1} \mathbf{1}_{b_{i1}} \otimes \mathbf{f}_1 \\ \sqrt{a_2} \mathbf{1}_{b_{i2}} \otimes \mathbf{f}_2 \\ \vdots \\ \sqrt{a_c} \mathbf{1}_{b_{ic}} \otimes \mathbf{f}_c \end{bmatrix} \begin{bmatrix} (\sqrt{a_1} \mathbf{1}'_{b_{i1}} \otimes \mathbf{f}'_1) & (\sqrt{a_2} \mathbf{1}'_{b_{i2}} \otimes \mathbf{f}'_2) & \cdots & (\sqrt{a_c} \mathbf{1}'_{b_{ic}} \otimes \mathbf{f}'_c) \end{bmatrix} \end{aligned}$$

where $\mathbf{f}_j' = (1, 0, ..., 0)$ is $a_j \times 1$ and $\mathbf{\Lambda}_j = \mathbf{\Gamma}_j \mathbf{C}_j \mathbf{\Gamma}_j = cov\left[(y_{ijk1}, \mathbf{z}_{ijk}')\right]$ is diagonal.

Put $M = \max_j(m_j)$. If $j \neq j'$, then δ_{jl} and $\delta_{j'l'}$ are different functions of the covariance parameters $\boldsymbol{\sigma}' = (\sigma_0, \sigma_1, \sigma_2, ..., \sigma_M)$, and so their within-family-size estimators (MLE, given normality) resulting from manipulating the $\mathbf{z}_{ijk}$ in the usual way, cannot be immediately combined across groups of families of different sizes to obtain overall estimators. The last part of this subsection describes a method of combining instead the within-family-size estimators of $\boldsymbol{\sigma}$.

Given normality, the MLE of V and of the first eigenvalues $(\delta_{11}, \delta_{21}, ...\delta_{c1})$ is obtained using the $\mathbf{w}_i$, each of which has covariance defined by

$$cov(w_{ijk}, w_{ij'k'}) = \begin{cases} \delta_{j1}, j = j' \text{ and } k = k', \\ V\sqrt{a_j a_{j'}}, j \neq j' \text{ or } k \neq k'. \end{cases}$$

Expressed in matrix form, this structure is

$$\begin{aligned} cov(\mathbf{w}_i) &= diag\left(f_1^{-1}\mathbf{1}_{b_{i1}}', f_2^{-1}\mathbf{1}_{b_{i2}}', ..., f_c^{-1}\mathbf{1}_{b_{ic}}'\right) \\ &\quad + V \begin{bmatrix} \sqrt{a_1}\mathbf{1}_{b_{i1}} \\ \sqrt{a_2}\mathbf{1}_{b_{i2}} \\ \vdots \\ \sqrt{a_c}\mathbf{1}_{b_{ic}} \end{bmatrix} \begin{bmatrix} \sqrt{a_1}\mathbf{1}_{b_{i1}}' & \sqrt{a_2}\mathbf{1}_{b_{i2}}' & \cdots & \sqrt{a_c}\mathbf{1}_{b_{ic}}' \end{bmatrix}, \end{aligned} \tag{29}$$

letting $f_j = (\delta_{j1} - a_j V)^{-1}$, $j = 1, 2, ..., c$. The loglikelihood given $\left(\mathbf{w}_1', \mathbf{w}_2', ..., \mathbf{w}_n'\right)$ is the sum of the loglikelihoods given each of the $\mathbf{w}_i$, the latter of which for each i involves the determinant and inverse of $cov(\mathbf{w}_i)$. Setting $t_i = \sum_{j=1}^{c} b_{ij} a_j f_j$, the results

$$\begin{aligned} \left|\mathbf{A} + \mathbf{u}\mathbf{u}'\right| &= |\mathbf{A}|\left(1 + \mathbf{u}'\mathbf{A}^{-1}\mathbf{u}\right) \text{ and} \\ \left[\mathbf{A} + V\mathbf{u}\mathbf{u}'\right]^{-1} &= \mathbf{A}^{-1} - \frac{V}{1 + V\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}}\mathbf{A}^{-1}\mathbf{u}\mathbf{u}'\mathbf{A}^{-1} \end{aligned}$$

(Henderson and Searle, 1981) give

$$\begin{aligned} |cov(\mathbf{w}_i)| &= \left[\prod_{j=1}^{c} f_j^{-b_{ij}}\right][1 + Vt_i] \text{ and} \\ [cov(\mathbf{w}_i)]^{-1} &= diag\left(f_1\mathbf{1}_{b_{i1}}', f_2\mathbf{1}_{b_{i2}}', ..., f_c\mathbf{1}_{b_{ic}}'\right) - \end{aligned}$$

$$\frac{V}{1+Vt_i}\begin{bmatrix} f_1\sqrt{a_1}\mathbf{1}_{b_{i1}} \\ f_2\sqrt{a_2}\mathbf{1}_{b_{i2}} \\ \vdots \\ f_c\sqrt{a_c}\mathbf{1}_{b_{ic}} \end{bmatrix}\begin{bmatrix} f_1\sqrt{a_1}\mathbf{1}'_{b_{i1}} & f_2\sqrt{a_2}\mathbf{1}'_{b_{i2}} & \cdots & f_c\sqrt{a_c}\mathbf{1}'_{b_{ic}} \end{bmatrix},$$

so that in the likelihood involving $\mathbf{w}_i$,

$$\begin{aligned} & \mathbf{w}_i' \left[cov(\mathbf{w}_i)\right]^{-1} \mathbf{w}_i \\ = & \sum_{j=1}^{c} f_j \sum_{k=1}^{b_{ij}} w_{ijk}^2 - \frac{V}{1+Vt_i}\left[\sum_{j=1}^{c} \sqrt{a_j} f_j w_{ij\bullet}\right]^2 . \end{aligned}$$

Assembling these elements, we have

$$\begin{aligned} l_i &= \ln L(\mathbf{w}_i) \\ &= const - \frac{1}{2}\sum_{j=1}^{c} b_{ij}\ln(f_j^{-1}) - \frac{1}{2}\ln\left[1+Vt_i\right] \\ & -\frac{1}{2}\sum_{j=1}^{c} f_j \sum_{k=1}^{b_{ij}} w_{ijk}^2 + \frac{V}{2(1+Vt_i)}\left[\sum_{j=1}^{c}\sqrt{a_j} f_j w_{ij\bullet}\right]^2 . \end{aligned}$$

We now prepare to obtain the score function and information of $\boldsymbol{\theta}' = (V, \delta_{11}, \delta_{21}, ..., \delta_{c1})$. Iterations of the Fisher Scoring algorithm $\hat{\boldsymbol{\theta}}^{(t-1)} = \hat{\boldsymbol{\theta}}^{(t)} + \mathbf{I}^{-1}\left(\hat{\boldsymbol{\theta}}^{(t)}\right) S\left(\hat{\boldsymbol{\theta}}^{(t)}\right)$ can then begin with suitable initial estimates $\hat{\boldsymbol{\theta}}^{(1)}$ and, assuming the uniqueness of the roots of the score equations, will converge with probability one to the MLE $\hat{\boldsymbol{\theta}}$. For shortness of notation, call the following random and nonrandom sums (conditional on the a_j and b_{ij})

$$\begin{aligned} s_{ij} &= \sum_{k=1}^{b_{ij}} w_{ijk}^2,\ p_{2i} = \sum_j \sqrt{a_j} w_{ij\bullet} f_j,\ p_{3i} = t_i + V\sum_j b_{ij} a_j^2 f_j^2,\ p_{4i} = \sum_j a_j^{3/2} w_{ij\bullet} f_j^2, \\ p_{5i} &= \sum_j a_j^{5/2} w_{ij\bullet} f_j^3,\ p_{6i} = \sum_j a_j^2 s_{ij} f_j^3,\ p_{7i} = \sum_j b_{ij} a_j^2 f_j^2,\ p_{8i} = \sum_j b_{ij} a_j^3 f_j^3, \\ p_{9i} &= \sum_j a_j s_{ij} f_j^2,\ t_i = \sum_j b_{ij} a_j f_j. \end{aligned}$$

Using these sums, the first and second partial derivatives of l_i can be specified as in Appendix A.

Before computing the expectation of the Hessian of csp i, let us find the expectations of the sums s_{ij}, and the required (random) squares and crossproducts among p_{1i} through p_{9i}. We first take advantage of the fact that the expected squares and cross products of p_{2i}, p_{4i} and p_{5i} are of the form

$$\mathbf{E}\left[\left(\sum_j a_j^e w_{ij\bullet} f_j^{e+1/2}\right)\left(\sum_j a_j^d w_{ij\bullet} f_j^{d+1/2}\right)\right]$$

for the values $d, e = 1/2$, $3/2$ and $5/2$. To find these expectations, notice that

$$\begin{aligned} \mathbf{E}\left(w_{ij\bullet}^2\right) &= b_{ij}(\delta_{j1} + Va_j(b_{ij}-1)) \text{ and, for } j \neq k, \\ \mathbf{E}\left(w_{ij\bullet}w_{ik\bullet}\right) &= Vb_{ij}b_{ik}\sqrt{a_j a_k}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathbf{E}\left[\left(\sum_j a_j^e w_{ij\bullet} f_j^{e+1/2}\right)\left(\sum_j a_j^d w_{ij\bullet} f_j^{d+1/2}\right)\right] \\ &= \sum_{k=1}^{c} a_k^{e+1/2} f_k^{e+1/2} b_{ik} \left[\begin{array}{c} a_k^{d-1/2} f_k^{d+1/2}(\delta_{k1} + Va_k(b_{ik}-1)) \\ +V\sum_{l\neq k} a_l^{d+1/2} f_l^{d+1/2} b_{il} \end{array}\right]. \end{aligned}$$

Using this general form, the expectations of the required squares and cross products of p_{2i}, p_{4i} and p_{5i} are

$$\begin{aligned} p_{22i} &= \mathbf{E}(p_{2i}^2) = \sum_{k=1}^{c} a_k f_k b_{ik}\left[f_k(\delta_{k1} + Va_k(b_{ik}-1)) + V\sum_{l\neq k} a_l f_l b_{il}\right], \\ p_{24i} &= \mathbf{E}(p_{2i}p_{4i}) = \sum_{k=1}^{c} a_k f_k b_{ik}\left[a_k f_k^2(\delta_{k1} + Va_k(b_{ik}-1)) + V\sum_{l\neq k} a_l^2 f_l^2 b_{il}\right], \\ p_{44i} &= \mathbf{E}(p_{4i}^2) = \sum_{k=1}^{c} a_k^2 f_k^2 b_{ik}\left[a_k f_k^2(\delta_{k1} + Va_k(b_{ik}-1)) + V\sum_{l\neq k} a_l^2 f_l^2 b_{il}\right], \\ p_{25i} &= \mathbf{E}(p_{2i}p_{5i}) = \sum_{k=1}^{c} a_k f_k b_{ik}\left[a_k^2 f_k^3(\delta_{k1} + Va_k(b_{ik}-1)) + V\sum_{l\neq k} a_l^3 f_l^3 b_{il}\right]. \end{aligned}$$

Next, $\mathbf{E}(w_{ij\bullet}p_{ki})$ for $k = 2, 4$ are of the form

$$\mathbf{E}\left(w_{ij\bullet}p_{ki}\right) = b_{ij}\left[a_j^{\frac{k-1}{2}} f_j^{k/2}\left(\delta_{j1} + Va_j(b_{ij}-1)\right) + V\sqrt{a_j}\sum_{l\neq j} a_l^{k/2} f_l^{k/2} b_{il}\right],$$

implying

$$\begin{aligned} p_{12ij} &= \mathbf{E}\left(w_{ij\bullet}p_{2i}\right) = b_{ij}\sqrt{a_j}\left[f_j\left(\delta_{j1} + Va_j(b_{ij}-1)\right) + V\sum_{k\neq j} a_k f_k b_{ik}\right] \text{ and} \\ p_{14ij} &= \mathbf{E}\left(w_{ij\bullet}p_{2i}\right) = b_{ij}\sqrt{a_j}\left[a_j f_j^2\left(\delta_{j1} + Va_j(b_{ij}-1)\right) + V\sum_{k\neq j} a_k^2 f_k^2 b_{ik}\right]. \end{aligned}$$

Also,

$$\mathbf{E}(s_{ij}) = b_{ij}\delta_{j1}, \text{ and } p_{16i} = \mathbf{E}(p_{6i}) = \sum_j a_j^2 f_j^3 b_{ij}\delta_{j1}.$$

Once initial estimates of the δ_{j1} and V are obtained, all these sums allow estimation of the elements of the expected values of the second partial derivatives that contribute to the information $\mathbf{I}(\boldsymbol{\theta})$ about $\boldsymbol{\theta}$. The elements of $-\mathbf{I}(\boldsymbol{\theta})$ are given by the terms in Appendix A, summing over all i.

Now that $\mathbf{I}(\boldsymbol{\theta}) = \mathbf{E}\left(-\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)$ and $S(\boldsymbol{\theta}) = \frac{\partial l}{\partial\boldsymbol{\theta}}$ have in effect been specified (since the entire loglikelihood is $l = \sum_{i=1}^n l_i$) it remains to suggest suitable initial estimates $\hat{\boldsymbol{\theta}}^{(1)}$ of $\boldsymbol{\theta}$. For a starting value of V, let $\mathbf{A}_i = \mathbf{J}_{b_{i\bullet}} - \mathbf{I}_{b_{i\bullet}}$. Using the expression (29) of $cov(\mathbf{w}_i)$ and the assumption that all sibs have mean zero, $\mathbf{E}(\mathbf{w}_i'\mathbf{A}_i\mathbf{w}_i) = V\left(\left(\sum_{j=1}^c b_{ij}\sqrt{a_j}\right)^2 - \sum_j a_j b_{ij}\right)$, which conveniently does not involve any of the δ_{j1}. So, a commonsense, unbiased initial estimate of V is, combining the data from all csp's,

$$v^{(1)} = \frac{\sum_i \mathbf{w}_i'(\mathbf{J}_{b_{i\bullet}} - \mathbf{I}_{b_{i\bullet}})\mathbf{w}_i}{\sum_i\left(\left(\sum_{j=1}^c b_{ij}\sqrt{a_j}\right)^2 - \sum_j a_j b_{ij}\right)}.$$

A reasonable initial estimate of each δ_{j1} is $\hat{\delta}_{j1}^{(1)} = b_{\bullet j}^{-1}\sum_{i=1}^n\sum_{l=1}^{b_{ij}} y_{ijl1}^2$, since each y_{ijl1}^2 is unbiased for δ_{j1}. We then start the FS iterations with $\hat{\boldsymbol{\theta}}^{(1)\prime} = \left(v^{(1)}, \hat{\delta}_{11}^{(1)}, \hat{\delta}_{21}^{(1)}, ..., \hat{\delta}_{c1}^{(1)}\right)$. Simulations showed that the iterations converge to the actual values $\boldsymbol{\theta}$, over a wide range of parameter values and sample sizes.

Grouping all families by size, the methods of Olkin and Press (1969) yield within-group (ML, given normality) estimates $\left(\hat{\delta}_{j2}, \hat{\delta}_{j3}, ..., \hat{\delta}_{j,m_j+1}\right)$ of the sets of family-size-specific $\boldsymbol{\delta}_j'$, $j = 1, 2, ..., c$ of the c distinct circular covariance matrices[7]. As already stated, for $j \neq j'$, δ_{jk} and $\delta_{j'k'}$ are different functions of the covariance parameters $\boldsymbol{\sigma}$. Therefore, within-group estimates of δ_{jk} and $\delta_{j'k'}$ cannot be directly combined to form overall estimates of the eigenvalues of any specific $\mathbf{C}_j$. The method we propose below for obtaining overall estimates of $\boldsymbol{\sigma}$ is to calculate instead

[7] Despite the segregation here of families according to family size, and thus the increase of parameters (i.e., eigenvalues) to be estimated, there is no chance the number of parameters will exceed the number of observations. Even if all $b_{\bullet j}$ are 1, meaning the number of families is equal to the number of family sizes and the number of observations is equal to $\sum_j a_j$, the number of parameters is only $1 + m_j + 1$ (the first 1 applying to v) in the j^{th} group and $1 + \sum_j (m_j + 1)$ overall. Also, we mention that the methods of Olkin and Press may be applied to each group of families because the canonical variables in $\mathbf{z}_{ijk}$ are mutually independent, even within the same csp.

within-group estimates of $\boldsymbol{\sigma}$, and combine these last estimates across groups according to the estimated information about each σ_l within each group. This procedure minimizes the asymptotic variance of the estimate of each σ_l; hence, it minimizes the trace of the asymptotic covariance of the $(M+1)$-variate estimate of $\boldsymbol{\sigma}$, letting $M = \max_j m_j$, where $m_j = m(a_j) = int(a_j/2)$, $j = 1, 2, ..., c$.

For the remainder of this subsection, assume normality and fix $l \in \{0, 1, 2, ..., M\}$, i.e., suppress the subscript of σ_l. Let $\hat{\boldsymbol{\delta}}_j'$ be the ML estimate of $\boldsymbol{\delta}_j'$, and $e_j = \hat{\boldsymbol{\delta}}_j'\mathbf{B}_{jl}^{-1}$ be the (unbiased) MLE of $\sigma = \sigma_l$ from all families of size a_j, where $\mathbf{B}_{jl}^{-1}$ is the l^{th} column of the inverse of the $\mathbf{B}$ matrix (as identified in Section 1) corresponding to family size a_j. Put $h = \min\{j : m_j \geq l, b_{\bullet j} > 0\}$, so that a_h is the smallest family size observed in the data large enough to provide any information about σ. We obtain the minimum asymptotic variance consistent estimator $\tilde{\sigma}$ as a linear combination of the within-group estimators in $\mathbf{e}' = (e_h, e_{h+1}, ..., e_c)$ of σ.

The e_j are within-group MLEs of σ. Minimizing the asymptotic variance of the linear combination $\mathbf{v}'\mathbf{e}$ of them through the choice of $\mathbf{v}$ while keeping $\mathbf{v}'\mathbf{1}$=1, is achieved at $\mathbf{v} = \frac{(cov(\mathbf{e}))^{-1}\mathbf{1}}{\mathbf{1}'(cov(\mathbf{e}))^{-1}\mathbf{1}}$. The diagonal elements of $cov(\mathbf{e})$ are, for $j = h, h+1, ..., c$,

$$\begin{aligned} var(e_j) &\doteq \left(\mathbf{B}_{jl}^{-1}\right)'\left[cov\left(\hat{\boldsymbol{\delta}}_j\right)\right]\mathbf{B}_{jl}^{-1} \\ &= \left(\mathbf{B}_{jl}^{-1}\right)' diag\left(\left(\mathbf{I}^{-1}(\boldsymbol{\theta})\right)_{j+1,j+1}, \frac{2\delta_{j2}^2}{b_{\bullet j}\alpha_{j2}}, \frac{2\delta_{j3}^2}{b_{\bullet j}\alpha_{j3}}, ..., \frac{2\delta_{j,m_j+1}^2}{b_{\bullet j}\alpha_{j,m_j+1}}\right)\mathbf{B}_{jl}^{-1}, \end{aligned}$$

letting α_{jk} be the multiplicity of δ_{jk} in $|\mathbf{C}_j|$ and continuing the use of the parameter $\boldsymbol{\theta}' = (V, \delta_{11}, \delta_{21}, ..., \delta_{c1})$, whose negative information was found in Appendix A.

If $j \neq k$ and $j, k \geq h$, the $(j-h+1, k-h+1)$ element of $cov(\mathbf{e})$ is

$$\begin{aligned} cov(e_j, e_k) &= cov\left(\hat{\boldsymbol{\delta}}_j'\mathbf{B}_{jl}^{-1}, \hat{\boldsymbol{\delta}}_k'\mathbf{B}_{kl}^{-1}\right) \\ &= \left(\mathbf{B}_{jl}^{-1}\right)' cov\left(\hat{\boldsymbol{\delta}}_j, \hat{\boldsymbol{\delta}}_k\right)\mathbf{B}_{kl}^{-1} \end{aligned}$$

where

$$cov\left(\hat{\delta}_j,\hat{\delta}_k\right)=\begin{bmatrix} cov\left(\hat{\delta}_{j1},\hat{\delta}_{k1}\right) & cov\left(\hat{\delta}_{j1},\hat{\delta}_{k2}\right) & \cdots & cov\left(\hat{\delta}_{j1},\hat{\delta}_{k,m_k-1}\right) \\ cov\left(\hat{\delta}_{j2},\hat{\delta}_{k1}\right) & cov\left(\hat{\delta}_{j2},\hat{\delta}_{k2}\right) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ cov\left(\hat{\delta}_{j,m_j-1},\hat{\delta}_{k1}\right) & cov\left(\hat{\delta}_{j,m_j-1},\hat{\delta}_{k2}\right) & \cdots & cov\left(\hat{\delta}_{j,m_j-1},\hat{\delta}_{k,m_k-1}\right) \end{bmatrix}.$$

However, $cov\left(\hat{\delta}_{jg},\hat{\delta}_{kg'}\right)=0$, unless $g=g'=1$ in which case $cov\left(\hat{\delta}_{jg},\hat{\delta}_{kg'}\right)\doteq\left(\mathbf{I}^{-1}(\theta)\right)_{j-1,k-1}$. Therefore, if $j\neq k$,

$$cov\left(e_j,e_k\right)=\left(\mathbf{B}_{jl}^{-1}\right)_1\left(\mathbf{B}_{kl}^{-1}\right)_1\left(\mathbf{I}^{-1}(\theta)\right)_{j-1,k-1}.$$

Combining the $cov\left(e_j,e_k\right)$ for all $j,k\geq h$ yields the elements of $cov(\mathbf{e})$. Concluding, we have that $\bar{\sigma}=\bar{\sigma}_l=\frac{\mathbf{1}'(cov(\mathbf{e}))^{-1}\mathbf{e}}{\mathbf{1}'(cov(\mathbf{e}))^{-1}\mathbf{1}}$, $l=0,1,...,M$, using the final estimate of $cov(\mathbf{e})$, is both consistent (though not unbiased, due to the biased estimators of the δ_{j1}) and asymptotically efficient, being the minimum variance linear combination of (asymptotically efficient) within-group MLEs $\mathbf{e}$.

4.4 Uniform Family and Cousinship Sizes with Scores of Parents. In this subsection, again invoke the restriction that each family has a sibs and each csp has c families, assuming normality throughout. To the $(i,j)^{th}$ family's vector $(x_{ij1},...,x_{ija})$ of sibs' scores, append a parent's score p_{ij} having mean zero and variance α. Each csp's scores can now be represented by a cA (with $A=a+1$) vector of observations

$$\mathbf{x}_i'=\left(p_{i1},x_{i11},x_{i12},...,x_{i1a},...,p_{ic},x_{ic1},x_{ic2},...,x_{ica}\right).$$

Let β be the parent-sib covariance, and let V be the covariance between any two observations from different families in the same csp. If

$$(\mathbf{C})_{jl}=\begin{cases}\sigma_{|j-l|}, |j-l|\leq a,\\ \sigma_{a-|j-l|}, |j-l|>a,\end{cases}$$

then the covariance structure in each family is $\mathbf{C}_{aug}=\begin{bmatrix}\alpha & \beta\mathbf{1}'\\ \beta\mathbf{1} & \mathbf{C}\end{bmatrix}$, and that in any csp is therefore $cov(\mathbf{x}_i)=\mathbf{I}_c\odot\left(\mathbf{C}_{aug}-V\mathbf{J}_A\right)+V\mathbf{J}_{cA}$.

Under these assumptions, if $\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}' = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Gamma} \end{bmatrix}$ where $\boldsymbol{\Gamma}$ is as specified in Subsection (4.2), then transformation of the data to canonical form $\mathbf{y}_i = [\mathbf{I}_c \otimes \boldsymbol{\Upsilon}]\,\mathbf{x}_i$ reduces the within-csp covariance to

$$\begin{aligned} cov(\mathbf{y}_i) &= [\mathbf{I}_c \otimes \boldsymbol{\Upsilon}]\,[\mathbf{I}_c \otimes (\mathbf{C}_{aug} - V\mathbf{J}_A) + V\mathbf{J}_{cA}]\,[\mathbf{I}_c \otimes \boldsymbol{\Upsilon}] \\ &= \mathbf{I}_c \otimes \left[\begin{bmatrix} \alpha & \beta\sqrt{a} & 0 & \cdots & 0 \\ \beta\sqrt{a} & \delta_1 & 0 & \vdots & \vdots \\ 0 & 0 & \delta_2 & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \delta_a \end{bmatrix} - V \begin{bmatrix} 1 & \sqrt{a} & 0 & \cdots & 0 \\ \sqrt{a} & a & 0 & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \right] \\ &\quad + V \left[\mathbf{1}_c \otimes \begin{bmatrix} 1 \\ \sqrt{a} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right] \left[\mathbf{1}_c' \otimes \begin{bmatrix} 1 & \sqrt{a} & 0 & \cdots & 0 \end{bmatrix} \right], \end{aligned}$$

making use of the fact that all rows (and columns) of $\boldsymbol{\Upsilon}$ but the first and second are orthogonal to $\mathbf{1}$. Therefore, for all (i,j), the third through A^{th} canonical variables $(y_{ij2}, y_{ij3}, \ldots, y_{ija})$ of the $(i,j)^{th}$ family have diagonal covariance, and are independent of each other and of $\mathbf{z}_{ij}' = (z_{ij1}, z_{ij2}) = \left(p_{ij}, a^{-1/2} y_{ij1}\right)$. On the other hand,

$$cov(\mathbf{z}_{ij}) = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \tag{30}$$

with $\gamma = \delta_1 / a$, whereas $cov(z_{ijk}, z_{ij'k'}) = V$ if $j \neq j'$. Defining $\mathbf{z}_i' = (\mathbf{z}_{i1}', \mathbf{z}_{i2}', \ldots, \mathbf{z}_{ic}')$, we have

$$cov\,(\mathbf{z}_i) = \mathbf{I}_c \otimes \left[\begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} - V\mathbf{J}_2 \right] + V\mathbf{J}_{2c}.$$

Thus, the MLE of α, β, V and δ_1 is found maximizing the likelihood involving $\mathbf{z}' = \left(\mathbf{z}_1', \mathbf{z}_2', \ldots, \mathbf{z}_n'\right)$ and, as in the last subsection, the MLE of the remaining δ_j is found maximizing the likelihood involving the canonical variables in $\mathbf{y}$ that were excluded from $\mathbf{z}$. Turning first to the likelihood involving each $\mathbf{z}_i$, the result $(\mathbf{A}+V\mathbf{uu}')^{-1} = \mathbf{A}^{-1} - (1+\mathbf{u}'\mathbf{A}^{-1}\mathbf{u})^{-1} V\mathbf{A}^{-1}\mathbf{uu}'\mathbf{A}^{-1}$ gives

$$
\begin{aligned}
(cov\,(\mathbf{z}_i))^{-1} &= d_v^{-1}\mathbf{I}_c\otimes \begin{bmatrix} \gamma - V & V-\beta \\ V-\beta & \alpha - V \end{bmatrix} \\
&\quad - \frac{Vd_v^{-2}}{1+Vcd_v^{-1}(\alpha+\gamma-2\beta)}\mathbf{J}_c\otimes \begin{bmatrix} (\gamma-\beta)^2 & (\gamma-\beta)(\alpha-\beta) \\ (\gamma-\beta)(\alpha-\beta) & (\alpha-\beta)^2 \end{bmatrix},
\end{aligned}
$$

where $d_v = (\alpha - V)(\gamma - V) - (V-\beta)^2$. We therefore have

$$
\begin{aligned}
\mathbf{z}_i'(cov\,(\mathbf{z}_i))^{-1}\mathbf{z}_i &= d_v^{-1}\mathbf{z}_i' \left\{ \begin{array}{c} \mathbf{I}_c\otimes \begin{bmatrix} \gamma - V & V-\beta \\ V-\beta & \alpha - V \end{bmatrix} - \frac{Vd_v^{-2}}{1+Vcd_v^{-1}(\alpha+\gamma-2\beta)}\times \\ \mathbf{J}_c\otimes \begin{bmatrix} (\gamma-\beta)^2 & (\gamma-\beta)(\alpha-\beta) \\ (\gamma-\beta)(\alpha-\beta) & (\alpha-\beta)^2 \end{bmatrix} \end{array} \right\} \mathbf{z}_i \\
&= d_v^{-1}\left[(\gamma-V)\sum_{j=1}^{c} z_{ij1}^2 + (\alpha-V)\sum_{j=1}^{c} z_{ij2}^2 + 2(V-\beta)\sum_{j=1}^{c} z_{ij1}z_{ij2}\right] \\
&\quad - \frac{V}{d_v^2+Vcd_v\theta}\left[(\gamma-\beta)^2 z_{i\bullet 1}^2 + (\alpha-\beta)^2 z_{i\bullet 2}^2 + 2(\gamma-\beta)(\alpha-\beta)z_{i\bullet 1}z_{i\bullet 2}\right].
\end{aligned}
$$

Next, it can be shown by induction that $|cov\,(\mathbf{z}_i)| = (d+(c-1)V\theta)(d-V\theta)^{c-1}$, where $d = \alpha\gamma - \beta^2$ and $\theta = \alpha+\gamma-2\beta$. Assembling the members of the loglikelihood involving $(\mathbf{z}_1', \mathbf{z}_2', ..., \mathbf{z}_n')$ gives

$$
\begin{aligned}
2l &= 2\ln L = \qquad\qquad (31) \\
&= const - n\ln\left\{(d_v + cV\theta)\, d_v^{c-1}\right\} \\
&\quad -d_v^{-1}\left\{ \begin{array}{c} (\gamma-V)\,s_{\bullet 11} + (\alpha-V)\,s_{\bullet 22} + 2(V-\beta)\,s_{\bullet 12} \\ -\frac{V}{d_v+Vc\theta}\left[(\gamma-\beta)^2 m_{\bullet 11} + (\alpha-\beta)^2 m_{\bullet 22} + 2(\gamma-\beta)(\alpha-\beta)\,m_{\bullet 12}\right] \end{array} \right\}
\end{aligned}
$$

where, shortening the notation, we set

$$
s_{\bullet kl} = \sum_{i,j} z_{ijk}z_{ijl} \text{ and } m_{\bullet kl} = \sum_{i} z_{i\bullet k}z_{i\bullet l},\ k,l = 1,2
$$

and reiterate that, in terms of the original parameters, the present parameters are

$$
\begin{aligned}
d &= \frac{\alpha\delta_1}{a} - \beta^2,\ \theta = \alpha + \frac{\delta_1}{a} - 2\beta,\ d_v = \frac{\alpha\delta_1}{a} - \beta^2 - V\left(\alpha + \frac{\delta_1}{a} - 2\beta\right), \\
\alpha, \beta \text{ and } \gamma &= \delta_1/a.
\end{aligned}
$$

(31) may be differentiated implicitly with respect to $(\alpha, \delta_1, \beta, V)$ to find the score and information of $(\alpha, \delta_1, \beta, V)$ required to perform Fisher iterations converging to the MLE of $(\alpha, \delta_1, \beta, V)$. However, the calculations are extremely detailed. Simplification is possible by transforming each $\mathbf{z}_{ij}$ to $\mathbf{u}_{ij} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{z}_{ij}$. If $\mathbf{u}_i' = (\mathbf{u}_{i1}', \mathbf{u}_{i2}', ..., \mathbf{u}_{ic}')$ for each i, the covariance of $\mathbf{u}_i$ is

$$\begin{aligned} cov(\mathbf{u}_i) &= \left\{ \mathbf{I}_c \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right\} \left\{ \mathbf{I}_c \otimes \left[\begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} - V\mathbf{J}_2 \right] + V\mathbf{J}_{2c} \right\} \left\{ \mathbf{I}_c \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right\} \\ &= \mathbf{I}_c \otimes \left\{ \begin{bmatrix} \tau & \sigma \\ \sigma & \phi \end{bmatrix} - \omega \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\} + \omega \mathbf{J}_c \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned}$$

putting $\begin{bmatrix} \tau & \sigma \\ \sigma & \phi \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $\omega = 4V$ so that the reverse transformation of parameters is

$$\begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \tau & \sigma \\ \sigma & \phi \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ and } V = \frac{\omega}{4}.$$

Thus, for each (i, j) and for $k, k' = 1, 2$,

$$cov(u_{ijk}, u_{ij'k'}) = \begin{cases} \tau, j = j', k = k' = 1, \\ \phi, j = j', k = k' = 2, \\ \sigma, j = j', k \neq k', \\ \omega, j \neq j', k = k' = 1, \\ 0, j \neq j', k = k' = 2, \\ 0, j \neq j', k \neq k'. \end{cases} \tag{32}$$

Deriving the likelihood involving each $\mathbf{u}_i$, we have that

$$|cov(\mathbf{u}_i)| = d^{c-1}(d + \omega c\phi) \text{ where } d = \left| \begin{bmatrix} \tau - \omega & \sigma \\ \sigma & \phi \end{bmatrix} \right| = (\tau - \omega)\phi - \sigma^2,$$

again using the result $|\mathbf{A}+\omega\mathbf{tt}'| = |\mathbf{A}|\left(1 + \omega\mathbf{t}'\mathbf{A}^{-1}\mathbf{t}\right)$. Also,

$$(\mathbf{A}+\omega\mathbf{tt}')^{-1} = \mathbf{A}^{-1} - \omega\left(1 + \mathbf{t}'\mathbf{A}^{-1}\mathbf{t}\right)^{-1}\mathbf{A}^{-1}\mathbf{tt}'\mathbf{A}^{-1}$$

gives

$$(cov(\mathbf{u}_i))^{-1} = \frac{1}{d}\left\{\mathbf{I}_c \otimes \begin{bmatrix} \phi & -\sigma \\ -\sigma & \tau-\omega \end{bmatrix} - \frac{\omega}{d+\omega c\phi}\mathbf{J}_c \otimes \begin{bmatrix} \phi^2 & -\phi\sigma \\ -\phi\sigma & \sigma^2 \end{bmatrix}\right\}.$$

Therefore, in the exponent of the likelihood,

$$\begin{aligned} & \mathbf{u}_i'(cov(\mathbf{u}_i))^{-1}\mathbf{u}_i \\ = \; & d^{-1}\left\{\mathbf{u}_i'\left[\mathbf{I}_c \otimes \begin{bmatrix} \phi & -\sigma \\ -\sigma & \tau-\omega \end{bmatrix}\right]\mathbf{u}_i' - \frac{\omega}{d+\omega c\phi}\mathbf{u}_i'\left[\mathbf{J}_c \otimes \begin{bmatrix} \phi^2 & -\phi\sigma \\ -\phi\sigma & \sigma^2 \end{bmatrix}\right]\mathbf{u}_i\right\} \\ = \; & d^{-1}\left\{\begin{array}{c} \sum_{j=1}^{c}\left[\phi u_{ij1}^2 + (\tau-\omega)u_{ij2}^2 - 2\sigma u_{ij1}u_{ij2}\right] \\ -\frac{\omega}{d+\omega c\phi}\left[\phi^2 u_{i\bullet 1}^2 + \sigma^2 u_{i\bullet 2}^2 - 2\phi\sigma u_{i\bullet 1}u_{i\bullet 2}\right] \end{array}\right\} \end{aligned}$$

and hence twice the entire loglikelihood is

$$\begin{aligned} 2l &= 2\sum_{i=1}^{n}\ln L(\mathbf{u}_i) \\ &= const - n\ln\left[d^{c-1}(d+\omega c\phi)\right] - d^{-1}\left\{\begin{array}{c} \phi q_{11} + (\tau-\omega)q_{22} - 2\sigma q_{12} \\ -\frac{\omega}{d+\omega c\phi}\left[\phi^2 r_{11} + \sigma^2 r_{22} - 2\phi\sigma r_{12}\right] \end{array}\right\}, \end{aligned}$$

where $q_{kl} = \sum_{i,j} u_{ijk}u_{ijl}$ and $r_{kl} = \sum_i u_{i\bullet k}u_{i\bullet l}$ for $k, l = 1, 2$.

Differentiating l with respect to $\boldsymbol{\theta}' = (\tau, \phi, \sigma, \omega)$ gives the elements of the Hessian as in Appendix A. To calculate the information about $\boldsymbol{\theta}$, we refer to the covariance structure identified in (32) and find the expectations of q_{kl} and r_{kl} :

$$\begin{aligned} \mathbf{E}(q_{11}) &= \mathbf{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij1}^2\right) = nc\tau,\ \mathbf{E}(q_{22}) = \mathbf{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij2}^2\right) = nc\phi, \\ \mathbf{E}(q_{12}) &= \mathbf{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij1}u_{ij2}\right) = nc\sigma,\ \mathbf{E}(r_{11}) = \mathbf{E}\sum_{i=1}^{n} u_{i\bullet 1}^2 = nc(\tau + (c-1)\omega), \\ \mathbf{E}(r_{22}) &= \mathbf{E}\left(\sum_{i=1}^{n} u_{i\bullet 2}^2\right) = nc\phi \text{ and } \mathbf{E}(r_{12}) = nc\sigma. \end{aligned}$$

Substituting these expectations for their respective random sums in the second derivatives in the Hessian and simplifying gives the elements of the information as specified in Appendix A.

Suitable unbiased initial estimators of $\boldsymbol{\theta}' = (\tau, \phi, \sigma, \omega)$ are

$$\hat{\tau}^{(1)} = (nc)^{-1}\sum_{i,j} u_{ij1}^2,\ \hat{\phi}^{(1)} = (nc)^{-1}\sum_{i,j} u_{ij2}^2,$$

$$\hat{\sigma}^{(1)} = (nc)^{-1}\sum_{i,j} u_{ij1}u_{ij2}, \text{ and } \hat{\omega}^{(1)} = \frac{2}{nc(c-1)}\sum_{i,j} u_{ij1}\sum_{l=j-1}^{c} u_{il1}.$$

Using the covariances specified in (32), it may be verified that each of the forgoing is unbiased for its corresponding parameter. Under the assumption that $V = 0$, the MLE of $\begin{bmatrix} \tau & \sigma \\ \sigma & \phi \end{bmatrix}$ would be $\begin{bmatrix} \hat{\tau}^{(1)} & \hat{\sigma}^{(1)} \\ \hat{\sigma}^{(1)} & \hat{\phi}^{(1)} \end{bmatrix}$; hence, $\begin{bmatrix} \hat{\tau}^{(1)} & \hat{\sigma}^{(1)} \\ \hat{\sigma}^{(1)} & \hat{\phi}^{(1)} \end{bmatrix}$ seem to be especially appropriate starting values. Simulations showed these estimation procedures produce consistent estimates of all the covariance parameters.

4.5 Uniform Family and Cousinship Sizes with Scores of Parents and Grandparents. In this subsection, most generally, we desire a setup that would be appropriate in modelling data from cousinships whose sibs' and parents' scores between families are stochastically related due to both common lineage and common environment. Consequently, let the assumptions of Subsection (4.4) hold, with the additional condition that a normally distributed grandparent's score g_i with mean zero and variance γ is appended to each csp's cA-vector (with $A = a+1$) of observations. Assume

$$cov\,(\mathbf{x}_{ij}) = cov\left([p_{ij}, x_{ij1}, x_{ij2}, ..., x_{ija}]'\right) = \mathbf{C}_{aug} = \begin{bmatrix} \alpha & \beta\mathbf{1}' \\ \beta\mathbf{1} & \mathbf{C} \end{bmatrix},$$

as in the last subsection, and $cov\,(g_i, \mathbf{x}_{ij}) = [\beta_2, \beta_1\mathbf{1}_a']$, so that the covariance in any csp can be represented by

$$cov\,(\mathbf{x}_i) = cov\left(\begin{bmatrix} g_i \\ \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \\ \vdots \\ \mathbf{x}_{ic} \end{bmatrix}\right) = \begin{bmatrix} \gamma & \mathbf{1}_c' \otimes \begin{bmatrix} \beta_2 & \beta_1\mathbf{1}_a' \end{bmatrix} \\ \mathbf{1}_c \otimes \begin{bmatrix} \beta_2 \\ \beta_1\mathbf{1}_a \end{bmatrix} & \mathbf{I}_c \otimes (\mathbf{C}_{aug} - V\mathbf{J}_A) + V\mathbf{J}_{cA} \end{bmatrix}.$$

Transformation of the $\mathbf{x}_i$ to $\mathbf{y}_i = \boldsymbol{\Upsilon}\mathbf{x}_i$, where

$$\boldsymbol{\Upsilon} = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_c \otimes \boldsymbol{\Psi} \end{bmatrix} \text{ and } \boldsymbol{\Psi} = \boldsymbol{\Psi}' = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Gamma} \end{bmatrix},$$

results in canonical variables $\mathbf{y}_i$ with a relatively simple covariance structure. Extending the results of Subsection (4.4), we have that the lower right and upper right submatrices of $cov(\mathbf{y}_i)$ are,

$$[\mathbf{I}_c \otimes \Psi][\mathbf{I}_c \otimes (\mathbf{C}_{aug} - V\mathbf{J}_A) + V\mathbf{J}_{cA}][\mathbf{I}_c \otimes \Psi]$$

$$= \left[\mathbf{I}_c \otimes \left[\begin{bmatrix} \alpha & \beta\sqrt{a} & 0 & \cdots & 0 \\ \beta\sqrt{a} & \delta_1 & 0 & \vdots & \vdots \\ 0 & 0 & \delta_2 & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \delta_a \end{bmatrix} - V \begin{bmatrix} 1 & \sqrt{a} & 0 & \cdots & 0 \\ \sqrt{a} & a & 0 & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}\right]\right.$$

$$\left. + V \left[\mathbf{1}_c \otimes \begin{bmatrix} 1 \\ \sqrt{a} \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right] \left[\mathbf{1}_c' \otimes \begin{bmatrix} 1 & \sqrt{a} & 0 & \cdots & 0 \end{bmatrix}\right]\right] \text{ and}$$

$$1\left\{\mathbf{1}_c' \otimes \begin{bmatrix} \beta_2 & \beta_1 \mathbf{1}_a' \end{bmatrix}\right\}[\mathbf{I}_c \otimes \Psi] = \mathbf{1}_c' \otimes \begin{bmatrix} \beta_2 & \beta_1\sqrt{a} & 0\mathbf{1}_{a-1}' \end{bmatrix}. \tag{33}$$

Consequently, as in previous subsections, the third through A^{th} canonical variables within each family have a diagonal covariance $diag(\delta_2, \delta_3, ..., \delta_a)$ and are independent of all other statistics and of each other, so that the methods of Olkin and Press (1969) may be used with them to find the MLE of the second through $(m+1)^{th}$ eigenvalues of $\mathbf{C}$.

Let $\mathbf{z}_{ij}' = (p_{ij}, a^{-1/2}y_{ij1})$ and $\mathbf{z}_i' = (g_i, \mathbf{z}_{i1}', \mathbf{z}_{i2}', ..., \mathbf{z}_{ic}')$. By extending the results of Subsection (4.4), it can be inferred that omitting the appropriate rows and columns from $cov(\mathbf{y}_i)$ and dividing the third, fifth, seventh,... rows and columns of the resulting, smaller matrix by $\sqrt{a}$ gives

$$cov(\mathbf{z}_i) = \begin{bmatrix} \gamma & \mathbf{1}_c' \otimes \begin{bmatrix} \beta_2 & \beta_1 \end{bmatrix} \\ \mathbf{1}_c \otimes \begin{bmatrix} \beta_2 \\ \beta_1 \end{bmatrix} & \mathbf{I}_c \otimes \left\{\begin{bmatrix} \alpha & \beta \\ \beta & \zeta \end{bmatrix} - V\mathbf{J}_2\right\} + V\mathbf{J}_{2c} \end{bmatrix}.$$

letting $\zeta = \delta_1/a$.

Further simplification of the covariance structure is possible through the transformations

$$\mathbf{u}_i = \begin{bmatrix} 1 & \mathbf{0}' \\ 0 & \mathbf{I}_c \otimes \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \end{bmatrix} \mathbf{z}_i,$$

because

$$cov(\mathbf{u}_i) = \begin{bmatrix} \gamma & \left\{ \mathbf{1}_c' \otimes \begin{bmatrix} \beta_2 - \beta_1 & \beta_1 \end{bmatrix} \right\} \\ \left\{ \mathbf{1}_c \otimes \begin{bmatrix} \beta_2 - \beta_1 \\ \beta_1 \end{bmatrix} \right\} & \mathbf{I}_c \otimes \left\{ \begin{bmatrix} \tau & \sigma \\ \sigma & \phi \end{bmatrix} \right\} + \mathbf{J}_c \otimes \begin{bmatrix} 0 & 0 \\ 0 & V \end{bmatrix} \end{bmatrix},$$

with

$$\begin{bmatrix} \tau & \sigma \\ \sigma & \phi \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha - V & \beta - V \\ \beta - V & \delta_1/a - V \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

so that the inverse transformation of parameters is

$$\begin{bmatrix} \alpha & \beta \\ \beta & \delta_1/a \end{bmatrix} = V\mathbf{J}_2 + \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tau & \sigma \\ \sigma & \phi \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Considerable simplification of $cov(\mathbf{u}_i)$ and therefore of the following calculations would be possible if $\beta_1 = \beta_2$; however, this condition seems excessively restrictive, as grandparents could seldom be expected to be related to their children in the same way they are related to their grandchildren.

Using the result $\begin{vmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{U}' & \mathbf{D} \end{vmatrix} = |\mathbf{D}|\,|\mathbf{A} - \mathbf{U}'\mathbf{D}^{-1}\mathbf{U}|$, and setting $d = \tau\phi - \sigma^2$ gives

$$|cov(\mathbf{u}_i)| = d^{c-1}(d + Vc\tau) \times \left\{ \begin{array}{c} \gamma - \frac{c}{d}\left[\phi(\beta_2 - \beta_1)^2 + \beta_1(2\sigma(\beta_1 - \beta_2) + \beta_1\tau)\right] \\ + \frac{c^2 V}{d(d+\omega c\phi)}\left[\sigma\beta_2 - \beta_1(\sigma + \tau)\right]^2 \end{array} \right\}.$$

The variation

$$\begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{U}' & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{U}\mathbf{D}^{-1}\mathbf{U}')^{-1} & -(\mathbf{A} - \mathbf{U}\mathbf{D}^{-1}\mathbf{U}')^{-1}\mathbf{U}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{U}'(\mathbf{A} - \mathbf{U}\mathbf{D}^{-1}\mathbf{U}')^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{U}'(\mathbf{A} - \mathbf{U}\mathbf{D}^{-1}\mathbf{U}')^{-1}\mathbf{U}\mathbf{D}^{-1} \end{bmatrix}$$

of an identity given by Henderson and Searle (1981) allows the derivation

$$[cov(\mathbf{u}_i)]^{-1} = \frac{f}{d(d+Vc\tau)} \times \tag{34}$$

$$\begin{bmatrix} d(d+Vc\tau) & -\mathbf{1}_c' \otimes \begin{bmatrix} e_1 & e_2 \end{bmatrix} \\ -\mathbf{1}_c \otimes \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} & \begin{matrix} \frac{d(d+Vc\tau)}{f}\mathbf{I}_c \otimes \begin{bmatrix} \phi & -\sigma \\ -\sigma & \tau \end{bmatrix} - \frac{V}{f}\mathbf{J}_c \otimes \begin{bmatrix} \sigma^2 & -\tau\sigma \\ -\tau\sigma & \tau^2 \end{bmatrix} \\ +\frac{1}{d(d+Vc\tau)}\mathbf{J}_c \otimes \begin{bmatrix} e_1^2 & e_1e_2 \\ e_1e_2 & e_2^2 \end{bmatrix} \end{matrix} \end{bmatrix},$$

setting

$$\begin{aligned} f &= \frac{d(d+Vc\tau)}{\gamma d(d+Vc\tau) - c\left\{ (d+Vc\tau) \begin{bmatrix} \phi(\beta_2-\beta_1)^2 \\ +\beta_1(\beta_1(\tau+2\sigma)-\beta_2) \end{bmatrix} - Vc(\sigma\beta_2 - \beta_1(\sigma+\tau))^2 \right\}}, \\ e_1 &= (\beta_2-\beta_1)\left[\phi d + Vc(\tau\phi - \sigma^2)\right] - \beta_1\sigma d, \\ e_2 &= d[\beta_1(\sigma+\tau) - \sigma\beta_2] \text{ and } d = \tau\phi - \sigma^2. \end{aligned}$$

In the unrestricted case under consideration now, iteratively maximizing the likelihood with respect to $(\gamma, \tau, \omega, \sigma, \phi, \beta_1, \beta_2)$ requires the inversion, at each step, of a 7×7 information matrix, having $\frac{7(7-1)}{2}$ =28 distinct terms, each of which is a complicated function of the estimates of these parameters.

The dimension of the calculations, though not their complexity, can be reduced with the restrictions that grandparents and parents have equal variances, and that the covariance between sibs and parents is the same as that between parents and grandparents, so that $\gamma = \alpha$ and $\beta_2 = \beta$. However, assuming that sibs share this variance (i.e., $\sigma_0 = \gamma = \alpha$) actually complicates the estimation, for then the likelihood involving the $\mathbf{u}_i$ no longer suffices in estimating δ_1.

Under the conditions $\gamma = \alpha$ and $\beta_2 = \beta$, assuming normality, five covariance parameters remain to be estimated by the information given by the $\mathbf{u}_i$ (in addition to the last m distinct eigenvalues of the circular covariance matrix $\mathbf{C}$), and the information matrix contains 15 distinct terms. In this case,

$$cov(\mathbf{u}_i) = \begin{bmatrix} \tau + 2\sigma + \phi + V & \mathbf{1}_c' \otimes \begin{bmatrix} \phi + V + \sigma - \beta_1 & \beta_1 \end{bmatrix} \\ \mathbf{1}_c \otimes \begin{bmatrix} \phi + V + \sigma - \beta_1 \\ \beta_1 \end{bmatrix} & \mathbf{I}_c \otimes \left\{ \begin{bmatrix} \tau & \sigma \\ \sigma & \phi \end{bmatrix} \right\} + \mathbf{J}_c \otimes \begin{bmatrix} 0 & 0 \\ 0 & V \end{bmatrix} \end{bmatrix},$$

with $\tau = \alpha - 2\beta + \delta_1/a$, $\sigma = \beta - \delta_1/a$, and $\phi = \delta_1/a - V$. Also,

$$\begin{aligned} |cov(\mathbf{u}_i)| &= d^{c-1}(d + Vc\tau) \times \\ & \left\{ \begin{array}{c} \tau + 2\sigma + \phi + V - \frac{c}{d} \begin{bmatrix} \phi(\phi + V + \sigma - \beta_1)^2 \\ +\beta_1(2\sigma(\beta_1 - \phi + V + \sigma) + \beta_1\tau) \end{bmatrix} \\ + \frac{c^2 V}{d(d - \omega c\phi)} [\sigma(\phi + V + \sigma) - \beta_1(\sigma + \tau)]^2 \end{array} \right\}, \end{aligned}$$

and $[cov(\mathbf{u}_i)]^{-1}$ is still given by (34), now setting

$$\begin{aligned} f &= \frac{d(d + Vc\tau)}{\left\{ \begin{array}{c} (\tau + 2\sigma + \phi + V)\, d(d + Vc\tau) \\ -c \left\{ \begin{array}{c} (d + Vc\tau)\left[\phi(\phi + V + \sigma - \beta_1)^2 + \beta_1(\beta_1(\tau + 2\sigma) - (\phi + V + \sigma))\right] \\ -Vc(\sigma(\phi + V + \sigma) - \beta_1(\sigma + \tau))^2 \end{array} \right\} \end{array} \right\}}, \\ e_1 &= (\phi + V + \sigma - \beta_1)\left[\phi d + Vc(\tau\phi - \sigma^2)\right] - \beta_1\sigma d, \\ e_2 &= d[\beta_1(\sigma + \tau) - \sigma(\phi + V + \sigma)] \text{ and } d = \tau\phi - \sigma^2. \end{aligned}$$

Unfortunately, while the conditions $\gamma = \alpha$ and $\beta_2 = \beta$ reduce the dimension of the parameter space, they increase the complexity of the calculations. In fact, whether or not γ and α, and β_2 and β are distinct, the score and information of $\boldsymbol{\theta}' = (\tau, \phi, \sigma, \beta_2, \beta_1, \gamma, V)$ are too involved to be of practical use, and hence maximum likelihood estimation is not a functional means of obtaining estimates.

A realistic special case which simplifies maximizing the likelihood is that which assumes families are related through lineage but not through environment. This condition is equivalent to setting $V = 0$, in which case (now with γ and α, and β_2 and β distinct)

$$|cov(\mathbf{u}_i)| = d^c \left\{ \gamma - \frac{c}{d} \left[\phi(\beta_2 - \beta_1)^2 + \beta_1(2\sigma(\beta_1 - \beta_2) + \beta_1\tau) \right] \right\},$$

and

$$[cov(\mathbf{u}_i)]^{-1}$$

$$= \frac{f}{d^2}\begin{bmatrix} d^2 & -\mathbf{1}_c'\otimes\begin{bmatrix} e_1 & e_2\end{bmatrix} \\ -\mathbf{1}_c\otimes\begin{bmatrix} e_1 \\ e_2\end{bmatrix} & \frac{d^2}{f}\mathbf{I}_c\otimes\begin{bmatrix}\phi & -\sigma \\ -\sigma & \tau\end{bmatrix} + \frac{1}{d^2}\mathbf{J}_c\otimes\begin{bmatrix} e_1^2 & e_1e_2 \\ e_1e_2 & e_2^2\end{bmatrix}\end{bmatrix},$$

where

$$f = \frac{d}{\gamma d - c\left[\phi(\beta_2-\beta_1)^2 + \beta_1(\beta_1(\tau+2\sigma)-\beta_2)\right]},$$

$$e_1 = [(\beta_2-\beta_1)\phi - \beta_1\sigma]\,d,$$

$$e_2 = d\,[\beta_1(\sigma+\tau)-\sigma\beta_2],\text{ and } d = \tau\phi-\sigma^2.$$

Here,

$$\mathbf{u}_i'[cov(\mathbf{u}_i)]^{-1}\mathbf{u}_i = \frac{f}{d^2}\left\{\begin{array}{c} d^2g_i^2 - 2g_i(e_1u_{i\bullet1}+e_2u_{i\bullet2}) + \left(\frac{d^2\phi}{f}+\frac{e_1^2}{d^2}\right)\sum_{j=1}^c u_{ij1}^2 \\ +2\left(-\frac{d^2\sigma}{f}+\frac{e_1e_2}{d^2}\right)\sum_{j=1}^c u_{ij1}u_{ij2} \\ +\left(\frac{d^2\tau}{f}+\frac{e_2^2}{d^2}\right)\sum_{j=1}^c u_{ij2}^2 \\ +\frac{e_1^2}{d^2}\left[u_{i\bullet1}^2-\sum_{j=1}^c u_{ij1}^2\right] \\ +2\frac{e_1e_2}{d^2}\left[u_{i\bullet1}u_{i\bullet2}-\sum_{j=1}^c u_{ij1}u_{ij2}\right] \\ +\frac{e_2^2}{d^2}\left[u_{i\bullet2}^2-\sum_{j=1}^c u_{ij2}^2\right]\end{array}\right\}.$$

However, even the restriction that $V=0$ does not make ML estimation practicable; the elements of the Hessian and the information of $\boldsymbol{\theta}$ are too complicated to be of any use. Some of these elements have more than 200 terms each in certain sufficient statistics in the loglikelihood of $\boldsymbol{\theta}$.

On the other hand, ANOVA-type estimators are available for the elements of $\boldsymbol{\theta}$ through simple manipulation of the sums of squares and cross products of the g_i and u_{ij}. Let γ and α, and β_2 and β be distinct, allow V to vary from 0, and generalize the mean structure of the data so that

$$\mathbf{E}([g_i, p_{ij}, x_{ijk}]) = [\mu_g, \mu_p, \mu_s].$$

The covariance parameters (excluding $(\delta_2, \ldots, \delta_a)$) may then be estimated without bias by

$$\tilde{\gamma} = \frac{\sum_i g_i^2 - n^{-1}g_\bullet^2}{n-1}.$$

$$
\begin{aligned}
\tilde{\tau} &= \frac{\sum_{ij} u_{ij1}^2 - c^{-1}\sum_i u_{i\bullet 1}^2}{n(c-1)}, \\
\tilde{\sigma} &= \frac{\sum_{ij} u_{ij1}u_{ij2} - c^{-1}\sum_i u_{i\bullet 1}u_{i\bullet 2}}{n(c-1)}, \\
\tilde{\phi} &= \frac{\sum_{ij} u_{ij2}^2 - c^{-1}\sum_i u_{i\bullet 2}^2}{n(c-1)}, \\
\tilde{V} &= \left\{\frac{(nc-1)\sum_i u_{i\bullet 2}^2 - c(n-1)\sum_{ij} u_{ij2}^2}{n^2c^2(c-1)} - \bar{u}_{\bullet\bullet 2}^2\right\}\frac{n}{n-1}, \\
\tilde{\beta}_1 &= \frac{\sum_i g_i u_{i\bullet 2} - nc\bar{g}_\bullet \bar{u}_{\bullet\bullet 2}}{c(n-1)} \text{ and} \\
\tilde{\beta}_2 &= \frac{\sum_i g_i\left(u_{i\bullet 1} + u_{i\bullet 2}\right) - nc\bar{g}_\bullet\left(\bar{u}_{\bullet\bullet 1} + \bar{u}_{\bullet\bullet 2}\right)}{c(n-1)}.
\end{aligned}
$$

Advantage is taken here of the facts that the g_i (as well as the u_{ij1}) are iid, and the u_{ij2} have compound symmetry covariance structure within each csp.

4.6 Unequal Family and Cousinship Sizes with Scores of Parents and Grandparents. In this subsection, again assume that in csp i there are b_{ij} families having a_j sibs, $j = 1, 2, \ldots, c$, $a_1 < \cdots < a_c$, $i = 1, 2, \ldots, n$, and let $A_j = a_j + 1$ so that $t_i = \sum_j b_{ij}A_j$ is the number of descendents of the i^{th} grandparent. Suspend the assumption of normality, except within sibships. In the k^{th} family having a_j sibs in csp i, let x_{ijkl} be the score on sib l and p_{ijk} be that on the parent. Also, adapt the less restrictive mean structure

$$
\mathbf{E}\left([g_i, p_{ijk}, x_{ijkl}]\right) = \left[\mu_g, \mu_p, \mu_s\right].
$$

of the last subsection. Denote by $\mathbf{C}_j$ the (circular) covariance within any sibship having a_j sibs so that, generalizing the covariance structure proposed in Subsection (4.5), the covariance of the observations in any such family is $cov\left(\mathbf{x}_{ijk}\right) = \mathbf{C}_{aug,j} = \begin{bmatrix} \alpha & \beta\mathbf{1}' \\ \beta\mathbf{1} & \mathbf{C}_j \end{bmatrix}$ and the covariance between the grandparent's score and the scores in any family of this size is $cov\left(g_i, \mathbf{x}_{ijk}\right) = \left[\beta_2, \beta_1\mathbf{1}'_{a_j}\right]$. If we were to retain V as the universal within-csp covariance parameter for sibs and parents (as is helpful for ML estimation, to limit the number of parameters), then assembling these suppositions would yield a within-csp covariance structure for each i of

$$
cov\left(\mathbf{x}_i\right) = cov\left(\left[g_i, \mathbf{x}'_{i11}, \mathbf{x}'_{i12}, \ldots, \mathbf{x}'_{icb_{ic}}\right]'\right) \tag{35}
$$

$$
= \begin{bmatrix} \gamma & \mathbf{1}'_{b,1} \otimes \begin{bmatrix} \beta_2 & \beta_1 \mathbf{1}'_{a_1} \end{bmatrix}, \ldots, \mathbf{1}'_{b,c} \otimes \begin{bmatrix} \beta_2 & \beta_1 \mathbf{1}'_{a_c} \end{bmatrix} \\ \mathbf{1}_{b,1} \otimes \begin{bmatrix} \beta_2 \\ \beta_1 \mathbf{1}_{a_1} \end{bmatrix} & \\ \mathbf{1}_{b,2} \otimes \begin{bmatrix} \beta_2 \\ \beta_1 \mathbf{1}_{a_2} \end{bmatrix} & block \left\{ \begin{matrix} \mathbf{I}_{b,1} \otimes [\mathbf{C}_{aug,1} - V\mathbf{J}_{A_1}], \\ \mathbf{I}_{b,2} \otimes [\mathbf{C}_{aug,2} - V\mathbf{J}_{A_2}], \\ \ldots, \mathbf{I}_{b,c} \otimes [\mathbf{C}_{aug,c} - V\mathbf{J}_{A_c}] \end{matrix} \right\} \\ \vdots & +V\mathbf{J}_{t_i} \\ \mathbf{1}_{b,c} \otimes \begin{bmatrix} \beta_2 \\ \beta_1 \mathbf{1}_{a_c} \end{bmatrix} & \end{bmatrix}.
$$

However, since in the present subsection we specify ANOVA-type estimators, it is both less computationally complex and, perhaps, more realistic in the majority of applications, to allow three separate parameters to represent the parent-parent, sib-sib and sib-parent covariances across families within csp's. We therefore set

$$
V_{pp} = cov\,(p_{ijk}, p_{ij'k'})\,,\ V_{ss} = cov\,(x_{ijkl}, x_{ij'k'l'}) \text{ and } V_{ps} = cov\,(p_{ijk}, x_{ij'k'l})
$$

for all $j \neq j'$ or $k \neq k'$, and define

$$
\begin{bmatrix} \eta \\ \psi \\ \phi_j \end{bmatrix} = \begin{bmatrix} \alpha - V_{pp} \\ \beta - V_{ps} \\ \delta_{j1}/a_j - V_{ss} \end{bmatrix}.
$$

That is to say, setting

$$
\mathbf{F}_{jk} = \begin{bmatrix} V_{pp} & V_{ps}\mathbf{1}'_{a_k} \\ V_{ps}\mathbf{1}_{a_j} & V_{ss}\mathbf{1}_{a_j}\mathbf{1}'_{a_k} \end{bmatrix},
$$

let the lower right submatrix of $cov\,(\mathbf{x}_i)$ (excluding the row and column corresponding to the grandparent's score) be

$$
(cov\,(\mathbf{x}_i))_{22} = block\,[\mathbf{I}_{b,1} \otimes (\mathbf{C}_{aug,1} - \mathbf{F}_{11})\,, \mathbf{I}_{b,2} \otimes (\mathbf{C}_{aug,2} - \mathbf{F}_{22})\,, \ldots, \mathbf{I}_{b,c} \otimes (\mathbf{C}_{aug,c} - \mathbf{F}_{cc})]
$$

$$+\begin{bmatrix} \mathbf{J}_{b_{\cdot 1}} \otimes \mathbf{F}_{11} & (\mathbf{1}_{b_{\cdot 1}}\mathbf{1}'_{b_{\cdot 2}}) \otimes \mathbf{F}_{12} & (\mathbf{1}_{b_{\cdot 1}}\mathbf{1}'_{b_{\cdot 3}}) \otimes \mathbf{F}_{13} & \cdots & (\mathbf{1}_{b_{\cdot 1}}\mathbf{1}'_{b_{\cdot c}}) \otimes \mathbf{F}_{1c} \\ & \mathbf{J}_{b_{\cdot 2}} \otimes \mathbf{F}_{22} & (\mathbf{1}_{b_{\cdot 2}}\mathbf{1}'_{b_{\cdot 3}}) \otimes \mathbf{F}_{23} & \cdots & (\mathbf{1}_{b_{\cdot 2}}\mathbf{1}'_{b_{\cdot c}}) \otimes \mathbf{F}_{2c} \\ & & \mathbf{J}_{b_{\cdot 3}} \otimes \mathbf{F}_{33} & \cdots & (\mathbf{1}_{b_{\cdot 3}}\mathbf{1}'_{b_{\cdot c}}) \otimes \mathbf{F}_{3c} \\ & symm. & & \ddots & \vdots \\ & & & & \mathbf{J}_{b_{\cdot c}} \otimes \mathbf{F}_{cc} \end{bmatrix}.$$

Later in this subsection, we return to the case in which $V_{pp} = V_{ps} = V_{ss} = V$.

Letting $\boldsymbol{\Psi}_j = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Gamma}_j \end{bmatrix}$ and $\boldsymbol{\Upsilon}_i = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & block\left(\mathbf{I}_{b_{\cdot 1}} \otimes \boldsymbol{\Psi}_1, \mathbf{I}_{b_{\cdot 2}} \otimes \boldsymbol{\Psi}_2, \ldots, \mathbf{I}_{b_{\cdot c}} \otimes \boldsymbol{\Psi}_c\right) \end{bmatrix}$, and making the transformations $\mathbf{y}_i = \boldsymbol{\Upsilon}_i \mathbf{x}_i$ gives

$$\begin{aligned} (cov\,(\mathbf{y}_i))_{22} \;=\; & block \begin{bmatrix} \mathbf{I}_{b_{\cdot 1}} \otimes (\mathbf{H}_1 - \mathbf{G}_{11}), \mathbf{I}_{b_{\cdot 2}} \otimes (\mathbf{H}_2 - \mathbf{G}_{22}), \\ \ldots, \mathbf{I}_{b_{\cdot c}} \otimes (\mathbf{H}_c - \mathbf{G}_{cc}) \end{bmatrix} + \\ & \begin{bmatrix} \mathbf{J}_{b_{\cdot 1}} \otimes \mathbf{G}_{11} & (\mathbf{1}_{b_{\cdot 1}}\mathbf{1}'_{b_{\cdot 2}}) \otimes \mathbf{G}_{12} & (\mathbf{1}_{b_{\cdot 1}}\mathbf{1}'_{b_{\cdot 3}}) \otimes \mathbf{G}_{13} & \cdots & (\mathbf{1}_{b_{\cdot 1}}\mathbf{1}'_{b_{\cdot c}}) \otimes \mathbf{G}_{1c} \\ & \mathbf{J}_{b_{\cdot 2}} \otimes \mathbf{G}_{22} & (\mathbf{1}_{b_{\cdot 2}}\mathbf{1}'_{b_{\cdot 3}}) \otimes \mathbf{G}_{23} & \cdots & (\mathbf{1}_{b_{\cdot 2}}\mathbf{1}'_{b_{\cdot c}}) \otimes \mathbf{G}_{2c} \\ & & \mathbf{J}_{b_{\cdot 3}} \otimes \mathbf{G}_{33} & \cdots & (\mathbf{1}_{b_{\cdot 3}}\mathbf{1}'_{b_{\cdot c}}) \otimes \mathbf{G}_{3c} \\ & symm. & & \ddots & \vdots \\ & & & & \mathbf{J}_{b_{\cdot c}} \otimes \mathbf{G}_{cc} \end{bmatrix}, \end{aligned}$$

where

$$\mathbf{G}_{jk} = \begin{bmatrix} V_{pp} & V_{ps}\left(\sqrt{a_k}, 0, 0, \ldots, 0\right) \\ V_{ps} \begin{bmatrix} \sqrt{a_j} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} & V_{ss} \begin{bmatrix} \sqrt{a_j a_k} & 0 & 0 & \ldots & 0 \\ 0 & 0 & & & \vdots \\ 0 & & 0 & & \\ \vdots & & & \ddots & \\ 0 & \ldots & & & 0 \end{bmatrix} \end{bmatrix}$$

and

$$\mathbf{H}_j = \begin{bmatrix} \alpha & \beta\left(\sqrt{a_j}, 0, 0, ..., 0\right) \\ \beta \begin{bmatrix} \sqrt{a_j} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} & \mathbf{\Lambda}_j \end{bmatrix}$$

again using $\mathbf{\Lambda}_j = diag\left(\delta_{j1}, \delta_{j2}, ..., \delta_{ja_j}\right)$. Consequently, once more the grandparents' and parents' scores and the y_{ijk1} are independent of the other canonical variables. The methods of Olkin and Press can thus be used with each group of sibships of uniform size a_j to estimate the second through $(m_j + 1)^{th}$ eigenvalues of each $\mathbf{C}_j$ without bias (these will be MLE given normality). Obtaining an estimate of the first eigenvalue δ_{j1} of each $\mathbf{C}_j$, as will be accomplished in what follows, allows within-group estimation of the circular covariance parameters $(\sigma_0, \sigma_1, ..., \sigma_{m_c})$. However, because ML estimates (assuming normality) of these first eigenvalues are effectively intractable, and we instead propose ANOVA-type estimators, the standard errors and covariances of the within-group estimates of $(\sigma_0, \sigma_1, ..., \sigma_{m_c})$ differ from those of the MLE of $(\sigma_0, \sigma_1, ..., \sigma_{m_c})$ presented in that subsection. This fact is important because it implies the within-group estimators of $(\sigma_0, \sigma_1, ..., \sigma_{m_c})$ are now combined across groups using weights that differ from those of Subsection (4.3). However, the adjustments are slight and not difficult to implement; we omit them here.

For each (i, j, k), put $m_{ijk} = a_j^{-1/2} y_{ijk1} = a_j^{-1} x_{ijk\bullet}$. Unbiased ANOVA-type estimators of γ, β_1 and β_2 can be given as

$$\tilde{\gamma} = \frac{\sum_i g_i^2 - n^{-1} g_\bullet^2}{n-1}, \; \tilde{\beta}_1 = \frac{n \sum_i g_i m_{i\bullet\bullet} - n b_{\bullet\bullet} \bar{g}_\bullet \bar{m}_{\bullet\bullet\bullet}}{b_{\bullet\bullet}(n-1)} \text{ and}$$
$$\tilde{\beta}_2 = \frac{n \sum_i g_i p_{i\bullet\bullet} - n b_{\bullet\bullet} \bar{g}_\bullet \bar{p}_{\bullet\bullet\bullet}}{b_{\bullet\bullet}(n-1)}.$$

The unbiasedness of these last expressions does not depend on any normality assumptions. Manipulating the p_{ijk} and m_{ijk} produces estimates of α, β, all the $\delta_{j1}, V_{pp}, V_{ps}$ and V_{ss}. The p_{ijk} have a compound symmetry covariance structure within csp's, and the m_{ijk} have a compound symmetry covariance structure within family sizes within csp's:

$$cov\left(\begin{bmatrix} p_{i11} \\ p_{i12} \\ \vdots \\ p_{icb_{ic}} \end{bmatrix}\right) = \eta \mathbf{I}_{b_{i\bullet}} + V_{pp}\mathbf{J}_{b_{i\bullet}}, \text{ and } cov\left(\begin{bmatrix} m_{ij1} \\ m_{ij2} \\ \vdots \\ m_{ijb_{ij}} \end{bmatrix}\right) = \phi_j \mathbf{I}_{b_{ij}} + V_{ss}\mathbf{J}_{b_{ij}},$$

where $\eta = \alpha - V_{pp}$ and $\phi_j = \delta_{j1}/a_j - V_{ss}$ for each family size a_j. Similarly, the p_{ijk} and m_{ijk} together have a covariance structure resembling compound symmetry:

$$cov\left(p_{ijk}, m_{ij'k'}\right) = \begin{cases} \psi + V_{ps}, & j = j' \text{ and } k = k', \\ V_{ps}, & \text{otherwise,} \end{cases}$$

where $\psi = \beta - V_{ps}$. Estimating α, β, all the $\delta_{j1}, V_{pp}, V_{ps}$ and V_{ss}, then, can be accomplished by treating the present setup as $2+c$ separate unbalanced compound symmetry arrangements (each of which can be modeled as a one-way random effects linear model) .

Optimal (in the sense of minimizing mean squared error) estimation of compound symmetry covariance parameters with unequal block sizes using reductions of sums of squares has been shown by C.A.B. Smith (1957) to require an iterative weighting scheme. In Smith's model, the correlation between observations in the same block is the ratio between the within-individual and the between-individual covariances $V_A + V_B$ and V_B. Without knowledge of this correlation, the ideal weight to be assigned to each block when computing weighted sums of squares for estimating V_B (though not V_A) cannot be determined exactly. If V_B is very small compared with V_A, then the units in the i^{th} block should be treated as independent observations, so that the block receives a weight proportional to its number of units n_i. On the other hand, when V_B is much larger than V_A, then each block should be treated as a single observation, meaning that all blocks are weighted (almost) equally, without respect to their sizes. Obviously, in most applications, the true balance between V_A and V_B is at some point between these two extremes. The usual one-way random effects linear model analysis of variance weights all blocks equally.

If V_A and V_B were known, the ideal weight of the i^{th} block in estimating V_B in any case could be expressed as

$$w_i = \left(\frac{V_A}{n_i} + V_B\right)^{-1}, \tag{36}$$

the reciprocal of the variance of its sample mean. The obvious procedure, then, is to estimate V_A by the usual sum of squared deviations v_A (which weights blocks according to their sizes) divided by the degrees of freedom, assign initial weights $w_i^{(1)}$, compute an initial estimate $v_B^{(1)}$ of V_B, and use v_A and $v_B^{(1)}$ to compute new weights $w_i^{(2)}$ through (36). Iterations between the $w_i^{(t)}$ and $v_B^{(t)}$ may continue in this way until convergence.

Adapting Smith's methods to the estimation of η, ϕ_j, $j = 1, 2, ..., c$, V_{pp} and V_{ss}, we estimate η and ϕ_j without bias (or iterations) by

$$\bar{\eta} = \frac{\sum_{ijk} p_{ijk}^2 - \sum_i p_{i\bullet\bullet}^2 / b_{i\bullet}}{b_{\bullet\bullet} - n} \text{ and } \bar{\phi}_j = \frac{\sum_{ik} m_{ijk}^2 - \sum_i m_{ij\bullet}^2 / b_{ij}}{b_{\bullet j} - n_j},$$

setting $n_j = \sum_i I_{\{b_{ij} > 0\}}$. Here and henceforth, sums over i specific to a particular family size a_j are over $i : b_{ij} > 0$. Assuming the parents' scores are normally distributed,[8]

$$var(\bar{\eta}) = \frac{2\eta^2}{b_{\bullet\bullet} - n} \text{ and } var\left(\bar{\phi}_j\right) = \frac{2\phi_j^2}{b_{\bullet j} - n_j}.$$

Next, set

$$q_\eta = \sum_i w_{\eta i} \bar{p}_{i\bullet\bullet}^2 - W_\eta^{-1}\left(\sum_i w_{\eta i} \bar{p}_{i\bullet\bullet}\right)^2 \text{ and } q_j = \sum_i w_{\phi ij} \bar{m}_{ij\bullet}^2 - W_{\phi j}^{-1}\left(\sum_i w_{\phi ij} \bar{m}_{ij\bullet}\right)^2,$$

where[9] $w_{\eta i} = (b_{i\bullet} + \bar{b}_{\bullet\bullet})/2$, $w_{\phi ij} = \begin{cases} (b_{ij} + \bar{b}_{\bullet j})/2, b_{ij} > 0, \\ 0, b_{ij} = 0, \end{cases}$ $W_\eta = \sum_i w_{\eta i}$ and $W_{\phi j} = \sum_i w_{\phi ij}$. Put

$$\begin{aligned} k_{\eta 1} &= \sum_i \frac{w_{\eta i}(W_\eta - w_{\eta i})}{b_{i\bullet} W_\eta}, \; k_{\eta 2} = W_\eta - W_\eta^{-1} \sum_i w_{\eta i}^2, \\ k_{\phi j 1} &= \sum_i \frac{w_{\phi ij}(W_{\phi j} - w_{\phi ij})}{b_{ij} W_{\phi j}} \text{ and } k_{\phi j 2} = W_{\phi j} - W_{\phi j}^{-1} \sum_i w_{\phi ij}^2. \end{aligned}$$

Initial estimates of V_{pp} and V_{ss} from the parents' and siblings' scores are subsequently found as

$$v_\eta^{(1)} = \frac{q_\eta - k_{\eta 1}\bar{\eta}}{k_{\eta 2}} \text{ and } v_{\phi j}^{(1)} = \frac{q_j - k_{\phi j 1}\bar{\phi}_j}{k_{\phi j 2}}, j = 1, 2, ..., c,$$

[8] The assumption of multivariate normality within sibships is global throughout the present subsection.

[9] The best initial estimates of the ideal weights actually depend on professional knowledge of the approximate values of the parent-parent and sib-sib correlations; the initial weights here merely take a middle ground between those that would be appropriate given no correlation, and those that would be appropriate given total correlation.

the subscripts of these estimates indicating that $\left(\bar{\eta}, v_{\eta}^{(t)}\right)$ jointly estimate (η, V_{pp}) and $\left(\bar{\phi}_j, v_{\phi j}^{(t)}\right)$ jointly estimate $\left(\phi_j, V_{ss}\right)$. Updated weights are then computed as

$$w_{\eta i} = \left(\frac{\bar{\eta}}{b_{i\bullet}} + v_{\eta}^{(1)}\right)^{-1} \text{ and } w_{\phi ij} = \left(\frac{\bar{\phi}_j}{b_{ij}} + v_{\phi j}^{(1)}\right)^{-1},$$

the appropriate adaptations of (36). From these new weights, updated estimates $v_{\eta}^{(2)}$ and $v_{\phi j}^{(2)}$ are available, and so on, iterating in each of these two cases. At each step t, $v_{\eta}^{(t)}$ and $v_{\phi j}^{(t)}$ are unbiased for V_{pp} and V_{ss} and (assuming the parents', like the sibs' scores, are normally distributed) have variances

$$var\left(v_{\eta}^{(t)}\right) = \frac{2n - 2 + k_{\eta 1}^2 \frac{2\eta^2}{b_{\bullet\bullet} - n}}{k_{\eta 2}^2} \text{ and } var\left(v_{\phi j}^{(t)}\right) = \frac{2n_j - 2 + k_{\phi j 1}^2 \frac{2\phi_j^2}{b_{\bullet j} - n_j}}{k_{\phi j 2}^2}. \tag{37}$$

In the same way, V_{ps} and the intra-family parent-sib covariance β may be estimated using the crossproducts of the parents' scores p_{ijk} and the sib means m_{ijk}, and the within-csp means of these statistics. We may set $\psi = \beta - V_{ps}$ and then its estimate

$$\bar{\psi} = \frac{\sum_{ijk} p_{ijk} m_{ijk} - \sum_i p_{i\bullet\bullet} m_{i\bullet\bullet}/b_{i\bullet}}{b_{\bullet\bullet} - n}$$

has expected value ψ and (assuming bivariate normality) variance $\frac{2\psi^2}{b_{\bullet\bullet} - n}$. Letting

$$q_{\psi} = \sum_i w_{\psi i} \bar{p}_{i\bullet\bullet} \bar{m}_{i\bullet\bullet} - \left(\sum_i w_{\psi i} \bar{p}_{i\bullet\bullet}\right)\left(\sum_i w_{\psi i} \bar{m}_{i\bullet\bullet}\right)/W_{\psi},$$

with $w_{\psi i} = \left(b_{i\bullet} + \bar{b}_{\bullet\bullet}\right)/2, W_{\psi} = \sum_i w_{\psi i}, k_{\psi 1} = \sum_i \frac{w_{\psi i}(W_{\psi} - w_{\psi i})}{b_{i\bullet} W_{\psi}}$ and $k_{\psi 2} = W_{\psi} - W_{\psi}^{-1} \sum_i w_{\psi i}^2$, the initial estimate of V_{ps} from the parent, sib-mean crossproducts is

$$v_{\psi}^{(1)} = \frac{q_{\psi} - k_{\psi 1}\bar{\psi}}{k_{\psi 2}}.$$

Iterations may then proceed between the weights $w_{\psi i}$ and the estimates $v_{\psi}^{(t)}$. At each step t after the first, the updated weights are

$$w_{\psi i} = \left(\frac{\bar{\psi}}{b_{i\bullet}} + v_{\psi}^{(t)}\right)^{-1}.$$

If all the p_{ijk}, m_{ijk} are together multivariate normal, the final estimate v_{ψ} of V_{ps} has variance $\frac{2n - 2 + k_{\psi 1}^2 \frac{2\psi^2}{b_{\bullet\bullet} - n}}{k_{\psi 2}^2}$.

The model under consideration demands that within any csp, the covariances between sibs of different families be represented by the same parameter V_{ss}, regardless of the size of the sibships. Thus, the estimates $v_{\phi j}, j = 1, 2, ..., c$, obtained above separately by convergence of the iterations with the $w_{\phi ij}$, must be combined to produce a final value v_{ss}. Setting $\mathbf{f}' = (v_{\phi 1}, v_{\phi 2}, ..., v_{\phi c})$, the optimum (unbiased minimum variance) linear combination is

$$v_{ss} = \frac{\mathbf{1}'(cov(\mathbf{f}))^{-1}\mathbf{f}}{\mathbf{1}'(cov(\mathbf{f}))^{-1}\mathbf{1}}. \tag{38}$$

The unbiasedness of the above estimates of $(\gamma, \eta, \psi, \phi_j, \beta_1, \beta_2)$ does not depend on any distributional assumptions beyond those of (35). However, providing an exact expression for $cov(\mathbf{f})$ does require assumptions about the fourth moments and expected squared cross products, etc., of the data within csps. If no such assumptions can be justified, then the elements of $\mathbf{f}$ may be combined into a single unbiased estimator of V_{ss} by ignoring all $cov(v_{\phi j}, v_{\phi k}), j \neq k$, and simply weighting each $v_{\phi j}$ according to the reciprocal of some estimate of its variance.

Henceforth, we retain the assumption of normality within sibships, and obtain $cov(\mathbf{f})$. Each $var(v_{\phi j}), j = 1, 2, ..., c$, was given by (37). Computing $cov(v_{\phi j}, v_{\phi k}), j \neq k$, could be done by directly expanding the sums of squares involved. Because terms of the form $\mathbf{E}(\bar{m}^2_{ij\bullet}\bar{m}^2_{ik\bullet})$ appear in the expansion, it would be helpful to note the fact that, for any bivariate

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathbf{N}\left(\mathbf{0}, \begin{bmatrix} \sigma^2_{\mathbf{x}} & \rho\sigma_{\mathbf{x}}\sigma_{\mathbf{y}} \\ \rho\sigma_{\mathbf{x}}\sigma_{\mathbf{y}} & \sigma^2_{\mathbf{y}} \end{bmatrix}\right),$$

we have (Lehmann, 1983, p.64)

$$\mathbf{E}(\mathbf{x}^2\mathbf{y}^2) = \sigma^2_{\mathbf{x}}\sigma^2_{\mathbf{y}} + 2(\rho\sigma_{\mathbf{x}}\sigma_{\mathbf{y}})^2. \tag{39}$$

However, the computations proceed more easily by expressing the q_j and the $\bar{\phi}_j$ as quadratic forms, and finding the covariances of the pairs of the four terms (as quadratic forms) in the expansion of $v_{\phi j}v_{\phi l} = \left(\frac{q_j - k_{\phi j1}\bar{\phi}_j}{k_{\phi j2}}\right)\left(\frac{q_l - k_{\phi l1}\bar{\phi}_l}{k_{\phi l2}}\right)$, i.e.,

$$cov(v_{\phi j}, v_{\phi l}) = \frac{1}{k_{\phi j2}k_{\phi l2}}\begin{bmatrix} cov(q_j, q_l) - k_{\phi l1}cov(q_j, \bar{\phi}_l) \\ -k_{\phi j1}cov(q_l, \bar{\phi}_j) + k_{\phi j1}k_{\phi l1}cov(\bar{\phi}_j, \bar{\phi}_l) \end{bmatrix}$$

Assume $\mathbf{t}_j$ are normal with zero means for $j = 1, 2$ (not necessarily of the same dimension), $\mathbf{E}(\mathbf{t}_1\mathbf{t}_2') = \mathbf{E}_{12}$, $\mathbf{A}_1 = \mathbf{A}_1'$ and $\mathbf{A}_2 = \mathbf{A}_2'$. Then the methods developed by Evans, as outlined by Searle (1971, p. 65), may be used to write the multivariate extension of (39) as

$$cov\left(\mathbf{t}_1'\mathbf{A}_1\mathbf{t}_1, \mathbf{t}_2'\mathbf{A}_2\mathbf{t}_2\right) = 2tr\left(\mathbf{A}_1\mathbf{E}_{12}\mathbf{A}_2\mathbf{E}_{12}'\right). \tag{40}$$

We first apply (40) to find $cov\left(q_j, q_l\right), j \neq l$. q_j is a weighted average of squared deviations of the within csp, within group sample sib means $\bar{m}_{ij\bullet}$ around a weighted mean of the $\bar{m}_{ij\bullet}$, and is therefore invariant with respect to the theoretical sib mean μ_s. We may simplify the calculations without loss of generality, therefore, by assuming $\mu_s = \mathbf{E}(m_{ijk}) = 0$. For each j, set $\bar{\mathbf{m}}_j' = \left(\bar{m}_{1j\bullet}, \bar{m}_{2j\bullet}, ..., \bar{m}_{nj\bullet}\right)$, $\mathbf{w}_j' = \left(w_{\phi 1j}, w_{\phi 2j}, ..., w_{\phi nj}\right)$ and $\left(b_{1j}^{-1}, b_{2j}^{-1}, ..., b_{nj}^{-1}\right)$ as $(1 \times n)$ vectors of means of sib sample means, weights and within-csp reciprocals of numbers of sibships of size a_j, each vector having zeros in the positions in which $b_{ij} = 0$, $i = 1, 2, ..., n$, and let the $(i, i)^{th}$ element of the $n \times n$ diagonal matrix $\mathbf{D}_j$ be $I_{\{b_{ij} > 0\}}$. With these conventions,

$$\begin{aligned} cov\left(\bar{\mathbf{m}}_j\right) &= V_{ss}\mathbf{D}_j + \phi_j diag\left(b_{1j}^{-1}, b_{2j}^{-1}, ..., b_{nj}^{-1}\right), \\ cov\left(\bar{\mathbf{m}}_j, \bar{\mathbf{m}}_l\right) &= V_{ss}\mathbf{D}_j\mathbf{D}_l \text{ and} \\ q_j &= \bar{\mathbf{m}}_j'\left[diag\mathbf{w}_j - W_{\phi j}^{-1}\mathbf{w}_j\mathbf{w}_j'\right]\bar{\mathbf{m}}_j. \end{aligned}$$

Thus (for $j \neq l$),

$$\begin{aligned} cov\left(q_j, q_l\right) &= 2V_{ss}^2 tr\left\{\begin{array}{c}\left[diag\mathbf{w}_j - W_{\phi j}^{-1}\mathbf{w}_j\mathbf{w}_j'\right]\mathbf{D}_j\mathbf{D}_l \times \\ \left[diag\mathbf{w}_l - W_{\phi l}^{-1}\mathbf{w}_l\mathbf{w}_l'\right]\mathbf{D}_l\mathbf{D}_j\end{array}\right\} = \\ & 2V_{ss}^2\left[\sum_i w_{\phi ij}w_{\phi il}\left(1 - \frac{w_{\phi ij}}{W_{\phi j}} - \frac{w_{\phi il}}{W_{\phi l}}\right) + \frac{\left(\sum_i w_{\phi ij}w_{\phi il}\right)^2}{W_{\phi j}W_{\phi l}}\right] \end{aligned}$$

which, surprisingly, is independent of any of the ϕ_j except insofar as the distributions of the estimates of the ideal weights $w_{\phi ij}$ depend on the ϕ_j and V_{ss}.

With respect to the $cov\left(q_j, \tilde{\phi}_k\right)$, $j \neq k$, each m_{ijp} is well known to be independent of ($\perp$) the deviations $m_{ijl} - \bar{m}_{ij\bullet}$ for all l, p, and hence $q_j \perp \tilde{\phi}_j$. Using matrix expressions for m_{ijp} and each $m_{ikl} - \bar{m}_{ik\bullet}$ it may be shown, too, that $q_j \perp \tilde{\phi}_k$, so that $cov\left(q_j, \tilde{\phi}_k\right) = 0$.

Lastly in the expansion of $cov(v_{\phi j}, v_{\phi l})$, if $\mathbf{m}_j' = (m_{1j1}, m_{1j2}, ..., m_{njb_{nj}})$ (no longer filling with zeros where $b_{ij} = 0$) is $(1 \times b_{\bullet j})$, we have (for $j \neq l$)

$$\begin{aligned} cov(\mathbf{m}_j, \mathbf{m}_l) &= V_{ss} block\left(\mathbf{1}_{b_{1j}}\mathbf{1}_{b_{1l}}', \mathbf{1}_{b_{2j}}\mathbf{1}_{b_{2l}}', ..., \mathbf{1}_{b_{nj}}\mathbf{1}_{b_{nl}}'\right) \text{ and} \\ \bar{\phi}_j &= \frac{1}{b_{\bullet j} - n_j}\mathbf{m}_j'\left[\mathbf{I} - block\left(b_{1j}^{-1}\mathbf{J}_{b_{1j}}, b_{2j}^{-1}\mathbf{J}_{b_{2j}}, ..., b_{nj}^{-1}\mathbf{J}_{b_{nj}}\right)\right]\mathbf{m}_j, \end{aligned}$$

so that

$$\begin{aligned} cov\left(\bar{\phi}_j, \bar{\phi}_l\right) &= \frac{2V_{ss}^2}{(b_{\bullet j} - n_j)(b_{\bullet l} - n_l)} \times \\ & tr\left\{\begin{array}{c} \left[\mathbf{I} - block\left(b_{1j}^{-1}\mathbf{J}_{b_{1j}}, b_{2j}^{-1}\mathbf{J}_{b_{2j}}, ..., b_{nj}^{-1}\mathbf{J}_{b_{nj}}\right)\right] \times \\ block\left(\mathbf{1}_{b_{1j}}\mathbf{1}_{b_{1l}}', \mathbf{1}_{b_{2j}}\mathbf{1}_{b_{2l}}', ..., \mathbf{1}_{b_{nj}}\mathbf{1}_{b_{nl}}'\right) \times \\ \left[\mathbf{I} - block\left(b_{1l}^{-1}\mathbf{J}_{b_{1l}}, b_{2l}^{-1}\mathbf{J}_{b_{2l}}, ..., b_{nl}^{-1}\mathbf{J}_{b_{nl}}\right)\right] \times \\ block\left(\mathbf{1}_{b_{1l}}\mathbf{1}_{b_{1j}}', \mathbf{1}_{b_{2l}}\mathbf{1}_{b_{2j}}', ..., \mathbf{1}_{b_{nl}}\mathbf{1}_{b_{nj}}'\right) \end{array}\right\} \\ &= 0 \end{aligned}$$

and therefore, for $j \neq l$,

$$cov(v_{\phi j}, v_{\phi l}) = \frac{1}{k_{\phi j2}k_{\phi l2}} \times 2V_{ss}^2\left[\begin{array}{c} \sum_i w_{\phi ij}w_{\phi il}\left(1 - \frac{w_{\phi ij}}{W_{\phi j}} - \frac{w_{\phi il}}{W_{\phi l}}\right) \\ + \frac{\left(\sum_i w_{\phi ij}w_{\phi il}\right)^2}{W_{\phi j}W_{\phi l}} \end{array}\right]. \quad (41)$$

Substituting $\bar{\phi}_j$ and $\bar{\phi}_l$ and an initial estimate of V_{ss} such as $c^{-1}\sum_j v_{\phi j}$, (41) for all combinations $\{(j, l) : j \neq l\}$, with (37), yields an estimate of the covariance of the minimum variance, sibship size specific, unbiased estimators $\mathbf{f}$ of V_{ss}. This estimated covariance is refined iteratively through (38) to produce an overall minimum variance unbiased estimator v_{ss} having approximate variance $\left(\mathbf{1}'(cov(\mathbf{f}))^{-1}\mathbf{1}\right)^{-1}$.

It is important and at the same time enlightening to note the inadequacy of $\left(v_{ss}^{(1)}\right)^2 = \left(c^{-1}\sum_j v_{\phi j}\right)^2$ and $\left(v_{ss}^{(t)}\right)^2$, $t > 1$, as initial and subsequent estimators of V_{ss}^2 in (41). Estimating covariance matrices through direct substitution of such quantities (as in Newton Raphson or Fisher Scoring iterations) is usually a functional technique. Nonetheless, in the present instance the tendency of these $\left(v_{ss}^{(t)}\right)^2$ to overestimate V_{ss}^2 (due to the fact that $\mathrm{E}\left[\left(v_{ss}^{(t)}\right)^2\right] = V_{ss}^2 +$

$var\left(v_{ss}^{(t)}\right)$ has unfortunate consequences since the estimates of the diagonal elements

$$\frac{2n_j - 2 + k_{\phi j1}^2 \frac{2\bar{\phi}_j^2}{b_{\bullet j} - n_j}}{k_{\phi j2}^2}$$

of $cov(\mathbf{f})$ do not depend on $v_{ss}^{(t)}$ in the same way as do the off-diagonal elements in (41). Further, the overestimation of each ϕ_j^2 by $\bar{\phi}_j^2$ does not adequately mitigate the overestimation of V_{ss}^2 by these $\left(v_{ss}^{(t)}\right)^2$ because of the tendency of the $k_{\phi j1}$ towards zero in the b_{ij}. Thus, except when both n and the $b_{i\bullet}$ are quite large, direct substitution of $\left(v_{ss}^{(t)}\right)^2$ for V_{ss}^2 often produces unacceptable estimates of $cov(\mathbf{f})$ (having correlations, in absolute value, too close to or greater than unity). The use of

$$\left(v_{ss}^{(t)}\right)^2 - \left(\mathbf{1}'(cov(\mathbf{f}))^{-1}\mathbf{1}\right)^{-1}$$

is recommended, instead, as an estimate of V_{ss}^2 in (41) at the $(t+1)^{th}$ step.

When it should be assumed that $V_{pp} = V_{ps} = V_{ss} = V$, as was the case in previous subsections of this section, then V is estimated by combining v_η and v_ψ with the $v_{\phi j}$ as

$$v = \frac{\mathbf{1}'\left(cov\left[(v_\eta, v_\psi, \mathbf{f}')'\right]\right)^{-1}(v_\eta, v_\psi, \mathbf{f}')'}{\mathbf{1}'\left(cov\left[(v_\eta, v_\psi, \mathbf{f}')'\right]\right)^{-1}\mathbf{1}}.$$

To identify $cov(v_\eta, v_\psi)$, $cov(v_\eta, v_{\phi j})$ and $cov(v_\psi, v_{\phi j})$ for each j, we assume normality throughout the data except in connection with the grandparents' scores and (for convenience and without loss of generality), as before, assume that $\mu_p = \mu_s = 0$. It is helpful to take advantage of a generalization of (40) to bilinear forms (Searle, 1971), because q_ψ and $\bar{\psi}$ can be expressed as bilinear but not quadratic forms: If $\mathbf{t}_i$ are multivariate normal with zero means and possibly unequal dimensions, $\mathbf{E}\left(\mathbf{t}_i\mathbf{t}_j'\right) = \mathbf{E}_{ij}, i, j = 1, 2, 3, 4$, $\mathbf{A}_{12} = \mathbf{A}_{12}'$ and $\mathbf{A}_{34} = \mathbf{A}_{34}'$, then

$$cov\left(\mathbf{t}_1'\mathbf{A}_{12}\mathbf{t}_2, \mathbf{t}_3'\mathbf{A}_{34}\mathbf{t}_4\right) = tr\left[\mathbf{A}_{12}\left(\mathbf{E}_{23}\mathbf{A}_{34}\mathbf{E}_{41} + \mathbf{E}_{24}\mathbf{A}_{34}\mathbf{E}_{31}\right)\right]. \qquad (42)$$

We turn first to the derivation of

$$cov(v_\eta, v_\psi) = k_{\eta 2}^{-1}k_{\psi 2}^{-1}\left[\begin{array}{c} cov(q_\eta, q_\psi) - k_{\psi 1}cov\left(q_\eta, \bar{\psi}\right) \\ -k_{\eta 1}cov(q_\psi, \bar{\eta}) + k_{\eta 1}k_{\psi 1}cov\left(\bar{\eta}, \bar{\psi}\right) \end{array}\right].$$

Set

$$\bar{\mathbf{p}}' = (\bar{p}_{1\bullet\bullet}, \ldots, \bar{p}_{n\bullet\bullet}), \ \bar{\mathbf{m}}' = (\bar{m}_{1\bullet\bullet}, \ldots, \bar{m}_{n\bullet\bullet}).$$

$$\begin{aligned} \mathbf{p}' &= (p_{111}, p_{112}, ..., p_{1cb_{11c}}, ..., p_{ncb_{nc}}) \text{ and} \\ \mathbf{m}' &= (m_{111}, m_{112}, ..., m_{1cbb_{11c}}, ..., m_{ncb_{nc}}); \\ \mathbf{w}_\eta' &= (w_{\eta 1}, ..., w_{\eta n}) \text{ and } \mathbf{w}_\psi' = (w_{\psi 1}, ..., w_{\psi n}). \end{aligned}$$

$\bar{\mathbf{p}}, \bar{\mathbf{m}}, \mathbf{w}_\eta$ and $\mathbf{w}_\psi$ are each $(n \times 1)$, and $\mathbf{p}$ and $\mathbf{m}$ are each $(b_{\bullet\bullet} \times 1)$. Then we may write

$$\begin{aligned} q_\eta &= \bar{\mathbf{p}}' \left[diag\mathbf{w}_\eta - W_\eta^{-1}\mathbf{w}_\eta\mathbf{w}_\eta'\right] \bar{\mathbf{p}}, \\ \tilde{\eta} &= \frac{\mathbf{p}' \left[\mathbf{I}_{b_{\bullet\bullet}} - block\left(b_{1\bullet}^{-1}\mathbf{J}_{b_{1\bullet}}, ..., b_{n\bullet}^{-1}\mathbf{J}_{b_{n\bullet}}\right)\right] \mathbf{p}}{b_{\bullet\bullet} - n}, \\ q_\psi &= \bar{\mathbf{p}}' \left[diag\mathbf{w}_\psi - W_\psi^{-1}\mathbf{w}_\psi\mathbf{w}_\psi'\right] \bar{\mathbf{m}} \text{ and} \\ \tilde{\psi} &= \frac{\mathbf{p}' \left[\mathbf{I}_{b_{\bullet\bullet}} - block\left(b_{1\bullet}^{-1}\mathbf{J}_{b_{1\bullet}}, ..., b_{n\bullet}^{-1}\mathbf{J}_{b_{n\bullet}}\right)\right] \mathbf{m}}{b_{\bullet\bullet} - n}; \\ \mathbf{E}(\bar{\mathbf{p}}\bar{\mathbf{p}}') &= diag\left(b_{1\bullet}^{-1}\eta + V, ..., b_{n\bullet}^{-1}\eta + V\right), \\ \mathbf{E}(\bar{\mathbf{p}}\bar{\mathbf{m}}') &= diag\left(b_{1\bullet}^{-1}\psi + V, ..., b_{n\bullet}^{-1}\psi + V\right), \\ \mathbf{E}(\bar{\mathbf{p}}\mathbf{p}') &= block\left[\left(b_{1\bullet}^{-1}\eta + V\right)\mathbf{1}_{b_{1\bullet}}', ..., \left(b_{n\bullet}^{-1}\eta + V\right)\mathbf{1}_{b_{n\bullet}}'\right], \\ \mathbf{E}(\bar{\mathbf{p}}\mathbf{m}') &= block\left[\left(b_{1\bullet}^{-1}\psi + V\right)\mathbf{1}_{b_{1\bullet}}', ..., \left(b_{n\bullet}^{-1}\psi + V\right)\mathbf{1}_{b_{n\bullet}}'\right], \\ \mathbf{E}(\mathbf{p}\mathbf{p}') &= block\left[\eta\mathbf{I}_{b_{1\bullet}} + V\mathbf{J}_{b_{1\bullet}}, ..., \eta\mathbf{I}_{b_{n\bullet}} + V\mathbf{J}_{b_{n\bullet}}\right] \text{ and} \\ \mathbf{E}(\mathbf{p}\mathbf{m}') &= block\left[\psi\mathbf{I}_{b_{1\bullet}} + V\mathbf{J}_{b_{1\bullet}}, ..., \psi\mathbf{I}_{b_{n\bullet}} + V\mathbf{J}_{b_{n\bullet}}\right]. \end{aligned}$$

Consequently,

$$\begin{aligned} cov(q_\eta, q_\psi) &= 2tr\left\{ \begin{array}{c} \left(diag\mathbf{w}_\eta - W_\eta^{-1}\mathbf{w}_\eta\mathbf{w}_\eta'\right) diag\left(b_{1\bullet}^{-1}\eta + V, ..., b_{n\bullet}^{-1}\eta + V\right) \times \\ \left(diag\mathbf{w}_\psi - W_\psi^{-1}\mathbf{w}_\psi\mathbf{w}_\psi'\right) diag\left(b_{1\bullet}^{-1}\psi + V, ..., b_{n\bullet}^{-1}\psi + V\right) \end{array} \right\} \\ & 2\sum_{i=1}^{n} w_{\eta i}w_{\psi i}\left(\frac{\eta}{b_{i\bullet}} + V\right)\left[\begin{array}{c} \left(\frac{\psi}{b_{i\bullet}} + V\right)\left(1 - \frac{w_{\eta i}}{W_\eta} - \frac{w_{\psi i}}{W_\psi}\right) \\ +W_\eta^{-1}W_\psi^{-1}\sum_{j=1}^{n} w_{\eta j}w_{\psi j}\left(\frac{\psi}{b_{j\bullet}} + V\right) \end{array} \right]. \end{aligned}$$

In the adaptation of (42) for $cov\left(q_\eta, \tilde{\psi}\right)$, we may take advantage of the fact that

$$\mathbf{E}(\bar{\mathbf{p}}\mathbf{p}')\left[\mathbf{I}_{b_{\bullet\bullet}} - block\left(b_{1\bullet}^{-1}\mathbf{J}_{b_{1\bullet}}, ..., b_{n\bullet}^{-1}\mathbf{J}_{b_{n\bullet}}\right)\right] = \mathbf{0}$$

appears twice, so that $cov\left(q_\eta, \tilde{\psi}\right) = 0$. $cov(q_\psi, \tilde{\eta}) = 0$ for analogous reasons.

Concluding the expansion of $cov\,(v_\eta, v_\psi)$,we have

$$cov\left(\bar{\eta},\bar{\psi}\right) = (b_{\bullet\bullet}-n)^{-2}\,tr\left\{\begin{array}{c}\left(\mathbf{I}_{b_{\bullet\bullet}} - block\left(b_{1\bullet}^{-1}\mathbf{J}_{b_{1\bullet}},...,b_{n\bullet}^{-1}\mathbf{J}_{b_{n\bullet}}\right)\right) \\ \times\left[\begin{array}{c} block\left(\eta\mathbf{I}_{b_{1\bullet}}+V\mathbf{J}_{b_{1\bullet}},...,\eta\mathbf{I}_{b_{1\bullet}}+V\mathbf{J}_{b_{1\bullet}}\right) \\ \times\left(\mathbf{I}_{b_{\bullet\bullet}} - block\left(b_{1\bullet}^{-1}\mathbf{J}_{b_{1\bullet}},...,b_{n\bullet}^{-1}\mathbf{J}_{b_{n\bullet}}\right)\right) \\ \times block\left(\psi\mathbf{I}_{b_{1\bullet}}+V\mathbf{J}_{b_{1\bullet}},...,\psi\mathbf{I}_{b_{1\bullet}}+V\mathbf{J}_{b_{1\bullet}}\right) \\ +block\left(\psi\mathbf{I}_{b_{1\bullet}}+V\mathbf{J}_{b_{1\bullet}},...,\psi\mathbf{I}_{b_{1\bullet}}+V\mathbf{J}_{b_{1\bullet}}\right) \\ \times\left(\mathbf{I}_{b_{\bullet\bullet}} - block\left(b_{1\bullet}^{-1}\mathbf{J}_{b_{1\bullet}},...,b_{n\bullet}^{-1}\mathbf{J}_{b_{n\bullet}}\right)\right) \\ \times block\left(\eta\mathbf{I}_{b_{1\bullet}}+V\mathbf{J}_{b_{1\bullet}},...,\eta\mathbf{I}_{b_{1\bullet}}+V\mathbf{J}_{b_{1\bullet}}\right)\end{array}\right]\end{array}\right\}$$

$$= \frac{2}{(b_{\bullet\bullet}-n)^2}\,tr\left\{\begin{array}{c}\eta block\left(\mathbf{I}_{b_{1\bullet}} - b_{1\bullet}^{-1}\mathbf{J}_{b_{1\bullet}},...,\mathbf{I}_{b_{n\bullet}} - b_{n\bullet}^{-1}\mathbf{J}_{b_{n\bullet}}\right) \\ \times\psi block\left(\mathbf{I}_{b_{1\bullet}} - b_{1\bullet}^{-1}\mathbf{J}_{b_{1\bullet}},...,\mathbf{I}_{b_{n\bullet}} - b_{n\bullet}^{-1}\mathbf{J}_{b_{n\bullet}}\right)\end{array}\right\}$$

$$= \frac{2\eta\psi}{(b_{\bullet\bullet}-n)^2}(b_{\bullet\bullet}-n) = \frac{2\eta\psi}{b_{\bullet\bullet}-n},$$

noting the idempotency of $block\left(\mathbf{I}_{b_{1\bullet}} - b_{1\bullet}^{-1}\mathbf{J}_{b_{1\bullet}},...,\mathbf{I}_{b_{n\bullet}} - b_{n\bullet}^{-1}\mathbf{J}_{b_{n\bullet}}\right)$. Summarizing and simplifying,

$$cov\,(v_\eta, v_\psi) = 2k_{\eta 2}^{-1}k_{\psi 2}^{-1}\left\{\begin{array}{c}\sum_{i=1}^{n} w_{\eta i}w_{\psi i}\left(\frac{\eta}{b_{i\bullet}}+V\right)\left[\begin{array}{c}\left(\frac{\psi}{b_{i\bullet}}+V\right)\left(1-\frac{w_{\eta i}}{W_\eta}-\frac{w_{\psi i}}{W_\psi}\right) \\ +W_\eta^{-1}W_\psi^{-1}\sum_{j=1}^{n} w_{\eta j}w_{\psi j}\left(\frac{\psi}{b_{j\bullet}}+V\right)\end{array}\right] \\ +k_{\eta 1}k_{\psi 1}\frac{\eta\psi}{b_{\bullet\bullet}-n}\end{array}\right\}.$$

The calculations of $cov\,(v_\eta, v_{\phi j})$ and $cov\,(v_\psi, v_{\phi j})$ are more difficult than those of $cov\,(v_\eta, v_\psi)$ but for the most part follow the same types of steps, and complete the derivation of $cov\,[(v_\eta, v_\psi, \mathbf{f}')]$. The added difficulty arises mainly from the possibility that some of the b_{ij} are zero, so that notation similar to that used in the derivation of $cov\,(v_{\phi j}, v_{\phi l})$ is necessary. With respect to $cov\,(v_\eta, v_{\phi j})$, we have

$$\mathbf{E}\left(\bar{\mathbf{p}}\bar{\mathbf{m}}_j'\right) = diag\,(c_{1j}, c_{2j}, ..., c_{nj}) \text{ and}$$

$$\mathbf{E}\left(\bar{\mathbf{p}}\mathbf{m}_j'\right) = block\left[\left(\frac{\psi}{b_{1\bullet}}+V\right)\mathbf{1}_{b_{1j}}',...,\left(\frac{\psi}{b_{n\bullet}}+V\right)\mathbf{1}_{b_{1j}}'\right],$$

in $\mathbf{E}\left(\bar{\mathbf{p}}\bar{\mathbf{m}}_j'\right)$ setting $c_{ij} = \begin{cases}\frac{\psi}{b_{i\bullet}}+V, b_{ij}>0, \\ 0, b_{ij}=0\end{cases}$ and, in $\mathbf{E}\left(\bar{\mathbf{p}}\mathbf{m}_j'\right)$ if any $b_{ij}=0$, making $\mathbf{1}_0$ a vector having no rows and no columns. Also, we can write

$$\mathbf{E}\left(\mathbf{p}\bar{\mathbf{m}}_j'\right) = \psi block\left(\mathbf{h}_{1j}, ..., \mathbf{h}_{nj}\right) + V block\left(\mathbf{1}_{b_{1\bullet}}, ..., \mathbf{1}_{b_{n\bullet}}\right)$$

where each $\mathbf{h}_{ij}' = \left(0, 0, ..., 0, b_{ij}^{-1}\mathbf{1}_{b_{ij}}', 0, 0, ..., 0\right)$ is $(1 \times b_{i\bullet})$ and has b_{ij}^{-1} in the positions corresponding to the families (if any exist) in csp i having a_j sibs each, and a zero vector replaces $\mathbf{1}_{b_{i\bullet}}$ in $block\left(\mathbf{1}_{b_{1\bullet}}, ..., \mathbf{1}_{b_{n\bullet}}\right)$ wherever $b_{ij} = 0$. Lastly

$$\mathbf{E}\left(\mathbf{p}\mathbf{m}_j'\right) = \psi block\left(\mathbf{l}_{1j}, ..., \mathbf{l}_{nj}\right) + V block\left(\mathbf{1}_{b_{1\bullet}}\mathbf{1}_{b_{1j}}', ..., \mathbf{1}_{b_{n\bullet}}\mathbf{1}_{b_{nj}}'\right)$$

where $\mathbf{l}_{ij}' = \left(0, \mathbf{I}_{b_{ij}}, 0\right)$ is $(b_{ij} \times b_{i\bullet})$ and has $\mathbf{I}_{b_{ij}}$ in the position corresponding to the families in csp i having a_j sibs each. This notation allows the derivations

$$\begin{aligned}
cov\left(q_\eta, q_j\right) &= 2tr\left\{\left(diag\mathbf{w}_\eta - W_\eta^{-1}\mathbf{w}_\eta\mathbf{w}_h'\right)\left[\mathbf{E}\left(\bar{\mathbf{p}}\bar{\mathbf{m}}_j'\right)\left(diag\mathbf{w}_j - W_{\phi j}^{-1}\mathbf{w}_j\mathbf{w}_j'\right)\mathbf{E}\left(\bar{\mathbf{m}}_j\bar{\mathbf{p}}'\right)\right]\right\} \\
&= 2\left\{\sum_i w_{\eta i}c_{ij}^2 w_{ij}\left(1 - \frac{w_{ij}}{W_{\phi j}} - \frac{w_{\eta i}}{W_\eta}\right) + \frac{\left(\sum_i w_{\eta i}c_{ij}w_{ij}\right)^2}{W_\eta W_{\phi j}}\right\}, \\
cov\left(q_\eta, \bar{\phi}_j\right) &= cov\left(q_j, \bar{\eta}\right) = 0 \text{ and} \\
cov\left(\bar{\eta}, \bar{\phi}_j\right) &= \frac{2\psi^2}{b_{\bullet\bullet} - n}, \text{ so that} \\
cov\left(v_\eta, v_{\phi j}\right) &= \frac{2}{k_{\eta 2}k_{\phi j 2}}\left[\begin{array}{c} \sum_i w_{\eta i}c_{ij}^2 w_{ij}\left(1 - \frac{w_{ij}}{W_{\phi j}} - \frac{w_{\eta i}}{W_\eta}\right) + \frac{\left(\sum_i w_{\eta i}c_{ij}w_{ij}\right)^2}{W_\eta W_{\phi j}} \\ +k_{\eta 1}k_{\phi j 1}\frac{\psi^2}{b_{\bullet\bullet} - n} \end{array}\right].
\end{aligned}$$

The remaining elements of $cov\left[\left(v_\eta, v_\psi, \mathbf{f}'\right)\right]$ to be specified are $cov\left(v_\psi, v_{\phi j}\right)$, $j = 1, 2, ..., c$, which are found using

$$\begin{aligned}
\mathbf{E}\left(\bar{\mathbf{m}}\bar{\mathbf{m}}_j'\right) &= diag\left(d_{1j}, d_{2j}, ..., d_{nj}\right), \\
\mathbf{E}\left(\bar{\mathbf{m}}_j\bar{\mathbf{p}}'\right) &= diag\left(c_{1j}, c_{2j}, ..., c_{nj}\right), \\
\mathbf{E}\left(\bar{\mathbf{m}}_j\mathbf{m}_j'\right) &= block\left(d_{1j}\mathbf{1}_{b_{1j}}', d_{2j}\mathbf{1}_{b_{2j}}', ..., d_{nj}\mathbf{1}_{b_{nj}}'\right) \text{ and} \\
\mathbf{E}\left(\mathbf{m}\mathbf{m}_j'\right) &= block\left(V\mathbf{1}_{b_{1\bullet}}\mathbf{1}_{b_{1j}}' + \phi_j\mathbf{l}_{1j}, ..., V\mathbf{1}_{b_{n\bullet}}\mathbf{1}_{b_{nj}}' + \phi_j\mathbf{l}_{nj}\right)
\end{aligned}$$

where $d_{ij} = \begin{cases} \frac{\phi_j}{b_{i\bullet}} + V, & b_{ij} > 0, \\ 0, & b_{ij} = 0. \end{cases}$ These expressions yield

$$cov\left(q_\psi, q_j\right) = 2\left[\sum_i w_{\psi i}w_{ij}c_{ij}d_{ij}\left(1 - \frac{w_{ij}}{W_{\phi j}} - \frac{w_{\psi i}}{W_\psi}\right) + \frac{\left(\sum_i w_{\psi i}c_{ij}w_{ij}\right)\left(\sum_i w_{\psi i}d_{ij}w_{ij}\right)}{W_\psi W_{\phi j}}\right],$$

$$
\begin{aligned}
cov\left(q_{\psi}, \bar{\phi}_j\right) &= cov\left(q_j, \bar{\psi}\right) = 0 \text{ and} \\
cov\left(\bar{\psi}, \bar{\phi}_j\right) &= \frac{2\psi\phi_j}{b_{\bullet\bullet} - n}, \text{ so that} \\
cov\left(v_{\psi}, v_{\phi j}\right) &= \frac{2}{k_{\psi 2} k_{\phi j 2}} \left[\begin{array}{c} \sum_i w_{\psi i} w_{ij} c_{ij} d_{ij} \left(1 - \frac{w_{ij}}{W_{\phi j}} - \frac{w_{\psi i}}{W_{\psi}}\right) + \frac{\left(\sum_i w_{\psi i} c_{ij} w_{ij}\right)\left(\sum_i w_{\psi i} d_{ij} w_{ij}\right)}{W_{\psi} W_{\phi j}} \\ + k_{\psi 1} k_{\phi j 1} \frac{\psi \phi_j}{b_{\bullet\bullet} - n} \end{array} \right].
\end{aligned}
$$

$cov\left(v_{\psi}, v_{\phi j}\right)$ completes the specification of the elements of $cov\left[(v_{\eta}, v_{\psi}, \mathbf{f}')\right]$. At each iteration t, the new estimate of V is

$$v^{(t)} = \frac{\mathbf{1}'\left(cov\left[(v_{\eta}, v_{\psi}, \mathbf{f}')\right]\right)^{-1}(v_{\eta}, v_{\psi}, \mathbf{f}')'}{\mathbf{1}'\left(cov\left[(v_{\eta}, v_{\psi}, \mathbf{f}')\right]\right)^{-1}\mathbf{1}},$$

where the current estimate of $cov\left[(v_{\eta}, v_{\psi}, \mathbf{f}')\right]$ depends on the previous value $v^{(t-1)}$. A suitable starting value for the iterations is

$$v^{(1)} = \frac{(v_{\eta}, v_{\psi}, \mathbf{f}')\,\mathbf{1}}{c+2},$$

the simple average of the elements of $(v_{\eta}, v_{\psi}, \mathbf{f}')$. Because of random variation, any current estimate of $cov\left[(v_{\eta}, v_{\psi}, \mathbf{f}')\right]$ can in some cases fail to be positive definite, giving undesirable $v^{(t)}$; in practice, therefore, $v^{(t)}$ should be bounded using

$$\min(v_{\eta}, v_{\psi}, \mathbf{f}') \leq v^{(t)} \leq \max(v_{\eta}, v_{\psi}, \mathbf{f}').$$

As noted above, $\left(v^{(t)}\right)^2$ tends to overestimate V^2 by $var\left(v^{(t)}\right)$; where V^2 appears in $cov\left(v_{\phi j} v_{\phi l}\right)$, $j \neq l$, it should therefore be estimated by $\left(v^{(t)}\right)^2 - \left(\mathbf{1}'\left(cov\left[(v_{\eta}, v_{\psi}, \mathbf{f}')\right]\right)^{-1}\mathbf{1}\right)^{-1}$.

5 RESULTS AND CONCLUSIONS

5.1 Introduction. This thesis presents a number of techniques which make circular covariance a more efficient and versatile modelling tool, facilitating new, useful applications of circular covariance. For instance, if we may assume that the correlations between disease rates in sectors of a city decay in an autoregressive fashion, then modelling the covariance structure using autoregressive circular covariance may improve significantly the efficiency of estimation. Alternatively, if the disease rates in these sectors in different cities can be expected to be correlated according to a compound symmetry structure, then the covariances between observations in different cities, and those around the same city, may be estimated under various sets of assumptions about the availability of covariates or the correlations between sectors of different cities. We may summarize the results and conclusions of the three main sections of this thesis as follows.

5.2 Missing Data. We have developed two classes of estimators for the circular covariance parameters and the parent-sib covariance for use when some observations are missing. Each of these classes has advantages under certain conditions, in terms of ease of calculation, unbiasedness and efficiency. We have shown that professional knowledge of the approximate expected sample sizes (na), the expected proportion of missing observations p_m and the value of unbiasedness relative to that of efficiency, will play a potentially important role in opting for one class of estimators over the other. Briefly, these considerations amount to the following: The unbiased LOO estimators and $\tilde{\sigma}_{ps}$ are usually more efficient than the (biased) "EM" estimators for large na or p_m, and are calculated without assumptions about the relation between the available and unavailable measurements. However, simulations demonstrated that the "EM" estimators are more efficient for a wide range of reasonable sample sizes assuming normality or any one of the t-distributions, especially when the parents and sibs' scores are highly correlated.

5.3 Autoregressive Circular Covariance. Only two parameters characterize the covariance structure of the "autoregressive" type of circular covariance we propose. Therefore, when this kind of covariance adequately explains the observed patterns of variation in the data, the advantages.

in terms of parsimony and efficiency of estimation, of fitting the model under the autoregressive assumption can be enormous, especially with a paucity of data or a large number of sibs in each family. We identified straightforward methods of finding the ML estimate of the autoregressive circular covariance parameters, adding to the utility of the model. We also showed that two methods (the Likelihood Ratio Test and Akaike's Information Criterion) of selecting between alternative covariance structures are easily performed and, under reasonable assumptions, have desirable sizes and powers. The alternative estimator we derived of the autoregressive parameter ρ is superior to the MLE when ρ and the sample size are small. The autoregressive assumption is easily extended to the case in which families are observed having different numbers of siblings; estimation of the circular covariance parameters and the parent-sib covariance can be accomplished by grouping families according to size, obtaining within-group estimators, and combining the within-group estimators so as to minimize the variances of the overall estimators. On the other hand, computer calculations of variances showed that the parent-sib correlation is, in most cases, more efficiently estimated by combining within-group estimates of the correlation, than by calculating the appropriate function of the overall estimates of its components (the sibs' variance, the parents' variance and the parent-sib covariance).

5.4 Compound Symmetry within Cousinships. Assuming members of different families may be correlated, we showed ML estimators of the circular covariance parameters, inter-generational covariances and inter-family covariances to be available under various sets of assumptions about the inter-family covariance structure. In the balanced case (all cousinships having the same number of families and all families having the same number of sibs) in which no parents' or grandparents' scores are available, we calculated ML estimates of all parameters through a simple extension of the work of Olkin and Press (1969). In other, unbalanced situations (including that in which parents' scores are available), iteratively maximizing the likelihood requires manipulating complicated information matrices and score functions, but does produce consistent estimates, as we have shown through simulations. However, when data from a third generation (grandparents) are added, the complexities of the resulting information matrices and score functions make ML estimation impractical. Given balanced data involving these three generations, in lieu of ML estimates, we calculated (without iterations) unbiased ANOVA types of estimators of the inter-generational and inter-family covariances. On the other hand, given unbalanced data involving all three generations, iterations, while not absolutely necessary, have been shown to improve the estimates of some of these covariance terms, through successively improved weightings for families and cousinships of different sizes. Different assumptions regarding the covariance structure between families, require different methods of combining parameter estimates into overall estimates. Computer simulations strengthened the claim that all the estimation procedures we describe (especially those based on ML, due to the possibilities of multiple roots to the likelihood equations) produce consistent estimators.

REFERENCES

Akaike, H. (1973a). "Information Theory and an Extension of the Maximum Likelihood Principle," *Second International Symposium on Information Theory,* (B.N. Petrov and F. Csaki, Eds.), Akademia Kaido, Budapest.

Akaike, H. (1973b). "Maximum Likelihood Identification of Bayesian Autoregressive Moving Average Models," *Biometrika* 60, 255-265.

Anderson, R.M. (1982). *Population Dynamics of Infectious Diseases: Theory and Applications.* Chapman and Hall, London.

Basilevsky, A. (1983). *Applied Matrix Algebra in the Statistical Sciences.* Elsevier Science, New York.

Gleser, L.J. (1992). "A Note on the Analysis of Familial Data," *Biometrika* 79, 412-415.

Hadi, A.S. (1988). "Diagnosing Collinearity and Influential Observations," *Comp. Stat. and Data Analysis* 7, 143-159.

Hamer, W.H. (1906). "Epidemic Disease in England," *Lancet* 1, 733-739.

Henderson, H.V. and Searle, S.R. (1981). "On Deriving the Inverse of a Sum of Matrices," *SIAM Review* 23, 53-59.

Johnson, R.A. and Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis.* Prentice-Hall, New Jersey.

Jones, R.H. (1994). *Longitudinal Data with Serial Correlation: A State-Space Approach.* Chapman and Hall, Suffolk, United Kingdom.

Khattree, R. and Naik, D.N. (1994a). "Estimation of Interclass Correlation under Circular Covariance," *Biometrika* 81, 612-617.

Khattree, R. and Naik, D.N. (1994b). "Optimal Tests for Nested Designs with Circular Stationary Dependence," *Journal of Statistical Planning and Inference* 41, 231-240.

Khattree, R. and Naik, D.N. (1995). *Applied Multivariate Statistics with SAS Software.* SAS Institute, North Carolina.

Leiberman, O. (1994). "A Laplace Approximation to the Moments of a Ratio of Quadratic Forms," *Biometrika* 81, 681-690.

Lehmann, E.L. (1983). *Theory of Point Estimation.* Wiley, New York.

Lilienfeld, D.E. and Stolley, P.D. (1994). *Foundations of Epidemiology.* Oxford University Press, New York.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* Wiley, New York.

Maxcy, K.F. (1926). "An Epidemiological Study of Endemic Typhus (Brill's Disease) in the Southeastern United States with Special Reference to its Mode of Transmission," *Pub. Health Reps.* 41, 2967-2995.

O'Connor, L. (1993a). "Setting the Shaft Straight," *Mech. Eng.* 115(4), 82-84.

O'Connor, L. (1993b). "Taming Vibrations in Turbines." *Mech. Eng.* 115(9), 49-52.

Olkin, I. and Press, S.J. (1969). "Testing and Estimation for a Circular Covariance Stationary Model," *Ann. Math. Statist.* 4, 1358-1373.

Searle, S.R. (1971). *Linear Models.* Wiley, New York.

Schwarz, G. (1978). "Estimating the Dimension of a Model," *Ann. Statist.* 6, 461-464.

Smith, C.A.B. (1957). "On the Estimation of Intraclass Correlation," *Ann. Human Genetics* 21, 363-373.

York, J.A. and London, W.P. (1973). "Recurrent Outbreaks of Measles, Chickenpox and Mumps. II. Systematic Differences in Contact Rates and Stochastic Effects," *Am. Jnl. of Epi.* 98, 469-482.

A. SOME LENGTHY HESSIANS AND INFORMATIONS

The elements of the Hessian of Subsection (4.3) for the i^{th} csp are

$$
\begin{aligned}
\frac{\partial l_i}{\partial V} &= \frac{t_i}{2} - \frac{p_{3i}}{2(1+Vt_i)} - \frac{p_{9i}}{2} + \frac{p_{2i}}{2(1+Vt_i)} \\
&\quad + \frac{Vp_{2i}p_{4i}}{1+Vt} - \frac{Vp_{2i}^2 p_{3i}}{2(1+Vt_i)^2}, \\
\frac{\partial^2 l_i}{\partial V^2} &= \frac{p_{7i}}{2} - \frac{p_{7i}+Vp_{8i}}{1+Vt_i} + \frac{p_{3i}^2}{2(1+Vt_i)^2} \\
&\quad - p_{6i} + \frac{2p_{2i}p_{4i}}{1+Vt_i} - \frac{p_{2i}^2 p_{3i}}{(1+Vt_i)^2} + \frac{Vp_{4i}^2}{1+Vt_i} \\
&\quad - \frac{2Vp_{2i}p_{3i}p_{4i}}{(1+Vt_i)^2} + \frac{2Vp_{2i}p_{5i}}{1+Vt_i} + \frac{Vp_{2i}^2 p_{3i}^2}{(1+Vt_i)^3} \\
&\quad - \frac{Vp_{2i}^2(p_{7i}+Vp_{8i})}{(1+Vt_i)^2}, \\
\frac{\partial^2 l_i}{\partial V \partial \delta_{j1}} &= -\frac{b_{ij}a_j f_j^2}{2} + \frac{b_{ij}a_j f_j^2 + 2Vb_{ij}a_j^2 f_j^3}{2(1+Vt_i)} - \frac{Vp_{3i}b_{ij}a_j f_j^2}{2(1+Vt_i)^2} \\
&\quad + a_j s_{ij} f_j^3 - \frac{p_{2i}\sqrt{a_j}\,w_{ij\bullet} f_j^2}{1+Vt_i} + \frac{p_{2i}^2 V b_{ij} a_j f_j^2}{2(1+Vt_i)^2} \\
&\quad - \frac{f_j^2 V\sqrt{a_j}\,w_{ij\bullet} p_{4i}}{1+Vt_i} + \frac{V^2 p_{2i}p_{4i}b_{ij}a_j f_j^2}{(1+Vt_i)^2} \\
&\quad - \frac{2Vp_{2i}a_j^{3/2} w_{ij\bullet} f_j^3}{1+Vt_i} + \frac{Vp_{2i}p_{3i}\sqrt{a_j}\,w_{ij\bullet} f_j^2}{(1+Vt_i)^2} \\
&\quad - \frac{V^2 p_{2i}^2 p_{3i} b_{ij} a_j f_j^2}{(1+Vt_i)^3} + \frac{Vp_{2i}^2(b_{ij}a_j f_j^2 + 2Vb_{ij}a_j^2 f_j^3)}{(1+Vt_i)^2}, \\
\frac{\partial l_i}{\partial \delta_{j1}} &= \frac{-b_{ij}f_j}{2} + \frac{Vb_{ij}a_j f_j^2}{2(1+Vt_i)} + \frac{s_{ij}f_j^2}{2} \\
&\quad - \frac{Vp_{2i}\sqrt{a_j}\,w_{ij\bullet} f_j^2}{1+Vt} + \frac{V^2 p_{2i}^2 b_{ij} a_j f_j^2}{2(1+Vt_i)^2}, \\
\frac{\partial^2 l_i}{\partial \delta_{j1}^2} &= \frac{b_{ij} f_j^2}{2} - \frac{Vb_{ij}a_j f_j^3}{1+Vt_i} + \frac{V^2 b_{ij}^2 a_j^2 f_j^4}{2(1+Vt_i)^2} \\
&\quad - s_{ij} f_j^3 + \frac{Va_j w_{ij\bullet}^2 f_j^4}{1+Vt_i} - \frac{2V^2 p_{2i} a_j^{3/2} w_{ij\bullet} b_{ij} f_{ij}^4}{(1+Vt_i)^2} \\
&\quad + \frac{2Vp_{2i}\sqrt{a_j}\,w_{ij\bullet} f_j^3}{(1+Vt_i)^3} + \frac{V^3 p_{2i}^2 b_{ij}^2 a_j^2 f_j^4}{(1+Vt_i)^3} - \frac{V^2 p_{2i}^2 b_{ij} a_j f_j^3}{(1+Vt_i)^2}, \\
\frac{\partial^2 l_i}{\partial \delta_{j1} \partial \delta_{k1}} &= \frac{Vf_j^2 f_k^2}{1+Vt_i} \left[\begin{array}{c} \frac{Vb_{ij}a_j b_{ik} a_k}{2(1+Vt_i)} + \frac{\sqrt{a_j}\,w_{ij\bullet}\sqrt{a_k}\,w_{ik\bullet}}{1} \\ -\frac{Vp_{2i}\left(\sqrt{a_j}\,w_{ij\bullet} b_{ik} a_k + \sqrt{a_k}\,w_{ik\bullet} b_{ij} a_j\right)}{1+Vt_i} + \frac{V^2 p_{2i}^2 b_{ij} b_{ik} a_j a_k}{(1+Vt_i)^2} \end{array} \right],
\end{aligned}
$$

for $k \neq j$. The expectations of the elements of the Hessian for that subsection are

$$
\mathrm{E}\left(\frac{\partial^2 l_i}{\partial V^2}\right) = \frac{p_{7i}}{2} - \frac{p_{7i}+Vp_{8i}}{1+Vt_i} + \frac{p_{3i}^2}{2(1+Vt_i)^2}
$$

$$-p_{16i}+\frac{2p_{24i}}{1+Vt_i}-\frac{p_{22i}p_{3i}}{(1+Vt_i)^2}+\frac{Vp_{44i}}{1+Vt_i}$$
$$-\frac{2Vp_{3i}p_{24i}}{(1+Vt_i)^2}+\frac{2Vp_{25i}}{1+Vt_i}+\frac{Vp_{22i}p_{3i}^2}{(1+Vt_i)^3}-\frac{Vp_{22i}(p_{7i}+Vp_{8i})}{(1+Vt_i)^2},$$

$$\mathbf{E}\left(\frac{\partial^2 l_i}{\partial V\partial\delta_{j1}}\right) = -\frac{b_{ij}a_jf_j^2}{2}+\frac{b_{ij}a_jf_j^2+2Vb_{ij}a_j^2f_j^3}{2(1+Vt_i)}-\frac{Vp_{3i}b_{ij}a_jf_j^2}{2(1+Vt_i)^2}$$
$$+a_jb_{ij}\delta_{j1}f_j^3-\frac{p_{12ij}\sqrt{a_j}f_j^2}{1+Vt_i}+\frac{p_{22i}Vb_{ij}a_jf_j^2}{2(1+Vt_i)^2}$$
$$-\frac{f_j^2V\sqrt{a_j}p_{14ij}}{1+Vt_i}+\frac{V^2p_{24i}b_{ij}a_jf_j^2}{(1+Vt_i)^2}-\frac{2Vp_{12ij}a_j^{3/2}f_j^3}{1+Vt_i}+\frac{Vp_{12ij}p_{3i}\sqrt{a_j}f_j^2}{(1+Vt_i)^2}$$
$$-\frac{V^2p_{22i}p_{3i}b_{ij}a_jf_j^2}{(1+Vt_i)^3}+\frac{Vp_{22i}(b_{ij}a_jf_j^2+2Vb_{ij}a_j^2f_j^3)}{2(1+Vt_i)^2},$$

$$\mathbf{E}\left(\frac{\partial^2 l_i}{\partial\delta_{j1}^2}\right) = \frac{b_{ij}f_j^2}{2}-\frac{Vb_{ij}a_jf_j^3}{1+Vt_i}+\frac{V^2b_{ij}^2a_j^2f_j^4}{2(1+Vt_i)^2}$$
$$-b_{ij}\delta_{j1}f_j^3+\frac{Va_jb_{ij}\left(\delta_{j1}+Va_j(b_{ij}-1)\right)f_j^4}{1+Vt_i}-\frac{2V^2p_{12ij}a_j^{3/2}b_{ij}f_{ij}^4}{(1+Vt_i)^2}$$
$$+\frac{2Vp_{12ij}\sqrt{a_j}f_j^3}{(1+Vt_i)^3}+\frac{V^3p_{22i}b_{ij}^2a_j^2f_j^4}{(1+Vt_i)^3}-\frac{V^2p_{22i}b_{ij}a_jf_j^3}{(1+Vt_i)^2},$$

$$\mathbf{E}\left(\frac{\partial^2 l_i}{\partial\delta_{j1}\partial\delta_{k1}}\right) = \frac{Vf_j^2f_k^2}{1+Vt_i}\left[\begin{array}{c}\frac{Vb_{ij}a_jb_{ik}a_k}{2(1-Vt_i)}+Va_ja_kb_{ij}b_{ik}\\ -\frac{V\left(p_{12ij}\sqrt{a_j}b_{ik}a_k-p_{12ik}\sqrt{a_k}b_{ij}a_j\right)}{1-Vt_i}+\frac{V^2p_{22i}b_{ij}b_{ik}a_ja_k}{(1-Vt_i)^2}\end{array}\right].$$

The Hessian of Subsection (4.4) is found using

$$\frac{\partial l}{\partial\tau} = \frac{-n(c-1)\phi}{d}-\frac{n\phi}{f}+\frac{\left(g-\frac{\omega(h)}{f}\right)\phi}{(d)^2}-\frac{q_{22}+\frac{\omega(h)\phi}{(f)^2}}{d},$$

$$\frac{\partial l}{\partial\phi} = \frac{-n(c-1)(\tau-\omega)}{d}-\frac{n(\tau-\omega+c\omega)}{f}$$
$$+\frac{\left(g-\frac{\omega(h)}{f}\right)(\tau-\omega)}{(d)^2}-\frac{\left(q_{11}+\frac{\omega(h)(\tau-\omega-\omega c)}{(f)^2}-\frac{2\omega(\phi\tau_{11}-\sigma\tau_{12})}{f}\right)}{d},$$

$$\frac{\partial l}{\partial\sigma} = \frac{2n(c-1)\sigma}{d}+\frac{2n\sigma}{f}-\frac{2\left(g-\frac{\omega(h)}{f}\right)\sigma}{(d)^2}$$
$$+\frac{2\left(q_{12}+\frac{\omega(h)\sigma}{(f)^2}+\frac{\omega(\sigma\tau_{22}-\phi\tau_{12})}{f}\right)}{d},$$

$$\frac{\partial l}{\partial\omega} = \frac{n(c-1)\phi}{d}-\frac{n\phi(c-1)}{f}-\frac{\left(g-\frac{\omega(h)}{f}\right)\phi}{(d)^2}$$
$$-\frac{-q_{22}-\frac{h}{f}+\frac{\omega(h)\phi(c-1)}{(f)^2}}{d},$$

$$\frac{\partial^2 l}{\partial\tau^2} = \frac{n(c-1)\phi^2}{(d)^2}+\frac{n\phi^2}{(f)^2}-2\frac{\left(g-\frac{\omega(h)}{f}\right)\phi^2}{(d)^3}$$
$$+2\frac{\left(q_{22}+\frac{\omega(h)\phi}{(f)^2}\right)\phi^2}{(d)^2}+2\frac{\omega(h)\phi^2}{(d)(f)^3}.$$

$$\begin{aligned}
\frac{\partial^2 l}{\partial \phi^2} &= \frac{n(c-1)(\tau-\omega)^2}{(d)^2} + \frac{n(\tau+\omega(c-1))^2}{(f)^2} - 2\frac{\left(g-\frac{\omega(h)}{f}\right)(\tau-\omega)^2}{(d)^3} \\
&\quad +2\frac{\left(q_{11}+\frac{\omega(h)(\tau-\omega(c-1))}{(f)^2}-\frac{2\omega(\phi r_{11}-\sigma r_{12})}{f}\right)(\tau-\omega)}{(d)^2} \\
&\quad -\frac{2}{d}\left[\frac{\omega(h)(\tau+\omega(c-1))^2}{(f)^3}+\frac{2\omega(\phi r_{11}-\sigma r_{12})}{(f)^2}-\frac{\omega r_{11}}{f}\right], \\
\frac{\partial^2 l}{\partial \sigma^2} &= 2\frac{n(c-1)}{d}+4\frac{n(c-1)\sigma^2}{(d)^2}+2\frac{n}{f}+4\frac{n\sigma^2}{(f)^2} \\
&\quad -8\frac{\left(g-\frac{\omega(h)}{f}\right)\sigma^2}{(d)^3}+8\frac{\left(q_{12}+\frac{\omega(h)\sigma}{(f)^2}+\frac{\omega(\sigma r_{22}-\phi r_{12})}{f}\right)\sigma}{(d)^2} \\
&\quad -2\frac{\phi q_{11}+(\tau-\omega)q_{22}-2\sigma q_{12}-\frac{\omega(h)}{f}}{(d)^2} \\
&\quad +\frac{2}{d}\left[4\frac{\omega(h)\sigma^2}{(f)^3}+4\frac{\omega(\sigma r_{22}-\phi r_{12})\sigma}{(f)^2}+\frac{\omega(h)}{(f)^2}+\frac{\omega r_{22}}{f}\right], \\
\frac{\partial^2 l}{\partial \omega^2} &= \frac{n(c-1)\phi^2}{(d)^2}+\frac{n\phi^2(c-1)^2}{(f)^2}-2\frac{\left(g-\frac{\omega(h)}{f}\right)\phi^2}{(d)^3} \\
&\quad -2\frac{\left(-q_{22}-\frac{h}{f}+\frac{\omega(h)\phi(c-1)}{(f)^2}\right)\phi}{(d)^2}-\frac{2\frac{(h)\phi(c-1)}{(f)^2}-2\frac{\omega(h)\phi^2(c-1)^2}{(f)^3}}{d}. \\
\frac{\partial^2 l}{\partial \tau \partial \phi} &= \frac{-n(c-1)}{d}+\frac{n(c-1)\phi(\tau-\omega)}{(d)^2}-\frac{n}{f} \\
&\quad +\frac{n\phi(\tau-\omega+\omega c)}{(f)^2}-\frac{2\left(g-\frac{\omega(h)}{f}\right)\phi(\tau-\omega)}{(d)^3} \\
&\quad +\frac{\left(q_{11}+\frac{\omega(h)(\tau-\omega-\omega c)}{(f)^2}-\frac{2\omega(\phi r_{11}-\sigma r_{12})}{f}\right)\phi}{(d)^2}+\frac{g-\frac{\omega(h)}{f}}{(d)^2} \\
&\quad +\frac{\left(q_{22}+\frac{\omega(h)\phi}{(f)^2}\right)(\tau-\omega)}{(d)^2} \\
&\quad -\frac{-2\frac{\omega(h)\phi(\tau-\omega+\omega c)}{(f)^3}+\frac{2\omega(\phi r_{11}-\sigma r_{12})\phi}{(f)^2}+\frac{\omega(h)}{(f)^2}}{d}, \\
\frac{\partial^2 l}{\partial \tau \partial \sigma} &= \frac{-2n(c-1)\phi\sigma}{(d)^2}-\frac{2n\phi\sigma}{(f)^2}+\frac{4\left(g-\frac{\omega(h)}{f}\right)\phi\sigma}{(d)^3} \\
&\quad -\frac{2\left(q_{12}+\frac{\omega(h)\sigma}{(f)^2}+\frac{\omega(\sigma r_{22}-\phi r_{12})}{f}\right)\phi}{(d)^2} \\
&\quad -\frac{4\frac{\omega(h)\phi\sigma}{(f)^3}+\frac{2\omega(\sigma r_{22}-\sigma r_{12})\phi}{(f)^2}}{d}, \\
\frac{\partial^2 l}{\partial \tau \partial \omega} &= -\frac{n(c-1)\phi^2}{(d)^2}+\frac{n\phi^2(c-1)}{(f)^2}+\frac{2\left(g-\frac{\omega(h)}{f}\right)\phi^2}{(d)^3}
\end{aligned}$$

$$+\frac{\left(-q_{22}-\frac{h}{f}+\frac{\omega(h)\phi(c-1)}{(f)^2}\right)\sigma}{(d)^2}$$

$$-\frac{\left(q_{22}+\frac{\omega(h)\phi}{(f)^2}\right)\phi}{(d)^2}-\frac{\frac{(h)\phi}{(f)^2}-2\frac{\omega(h)\phi^2(c-1)}{(f)^3}}{d},$$

$$\begin{aligned}
\frac{\partial^2 l}{\partial\phi\partial\sigma} = {} & -2\frac{n(c-1)(\tau-\omega)\sigma}{(d)^2}-2\frac{n(\tau-\omega+\omega c)\sigma}{(f)^2}+ \\
& 4\frac{\left(g-\frac{\omega(h)}{f}\right)(\tau-\omega)\sigma}{(d)^3}-2\frac{\left(q_{12}+\frac{\omega(h)\sigma}{(f)^2}+\frac{\omega(\sigma\tau_{22}-\phi\tau_{12})}{f}\right)(\tau-\omega)}{(d)^2} \\
& -2\frac{\left(q_{11}+\frac{\omega(h)(\tau-\omega-\omega c)}{(f)^2}-\frac{2\omega(\phi\tau_{11}-\sigma\tau_{12})}{f}\right)\sigma}{(d)^3}- \\
& 4\frac{\omega(h)(\tau+(c-1)\omega)\sigma}{(f)^3(d)}+2\frac{\omega(\sigma\tau_{22}-\phi\tau_{12})(\tau+(c-1)\omega)}{(f)^2(d)} \\
& -4\frac{\omega(\phi\tau_{11}-\sigma\tau_{12})\sigma}{(f)^2(d)}+2\frac{\omega\tau_{12}}{(f)(d)},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l}{\partial\phi\partial\omega} = {} & \frac{n(c-1)}{d}-\frac{n(c-1)\phi(\tau-\omega)}{(d)^2}-\frac{n(c-1)}{f} \\
& +\frac{n(\tau+(c-1)\omega)\phi(c-1)}{(f)^2}+2\frac{\left(g-\frac{\omega(h)}{f}\right)\phi(\tau-\omega)}{(d)^3} \\
& +\frac{\left(-q_{22}-\frac{h}{f}+\frac{\omega(h)\phi(c-1)}{(f)^2}\right)(\tau-\omega)}{(d)^2} \\
& -\frac{g-\frac{\omega(h)}{f}}{(d)^2}-\frac{\left(q_{11}+\frac{\omega(h)(\tau+(c-1)\omega)}{(f)^2}-\frac{2\omega(\phi\tau_{11}-\sigma\tau_{12})}{f}\right)\phi}{(d)^2} \\
& -\frac{(h)(\tau+(c-1)\omega)}{(f)^2(d)}+2\frac{\omega(h)(\tau+(c-1)\omega)\phi(c-1)}{(f)^3(d)} \\
& -\frac{\omega(h)(c-1)}{(f)^2(d)}+2\frac{\phi\tau_{11}-\sigma\tau_{12}}{(f)(d)}-2\frac{\omega(\sigma\tau_{11}-\sigma\tau_{12})\phi(c-1)}{(f)^2(d)},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l}{\partial\sigma\partial\omega} = {} & 2\frac{n(c-1)\phi\sigma}{(d)^2}-2\frac{n\sigma\phi(c-1)}{(f)^2}-4\frac{\left(g-\frac{\omega(h)}{f}\right)\phi\sigma}{(d)^3} \\
& -2\frac{\left(-q_{22}-\frac{h}{f}+\frac{\omega(h)\phi(c-1)}{(f)^2}\right)\sigma}{(d)^2}+2\frac{\left(q_{12}+\frac{\omega(h)\sigma}{(f)^2}+\frac{\omega(\sigma\tau_{22}-\phi\tau_{12})}{f}\right)}{(d)^2} \\
& -\frac{2}{d}\left[\begin{array}{c}-\frac{(h)\sigma}{(f)^2}+2\frac{\omega(h)\sigma\phi(c-1)}{(f)^3}-\frac{\sigma\tau_{22}-\phi\tau_{12}}{f} \\ \frac{\omega(\sigma\tau_{22}-\phi\tau_{12})\phi(c-1)}{(f)^2}\end{array}\right].
\end{aligned}$$

Twice the expectations of the above second derivatives are

$$2\mathrm{E}\left(\frac{\partial^2 l}{\partial\tau^2}\right) = \frac{-\sigma^2 nc\left(\begin{array}{c}\phi^2\tau^2+3\phi^2\omega^2(c-1)-4\phi^2\tau\omega+4\phi\omega\sigma^2 \\ +\omega^2c^2\phi^2+\sigma^4+2\phi^2\tau\omega c-2\sigma^2\omega c\phi-2\phi\tau\sigma^2\end{array}\right)}{(d)^2(f)^2}.$$

$$2\mathrm{E}\left(\frac{\partial^2 l}{\partial\tau\partial\phi}\right) = \frac{-\sigma^2 nc\left(\begin{array}{c}\phi^2\tau^2 - 3\phi^2\omega^2(c-1) - 4\phi^2\tau\omega + 4\phi\omega\sigma^2 \\ +\omega^2c^2\phi^2 + \sigma^4 + 2\phi^2\tau\omega c - 2\sigma^2\omega c\phi - 2\phi\tau\sigma^2\end{array}\right)}{(d)^2(f)^2},$$

$$2\mathrm{E}\left(\frac{\partial^2 l}{\partial\tau\partial\sigma}\right) = \frac{2\sigma\phi nc\left(\begin{array}{c}\phi^2\tau^2 - 3\phi^2\omega^2(c-1) - 4\phi^2\tau\omega + 4\phi\omega\sigma^2 \\ +\omega^2c^2\phi^2 + \sigma^4 + 2\phi^2\tau\omega c - 2\sigma^2\omega c\phi - 2\phi\tau\sigma^2\end{array}\right)}{(d)^2(f)^2},$$

$$2\mathrm{E}\left(\frac{\partial^2 l}{\partial\tau\partial\omega}\right) = \frac{\phi^3 nc\omega\left(2\phi\omega - 3\omega c\phi - 2\phi\tau + 2\sigma^2 + \omega c^2\phi + 2c\left(\phi\tau - \sigma^2\right)\right)}{(d)^2(f)^2},$$

$$2\mathrm{E}\left(\frac{\partial^2 l}{\partial\phi^2}\right) = \frac{-nc\left(\begin{array}{c}\tau^4\phi^2 + \tau^2\sigma^4 + \omega^4\phi^2 - 2c\omega^4\phi^2 - \omega^2\sigma^4 + c^2\omega^4\phi^2 \\ +c\omega^2\sigma^4 - 4\tau^3\phi^2\omega - 2\tau\phi\sigma^2 + 2c\tau^3\phi^2\omega + 6\tau^2\phi^2\omega^2 \\ +4\tau^2\phi\omega\sigma^2 - 6c\tau^2\phi^2\omega^2 - 2c\tau^2\phi\omega\sigma^2 + c^2\tau^2\phi^2\omega^2 \\ -4\tau\omega^3\phi^2 - 2\tau\omega^2\phi\sigma^2 + 6c\tau\omega^3\phi^2 + 2c\tau\omega^2\phi\sigma^2 - 2c^2\tau\omega^3\phi^2\end{array}\right)}{(d)^2(f)^2},$$

$$2\mathrm{E}\left(\frac{\partial^2 l}{\partial\phi\partial\sigma}\right) = \frac{-2\sigma nc\left(\begin{array}{c}-\phi^2\tau^3 - \tau\sigma^4 + 2\phi^2\omega^3 - 3\omega^3c\phi^2 + 4\phi^2\tau^2\omega \\ +2\phi\tau^2\sigma^2 - 2c\tau^2\phi^2\omega - 5\tau\phi^2\omega^2 - 4\tau\phi\omega\sigma^2 \\ +5c\tau\phi^2\omega^2 + 2c\tau\phi\omega\sigma^2 - c^2\tau\phi^2\omega^2 + 2\phi\omega^2\sigma^2 \\ +\omega^3c^2\phi^2 - 2\omega^2c\phi\sigma^2\end{array}\right)}{(d)^2(f)^2},$$

$$2\mathrm{E}\left(\frac{\partial^2 l}{\partial\phi\partial\omega}\right) = \frac{\sigma^2\omega\phi nc\left(2\phi\omega - 3\omega c\phi - 2\phi\tau + 2\sigma^2 + \omega c^2\phi + 2c\phi\tau - 2c\sigma^2\right)}{(d)^2(f)^2},$$

$$2\mathrm{E}\left(\frac{\partial^2 l}{\partial\sigma^2}\right) = \frac{2nc\left(\begin{array}{c}2\phi^3\omega^3 - 3\phi^3\omega^3c + \phi^3\omega^3c^2 + 4\phi^3\tau^2\omega - 5\phi^3\tau\omega^2 \\ +5\phi^3\tau\omega^2c - \phi^2\omega^2\sigma^2 - 4\phi\omega\sigma^4 + \phi^2\omega^2\sigma^2c - \phi^3\tau^3 - \sigma^6 \\ +\phi^2\tau^2\sigma^2 - 2\phi^3\tau^2\omega c + \phi\tau\sigma^4 - \phi^3\tau\omega^2c^2 + 2\sigma^4\omega c\phi - \sigma^2\omega^2c^2\phi^2\end{array}\right)}{(d)^2(f)^2},$$

$$2\mathrm{E}\left(\frac{\partial^2 l}{\partial\sigma\partial\omega}\right) = \frac{-2\sigma\phi^2 nc\omega\left(2\phi\omega - 3\omega c\phi - 2\phi\tau + 2\sigma^2 + \omega c^2\phi + 2c\phi\tau - 2c\sigma^2\right)}{(d)^2(f)^2},$$

$$2\mathrm{E}\left(\frac{\partial^2 l}{\partial\omega^2}\right) = \frac{-\phi^2 nc\left(\begin{array}{c}-\phi^2\tau^2 + \phi^2\omega^2 - 2\phi^2\omega^2c + \omega^2c^2\phi^2 + c\phi^2\tau^2 \\ +\left(\sigma^4 - 2\phi\tau\sigma^2\right)(c-1)\end{array}\right)}{(d)^2(f)^2},$$

giving the information in the csp for these parameters. In these expressions,

$$\begin{aligned} f &= d + \omega c \phi = (\tau - \omega)\,\phi - \sigma^2 + \omega c \phi, \\ g &= \phi q_{11} + (\tau - \omega)\, q_{22} - 2\sigma q_{12} \text{ and} \\ h &= \phi^2 r_{11} + \sigma^2 r_{22} - 2\phi\sigma r_{12}. \end{aligned}$$

VITA

Name: Andrew Montgomery Hartley

Address: Old Dominion University, Department of Mathematics and Statistics
Hampton Boulevard, Norfolk, Virginia 23529

Education: 1994 M.S. Old Dominion University, Norfolk, Virginia.

1987 Naval Supply Corps Basic Qualification.
Naval Supply Corps School, Athens, Georgia.

1986 B.S. Mathematics, Allegheny College, Meadville, Pennsylvania.

Awards: 1997 Second Place, Virginia Academy of Sciences Competition for Submission and Presentation of Papers.

1996 Navy Achievement Medal, U.S. Navy.

1992 Navy Expeditionary Medal, U.S. Navy.

1983-1985 Alden Scholar, Allegheny College.

1980 Eagle Scout, Boy Scouts of America.

1978-1980 National Honor Society.

Hartley, A. Ph.D. Math December 3, 1997

Table of Contents	-	"Page" should be right justified with page numbers
List of Tables	-	"Page" should be right justified with page numbers
	-	add "Table" as left heading
	-	use indent or spacing to left align title captions
References	-	remove bold from volume numbers to match Journal Model

JHR