Summer 2006

Efficient Unbiased Estimating Equations for Analyzing Structured Correlation Matrices

Yihao Deng
Old Dominion University

EFFICIENT UNBIASED ESTIMATING EQUATIONS FOR ANALYZING STRUCTURED CORRELATION MATRICES

by

Yihao Deng
B.S. 1998, Dongbei University of Finance & Economics
M.S. 2001, Dalian University of Technology
M.S. 2004, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

MATHEMATICS AND STATISTICS

OLD DOMINION UNIVERSITY
August 2006

Approved by:

N. Rao Chaganty (Director)

Dayanand Naik

Michael Doviak

Larry Filer

ABSTRACT

EFFICIENT UNBIASED ESTIMATING EQUATIONS FOR ANALYZING STRUCTURED CORRELATION MATRICES

Yihao Deng
Old Dominion University, 2006
Director: Dr. N. Rao Chaganty

Analysis of dependent continuous and discrete data has become an active area of research. For normal data, correlations fully quantify the dependence. And historically, maximum likelihood method has been very successful to estimate the correlations and unbiased estimating equation approach has become a popular alternative when there maybe a departure from normality. In this thesis we show that the optimal unbiased estimating equation coincides with the likelihood equations for normal data. We then introduce a general class of weighted unbiased estimating equations to estimate parameters in a structured correlation matrix. We derive expressions for asymptotic covariance of the estimates, and use those expressions to determine the optimal weights. We also study an important subclass of unbiased estimating equations. The optimal weights for this subclass are not tractable, especially for the familial correlation structure. We suggest approximations and study performance of these approximate weights using simulations.

For familial binary responses we first investigate ranges of associations measures, which include odds ratios, kappa statistics, and relative risks besides correlations. Knowing and understanding these ranges is important for developing efficient estimation methods. We study estimation of the familial correlations using a probit model and stochastic representation of the latent variables. We discuss some extensions of our results to nuclear families. Some real life examples are presented to illustrate the estimation methods.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. N. Rao Chaganty, for his continual guidance and support throughout the entire process of my dissertation. His comments and suggestions lead to many important results in this research endeavor. Had it not been for his intellect and patience, I would not have been able to accomplish this. Dr. Chaganty was my major source of inspiration during my graduate studies. His persistence for scientific precision helped me establish such a philosophy in my own academic investigations. Under his supervision, the teaching and consulting experience I gained were both enjoyable, rewardable, and helped to shape my future career development.

I would also like to thank Dr. Dayanand Naik, Dr. Michael Doviak, and Dr. Larry Filer for their roles on my dissertation committee. I especially want to recognize Dr. Naik for his many helpful constructive criticisms, which proved to be invaluable. I am very much indebted to the statistics group in the Department of Mathematics and Statistics, which through its experienced instructors, including Dr. Ram Dahiya and Dr. Larry Lee, have deepened my insights in the field of statistics.

And last but not the least, I would like to thank all the professors and staff members in the Department of Mathematics and Statistics for providing me with a rich, academic atmosphere and a pleasant learning environment. My professors and fellow graduate students have made my graduate study an unforgettable experience.

I dedicate this thesis to my parents.

TABLE OF CONTENTS

CHAPTERS Page

CHAPTERS Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Continuous and discrete repeated measurements data naturally arise in many research studies in biomedicine, psychology, health and social sciences. Hence there is a strong need for developing efficient and easy-to-implement statistical estimation methods for analyzing such data. In a seminal paper Godambe (1960) introduced the theory of unbiased estimating equations for independent observations. This theory was extended for correlated and dependent data by Liang and Zeger (1986), who introduced generalized estimating equations. In recent years generalized estimation equations and methods based on unbiased estimating equations have become a popular alternative to the traditional maximum likelihood estimation. However, unbiased estimating equations for structured correlation matrices have not been explored systematically. The main goal of this thesis is to adequately address unbiased estimating equations, derive important properties, and develop efficient methods for estimating the correlation parameter in structured correlation matrices.

I.1 The General Setup

The classical setup for the longitudinal data analysis is as follows. Suppose that we have n independent subjects or clusters in a sample. On subject i or cluster i, we observe response $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{it_i})'$, where y_{ij} for $j = 1, 2, \ldots, t_i$ could be a continuous measurement or binary (for example, it could be an indicator of yes/no, success/failure or present/absent). The expected value of $\mathbf{y}_i$ is given by $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{it_i})'$ and the variance covariance matrix of $\mathbf{y}_i$ is assumed to be $\mathbf{V}_i = (\sigma_{ij}\sigma_{ik}\alpha_{ijk})$ for $j, k = 1, 2, \ldots, t_i$ and $\alpha_{ijk} = 1$ if $j = k$. The general setup is presented in Table 1.1.

In addition to the response variable $\mathbf{y}_i$, we have a covariate matrix $\mathbf{X}_i'$ containing measurements on some covariates for subject or cluster i. Here $\mathbf{X}_i'$ is a matrix with t_i rows and p columns. We assume that the mean $\boldsymbol{\mu}_i$ of response $\mathbf{y}_i$ is a function of $\mathbf{X}_i'$ and an unknown regression parameter $\boldsymbol{\beta}$, that is, $\boldsymbol{\mu}_i = g(\mathbf{X}_i'\boldsymbol{\beta})$ for some known

This dissertation follows the style of *Journal of the American Statistical Association.*

Table 1.1: General setup

subject	response	mean	covariance matrix			
i	y_{i1}	μ_{i1}	σ_{i1}^2	$\sigma_{i1}\sigma_{i2}\alpha_{i12}$	$\cdots$	$\sigma_{i1}\sigma_{it_i}\alpha_{i1t_i}$
	y_{i2}	μ_{i2}	$\sigma_{i1}\sigma_{i2}\alpha_{i21}$	σ_{i2}^2	$\cdots$	$\sigma_{i2}\sigma_{it_i}\alpha_{i2t_i}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	y_{it_i}	μ_{it_i}	$\sigma_{i1}\sigma_{it_i}\alpha_{it_i1}$	$\sigma_{i2}\sigma_{it_i}\alpha_{it_i2}$	$\cdots$	$\sigma_{it_i}^2$

function $g(\cdot)$.

Generally, the goal is to estimate $\boldsymbol{\beta}$ efficiently. However, in some analysis, the correlation parameters $\boldsymbol{\alpha}_i = (\alpha_{i12}, \alpha_{i13}, \ldots, \alpha_{i\{t_i-1\}t_i})'$ for $i = 1, 2, \ldots, n$ maybe of research interest, and we need to efficiently estimate them as well. Many books focusing on the analysis of longitudinal data are available, for example, Diggle et al. (2002) and Fitzmaurice et al. (2004). A comparison and review of various estimation methods can be found in Zeger and Liang (1992) and Wu et al. (2001). A likelihood approach for efficient parameter estimation is facilitated by general linear models, which we discuss in the next section.

I.2 Generalized Linear Models

Generalized linear models (GLMs) were introduced by Nelder and Wedderburn (1972). GLMs provide a unified class of models for regression analysis of independent observations, which could be continuous or discrete. Possible applications of GLMs in various fields of study can be found in McCullagh and Nelder (1989) and Myers et al. (2002). There are three key components in generalized linear models: the marginal response distribution, a linear predictor, and a link function. The detailed description of these components is as follows.

(1) The first component is the distribution of the response variable y_{ij}. In GLMs it is assumed to belong to the exponential family. The probability density function or probability mass function can be written as

$$f(y_{ij}; \theta_i, \phi) = \exp\left\{\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi)\right\} \qquad (1.2.1)$$

where the functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are known. From (1.2.1), we can check that the mean and variance of y_{ij} are $\mathrm{E}(y_{ij}) = \mu_{ij} = b'(\theta_{ij})$ and $\mathrm{Var}(y_{ij}) = b''(\theta_{ij})a(\phi)$, respectively. Thus, the variance of y_{ij} is a function of its mean, and this function is often referred to as the "variance function". In some specific examples, $a(\phi) = \phi/w$, where w is a known *prior weight.*

(2) The second component in GLMs is the linear predictor η_{ij}. This is simply a linear combination of the covariates and the regression parameters, that is, $\eta_{ij} = \mathbf{X}'_{ij}\boldsymbol{\beta} = \sum_{k=1}^{p} \beta_k x_{ijk}$.

(3) The third component is the link function $g(\cdot)$. This function specifies the relationship between the linear predictor η_{ij} and the expected value μ_{ij} of y_{ij} as $\eta_{ij} = g(\mu_{ij})$. The link function $g(\cdot)$ is known as the canonical link function when $g(\mu_{ij}) = \theta_{ij}$.

Below are two common examples of GLMs.

Example 1.1 *Suppose y_{ij} are independently distributed as univariate normal. The probability density function of y_{ij} is*

$$\begin{aligned} f(y_{ij}, \theta_{ij}, \phi) &= \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left\{\frac{-(y_{ij}-\mu_{ij})^2}{2\sigma_{ij}^2}\right\} \\ &= \exp\left\{\frac{y_{ij}\mu_{ij} - \mu_{ij}^2/2}{\sigma_{ij}^2} - \frac{1}{2}\left(y_{ij}^2/\sigma_{ij}^2 + \log(2\pi\sigma_{ij}^2)\right)\right\} \end{aligned}$$

so that $\theta_{ij} = \mu_{ij}$, $\phi = \sigma_{ij}^2$. In this case $a(\phi) = \phi$, $b(\theta_{ij}) = \theta_{ij}^2/2$ and $c(y_{ij}, \phi) = -\frac{1}{2}\left\{y_{ij}^2/\sigma_{ij}^2 + \log(2\pi\sigma_{ij}^2)\right\}$. It is easy to check that the link function is the identity function, $g(\mu_{ij}) = \mu_{ij}$.

Example 1.2 *Suppose y_{ij} are independently distributed as binomial with parameters n_{ij} and p_{ij}. The probability mass function of y_{ij} is*

$$\begin{aligned} f(y_{ij}, \theta_{ij}, \phi) &= \frac{n_{ij}!}{(n_{ij}-y_{ij})!\,y_{ij}!} p_{ij}^{y_{ij}} (1-p_{ij})^{n_{ij}-y_{ij}} \\ &= \exp\left\{y_{ij}\log\left(\frac{p_{ij}}{1-p_{ij}}\right) + n_{ij}\log(1-p_{ij}) + \log\left(\frac{n_{ij}!}{(n_{ij}-y_{ij})!\,y_{ij}!}\right)\right\} \end{aligned}$$

so that $\theta_{ij} = \log\left(\frac{p_{ij}}{1-p_{ij}}\right)$, $\phi = 1$ and thus $a(\phi) = \phi$, $b(\theta_{ij}) = n_{ij}\log\{1+\exp(\theta_{ij})\}$ and $c(y_{ij}, \phi) = \log\left(\frac{n_{ij}!}{(n_{ij}-y_{ij})!\,y_{ij}!}\right)$. The link function is $g(\mu_{ij}) = \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right)$.

Table 1.2 contains the canonical link and variance functions for some common univariate distributions with mean μ.

Table 1.2: Canonical link and variance functions

Distribution	Canonical link function		Variance function
Normal	Identity	$\eta = \mu$	1
Binomial	Logit	$\eta = \log\left(\frac{\mu}{1-\mu}\right)$	$\mu\left(1-\frac{\mu}{n}\right)$
Poisson	Log	$\eta = \log(\mu)$	μ
Gamma	Inverse	$\eta = 1/\mu$	μ^2
Inverse Gaussian	Inverse square	$\eta = 1/\mu^2$	μ^3

In GLMs the estimation of the regression parameter $\boldsymbol{\beta}$ is carried out by the principle of maximum likelihood. We illustrate this method and present some details on how to compute the standard errors with an example.

I.3 Maximum Likelihood Estimate

In this section, we present some details of maximum likelihood estimation for the longitudinal setup described in Section I.1. Assume that the distribution of $\mathbf{y}_i$ is multivariate normal with mean $\boldsymbol{\mu}_i = \mathbf{X}_i'\boldsymbol{\beta}$ and covariance matrix $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\sigma}_i^2, \boldsymbol{\alpha}_i)$. The likelihood function is

$$L(\boldsymbol{\beta}, \mathbf{V}_i) = \frac{1}{(2\pi)^{\frac{1}{2}\sum_{i=1}^n t_i} \prod_{i=1}^n \sqrt{|\mathbf{V}_i|}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})\right\}, \tag{1.3.1}$$

and the log-likelihood is

$$\ell(\boldsymbol{\beta}, \mathbf{V}_i) = -\frac{1}{2}\left\{\sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) + \sum_{i=1}^n \log|\mathbf{V}_i| + \log(2\pi)\sum_{i=1}^n t_i\right\}. \tag{1.3.2}$$

Taking the partial derivative of (1.3.2) with respect to $\boldsymbol{\beta}$ and equating to zero, we get

$$\sum_{i=1}^n \mathbf{X}_i\mathbf{V}_i^{-1}\mathbf{y}_i - \left(\sum_{i=1}^n \mathbf{X}_i\mathbf{V}_i^{-1}\mathbf{X}_i'\right)\boldsymbol{\beta} = \mathbf{0}.$$

Solving this equation gives the maximum likelihood (ML) estimate of $\boldsymbol{\beta}$ as

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{X}_i \mathbf{V}_i^{-1} \mathbf{X}_i'\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{V}_i^{-1} \mathbf{y}_i . \tag{1.3.3}$$

Similarly, differentiating with respect to $\boldsymbol{\sigma}_i^2$ and $\boldsymbol{\alpha}_i$, and equating to zero, we have

$$\sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})' \frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\sigma}_i^2} (\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) + \sum_{i=1}^{n} \frac{\log|\mathbf{V}_i|}{\partial \boldsymbol{\sigma}_i^2} = \mathbf{0}, \tag{1.3.4}$$

$$\sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})' \frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\alpha}_i} (\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) + \sum_{i=1}^{n} \frac{\log|\mathbf{V}_i|}{\partial \boldsymbol{\alpha}_i} = \mathbf{0}. \tag{1.3.5}$$

Using the identities

$$\begin{aligned} \frac{\partial \log|\mathbf{V}_i|}{\partial \boldsymbol{\alpha}_i} &= \frac{1}{|\mathbf{V}_i|} \frac{\partial |\mathbf{V}_i|}{\partial \boldsymbol{\alpha}_i} \\ \frac{\partial |\mathbf{V}_i|}{\partial \boldsymbol{\alpha}_i} &= |\mathbf{V}_i| \operatorname{tr}\left(\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\alpha}_i}\right) \\ \frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\alpha}_i} &= -\mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\alpha}_i} \mathbf{V}_i^{-1}, \end{aligned}$$

equation (1.3.5) can be written as

$$\sum_{i=1}^{n} \operatorname{tr}\left\{ \frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\alpha}_i} (\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' - \mathbf{V}_i) \right\} = \mathbf{0}, \tag{1.3.6}$$

where $\boldsymbol{\varepsilon}_i = \mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}$ is the residual. Later, we will use this form (1.3.6) of the ML equation to make comparisons with the unbiased estimating equation approaches. The solutions to the equations (1.3.4) and (1.3.6) are usually not in the closed form, and a nonlinear optimization algorithm is often used to solve (1.3.3), (1.3.4) and (1.3.6) simultaneously.

The covariance matrix of regression parameter estimate $\widehat{\boldsymbol{\beta}}$ is obtained by finding the inverse of the Fisher information matrix, which is given by

$$\mathbf{I}_{\beta} = \left[\mathrm{E}\left\{-\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}^2}\right\}\right] = \left\{\sum_{i=1}^{n} \mathbf{X}_i \mathbf{V}_i^{-1} \mathbf{X}_i'\right\}.$$

An estimate of $\mathbf{I}_{\beta}$ is $\sum_{i=1}^{n} \mathbf{X}_i \widehat{\mathbf{V}}_i^{-1} \mathbf{X}_i'$ where $\widehat{\mathbf{V}}_i = \mathbf{V}_i(\widehat{\boldsymbol{\sigma}}_i^2, \widehat{\boldsymbol{\alpha}}_i)$ is the covariance matrix evaluated at the estimated parameters $\widehat{\boldsymbol{\sigma}}_i$ and $\widehat{\boldsymbol{\alpha}}_i$. More generally, the covariance matrix of the parameter $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}^2, \widehat{\boldsymbol{\alpha}})'$ with $\widehat{\boldsymbol{\sigma}}^2 = (\widehat{\boldsymbol{\sigma}}_1^2, \ldots, \widehat{\boldsymbol{\sigma}}_n^2)'$ and $\widehat{\boldsymbol{\alpha}} = (\widehat{\boldsymbol{\alpha}}_1, \ldots, \widehat{\boldsymbol{\alpha}}_n)'$

though complicated can be obtained by taking the inverse of the Fisher information matrix

$$-\begin{bmatrix} \mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \beta^2}\right\} & \mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \beta \partial \sigma^2}\right\} & \mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \beta \partial \boldsymbol{\alpha}}\right\} \\ \mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \sigma^2 \partial \beta}\right\} & \mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial (\sigma^2)^2}\right\} & \mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \sigma^2 \partial \boldsymbol{\alpha}}\right\} \\ \mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \boldsymbol{\alpha} \partial \beta}\right\} & \mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \boldsymbol{\alpha} \partial \sigma^2}\right\} & \mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \boldsymbol{\alpha}^2}\right\} \end{bmatrix} \tag{1.3.7}$$

evaluated at $\widehat{\boldsymbol{\theta}}$.

For non-normal correlated models the likelihood function is intractable and calculating the ML estimates pose computationally challenging problems. An alternative method is the generalized estimating equation approach which we briefly discuss in the next section.

I.4 Generalized Estimating Equations

As an alternative to the full likelihood approach Liang and Zeger (1986) introduced the generalized estimating equations (GEEs). Their approach does not require complete specification of the likelihood function and can be thought as an extension of GLMs for correlated observations. The general form of the generalized estimation equation is

$$\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1}\Big(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\Big) = 0\,, \tag{1.4.1}$$

where $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$ and $\mathbf{V}_i = \phi\, \mathbf{A}_i^{-1/2} \mathbf{R}_i^{-1} \mathbf{A}_i^{-1/2}$ is the so-called "working" covariance matrix, ϕ is a scale parameter, $\mathbf{R}_i$ is the "working" correlation matrix depending on $\boldsymbol{\alpha}_i$, and $\mathbf{A}_i = \mathrm{diag}(\sigma_{ij}^2)$ for $j = 1, 2, \ldots, t_i$. Since $\boldsymbol{\mu}_i = g^{-1}(\boldsymbol{\eta}_i) = g^{-1}(\mathbf{X}_i' \boldsymbol{\beta})$ and usually $g(\cdot)$ is a nonlinear function, an iterative algorithm described below, is required to get the solution to (1.4.1).

(1) With initial values of $\boldsymbol{\alpha}^{(0)}$, $\sigma_{ij}^{(0)}$, solve (1.4.1) for $\boldsymbol{\beta}$. Denote the solution as $\boldsymbol{\beta}^{(1)}$.

At the kth step for $k = 1, 2, \ldots$

(2) Compute the estimate of σ_{ij} using the residuals $\boldsymbol{\varepsilon}_i^{(k)} = \mathbf{y}_i - \boldsymbol{\mu}_i\left(\boldsymbol{\beta}^{(k)}\right)$.

(3) Compute the Pearson residuals as $\mathbf{z}_i^{(k)} = \mathbf{A}_i^{-\frac{1}{2}}(\boldsymbol{\mu}_i)\left(\mathbf{y}_i - \boldsymbol{\mu}_i\left(\boldsymbol{\beta}^{(k)}\right)\right)$.

(4) Update $\boldsymbol{\alpha}$ using $\mathbf{z}_i^{(k)}$ to get $\boldsymbol{\alpha}^{(k)}$, and $\mathbf{V}_i^{(k)}$.

(5) Update $\boldsymbol{\beta}$ using the equation

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \left[\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \left(\mathbf{V}_i^{(k)}\right)^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right]^{-1} \left[\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \left(\mathbf{V}_i^{(k)}\right)^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)\right].$$

(6) Check convergence criteria for $|\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}|$ and $|\boldsymbol{\alpha}^{(k+1)} - \boldsymbol{\alpha}^{(k)}|$. If the criteria are met, then stop; otherwise, go to (2) and repeat steps (2) – (6).

For the GEE method the updating σ_{ij} and $\boldsymbol{\alpha}$ in steps (2) and (4) is done using the method of moments. The variance σ_{ij} is estimated by $\boldsymbol{\varepsilon}_i^{(k)}\boldsymbol{\varepsilon}_i^{(k)\prime}$ and the estimate of $\boldsymbol{\alpha}$ is computed using the residuals $\mathbf{z}_i$. For example, if the "working" covariance matrix is of the form $\sigma^2\mathbf{R}_i$, and $\mathbf{R}_i$ has an exchangeable structure, the estimates are given by

$$\hat{\alpha} = \frac{1}{2\phi(N^* - p)} \sum_{i=1}^{n} \mathbf{z}_i'(\mathbf{J}_i - \mathbf{I}_i)\mathbf{z}_i$$

where

$$N^* = \frac{1}{2}\sum_{i=1}^{n} t_i(t_i - 1)$$

$$\hat{\phi} = \frac{1}{N-p}\sum_{i=1}^{n} \mathbf{z}_i'\mathbf{z}_i$$

with $N = \sum_{i=1}^{n} t_i$ and p is the number of regression parameters to be estimated.

The covariance matrix of $\widehat{\boldsymbol{\beta}}$ is given by the sandwich estimator $\mathbf{I}_0^{-1}\,\mathbf{I}_1\,\mathbf{I}_0^{-1}$ where

$$\mathbf{I}_0 = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}},$$

$$\mathbf{I}_1 = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} \mathrm{Cov}(\mathbf{y}_i) \mathbf{V}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}.$$

In the above $\mathrm{Cov}(\mathbf{y}_i)$ is estimated by $\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'$, which is consistent even if the "working" correlation matrix is misspecified. The GEE estimation method has been implemented in popular commercial statistical packages like SAS and S-Plus.

An important property of the GEE approach is that, even the "working" correlation matrix is misspecified, the estimate of $\boldsymbol{\beta}$ is still consistent, but there could be some loss in efficiency (Hardin and Hilbe, 2003). Since the introduction of the GEEs, numerous authors have addressed this loss in efficiency issue and suggested many alternative approaches to overcome this problem (for example, Prentice and Zhao, 1991; Qu et al., 2000). In particular, Prentice and Zhao (1991) suggested to estimate the correlation parameters simultaneously with the regression parameter, taking into account the covariance of $\mathbf{y}_i$ and $\mathbf{v}_i$. Their estimating equation, known as GEE2, can be written as

$$\sum_{i=1}^{n} \begin{pmatrix} \dfrac{\partial \mathrm{E}(\mathbf{y}_i)}{\partial \boldsymbol{\beta}} & \mathbf{0} \\ \dfrac{\partial \mathrm{E}(\mathbf{v}_i)}{\partial \boldsymbol{\beta}} & \dfrac{\partial \mathrm{E}(\mathbf{v}_i)}{\partial \boldsymbol{\alpha}} \end{pmatrix}' \begin{pmatrix} \mathbf{V}_i & \mathrm{Cov}(\mathbf{y}_i, \mathbf{v}_i) \\ \mathrm{Cov}(\mathbf{y}_i, \mathbf{v}_i) & \mathbf{W}_i \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y}_i - \mathrm{E}(\mathbf{y}_i) \\ \mathbf{v}_i - \mathrm{E}(\mathbf{v}_i) \end{pmatrix} = \mathbf{0}$$

where $\mathbf{v}_i = \mathrm{vech}\{(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'\}$, vech being the usual vech operator, and $\mathbf{W}_i$ is the covariance matrix of $\mathbf{v}_i$. Prentice and Zhao (1991) have shown that use of GEE2 could improve efficiency of the regression parameter.

I.5 Overview of the Thesis

The overall goal of the thesis is to study alternative, robust and efficient methods for estimating the correlation parameter in structured matrices. These methods will also improve the efficiency of the regression parameter. Apart from this introductory chapter, wherein we have briefly reviewed generalized linear models, maximum likelihood method and the generalized estimating equations approach, this thesis consists of five additional chapters.

In Chapter II, we consider the simple case where the responses in the longitudinal data are continuous and the number of repeated measurements per subject are equal (balanced data). We first explore properties of the common correlation structures in longitudinal analysis, and then present details for the maximum likelihood estimates and their asymptotic distributions under the assumption of normality. In a recent paper Wang and Carey (2004) introduced unbiased estimating equations based on Cholesky decomposition of the inverse of the structured correlation matrix. Extending their idea we introduce a number of unbiased estimating equations for estimating the correlation parameter. In order to gain insight for selecting the best among these unbiased equations, we derive the asymptotic variance of the estimates and study the asymptotic relative efficiencies in the special case where $t_i = 3$ for all i. We derive closed form expressions for the exchangeable and the first order autoregressive structures.

In Chapter III, we generalize the results of Chapter II to the unbalanced case, that is, to the case where t_i's are unequal. It is well known that for Gaussian variables the likelihood equation for regression parameter (β) is the optimal estimating equation in the sense of Godambe (1960). As an important result, we show that for Gaussian variables, the likelihood equations for estimating the correlation are also optimal in the sense of Godambe (1960). We then introduce a general class of unbiased estimating equations for estimating the correlation parameter. We derive the asymptotic variances of the estimates for several common structures. These asymptotic variance expressions were used to derive the optimal weights for common structured correlation matrices including exchangeable and AR(1). We simplify the variance expressions for Gaussian variables. For the AR(1) structure, since the optimal weight matrices are complicated we provide simpler approximations which will yield nearly optimal estimates. We present some simulation results to study the efficiency of the estimates obtained as solutions to the various unbiased estimating equations. A real life data analysis is presented to contrast the various estimates.

In Chapter IV, we focus our attention on a correlation structure that has been widely used to model intra-family correlations. This correlation structure, known as the familial structure, has one parameter representing the correlation between the parent and siblings (parent-sibling correlation), and another parameter which measures the intra-correlation between the siblings (sibling-sibling correlation). We first study some properties of the familial correlation structure, and then discuss ML estimation of the unknown parameters, under the assumption of normality. As an alternative to the ML estimation, we construct a general class of weighted unbiased estimating equations, and discuss the selection of optimal weights. It turns out that the optimal weights are the same whether we minimize the determinant or the trace of the asymptotic covariance matrix. However, these optimal weights are complicated. We suggest simple approximations that are straightforward to construct. Simulation results and real life data analysis are also presented in this chapter.

In Chapter V, we generalize the results of the previous chapter to a nuclear family consisting of two parents and siblings. After discussing the general properties of the familial structure and ML estimation of the parameters, we turn our attention to the unbiased estimating equation approach. As before we derive optimal weights which minimize the determinant or the trace of the asymptotic covariance matrix. We also suggest some approximate weights that are nearly optimal.

The results of the previous chapters are mainly applicable for continuous data. However, in medical, biological and social studies the outcomes are binary in several research studies. Unlike Gaussian variables, the ranges of the familial correlations and other measures of associations depend on the marginal means. Knowing and understanding the ranges of these association measures is crucial for developing efficient methods of estimation. In Chapter VI, we study the ranges of familial correlations, odds ratios, kappa statistics, and relative risks for binary variables. We also study stochastic representations of some latent variable models for familial binary data, and explore possible efficient methods of estimation.

Lastly, we give a summary of our methods for analyzing structured correlation matrices with continuous or binary outcomes. Our investigations show that the weighted unbiased estimating equations are a good alternative to maximum likelihood. For common correlation structures, weighted unbiased estimating equations give rise to efficient or nearly efficient estimates, in the sense of the asymptotic variances. These weighted estimating equations are easy to implement, reduce the computational burden, and have less convergence problems when compared with ML estimates.

CHAPTER II

UNBIASED ESTIMATING EQUATIONS

In this chapter, we will discuss simple and popular correlation structures parametrized by a single parameter. We assume the number of observations on each subject are the same, that is, the longitudinal data is balanced. As is customary in generalized linear models, we assume the variance is constant within subject measurements, or in a cluster. Special attention is given to multivariate normal distribution in this chapter.

The organization of this chapter is as follows. We first present some properties of the common correlation structures, namely exchangeable and first order autoregressive structures. In Section II.2 we present details of ML estimation for estimating the regression and the correlation parameters for Gaussian data. The asymptotic covariance matrix of ML estimates are also given explicitly for these two correlation structures. In Section II.3, using the approach based on Cholesky decompositions of the correlation matrices, first suggested by Wang and Carey (2004), we construct several unbiased estimating equations for the correlation parameter. We derive asymptotic properties of the estimates as well. In Section II.4, we consider a general class of estimating equations. This general class contains as special cases the equations in Section II.3 which are based on the Cholesky decompositions. To further understand the unbiased estimating equation approach, we illustrate the method in a special case when $t = 3$, in Section II.5. Explicit expressions for the Cholesky decomposition matrices, estimating equations, and asymptotic variances are given. We compare the estimates of the correlation parameter via asymptotic relative efficiency.

II.1 Common Correlation Structures

The most common correlation structures in longitudinal analysis are exchangeable (EXCH) and first order autoregressive (AR(1)). In this section, we will study properties of these correlation matrices.

II.1.1 Exchangeable Correlation Structure

The exchangeable correlation structure is defined as

$$\mathbf{R} = (1-\alpha)\mathbf{I} + \alpha\mathbf{J} = \begin{bmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \dots & 1 \end{bmatrix}_{t \times t} \tag{2.1.1}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{J}$ is a matrix of ones and α is the common correlation. The eigenvalues of $\mathbf{R}$ are $\lambda_1 = 1 + (t-1)\alpha$ and $\lambda_i = (1-\alpha)$ for $i = 2, 3, \dots, t$. The necessary and sufficient condition for $\mathbf{R}$ to be positive definite is

$$-\frac{1}{t-1} < \alpha < 1 \,. \tag{2.1.2}$$

The determinant of $\mathbf{R}$ is $(1-\alpha)^{t-1}[1+(t-1)\alpha]$, and the inverse is given by

$$\begin{aligned} \mathbf{R}^{-1} &= \frac{1}{1-\alpha}\mathbf{I} - \frac{\alpha}{(1-\alpha)[1+(t-1)\alpha]}\mathbf{J} \\ &= \frac{1}{1-\alpha}\begin{bmatrix} \frac{1+(t-2)\alpha}{1+(t-1)\alpha} & \frac{-\alpha}{1+(t-1)\alpha} & \cdots & \frac{-\alpha}{1+(t-1)\alpha} \\ \frac{-\alpha}{1+(t-1)\alpha} & \frac{1+(t-2)\alpha}{1+(t-1)\alpha} & \cdots & \frac{-\alpha}{1+(t-1)\alpha} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-\alpha}{1+(t-1)\alpha} & \frac{-\alpha}{1+(t-1)\alpha} & \cdots & \frac{1+(t-2)\alpha}{1+(t-1)\alpha} \end{bmatrix} \end{aligned} \tag{2.1.3}$$

In Appendix A.1, the Cholesky decomposition matrices of $\mathbf{R}$ and $\mathbf{R}^{-1}$ are presented.

II.1.2 First Order Autoregressive Correlation Structure

The first order autoregressive correlation structure is defined as

$$\mathbf{R} = \begin{bmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{t-1} \\ \alpha & 1 & \alpha & \dots & \alpha^{t-2} \\ \alpha^2 & \alpha & 1 & \dots & \alpha^{t-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{t-1} & \alpha^{t-2} & \alpha^{t-3} & \dots & 1 \end{bmatrix}_{t \times t} \tag{2.1.4}$$

where α is the correlation between adjacent variables. The determinant of $\mathbf{R}$ is $(1-\alpha^2)^{t-1}$. And the necessary and sufficient conditions for $\mathbf{R}$ to be positive definite is

$$-1 < \alpha < 1\,. \tag{2.1.5}$$

The inverse of $\mathbf{R}$ is given by

$$\begin{aligned}\mathbf{R}^{-1} &= \frac{1}{1-\alpha^2}\left(\mathbf{I}+\alpha^2\mathbf{C}_0-\alpha\mathbf{C}_1\right)\\ &= \frac{1}{1-\alpha^2}\begin{bmatrix} 1 & -\alpha & 0 & \dots & 0 & 0\\ -\alpha & 1+\alpha^2 & -\alpha & \dots & 0 & 0\\ 0 & -\alpha & 1+\alpha^2 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & 1+\alpha^2 & -\alpha\\ 0 & 0 & 0 & \dots & -\alpha & 1\end{bmatrix},\end{aligned} \tag{2.1.6}$$

where $\mathbf{C}_0 = \text{diag}(0,1,1,\dots,1,0)$ and $\mathbf{C}_1$ is a tridiagonal matrix with 0 on the main diagonal and 1 on the upper and lower diagonals. Appendix A.2 contains the Cholesky decomposition matrices of $\mathbf{R}$ and $\mathbf{R}^{-1}$.

II.2 Maximum Likelihood Estimate

For Gaussian model the log-likelihood function can be written as

$$\ell = -\frac{1}{2}\left\{\frac{1}{\sigma^2}\sum_{i=1}^{n}(\mathbf{y}_i-\mathbf{X}_i'\boldsymbol{\beta})'\mathbf{R}^{-1}(\mathbf{y}_i-\mathbf{X}_i'\boldsymbol{\beta}) + n\log|\mathbf{R}| + nt\log(\sigma^2) + nt\log(2\pi)\right\}.$$

Recall that the maximum likelihood estimate of the unknown parameters under multivariate normality assumption can be obtained by solving the system of equations (1.3.3), (1.3.4) and (1.3.6) simultaneously. Further simplifications of those equations give us

$$\begin{aligned}\boldsymbol{\beta} &= \left(\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}_i^{-1}\mathbf{X}_i'\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}_i^{-1}\mathbf{y}_i\,,\\ \sigma^2 &= \frac{1}{nt}\sum_{i=1}^{n}(\mathbf{y}_i-\mathbf{X}_i'\boldsymbol{\beta})'\mathbf{R}^{-1}(\mathbf{y}_i-\mathbf{X}_i'\boldsymbol{\beta})\,,\end{aligned} \tag{2.2.1}$$

and the estimate of α is obtained by solving the equation

$$\sum_{i=1}^{n} \mathrm{tr}\left\{\frac{\partial \mathbf{R}^{-1}}{\partial \alpha}\left((\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})' - \sigma^2\mathbf{R}\right)\right\} \;=\; 0\,. \tag{2.2.2}$$

The following identities are useful to find the asymptotic variances of the ML estimates via Fisher Information.

$$\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}^{-1}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) \\
\frac{\partial \ell}{\partial \sigma^2} &= -\frac{nt}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\mathbf{R}^{-1}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) \\
\frac{\partial \ell}{\partial \alpha} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\frac{\partial \mathbf{R}^{-1}}{\partial \alpha}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) - n\frac{\partial \log|\mathbf{R}|}{2\,\partial \alpha} \\
\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}^2} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}^{-1}\mathbf{X}_i' \\
\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}^{-1}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) \\
\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \alpha} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\frac{\partial \mathbf{R}^{-1}}{\partial \alpha}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) \\
\frac{\partial^2 \ell}{\partial (\sigma^2)^2} &= \frac{nt}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\mathbf{R}^{-1}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) \\
\frac{\partial^2 \ell}{\partial \sigma^2\partial \alpha} &= \frac{1}{2\sigma^4}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\frac{\partial \mathbf{R}^{-1}}{\partial \alpha}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) \\
\frac{\partial^2 \ell}{\partial \alpha^2} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\frac{\partial^2 \mathbf{R}^{-1}}{\partial \alpha^2}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) - \frac{n}{2}\frac{\partial^2 \log|\mathbf{R}|}{\partial \alpha^2}\,.
\end{aligned}$$

Since $\mathrm{E}(\mathbf{y}_i) = \mathbf{X}_i'\boldsymbol{\beta}$ and $\mathrm{E}\left\{(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\right\} = \mathrm{Cov}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) = \sigma^2\mathbf{R}$, we have

$$\begin{aligned}
\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}^2}\right\} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}^{-1}\mathbf{X}_i' \\
\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \sigma^2}\right\} &= \mathbf{0} \qquad\qquad \mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \alpha}\right\} = \mathbf{0} \\
\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial (\sigma^2)^2}\right\} &= -\frac{nt}{2\sigma^4} \\
\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \sigma^2\partial \alpha}\right\} &= \frac{n}{2\sigma^2}\,\mathrm{tr}\left(\frac{\partial \mathbf{R}^{-1}}{\partial \alpha}\mathbf{R}\right) \\
\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \alpha^2}\right\} &= -\frac{n}{2}\left\{\mathrm{tr}\left(\frac{\partial^2 \mathbf{R}^{-1}}{\partial \alpha^2}\mathbf{R}\right) + \frac{\partial^2 \log|\mathbf{R}|}{\partial \alpha^2}\right\}.
\end{aligned}$$

Thus the Fisher information matrix is given by

$$\begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}_i^{-1}\mathbf{X}_i' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{nt}{2\sigma^4} & -\frac{n}{2\sigma^2}\operatorname{tr}\left(\frac{\partial\mathbf{R}^{-1}}{\partial\alpha}\mathbf{R}\right) \\ \mathbf{0} & -\frac{n}{2\sigma^2}\operatorname{tr}\left(\frac{\partial\mathbf{R}^{-1}}{\partial\alpha}\mathbf{R}\right) & \frac{n}{2}\left\{\operatorname{tr}\left(\frac{\partial^2\mathbf{R}^{-1}}{\partial\alpha^2}\mathbf{R}\right)+\frac{\partial^2\log|\mathbf{R}|}{\partial\alpha^2}\right\} \end{bmatrix} \quad (2.2.3)$$

When $\mathbf{R}$ has an exchangeable structure, Lee (1988) has shown that maximum likelihood estimates of α and σ^2 are

$$\begin{aligned} \hat{\alpha}_L &= \frac{\sum_{i=1}^{n}\operatorname{tr}\left((\mathbf{J}-\mathbf{I})\varepsilon_i\varepsilon_i'\right)}{(t-1)\sum_{i=1}^{n}\operatorname{tr}(\varepsilon_i\varepsilon_i')} \\ \hat{\sigma}_L^2 &= \frac{1}{nt}\sum_{i=1}^{n}\operatorname{tr}(\varepsilon_i\varepsilon_i'). \end{aligned}$$

For the exchangeable structure we also have

$$\begin{aligned} \frac{\partial\mathbf{R}^{-1}}{\partial\alpha}\mathbf{R} &= \frac{1}{1-\alpha}\left[\mathbf{I}-\frac{1}{1+(t-1)\alpha}\mathbf{J}\right] \\ \frac{\partial^2\mathbf{R}^{-1}}{\partial\alpha^2}\mathbf{R} &= \frac{2}{(1-\alpha)^2}\left[\mathbf{I}+\frac{t-2[1+(t-1)\alpha]}{[1+(t-1)\alpha]^2}\mathbf{J}\right] \end{aligned}$$

therefore,

$$\begin{aligned} \operatorname{tr}\left(\frac{\partial\mathbf{R}^{-1}}{\partial\alpha}\mathbf{R}\right) &= \frac{t(t-1)\alpha}{(1-\alpha)[1+(t-1)\alpha]} \\ \operatorname{tr}\left(\frac{\partial^2\mathbf{R}^{-1}}{\partial\alpha^2}\mathbf{R}\right) &= \frac{2t(t-1)[1+(t-1)\alpha^2]}{(1-\alpha)^2[1+(t-1)\alpha]^2} \\ \frac{\partial^2\log|\mathbf{R}|}{\partial\alpha^2} &= \frac{-t(t-1)[1+(t-1)\alpha^2]}{(1-\alpha)^2[1+(t-1)\alpha]^2}. \end{aligned}$$

Thus the lower right partition of the Fisher information is given by

$$\begin{bmatrix} \frac{nt}{2\sigma^4} & \frac{-nt(t-1)\alpha}{2\sigma^2(1-\alpha)[1+(t-1)\alpha]} \\ \frac{-nt(t-1)\alpha}{2\sigma^2(1-\alpha)[1+(t-1)\alpha]} & \frac{nt(t-1)[1+(t-1)\alpha^2]}{2(1-\alpha)^2[1+(t-1)\alpha]^2} \end{bmatrix},$$

and therefore the asymptotic variances of $\hat{\sigma}_L^2$ and $\hat{\alpha}_L$ are

$$\frac{2\sigma^4}{t}\left[1+(t-1)\alpha^2\right] \quad\text{and}\quad \frac{2}{t(t-1)}\left[(1-\alpha)[1+(t-1)\alpha]\right]^2$$

respectively, as in Chaganty (2003).

When $\mathbf{R}$ has an AR(1) structure, the maximum likelihood estimate of α is not in a closed form. In this case we have

$$\begin{aligned}\frac{\partial \mathbf{R}^{-1}}{\partial \alpha}\mathbf{R} &= \frac{1}{(1-\alpha^2)^2}\Big[2\alpha(\mathbf{I}+\mathbf{C}_0) - (1+\alpha^2)\mathbf{C}_1\Big]\mathbf{R} \\ \frac{\partial^2 \mathbf{R}^{-1}}{\partial \alpha^2}\mathbf{R} &= \frac{2}{(1-\alpha^2)^3}\Big[(1+3\alpha^2)(\mathbf{I}+\mathbf{C}_0) - \alpha(3+\alpha^2)\mathbf{C}_1\Big]\mathbf{R}\end{aligned}$$

and therefore,

$$\begin{aligned}\operatorname{tr}\left(\frac{\partial \mathbf{R}^{-1}}{\partial \alpha}\mathbf{R}\right) &= \frac{2(t-1)\alpha}{1-\alpha^2} \\ \operatorname{tr}\left(\frac{\partial^2 \mathbf{R}^{-1}}{\partial \alpha^2}\mathbf{R}\right) &= \frac{4(t-1)(1+\alpha^2)}{(1-\alpha^2)^2} \\ \frac{\partial^2 \log|\mathbf{R}|}{\partial \alpha^2} &= \frac{-2(t-1)(1+\alpha^2)}{(1-\alpha^2)^2}\end{aligned}$$

The lower right partition of the Fisher information is given by

$$\begin{bmatrix} \dfrac{nt}{2\sigma^4} & \dfrac{-n(t-1)\alpha}{\sigma^2(1-\alpha^2)} \\ \dfrac{-n(t-1)\alpha}{\sigma^2(1-\alpha^2)} & \dfrac{n(t-1)(1+\alpha^2)}{(1-\alpha^2)^2} \end{bmatrix},$$

and thus the asymptotic variances of $\hat{\sigma}_L^2$ and $\hat{\alpha}_L$ are

$$\frac{2\sigma^4(1+\alpha^2)}{2\alpha^2+t(1-\alpha^2)} \quad \text{and} \quad \frac{t(1-\alpha^2)^2}{(t-1)\left[2\alpha^2+t(1-\alpha^2)\right]}$$

respectively, as in Chaganty (2003).

II.3 Unbiased Estimating Equations Based on Cholesky Decompositions

Unbiased estimating equations provide an alternative method of estimating correlation parameters. The unbiasedness guarantees that the estimates are consistent. Recently, Wang and Carey (2004) introduced an unbiased estimating equation that is based on Cholesky decomposition of the inverse of the correlation matrix. Suppose that $\mathbf{R}^{-1} = \mathbf{B}_l\mathbf{\Lambda}_l\mathbf{B}_l' = \mathbf{B}_u\mathbf{\Lambda}_u\mathbf{B}_u'$, where $\mathbf{B}_l$ is a lower triangular matrix with unit diagonal elements, $\mathbf{B}_u$ is an upper triangular matrix with unit diagonal elements, and $\mathbf{\Lambda}_l$ and $\mathbf{\Lambda}_u$ are corresponding diagonal matrices. See the Appendix for details on Cholesky decompositions of different correlation structures.

Let $\bar{\mathbf{Z}} = \frac{1}{n\sigma^2}\sum_{i=1}^{n} \boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'$, where $\mathbf{z}_i$ is the Pearson residual. Then Wang and Carey (2004)'s estimating equation can be written as

$$U_1: \quad \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\bar{\mathbf{Z}}\right\} = 0. \tag{2.3.1}$$

The equation U_1 is unbiased since

$$\begin{aligned}
\mathrm{E}(U_1) &= \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\mathbf{R}\right\} \\
&= \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{B}_l^{-1}\mathbf{B}_l\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{B}_u^{-1}\mathbf{B}_u\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\mathbf{R}\right\} \\
&= \mathrm{tr}\left\{\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{B}_l^{-1} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{B}_u^{-1}\right\} = 0.
\end{aligned}$$

The last equality holds because $\partial\mathbf{B}_l/\partial\alpha$ (or $\partial\mathbf{B}_u/\partial\alpha$) is a lower (or upper) triangular matrix with diagonal elements 0 and also $\mathbf{B}_l^{-1}$ (or $\mathbf{B}_u^{-1}$) is a lower (or upper) triangular matrix with 0 on the diagonal.

Using the idea of Wang and Carey (2004), we can construct many other estimating equations based on the Cholesky decomposition matrices similar to U_1. Some examples are

$$U_2: \quad \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{R}^{-1}\right)\bar{\mathbf{Z}}\right\} = 0 \tag{2.3.2}$$

$$U_3: \quad \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{B}_u'\right)\bar{\mathbf{Z}}\right\} = 0 \tag{2.3.3}$$

$$U_4: \quad \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial \alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial \alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\bar{\mathbf{Z}}\right\} = 0 \tag{2.3.4}$$

$$U_5: \quad \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial \alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial \alpha}\mathbf{R}^{-1}\right)\bar{\mathbf{Z}}\right\} = 0 \tag{2.3.5}$$

$$U_6: \quad \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial \alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial \alpha}\mathbf{B}_u'\right)\bar{\mathbf{Z}}\right\} = 0. \tag{2.3.6}$$

It is easy to verify that the estimating equations U_i, for $i = 2, \ldots, 6$ are unbiased. Therefore the solutions $\hat{\alpha}_i$, $i = 1, \ldots, 6$ to these equations are consistent estimates of α. It follows from the general theory of unbiased estimating equations, these estimates are asymptotically normal. The following lemma facilitates the computation of asymptotic variances under the assumption of normality.

Lemma 2.1 *Let* $\mathbf{z}_{t\times 1}$ *be a multivariate normal vector with mean* $\mathbf{0}$ *and covariance matrix* $\mathbf{\Sigma}$. *Let* $\mathbf{A}$, $\mathbf{B}$ *be two square matrices of order* t *and* $\mathbf{Z} = \mathbf{z}\mathbf{z}'$. *Then*

$$\mathrm{Cov}\left\{\mathrm{tr}(\mathbf{AZ}), \mathrm{tr}(\mathbf{BZ})\right\} = \mathrm{tr}(\mathbf{B\Sigma A\Sigma}) + \mathrm{tr}(\mathbf{B}'\mathbf{\Sigma A\Sigma}).$$

Proof. Note that

$$(1)\ \mathrm{Cov}\left\{\mathrm{vec}(\mathbf{Z})\right\} = (\mathbf{I}_{t^2} + \mathbf{I}_{(t,t)})(\mathbf{\Sigma}\otimes\mathbf{\Sigma})$$

$$(2)\ \mathrm{vec}'(\mathbf{A})(\mathbf{D}\otimes\mathbf{B})\mathrm{vec}(\mathbf{C}) = \mathrm{tr}(\mathbf{A}'\mathbf{BCD}') \text{ (Harville, 1997, page 342)}$$

$$(3)\ \mathrm{vec}'(\mathbf{B}')\mathbf{I}_{(t,t)} = \mathrm{vec}'(\mathbf{B})$$

where $\mathbf{I}_{t^2}$ is the $t^2\times t^2$ identity matrix and $\mathbf{I}_{(t,t)}$ is the $t^2\times t^2$ permuted identity matrix given by

$$\mathbf{I}_{(t,t)} = \begin{bmatrix} \mathbf{E}'_{11} & \mathbf{E}'_{12} & \cdots & \mathbf{E}'_{1t} \\ \mathbf{E}'_{21} & \mathbf{E}'_{22} & \cdots & \mathbf{E}'_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}'_{t1} & \mathbf{E}'_{t2} & \cdots & \mathbf{E}'_{tt} \end{bmatrix}$$

where $\mathbf{E}_{rs}$ is a $t\times t$ matrix of zeros except for the (r,s)th element being 1 for $r,s = 1,2,\ldots,t$ (Vonesh and Chinchilli, 1997, page 22). Now

$$\begin{aligned}
\mathrm{Cov}\left\{\mathrm{tr}(\mathbf{AZ}), \mathrm{tr}(\mathbf{BZ})\right\} &= \mathrm{E}\left\{\mathrm{tr}(\mathbf{AZ})\mathrm{tr}(\mathbf{BZ})\right\} - \mathrm{E}\left\{\mathrm{tr}(\mathbf{AZ})\right\}\mathrm{E}\left\{\mathrm{tr}(\mathbf{BZ})\right\} \\
&= \mathrm{E}\left\{\mathrm{vec}'(\mathbf{Z}')\mathrm{vec}(\mathbf{A})\mathrm{vec}'(\mathbf{B}')\mathrm{vec}(\mathbf{Z})\right\} - \mathrm{tr}(\mathbf{A\Sigma})\mathrm{tr}(\mathbf{B\Sigma}) \\
&= \mathrm{tr}\left\{\mathrm{vec}(\mathbf{A})\mathrm{vec}'(\mathbf{B}')\mathrm{Cov}\{\mathrm{vec}(\mathbf{Z})\}\right\} \\
&\qquad + \mathrm{vec}'(\mathbf{\Sigma})\mathrm{vec}(\mathbf{A})\mathrm{vec}'(\mathbf{B}')\mathrm{vec}(\mathbf{\Sigma}) - \mathrm{tr}(\mathbf{A\Sigma})\mathrm{tr}(\mathbf{B\Sigma}) \\
&= \mathrm{tr}\left\{\mathrm{vec}'(\mathbf{B}')\mathrm{Cov}\{\mathrm{vec}(\mathbf{Z})\}\mathrm{vec}(\mathbf{A})\right\} \\
&= \mathrm{vec}'(\mathbf{B}')\left(\mathbf{I}_{t^2} + \mathbf{I}_{(t,t)}\right)(\mathbf{\Sigma}\otimes\mathbf{\Sigma})\mathrm{vec}(\mathbf{A}) \\
&= \mathrm{vec}'(\mathbf{B}')(\mathbf{\Sigma}\otimes\mathbf{\Sigma})\mathrm{vec}(\mathbf{A}) + \mathrm{vec}'(\mathbf{B})(\mathbf{\Sigma}\otimes\mathbf{\Sigma})\mathrm{vec}(\mathbf{A}) \\
&= \mathrm{tr}(\mathbf{B\Sigma A\Sigma}') + \mathrm{tr}(\mathbf{B}'\mathbf{\Sigma A\Sigma}') \\
&= \mathrm{tr}(\mathbf{B\Sigma A\Sigma}) + \mathrm{tr}(\mathbf{B}'\mathbf{\Sigma A\Sigma})
\end{aligned}$$

and this complete the proof of the lemma. ◇

The next theorem, due to Chaganty and Shi (2004), is useful to derive the asymptotic variances of the parameter estimates obtained from unbiased estimating equations.

Theorem 2.1 *Let* $\mathbf{z}_i$ *be independent random vectors of dimension* t_i, $i = 1, 2, \ldots, n$. *Assume that* $t_i \leq t$ *for all* i. *Let* $\boldsymbol{\alpha}$ *be a parameter of fixed dimension, and the multivariate functions* $h(\mathbf{z}_i, \boldsymbol{\alpha})$ *be such that*

$$\sum_{i=1}^{n} \mathrm{E}\left\{h_i(\mathbf{z}_i, \boldsymbol{\alpha})\right\} \;=\; \mathbf{0}$$

Define $\mathbf{M}_n(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}\left\{h_i(\mathbf{z}_i, \boldsymbol{\alpha})\right\}$ *and* $\mathbf{I}_n(\boldsymbol{\alpha}) = -\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left\{\partial h(\mathbf{z}_i, \boldsymbol{\alpha})/\partial \boldsymbol{\alpha}'\right\}$. *Suppose* $\hat{\boldsymbol{\alpha}}$ *is the solution of the unbiased estimating equation*

$$\frac{1}{n}\sum_{i=1}^{n} h_i(\mathbf{z}_i, \boldsymbol{\alpha}) \;=\; \mathbf{0}\,.$$

Then under usual regularity conditions we have

$$\sqrt{n}\,\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\right) \;\sim\; \mathrm{AMVN}\left(\mathbf{0}\,, \mathbf{I}_n^{-1}\mathbf{M}_n \mathbf{I}_n'^{\,-1}\right). \tag{2.3.7}$$

Here AMVN *stands for Asymptotically Multi-Variate Normal.*

Using Lemma 2.1 and Theorem 2.1, we can get the asymptotic variances of $\hat{\alpha}_i$, $1 \leq i \leq 6$. These are given by

$$\sigma^2_{\hat{\alpha}_1} \;=\; \frac{\begin{array}{l}\mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\mathbf{R}\right\} \\ \quad + \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)'\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\mathbf{R}\right\}\end{array}}{\left[\mathrm{tr}\left\{\frac{\partial}{\partial \alpha}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\mathbf{R}\right\}\right]^2}$$

$$\begin{aligned}\sigma^2_{\hat{\alpha}_2} \;&=\; \frac{\begin{array}{l}\mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{R}^{-1}\right)\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{R}^{-1}\right)\mathbf{R}\right\} \\ \quad + \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{R}^{-1}\right)'\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{R}^{-1}\right)\mathbf{R}\right\}\end{array}}{\left[\mathrm{tr}\left\{\frac{\partial}{\partial \alpha}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{R}^{-1}\right)\mathbf{R}\right\}\right]^2} \\ &=\; \frac{\begin{array}{l}\mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\right)\left(\frac{\partial \mathbf{B}_l}{\partial \alpha} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\right)\right\} \\ \quad + \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{R}^{-1}\right)'\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\right)\right\}\end{array}}{\left[\mathrm{tr}\left\{\frac{\partial}{\partial \alpha}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{R}^{-1}\right)\mathbf{R}\right\}\right]^2}\end{aligned}$$

$$\sigma^2_{\hat{\alpha}_3} \;=\; \frac{\begin{array}{l}\mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{B}_u'\right)\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{B}_u'\right)\mathbf{R}\right\} \\ \quad + \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{B}_u'\right)'\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{B}_u'\right)\mathbf{R}\right\}\end{array}}{\left[\mathrm{tr}\left\{\frac{\partial}{\partial \alpha}\left(\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{B}_u'\right)\mathbf{R}\right\}\right]^2}$$

$$\sigma^2_{\hat{\alpha}_4} = \frac{\begin{array}{l} \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\mathbf{R}\right\} \\ \quad + \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)'\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\mathbf{R}\right\} \end{array}}{\left[\mathrm{tr}\left\{\frac{\partial}{\partial\alpha}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\boldsymbol{\Lambda}_l\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\boldsymbol{\Lambda}_u\mathbf{B}_u'\right)\mathbf{R}\right\}\right]^2}$$

$$\begin{aligned} \sigma^2_{\hat{\alpha}_5} &= \frac{\begin{array}{l} \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{R}^{-1}\right)\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{R}^{-1}\right)\mathbf{R}\right\} \\ \quad + \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{R}^{-1}\right)'\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{R}^{-1}\right)\mathbf{R}\right\} \end{array}}{\left[\mathrm{tr}\left\{\frac{\partial}{\partial\alpha}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{R}^{-1}\right)\mathbf{R}\right\}\right]^2} \\ &= \frac{\begin{array}{l} \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha} + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\right)\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha} + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\right)\right\} \\ \quad + \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{R}^{-1}\right)'\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha} + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\right)\right\} \end{array}}{\left[\mathrm{tr}\left\{\frac{\partial}{\partial\alpha}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{R}^{-1}\right)\mathbf{R}\right\}\right]^2} \end{aligned}$$

$$\sigma^2_{\hat{\alpha}_6} = \frac{\begin{array}{l} \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{B}_u'\right)\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{B}_u'\right)\mathbf{R}\right\} \\ \quad + \mathrm{tr}\left\{\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{B}_u'\right)'\mathbf{R}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\mathbf{B}_u'\right)\mathbf{R}\right\} \end{array}}{\left[\mathrm{tr}\left\{\frac{\partial}{\partial\alpha}\left(\frac{\partial \mathbf{B}_l}{\partial\alpha}\frac{\partial \boldsymbol{\Lambda}_l}{\partial\alpha}\mathbf{B}_l' + \frac{\partial \mathbf{B}_u}{\partial\alpha}\mathbf{B}_u'\frac{\partial \boldsymbol{\Lambda}_u}{\partial\alpha}\right)\mathbf{R}\right\}\right]^2} \qquad (2.3.8)$$

II.4 Classes of Unbiased Estimating Equations

A general form for the six unbiased estimating equations that we discussed in the previous section is given by

$$\mathrm{tr}\left\{\mathbf{W}\mathbf{R}^{-1}\left(\bar{\mathbf{Z}} - \mathbf{R}\right)\right\} = 0, \qquad (2.4.1)$$

where $\mathbf{W}$ is some weighting matrix. Equation (2.4.1) is unbiased since

$$\mathrm{E}\left[\mathrm{tr}\left\{\mathbf{W}\mathbf{R}^{-1}\left(\bar{\mathbf{Z}} - \mathbf{R}\right)\right\}\right] = \mathrm{tr}\left\{\mathbf{W}\mathbf{R}^{-1}\left(\mathrm{E}(\bar{\mathbf{Z}}) - \mathbf{R}\right)\right\} = 0.$$

Using Lemma 2.1 and Theorem 2.1, we can see that the asymptotic covariance of the estimate of the correlation parameter obtained from solving (2.4.1) is given by

$$\frac{\mathrm{Cov}\left\{\mathrm{tr}\left(\mathbf{W}\mathbf{R}^{-1}\left(\bar{\mathbf{Z}} - \mathbf{R}\right)\right)\right\}}{\left[\mathrm{E}\left\{\frac{\partial}{\partial\alpha}\mathrm{tr}\left(\mathbf{W}\mathbf{R}^{-1}\left(\bar{\mathbf{Z}} - \mathbf{R}\right)\right)\right\}\right]^2} = \frac{\mathrm{Cov}\left\{\mathrm{tr}\left(\mathbf{W}\mathbf{R}^{-1}\bar{\mathbf{Z}}\right) - \mathrm{tr}(\mathbf{W})\right\}}{\left[\mathrm{E}\left\{\frac{\partial}{\partial\alpha}\left(\mathrm{tr}\left(\mathbf{W}\mathbf{R}^{-1}\bar{\mathbf{Z}}\right) + \mathrm{tr}(\mathbf{W})\right)\right\}\right]^2}$$

$$= \frac{\mathrm{Cov}\left\{\mathrm{tr}\left(\mathbf{W}\mathbf{R}^{-1}\bar{\mathbf{Z}}\right)\right\}}{\left[\mathrm{E}\left\{\frac{\partial}{\partial\alpha}\mathrm{tr}\left(\mathbf{W}\mathbf{R}^{-1}\bar{\mathbf{Z}}\right)\right\}\right]^2}$$

$$= \frac{\mathrm{tr}\left(\mathbf{W}\mathbf{R}^{-1}\mathbf{W}'\mathbf{R}+\mathbf{W}^2\right)}{\left[\mathrm{tr}\left(\mathbf{W}\frac{\partial\mathbf{R}^{-1}}{\partial\alpha}\mathbf{R}\right)\right]^2}. \tag{2.4.2}$$

An interesting subclass of (2.4.1) is obtained if we place the constraint $\mathrm{tr}(\mathbf{W}) = 0$, that is, the subclass consists of equations of the form

$$\mathrm{tr}\left\{\mathbf{W}\mathbf{R}^{-1}\bar{\mathbf{Z}}\right\} = 0 \quad \text{subject to} \quad \mathrm{tr}(\mathbf{W}) = 0. \tag{2.4.3}$$

This subclass of estimating equations is useful when the parameter space is constrained. Unlike the general class (2.4.1) of estimating equations, which may pose computational problems, the subclass can yield relatively simple estimates.

The asymptotic variance expressions for the subclass (2.4.3) are the same as that of the general class (2.4.2) since in the derivation of the asymptotic variance, the extra term $\mathrm{tr}(\mathbf{W})$ can be taken as a constant. It drops out when we take the first order partial derivative of the estimating equation.

A closer look at the six unbiased estimating equations ($U_i, 1 \le i \le 6$) that we discussed in the previous chapter reveals that those equations fall in the subclass. The weighting matrices $\mathbf{W}$'s for the different estimating equations are as follows:

$$\begin{aligned}
U_1 :\ \mathbf{W} &= \frac{\partial\mathbf{B}_l}{\partial\alpha}\mathbf{B}_l^{-1} + \frac{\partial\mathbf{B}_u}{\partial\alpha}\mathbf{B}_u^{-1}\\
U_2 :\ \mathbf{W} &= \frac{\partial\mathbf{B}_l}{\partial\alpha} + \frac{\partial\mathbf{B}_u}{\partial\alpha}\\
U_3 :\ \mathbf{W} &= \frac{\partial\mathbf{B}_l}{\partial\alpha}\mathbf{\Lambda}_l^{-1}\mathbf{B}_l^{-1} + \frac{\partial\mathbf{B}_u}{\partial\alpha}\mathbf{\Lambda}_u^{-1}\mathbf{B}_l^{-1}\\
U_4 :\ \mathbf{W} &= \frac{\partial\mathbf{B}_l}{\partial\alpha}\frac{\partial\mathbf{\Lambda}_l}{\partial\alpha}\mathbf{B}_l^{-1} + \frac{\partial\mathbf{B}_u}{\partial\alpha}\frac{\partial\mathbf{\Lambda}_u}{\partial\alpha}\mathbf{B}_u^{-1}\\
U_5 :\ \mathbf{W} &= \frac{\partial\mathbf{B}_l}{\partial\alpha}\frac{\partial\mathbf{\Lambda}_l}{\partial\alpha} + \frac{\partial\mathbf{B}_u}{\partial\alpha}\frac{\partial\mathbf{\Lambda}_u}{\partial\alpha}\\
U_6 :\ \mathbf{W} &= \frac{\partial\mathbf{B}_l}{\partial\alpha}\frac{\partial\mathbf{\Lambda}_l}{\partial\alpha}\mathbf{\Lambda}_l^{-1}\mathbf{B}_l^{-1} + \frac{\partial\mathbf{B}_u}{\partial\alpha}\frac{\partial\mathbf{\Lambda}_u}{\partial\alpha}\mathbf{\Lambda}_u^{-1}\mathbf{B}_u^{-1}
\end{aligned}$$

II.5 Special Cases

In this section we will study asymptotic properties of the six unbiased estimating equations in the special case where $t_i = 3$ for all i. Suppose first $\mathbf{R}$ is exchangeable, that is, $\mathbf{R} = (1-\alpha)\mathbf{I} + \alpha\mathbf{J}$. The $\mathbf{B}$ and $\mathbf{\Lambda}$ matrices involved in the Cholesky decomposition of $\mathbf{R}^{-1}$ are

$$\mathbf{B}_l = \begin{bmatrix} 1 & 0 & 0 \\ \frac{-\alpha}{1+\alpha} & 1 & 0 \\ \frac{-\alpha}{1+\alpha} & -\alpha & 1 \end{bmatrix} \qquad \mathbf{\Lambda}_l = \begin{bmatrix} \frac{1+\alpha}{(1-\alpha)(1+2\alpha)} & 0 & 0 \\ 0 & \frac{1}{1-\alpha^2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{B}_u = \begin{bmatrix} 1 & -\alpha & \frac{-\alpha}{1+\alpha} \\ 0 & 1 & \frac{-\alpha}{1+\alpha} \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{\Lambda}_u = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{1-\alpha^2} & 0 \\ 0 & 0 & \frac{1+\alpha}{(1-\alpha)(1+2\alpha)} \end{bmatrix}.$$

Thus we have

$$\frac{\partial \mathbf{B}_l}{\partial\alpha}\mathbf{\Lambda}_l\mathbf{B}_l' = c_1 \begin{bmatrix} 0 & 0 & 0 \\ -(1+\alpha) & \alpha & \alpha \\ -(1+\alpha) & -(1+2\alpha+2\alpha^2) & \alpha(2+3\alpha+2\alpha^2) \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial\alpha}\mathbf{\Lambda}_u\mathbf{B}_u' = c_1 \begin{bmatrix} \alpha(2+3\alpha+2\alpha^2) & -(1+2\alpha+2\alpha^2) & -(1+\alpha) \\ \alpha & \alpha & -(1+\alpha) \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_l}{\partial\alpha}\mathbf{R}^{-1} = c_1 \begin{bmatrix} 0 & 0 & 0 \\ -(1+\alpha) & \alpha & \alpha \\ -1+2\alpha^2+\alpha^3 & -(1+2\alpha+3\alpha^2+\alpha^3) & \alpha(2+2\alpha+\alpha^2) \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial\alpha}\mathbf{R}^{-1} = c_1 \begin{bmatrix} \alpha(2+2\alpha+\alpha^2) & -(1+2\alpha+3\alpha^2+\alpha^3) & -1+2\alpha^2+\alpha^3 \\ \alpha & \alpha & -(1+\alpha) \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_l}{\partial\alpha}\mathbf{B}_l' = c_2 \begin{bmatrix} 0 & 0 & 0 \\ -(1+\alpha) & \alpha & \alpha \\ -(1+\alpha) & -(1+2\alpha+3\alpha^2+\alpha^3) & \alpha(2+3\alpha+3\alpha^2+\alpha^3) \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial\alpha}\mathbf{B}_u' = c_2 \begin{bmatrix} \alpha(2+3\alpha+3\alpha^2+\alpha^3) & -(1+2\alpha+3\alpha^2+\alpha^3) & -(1+\alpha) \\ \alpha & \alpha & -(1+\alpha) \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$\frac{\partial \mathbf{B}_l}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_l}{\partial \alpha}\mathbf{\Lambda}_l\mathbf{B}_l' = c_3 \begin{bmatrix} 0 & 0 & 0 \\ -(1+\alpha)^2(2+\alpha) & \alpha(1+\alpha)(2+\alpha) & \alpha(1+\alpha)(2+\alpha) \\ -(1+\alpha)^2(2+\alpha) & -(1+4\alpha+9\alpha^2+7\alpha^3) & \alpha(3+9\alpha+13\alpha^2+8\alpha^3) \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_u}{\partial \alpha}\mathbf{\Lambda}_u\mathbf{B}_u' = c_3 \begin{bmatrix} \alpha(3+9\alpha+13\alpha^2+8\alpha^3) & -(1+4\alpha+9\alpha^2+7\alpha^3) & -(1+\alpha)^2(2+\alpha) \\ \alpha(1+\alpha)(2+\alpha) & \alpha(1+\alpha)(2+\alpha) & -(1+\alpha)^2(2+\alpha) \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_l}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_l}{\partial \alpha}\mathbf{R}^{-1} = c_4 \begin{bmatrix} 0 & 0 & 0 \\ -(1+\alpha)(2+\alpha) & \alpha(2+\alpha) & \alpha(2+\alpha) \\ -(2+2\alpha-3\alpha^2-4\alpha^3) & -(1+3\alpha+7\alpha^2+4\alpha^3) & \alpha(3+5\alpha+4\alpha^2) \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_u}{\partial \alpha}\mathbf{R}^{-1} = c_4 \begin{bmatrix} \alpha(3+5\alpha+4\alpha^2) & -(1+3\alpha+7\alpha^2+4\alpha^3) & -(2+2\alpha-3\alpha^2-4\alpha^3) \\ \alpha(2+\alpha) & \alpha(2+\alpha) & -(1+\alpha)(2+\alpha) \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_l}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_l}{\partial \alpha}\mathbf{B}_l' = c_5 \begin{bmatrix} 0 & 0 & 0 \\ -(1+\alpha)(2+\alpha) & \alpha(2+\alpha) & \alpha(2+\alpha) \\ -(1+\alpha)(2+\alpha) & -(1+3\alpha+7\alpha^2+4\alpha^3) & \alpha(3+6\alpha+8\alpha^2+4\alpha^3) \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_u}{\partial \alpha}\mathbf{B}_u' = c_5 \begin{bmatrix} \alpha(3+6\alpha+8\alpha^2+4\alpha^3) & -(1+3\alpha+7\alpha^2+4\alpha^3) & -(1+\alpha)(2+\alpha) \\ \alpha(2+\alpha) & \alpha(2+\alpha) & -(1+\alpha)(2+\alpha) \\ 0 & 0 & 0 \end{bmatrix}$$

where the constants c_i's are

$$\begin{aligned} c_1 &= \frac{1}{(1-\alpha)(1+\alpha)^2(1+2\alpha)} \\ c_2 &= \frac{1}{(1+\alpha)^3} \\ c_3 &= \frac{2\alpha}{(1-\alpha)^3(1+\alpha)^3(1+2\alpha)^3} \\ c_4 &= \frac{2\alpha}{(1-\alpha)^3(1+\alpha)^2(1+2\alpha)^3} \\ c_5 &= \frac{2\alpha}{(1-\alpha)^2(1+\alpha)^3(1+2\alpha)^2}. \end{aligned}$$

Let z_{rs} be the (r, s) element of $\bar{\mathbf{Z}}$. Ignoring the constants c_i, $1 \leq i \leq 5$, we can see that the six unbiased estimating equations reduce to six polynomial equations. They are given by

$$
\begin{aligned}
U_1: \quad & (2z_{11}+2z_{33})\alpha^3+(3z_{11}-2z_{12}-2z_{23}+3z_{33})\alpha^2 \\
& +(2z_{11}-2z_{12}-2z_{13}+2z_{22}-2z_{23}+2z_{33})\alpha-(2z_{12}+2z_{13}+2z_{23}) \;=\; 0 \\
U_2: \quad & (z_{11}-z_{12}+2z_{13}-z_{23}+z_{33})\alpha^3+(2z_{11}-3z_{12}+4z_{13}-3z_{23}+2z_{33})\alpha^2 \\
& +(2z_{11}-2z_{12}+2z_{22}-2z_{23}+2z_{33})\alpha-(2z_{12}+2z_{13}+2z_{23}) \;=\; 0 \\
U_3: \quad & (z_{11}+z_{33})\alpha^4+(3z_{11}-z_{12}-z_{23}+3z_{33})\alpha^3+(3z_{11}-3z_{12}-3z_{23}+3z_{33})\alpha^2 \\
& +(2z_{11}-2z_{12}-2z_{13}+2z_{22}-2z_{23}+2z_{33})\alpha-(2z_{12}+2z_{13}+2z_{23}) \;=\; 0 \\
U_4: \quad & (8z_{11}+8z_{33})\alpha^4+(13z_{11}-7z_{12}-2z_{13}+2z_{22}-7z_{23}+13z_{33})\alpha^3 \\
& +(9z_{11}-10z_{12}-8z_{13}+6z_{22}-10z_{23}+9z_{33})\alpha^2 \\
& +(3z_{11}-7z_{12}-10z_{13}+4z_{22}-7z_{23}+3z_{33})\alpha-(3z_{12}+4z_{13}+3z_{23}) \;=\; 0 \\
U_5: \quad & (4z_{11}-4z_{12}+8z_{13}-4z_{23}+4z_{33})\alpha^3+(5z_{11}-7z_{12}+6z_{13}+2z_{22}-7z_{23}+5z_{33})\alpha^2 \\
& +(3z_{11}-4z_{12}-4z_{13}+4z_{22}-4z_{23}+3z_{33})\alpha-(3z_{12}+4z_{13}+3z_{23}) \;=\; 0 \\
U_6: \quad & (4z_{11}+4z_{33})\alpha^4+(8z_{11}-4z_{12}-4z_{23}+8z_{33})\alpha^3 \\
& +(6z_{11}-7z_{12}-2z_{13}+2z_{22}-7z_{23}+6z_{33})\alpha^2 \\
& +(3z_{11}-4z_{12}-6z_{13}+4z_{22}-4z_{23}+3z_{33})\alpha-(3z_{12}+4z_{13}+3z_{23}) \;=\; 0\,.
\end{aligned}
$$

Solving these equations, we get six estimates of α, which we denote them as $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_6$, respectively. Using the formula (2.3.8), we can check that the asymptotic variances of these six estimates are

$$
\begin{aligned}
\sigma^2_{\hat{\alpha}_1} &= \frac{3+12\alpha+10\alpha^2-12\alpha^3-20\alpha^4-6\alpha^5+5\alpha^6+4\alpha^7+4\alpha^8}{(3+3\alpha+2\alpha^2)^2} \\
\sigma^2_{\hat{\alpha}_2} &= \frac{6+20\alpha+6\alpha^2-24\alpha^3-3\alpha^4+11\alpha^5-3\alpha^6-11\alpha^7-2\alpha^8}{2(3+2\alpha+\alpha^2)^2} \\
\sigma^2_{\hat{\alpha}_3} &= \frac{\begin{array}{l}6+24\alpha+24\alpha^2-8\alpha^3-27\alpha^4-25\alpha^5 \\ \qquad -7\alpha^6-3\alpha^7+6\alpha^8+8\alpha^9+2\alpha^{10}\end{array}}{2(3+3\alpha+3\alpha^2+\alpha^3)^2} \\
\sigma^2_{\hat{\alpha}_4} &= \frac{\begin{array}{l}17+115\alpha+301\alpha^2+335\alpha^3-48\alpha^4-574\alpha^5 \\ \qquad -593\alpha^6-135\alpha^7+198\alpha^8+256\alpha^9+128\alpha^{10}\end{array}}{2(5+12\alpha+14\alpha^2+8\alpha^3)^2} \\
\sigma^2_{\hat{\alpha}_5} &= \frac{17+73\alpha+74\alpha^2-62\alpha^3-125\alpha^4+9\alpha^5+78\alpha^6-32\alpha^7-32\alpha^8}{2(5+6\alpha+4\alpha^2)^2} \\
\sigma^2_{\hat{\alpha}_6} &= \frac{\begin{array}{l}17+81\alpha+128\alpha^2+50\alpha^3-109\alpha^4-183\alpha^5 \\ \qquad -80\alpha^6-16\alpha^7+16\alpha^8+64\alpha^9+32\alpha^{10}\end{array}}{2(5+7\alpha+8\alpha^2+4\alpha^3)^2}\,.
\end{aligned}
$$

Figure 2.1 shows the plot of the above asymptotic variances for values of α in the feasible region. The plot also contains the asymptotic variance of the ML estimate. Since ML estimate is the optimal, we computed the asymptotic relative efficiencies

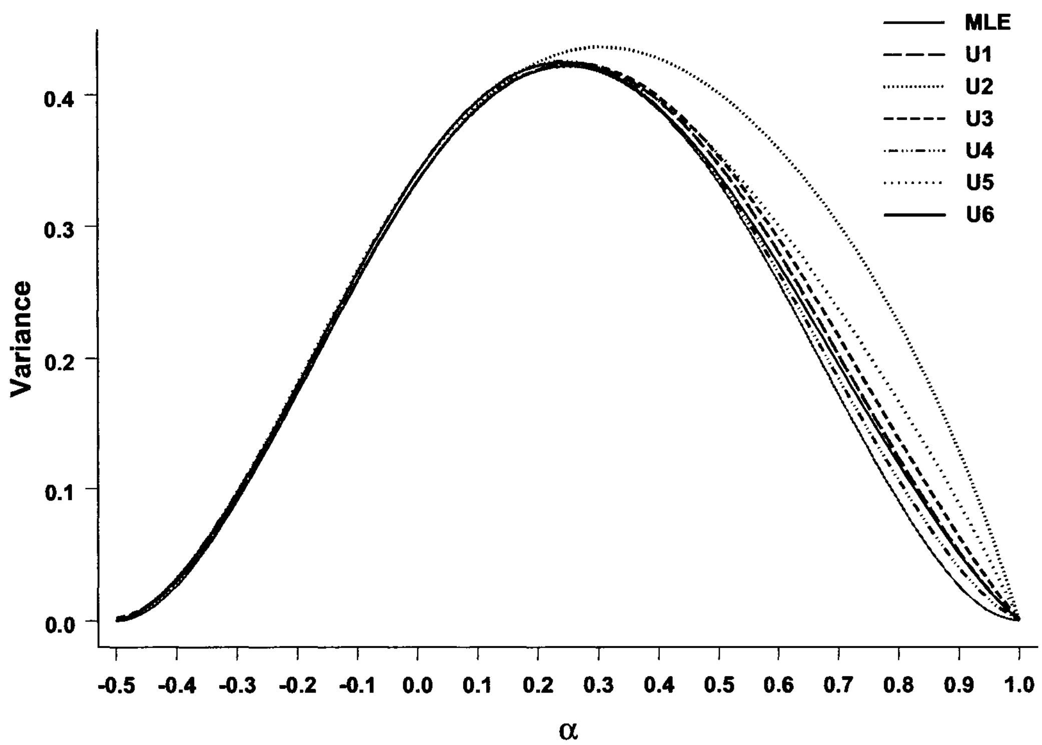


Figure 2.1: EXCH: Variances of $\hat{\alpha}_1$ through $\hat{\alpha}_6$ and MLE.

(ARE) of the six estimates as the ratio of the asymptotic variances with respect to the variance of the ML estimate, that is, $\text{ARE}(\hat{\alpha}_i, \hat{\alpha}_L) = \sigma^2_{\hat{\alpha}_L}/\sigma^2_{\hat{\alpha}_i}$ for $i = 1, 2, \ldots, 6$. Figure 2.2 shows the relative efficiencies of the six estimates. We can see from Figure 2.2, U_4 yields an estimate that has high efficiency when α takes moderate or large positive values. And equation U_2 yields an estimate that is highly efficient when α takes negative values. In general, there is no uniformly "best" equation among the six.

We now consider the case where the correlation structure is AR(1). Here we have

$$\mathbf{R} = \begin{bmatrix} 1 & \alpha & \alpha^2 \\ \alpha & 1 & \alpha \\ \alpha^2 & \alpha & 1 \end{bmatrix}, \quad \mathbf{R}^{-1} = \frac{1}{1-\alpha^2} \begin{bmatrix} 1 & -\alpha & 0 \\ -\alpha & 1+\alpha^2 & -\alpha \\ 0 & -\alpha & 1 \end{bmatrix}.$$

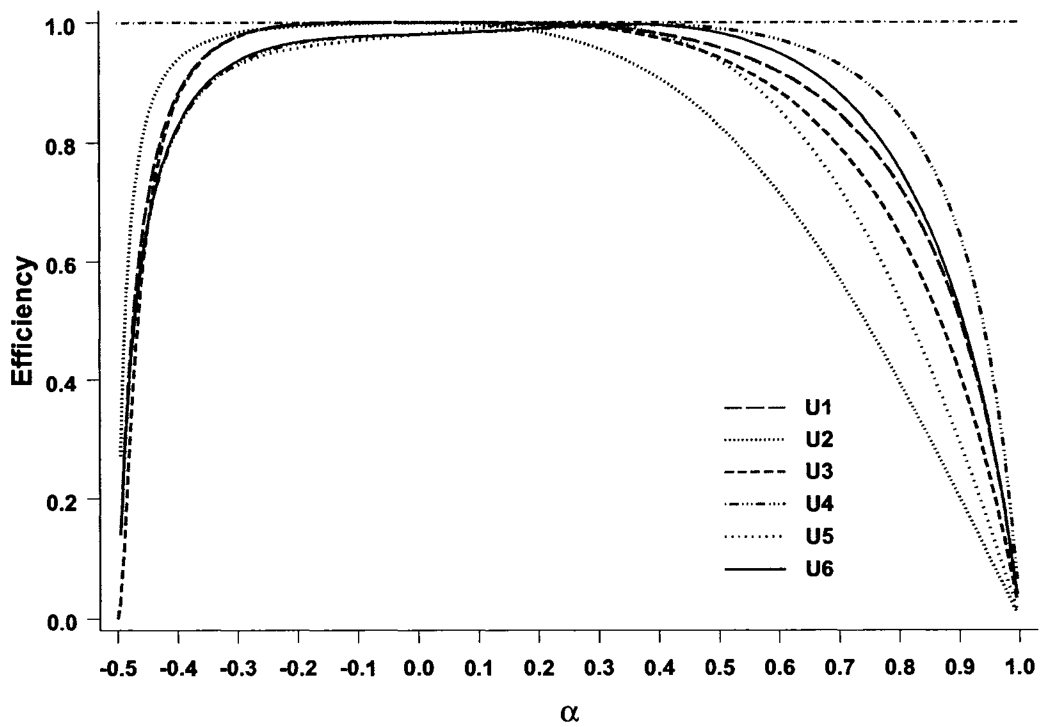


Figure 2.2: EXCH: AREs of $\hat{\alpha}_1$ through $\hat{\alpha}_6$.

The matrices involved in the Cholesky decomposition of $\mathbf{R}^{-1}$ are

$$\mathbf{B}_l = \begin{bmatrix} 1 & 0 & 0 \\ -\alpha & 1 & 0 \\ 0 & -\alpha & 1 \end{bmatrix} \qquad \mathbf{\Lambda}_l = \begin{bmatrix} \frac{1}{1-\alpha^2} & 0 & 0 \\ 0 & \frac{1}{1-\alpha^2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{B}_u = \begin{bmatrix} 1 & -\alpha & 0 \\ 0 & 1 & -\alpha \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{\Lambda}_u = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{1-\alpha^2} & 0 \\ 0 & 0 & \frac{1}{1-\alpha^2} \end{bmatrix}.$$

Therefore,

$$\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{\Lambda}_l\mathbf{B}_l' = d_1 \begin{bmatrix} 0 & 0 & 0 \\ -1 & \alpha & 0 \\ 0 & -1 & \alpha \end{bmatrix} \qquad \frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{\Lambda}_u\mathbf{B}_u' = d_1 \begin{bmatrix} \alpha & -1 & 0 \\ 0 & \alpha & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{R}^{-1} = d_1 \begin{bmatrix} 0 & 0 & 0 \\ -1 & \alpha & 0 \\ \alpha & -(1+\alpha^2) & \alpha \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{R}^{-1} = d_1 \begin{bmatrix} \alpha & -(1+\alpha^2) & \alpha \\ 0 & \alpha & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_l}{\partial \alpha}\mathbf{B}_l' = \begin{bmatrix} 0 & 0 & 0 \\ -1 & \alpha & 0 \\ 0 & -1 & \alpha \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial \alpha}\mathbf{B}_u' = \begin{bmatrix} \alpha & -1 & 0 \\ 0 & \alpha & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_l}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_l}{\partial \alpha}\mathbf{\Lambda}_l\mathbf{B}_l' = d_2 \begin{bmatrix} 0 & 0 & 0 \\ -1 & \alpha & 0 \\ 0 & -1 & \alpha \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_u}{\partial \alpha}\mathbf{\Lambda}_u\mathbf{B}_u' = d_2 \begin{bmatrix} \alpha & -1 & 0 \\ 0 & \alpha & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_l}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_l}{\partial \alpha}\mathbf{R}^{-1} = d_2 \begin{bmatrix} 0 & 0 & 0 \\ -1 & \alpha & 0 \\ \alpha & -(1+\alpha^2) & \alpha \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_u}{\partial \alpha}\mathbf{R}^{-1} = d_2 \begin{bmatrix} \alpha & -(1+\alpha^2) & \alpha \\ 0 & \alpha & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_l}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_l}{\partial \alpha}\mathbf{B}_l' = d_3 \begin{bmatrix} 0 & 0 & 0 \\ -1 & \alpha & 0 \\ 0 & -1 & \alpha \end{bmatrix}$$

$$\frac{\partial \mathbf{B}_u}{\partial \alpha}\frac{\partial \mathbf{\Lambda}_u}{\partial \alpha}\mathbf{B}_u' = d_3 \begin{bmatrix} \alpha & -1 & 0 \\ 0 & \alpha & -1 \\ 0 & 0 & 0 \end{bmatrix},$$

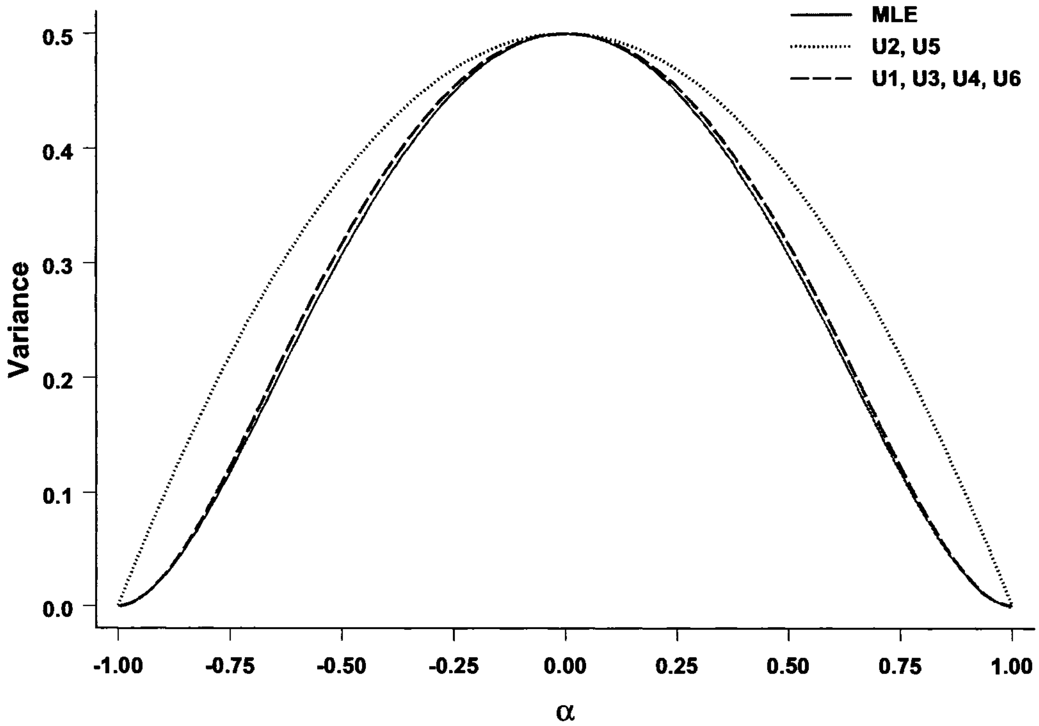


Figure 2.3: AR(1): Variances of $\hat{\alpha}_1$ through $\hat{\alpha}_6$ and MLE.

where the constants are

$$d_1 = \frac{1}{1-\alpha^2} \qquad d_2 = \frac{2\alpha}{(1-\alpha^2)^3} \qquad d_3 = \frac{2\alpha}{(1-\alpha^2)^2}.$$

In this case the four estimating equations U_1, U_3, U_4 and U_6 are equivalent and simplify to

$$(z_{11} + 2z_{22} + z_{33})\alpha - 2(z_{12} + z_{23}) = 0.$$

Further, the two equations U_2 and U_5 are equivalent and reduce to

$$(z_{12} + z_{23})\alpha^2 - (z_{11} + 2z_{13} + 2z_{22} + z_{33})\alpha + 2(z_{12} + z_{23}) = 0.$$

Hence,

$$\hat{\alpha}_k = \frac{2(z_{12} + z_{23})}{z_{11} + z_{22} + z_{33}} \quad \text{for} \quad k = 1,3,4,6$$

and

$$\hat{\alpha}_k = \frac{(z_{11} + 2z_{13} + 2z_{22} + z_{33}) - \sqrt{(z_{11} + 2z_{13} + 2z_{22} + z_{33})^2 - 8(z_{12} + z_{23})^2}}{2(z_{12} + z_{23})} \quad \text{for} \quad k = 2,5.$$

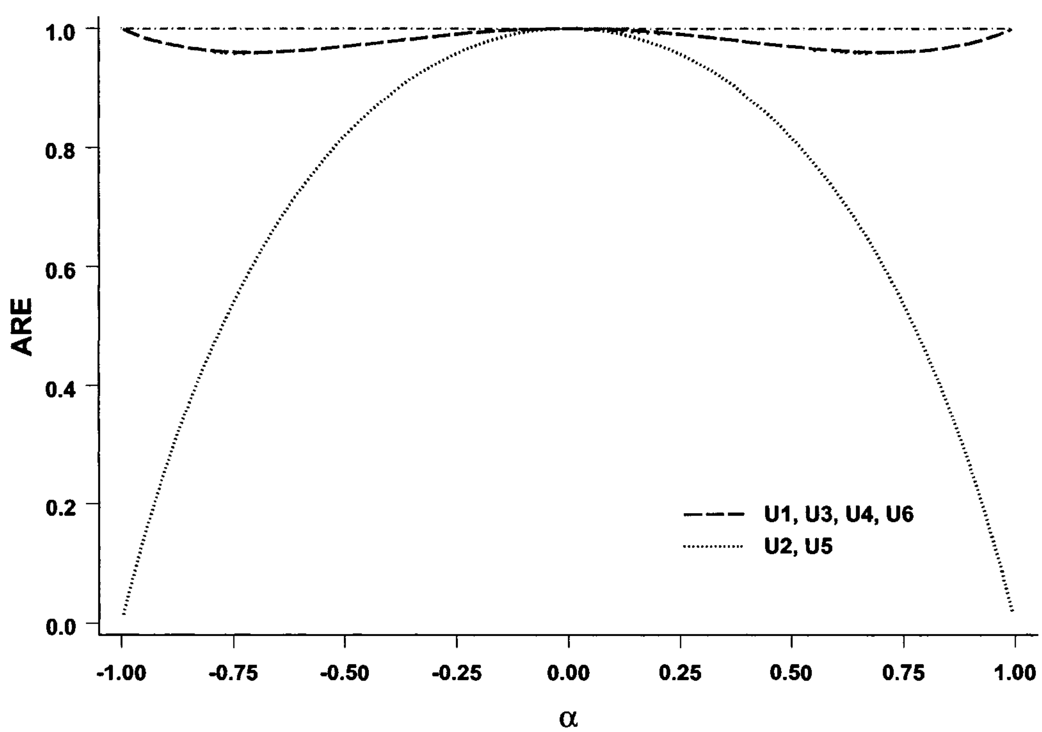


Figure 2.4: AR(1): AREs of $\hat{\alpha}_1$ through $\hat{\alpha}_6$.

The asymptotic variances of these estimates are obtained from (2.3.8) as

$$\sigma^2_{\hat{\alpha}_k} = \begin{cases} \dfrac{\alpha^6 - 3\alpha^2 + 2}{4} & \text{if } k = 1, 3, 4 \text{ or } 6 \\ \dfrac{1-\alpha^2}{2} & \text{if } k = 2 \text{ or } 5\,. \end{cases}$$

Figure 2.3 contains the plot of these variances and the variance of the ML estimate. As before we calculated the asymptotic relative efficiences as the ratios $\left(\sigma^2_{\hat{\alpha}_L}/\sigma^2_{\hat{\alpha}_i}\right)$ of the asymptotic variances with respect to the variance of the ML estimate. Figure 2.4 shows the graphs of these relative efficiencies.

It is clear from the graph, equation U_4 (equivalently U_1, U_3 and U_6) is better than U_2 (or U_5). Further U_4 is almost as good as the likelihood estimating equation.

In summary we can conclude that $\hat{\alpha}_4$ is a good estimate for both the exchangeable and AR(1) structures when the underlying correlation is positive. It is a good competitor to the ML estimate.

Since all these six estimating equations belong to the subclass of weighted estimating equations (2.4.3), we wonder whether there exists an efficient estimating equation within this subclass. We explore this issue in the next chapter.

CHAPTER III

EXTENSIONS TO UNBALANCED DATA

In Chapter II, we have studied the behavior of unbiased estimating equations in the balanced case. Even though in practice the data is unbalanced, the results throw some light on what we can expect in general. We now extend the results of the previous chapter to more general situations. We assume that the longitudinal data is unbalanced but the within cluster or subject variance is constant, that is, the covariance matrix of the repeated measurements on subject i is $\mathbf{V}_i = \sigma^2\mathbf{R}_i(\alpha)$ where $\mathbf{R}_i(\alpha)$ is a function of the unknown parameter α.

The organization of this chapter is as follows. In Section III.1, we study ML estimation under the multivariate normality assumption, derive asymptotic distributions for common correlation structures. As an important result we show that ML estimating equation for the correlation parameter is Godambe optimal. In Section III.2, we extend the general class and the subclass of unbiased estimating equations studied in the previous chapter, to the unbalanced situation. We derive asymptotic properties of the estimates under the assumption of normality. Several expressions are simplified further under special correlation structures. In Section III.3, we discuss the relative efficiencies under the normality assumption, and when there is a violation of the normality assumption. Finally, in Section III.4, we illustrate the estimation methods using a real life data set and contrast them with ML estimates.

III.1 Maximum Likelihood Estimate

Suppose that $\mathbf{y}_i$ is multivariate normal with mean $\mathbf{X}_i'\boldsymbol{\beta}$ and covariance matrix $\mathbf{V}_i = \sigma^2\mathbf{R}_i$, for $1 \le i \le n$. The log-likelihood function is

$$\ell = -\frac{1}{2}\left\{\frac{1}{\sigma^2}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) + \sum_{i=1}^{n}\log|\mathbf{R}_i| + \log(\sigma^2)\sum_{i=1}^{n}t_i \right.$$
$$\left. + \log(2\pi)\sum_{i=1}^{n}t_i\right\}.$$

We assume that $\mathbf{R}_i$ is a function of an unknown parameter α. Taking derivatives we can see that the ML estimating equation for α reduces to

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathrm{tr}\left\{\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' - \sigma^2\mathbf{R}_i\right)\right\} = 0$$

or equivalently,

$$\sum_{i=1}^{n}\mathrm{tr}\left\{\mathbf{R}_i^{-1}\frac{\partial\mathbf{R}_i}{\partial\alpha}\mathbf{R}_i^{-1}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' - \sigma^2\mathbf{R}_i\right)\right\} = 0, \qquad (3.1.1)$$

where $\boldsymbol{\varepsilon}_i = \mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}$. The ML estimate $\hat{\alpha}_L$ of α is simply the root of the above equation (3.1.1). The following derivatives are useful to simplify the Fisher information matrix.

$$\begin{aligned}
\frac{\partial^2\ell}{\partial\boldsymbol{\beta}^2} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}_i^{-1}\mathbf{X}_i' \\
\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) \\
\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\alpha} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) \\
\frac{\partial^2\ell}{\partial(\sigma^2)^2} &= \frac{1}{2\sigma^4}\sum_{i=1}^{n}t_i - \frac{1}{\sigma^6}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) \\
\frac{\partial^2\ell}{\partial\sigma^2\partial\alpha} &= \frac{1}{2\sigma^4}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) \\
\frac{\partial^2\ell}{\partial\alpha^2} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta})'\frac{\partial^2\mathbf{R}_i^{-1}}{\partial\alpha^2}(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}) - \frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2\log|\mathbf{R}_i|}{\partial\alpha^2}.
\end{aligned}$$

Thus the elements of the Fisher information are

$$\begin{aligned}
\mathrm{E}\left\{\frac{\partial^2\ell}{\partial\boldsymbol{\beta}^2}\right\} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}_i^{-1}\mathbf{X}_i' \\
\mathrm{E}\left\{\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\sigma^2}\right\} &= \mathbf{0} \qquad \mathrm{E}\left\{\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\partial\alpha}\right\} = \mathbf{0} \\
\mathrm{E}\left\{\frac{\partial^2\ell}{\partial(\sigma^2)^2}\right\} &= -\frac{1}{2\sigma^4}\sum_{i=1}t_i \\
\mathrm{E}\left\{\frac{\partial^2\ell}{\partial\sigma^2\partial\alpha}\right\} &= \frac{1}{2\sigma^2}\sum_{i=1}^{n}\mathrm{tr}\left(\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right) \\
\mathrm{E}\left\{\frac{\partial^2\ell}{\partial\alpha^2}\right\} &= -\frac{1}{2}\left\{\sum_{i=1}^{n}\mathrm{tr}\left(\frac{\partial^2\mathbf{R}_i^{-1}}{\partial\alpha^2}\mathbf{R}_i\right) + \sum_{i=1}^{n}\frac{\partial^2\log|\mathbf{R}_i|}{\partial\alpha^2}\right\}.
\end{aligned}$$

In matrix notation the Fisher information can be written as

$$\begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{R}_i^{-1}\mathbf{X}_i' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \frac{1}{2\sigma^4}\sum t_i & -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\mathrm{tr}\left(\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right) \\ \mathbf{0} & -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\mathrm{tr}\left(\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right) & \frac{1}{2}\left\{\sum_{i=1}^{n}\mathrm{tr}\left(\frac{\partial^2\mathbf{R}_i^{-1}}{\partial\alpha^2}\mathbf{R}_i\right)+\sum_{i=1}^{n}\frac{\partial^2\log|\mathbf{R}_i|}{\partial\alpha^2}\right\} \end{bmatrix}.$$

We now further simplify the above matrix under common correlation structures. Suppose that $\mathbf{R}_i$ is exchangeable. Then

$$\begin{aligned} \mathrm{tr}\left(\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right) &= \frac{t_i(t_i-1)\alpha}{(1-\alpha)[1+(t_i-1)\alpha]} \\ \mathrm{tr}\left(\frac{\partial^2\mathbf{R}_i^{-1}}{\partial\alpha^2}\mathbf{R}_i\right) &= \frac{2t_i(t_i-1)[1+(t_i-1)\alpha^2]}{(1-\alpha)^2[1+(t_i-1)\alpha]^2} \\ \frac{\partial^2\log|\mathbf{R}_i|}{\partial\alpha^2} &= \frac{-t_i(t_i-1)[1+(t_i-1)\alpha^2]}{(1-\alpha)^2[1+(t_i-1)\alpha]^2}. \end{aligned}$$

The lower right part of the Fisher information matrix is

$$\begin{bmatrix} \frac{1}{2\sigma^4}\sum_{i=1} t_i & \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{-t_i(t_i-1)\alpha}{(1-\alpha)[1+(t_i-1)\alpha]} \\ \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{-t_i(t_i-1)\alpha}{(1-\alpha)[1+(t_i-1)\alpha]} & \sum_{i=1}^{n}\frac{t_i(t_i-1)[1+(t_i-1)\alpha^2]}{(1-\alpha)^2[1+(t_i-1)\alpha]^2} \end{bmatrix}.$$

Taking the inverse of the above matrix we get the asymptotic covariance matrix of the ML estimates of $(\sigma^2, \alpha)'$ as

$$\frac{2}{\omega_E}\begin{bmatrix} \sigma^4\sum_{i=1}^{n}\frac{t_i(t_i-1)[1+(t_i-1)\alpha^2]}{[1+(t_i-1)\alpha]^2} & (1-\alpha)\sigma^2\sum_{i=1}^{n}\frac{t_i(t_i-1)\alpha}{1+(t_i-1)\alpha} \\ (1-\alpha)\sigma^2\sum_{i=1}^{n}\frac{t_i(t_i-1)\alpha}{1+(t_i-1)\alpha} & (1-\alpha)^2\sum_{i=1} t_i \end{bmatrix},$$

where $\omega_E = \left\{\sum_{i=1} t_i\right\}\left\{\sum_{i=1}^{n}\frac{t_i(t_i-1)[1+(t_i-1)\alpha^2]}{[1+(t_i-1)\alpha]^2}\right\} - \left\{\sum_{i=1}^{n}\frac{t_i(t_i-1)\alpha}{1+(t_i-1)\alpha}\right\}^2.$

Suppose now that $\mathbf{R}_i$ has an AR(1) structure. In this case we have

$$\begin{aligned} \mathrm{tr}\left(\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right) &= \frac{2(t_i-1)\alpha}{1-\alpha^2} \\ \mathrm{tr}\left(\frac{\partial^2\mathbf{R}_i^{-1}}{\partial\alpha^2}\mathbf{R}_i\right) &= \frac{4(t_i-1)(1+\alpha^2)}{(1-\alpha^2)^2} \\ \frac{\partial^2\log|\mathbf{R}_i|}{\partial\alpha^2} &= \frac{-2(t_i-1)(1+\alpha^2)}{(1-\alpha^2)^2}. \end{aligned}$$

The lower right part of the Fisher information simplifies to

$$\begin{bmatrix} \dfrac{1}{2\sigma^4}\sum_{i=1} t_i & \dfrac{-\alpha}{\sigma^2(1-\alpha^2)}\left\{\sum_{i=1}^{n} t_i - n\right\} \\ \dfrac{-\alpha}{\sigma^2(1-\alpha^2)}\left\{\sum_{i=1}^{n} t_i - n\right\} & \dfrac{1+\alpha^2}{(1-\alpha^2)^2}\left\{\sum_{i=1}^{n} t_i - n\right\} \end{bmatrix}.$$

And again taking the inverse of the above matrix we get asymptotic covariance matrix of the ML estimates of $(\sigma^2, \alpha)'$ as

$$\frac{1}{\omega_A}\begin{bmatrix} 2(1+\alpha^2)\sigma^4\left\{\sum_{i=1}^{n} t_i - n\right\} & \alpha(1-\alpha^2)\sigma^2\left\{\sum_{i=1}^{n} t_i - n\right\} \\ \alpha(1-\alpha^2)\sigma^2\left\{\sum_{i=1}^{n} t_i - n\right\} & (1-\alpha^2)^2\sum_{i=1} t_i \end{bmatrix},$$

where $\omega_A = (1+\alpha^2)\left\{\sum_{i=1}^{n} t_i\right\}\left\{\sum_{i=1}^{n} t_i - n\right\} - 2\alpha^2\left\{\sum_{i=1}^{n} t_i - n\right\}^2$.

We now show that the ML equation (3.1.1) is the optimal estimating equation in the sense of Godambe (1960). First it is easy to check that we can write equation (3.1.1) as

$$\sum_{i=1}^{n} \operatorname{tr}\left\{\frac{\partial \mathbf{R}_i}{\partial \alpha}\mathbf{R}_i^{-1}\left(\varepsilon_i\varepsilon_i' - \sigma^2\mathbf{R}_i\right)\mathbf{R}_i^{-1}\right\} = 0. \qquad (3.1.2)$$

Using the identity (Rao and Rao, 1998)

$$\operatorname{tr}(\mathbf{A}'\mathbf{B}\mathbf{C}\mathbf{D}') = \operatorname{tr}(\mathbf{B}\mathbf{C}\mathbf{D}'\mathbf{A}') = \operatorname{vec}'(\mathbf{A})(\mathbf{D}\otimes\mathbf{B})\operatorname{vec}(\mathbf{C})$$

for any matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ of appropriate order, we can rewrite equation (3.1.2) as

$$\sum_{i=1}^{n}\left\{\frac{\partial \operatorname{vec}(\mathbf{R}_i)}{\partial \alpha}\right\}'\left(\mathbf{R}_i^{-1}\otimes\mathbf{R}_i^{-1}\right)\left\{\operatorname{vec}(\varepsilon_i\varepsilon_i') - \sigma^2\operatorname{vec}(\mathbf{R}_i)\right\} = 0,$$

which is equivalent to

$$\sum_{i=1}^{n}\left\{\frac{\partial \operatorname{vech}(\mathbf{R}_i)}{\partial \alpha}\right\}'\mathbf{G}_i'(\mathbf{R}_i^{-1}\otimes\mathbf{R}_i^{-1})\mathbf{G}_i\left\{\operatorname{vech}(\varepsilon_i\varepsilon_i') - \sigma^2\operatorname{vech}(\mathbf{R}_i)\right\} = 0. \quad (3.1.3)$$

Here $\mathbf{G}_i$ is the duplication matrix defined in Harville (1997, page 352). Now if $\mathbf{H}_i = (\mathbf{G}_i'\mathbf{G}_i)^{-1}\mathbf{G}_i'$, we have $\operatorname{vech}(\varepsilon_i\varepsilon_i') = \mathbf{H}_i\operatorname{vec}(\varepsilon_i\varepsilon_i') = (\mathbf{G}_i'\mathbf{G}_i)^{-1}\mathbf{G}_i'\operatorname{vec}(\varepsilon_i\varepsilon_i')$. The covariance matrix of $\operatorname{vech}(\varepsilon_i\varepsilon_i')$ is

$$\operatorname{Cov}\left\{\operatorname{vech}(\varepsilon_i\varepsilon_i')\right\} = \mathbf{H}_i\operatorname{Cov}\left\{\operatorname{vec}(\varepsilon_i\varepsilon_i')\right\}\mathbf{H}_i'$$

$$
\begin{aligned}
&= (\mathbf{G}_i'\mathbf{G}_i)^{-1}\mathbf{G}_i'\mathrm{Cov}\{\mathrm{vec}(\varepsilon_i\varepsilon_i')\}\mathbf{G}_i(\mathbf{G}_i'\mathbf{G}_i)^{-1} \\
&= \sigma^4(\mathbf{G}_i'\mathbf{G}_i)^{-1}\mathbf{G}_i'\left(\mathbf{I}_{t_i^2}+\mathbf{I}_{(t_i,t_i)}\right)(\mathbf{R}_i\otimes\mathbf{R}_i)\mathbf{G}_i(\mathbf{G}_i'\mathbf{G}_i)^{-1} \\
&= 2\sigma^4(\mathbf{G}_i'\mathbf{G}_i)^{-1}\mathbf{G}_i'(\mathbf{R}_i\otimes\mathbf{R}_i)\mathbf{G}_i(\mathbf{G}_i'\mathbf{G}_i)^{-1},
\end{aligned}
$$

where $\mathbf{I}_{t_i^2}$ and $\mathbf{I}_{(t_i,t_i)}$ are defined in proof of Lemma 2.1. Since $\mathrm{E}\{\mathrm{vech}(\varepsilon_i\varepsilon_i')\} = \sigma^2\mathrm{vech}(\mathbf{R}_i)$, the optimal estimating equation according to Godambe (1960) is

$$
\sum_{i=1}^{n}\left\{\frac{\partial\mathrm{vech}(\mathbf{V}_i)}{\partial\alpha}\right\}'\mathbf{G}_i'\mathbf{G}_i(\mathbf{G}_i'\mathbf{R}_i\otimes\mathbf{R}_i\mathbf{G}_i)^{-1}\mathbf{G}_i'\mathbf{G}_i\{\mathrm{vech}(\varepsilon_i\varepsilon_i')-\sigma^2\mathrm{vech}(\mathbf{R}_i)\} = 0\,. \tag{3.1.4}
$$

Comparing the terms in equations (3.1.3) and (3.1.4), we find that the only difference is the middle term. We will show that equation (3.1.3) is equivalent to (3.1.4) by proving that the middle terms are equivalent. Notice that $\mathbf{H}_i$ is defined as $(\mathbf{G}_i'\mathbf{G}_i)^{-1}\mathbf{G}_i'$, and $(\mathbf{R}_i\otimes\mathbf{R}_i)^{-1} = \mathbf{R}_i^{-1}\otimes\mathbf{R}_i^{-1}$, using the identities below (Harville, 1997, page 358),

(1) $\mathbf{R}_i^{-1}\otimes\mathbf{R}_i^{-1}\mathbf{G}_i = \mathbf{G}_i\mathbf{H}_i(\mathbf{R}_i^{-1}\otimes\mathbf{R}_i^{-1})\mathbf{G}_i$

(2) $\{\mathbf{H}_i(\mathbf{R}_i\otimes\mathbf{R}_i)\mathbf{G}_i\}^{-1} = \mathbf{H}_i(\mathbf{R}_i^{-1}\otimes\mathbf{R}_i^{-1})\mathbf{G}_i$

we have

$$
\begin{aligned}
\mathbf{G}_i'(\mathbf{R}_i^{-1}\otimes\mathbf{R}_i^{-1})\mathbf{G}_i &= \mathbf{G}_i'\mathbf{G}_i\mathbf{H}_i(\mathbf{R}_i^{-1}\otimes\mathbf{R}_i^{-1})\mathbf{G}_i \\
&= \mathbf{G}_i'\mathbf{G}_i\{\mathbf{H}_i(\mathbf{R}_i\otimes\mathbf{R}_i)\mathbf{G}_i\}^{-1} \\
&= \mathbf{G}_i'\mathbf{G}_i\{(\mathbf{G}_i'\mathbf{G}_i)^{-1}\mathbf{G}_i'(\mathbf{R}_i\otimes\mathbf{R}_i)\mathbf{G}_i\}^{-1} \\
&= \mathbf{G}_i'\mathbf{G}_i\{\mathbf{G}_i'\mathbf{R}_i\otimes\mathbf{R}_i\mathbf{G}_i\}^{-1}\mathbf{G}_i'\mathbf{G}_i.
\end{aligned}
$$

Thus the middle terms of (3.1.3) and (3.1.4) are equivalent. Therefore, the likelihood estimation equation for the correlation parameter and the Godambe's optimal estimating equation are identical.

III.2 Classes of Unbiased Estimating Equations

In general, a class of unbiased estimating equations for estimating correlation parameter can be written as

$$
U_G = \sum_{i=1}^{n}\mathrm{tr}\left\{\mathbf{W}_i\mathbf{R}_i^{-1}\left(\mathbf{z}_i\mathbf{z}_i'-\mathbf{R}_i\right)\right\} = 0\,, \tag{3.2.1}
$$

where $\mathbf{W}_i$ is a weighting matrix and $\mathbf{z}_i$ is the Pearson residual of $\mathbf{y}_i$. It is easy to verify that $\mathrm{E}(U_G) = 0$ since $\mathrm{E}(\mathbf{z}_i\mathbf{z}_i') = \mathbf{R}_i$. Hence U_G is an unbiased estimating equation and the solution of (3.2.1) is a consistent estimate of α. Notice that likelihood equation of α for normal data is of the form U_G. By Lemma 2.1 and Theorem 2.1, the asymptotic variance of the root $\hat{\alpha}$ of equation (3.2.1), is given by

$$\begin{aligned}
\frac{\mathrm{Cov}\,(U_G)}{\left[\mathrm{E}\left(\frac{\partial U_G}{\partial \alpha}\right)\right]^2} &= \frac{\mathrm{Cov}\left\{\sum_{i=1}^{n} \mathrm{tr}\left\{\mathbf{W}_i\mathbf{R}_i^{-1}\left(\mathbf{z}_i\mathbf{z}_i' - \mathbf{R}_i\right)\right\}\right\}}{\left[\mathrm{E}\left\{\frac{\partial}{\partial \alpha}\sum_{i=1}^{n} \mathrm{tr}\left(\mathbf{W}_i\mathbf{R}_i^{-1}\left(\mathbf{z}_i\mathbf{z}_i' - \mathbf{R}_i\right)\right)\right\}\right]^2} \\
&= \frac{\mathrm{Cov}\left\{\sum_{i=1}^{n} \mathrm{tr}\left(\mathbf{W}_i\mathbf{R}_i^{-1}\mathbf{z}_i\mathbf{z}_i'\right) - \sum_{i=1}^{n}\mathrm{tr}(\mathbf{W}_i)\right\}}{\left[\mathrm{E}\left\{\frac{\partial}{\partial \alpha}\left(\sum_{i=1}^{n} \mathrm{tr}\left(\mathbf{W}_i\mathbf{R}_i^{-1}\mathbf{z}_i\mathbf{z}_i'\right) + \sum_{i=1}^{n}\mathrm{tr}(\mathbf{W}_i)\right)\right\}\right]^2} \\
&= \frac{\mathrm{Cov}\left\{\sum_{i=1}^{n} \mathrm{tr}\left(\mathbf{W}_i\mathbf{R}_i^{-1}\mathbf{z}_i\mathbf{z}_i'\right)\right\}}{\left[\mathrm{E}\left\{\frac{\partial}{\partial \alpha}\sum_{i=1}^{n} \mathrm{tr}\left(\mathbf{W}_i\mathbf{R}_i^{-1}\mathbf{z}_i\mathbf{z}_i'\right)\right\}\right]^2} \\
&= \frac{\sum_{i=1}^{n} \mathrm{tr}\left(\mathbf{W}_i\mathbf{R}_i^{-1}\mathbf{W}_i'\mathbf{R}_i + \mathbf{W}_i^2\right)}{\left[\sum_{i=1}^{n} \mathrm{tr}\left(\mathbf{W}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right)\right]^2}. \qquad (3.2.2)
\end{aligned}$$

We will now study the problem of choosing the optimal weights which minimize the asymptotic variance (3.2.2). Using the identities (Magnus and Neudecker, 1999)

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{W}_i}\left\{\mathrm{tr}\left(\mathbf{W}_i\mathbf{R}_i^{-1}\mathbf{W}_i'\mathbf{R}_i\right)\right\} &= 2\,\mathbf{R}_i\mathbf{W}_i\mathbf{R}_i^{-1} \\
\frac{\partial}{\partial \mathbf{W}_i}\left\{\mathrm{tr}\left(\mathbf{W}_i^2\right)\right\} &= 2\,\mathbf{W}_i' \\
\frac{\partial}{\partial \mathbf{W}_i}\left\{\mathrm{tr}\left(\mathbf{W}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right)\right\} &= \mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}
\end{aligned}$$

and equating the derivative of (3.2.2) with respect to $\mathbf{W}_i$ to zero we get

$$\left\{\sum_{j=1}^{n} \mathrm{tr}\left(\mathbf{W}_j\frac{\partial \mathbf{R}_j^{-1}}{\partial \alpha}\mathbf{R}\right)\right\}\left\{\mathbf{R}_i\mathbf{W}_i\mathbf{R}_i^{-1} + \mathbf{W}_i'\right\}$$

$$-\left\{\sum_{j=1}^{n}\mathrm{tr}(\mathbf{W}_j\mathbf{R}_j^{-1}\mathbf{W}_j'\mathbf{R}_j+\mathbf{W}_j^2)\right\}\left\{\mathbf{R}_i\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\right\} = 0 \qquad (3.2.3)$$

for $i = 1, 2, \ldots, n$. Note that all the traces are scalars. The above equation can be written as

$$\{\mathbf{R}_i\mathbf{W}_i\mathbf{R}_i^{-1}+\mathbf{W}_i'\} + c\left\{\mathbf{R}_i\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\right\} = 0$$

where the constant

$$c = \frac{\sum_{j=1}^{n}\mathrm{tr}\left(\mathbf{W}_j\mathbf{R}_j^{-1}\mathbf{W}_j'\mathbf{R}_j+\mathbf{W}_j^2\right)}{\sum_{j=1}^{n}\mathrm{tr}\left(\mathbf{W}_j\frac{\partial\mathbf{R}_j^{-1}}{\partial\alpha}\mathbf{R}\right)}.$$

Since $\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}$ equals $-\mathbf{R}_i^{-1}\frac{\partial\mathbf{R}_i}{\partial\alpha}\mathbf{R}_i^{-1}$, we have

$$\{\mathbf{R}_i\mathbf{W}_i\mathbf{R}_i^{-1}+\mathbf{W}_i'\} = c\left\{\frac{\partial\mathbf{R}_i}{\partial\alpha}\mathbf{R}_i^{-1}\right\}.$$

Post multiplying by $\mathbf{R}_i$ we get

$$\{\mathbf{R}_i\mathbf{W}_i+(\mathbf{R}_i\mathbf{W}_i)'\} = c\left\{\frac{\partial\mathbf{R}_i}{\partial\alpha}\right\}. \qquad (3.2.4)$$

Under the additional assumption $\mathbf{R}_i\mathbf{W}_i$ is symmetric, a solution to equation (3.2.4) is $\mathbf{W}_i = \mathbf{R}_i^{-1}\frac{\partial\mathbf{R}_i}{\partial\alpha}$. Interestingly, the ML equation (3.1.1) uses this weighting matrix. Thus we have an alternative derivation of the optimality of the ML equation.

We now look at some specific cases. Suppose that $\mathbf{R}_i = (1-\alpha)\mathbf{I}_i + \alpha\mathbf{J}_i$. The optimal weighting matrix is given by

$$\begin{aligned}\mathbf{W}_i = \mathbf{R}_i^{-1}\frac{\partial\mathbf{R}_i}{\partial\alpha} &= \left\{\frac{1}{1-\alpha}\mathbf{I}_i - \frac{\alpha}{(1-\alpha)[1+(t_i-1)\alpha)]}\mathbf{J}_i\right\}\{\mathbf{J}_i-\mathbf{I}_i\} \\ &= \frac{1}{1-\alpha}\left\{\frac{1}{1+(t_i-1)\alpha}\mathbf{J}_i-\mathbf{I}_i\right\}.\end{aligned}$$

Substituting this into the estimating equation (3.2.1), gives us the best estimating equation:

$$\sum_{i=1}^{n}\mathrm{tr}\left\{\frac{1+(t_i-1)\alpha^2}{\{1+(t_i-1)\alpha\}^2}\mathbf{J}_i-\mathbf{I}_i\right\}\mathbf{z}_i\mathbf{z}_i' + \alpha(1-\alpha)\sum_{i=1}^{n}\frac{t_i(t_i-1)}{1+(t_i-1)\alpha} = 0,$$

and this coincides with the ML equation. Now suppose $\mathbf{R}_i$ has an AR(1) structure. In this case the optimal weighting matrix is given by

$$\mathbf{R}_i^{-1}\frac{\partial \mathbf{R}_i}{\partial \alpha} = -\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i = \frac{1}{(1-\alpha^2)^2}\left\{(1-\alpha^2)\mathbf{C}_{1i} - 2\alpha(\mathbf{I}_i + \mathbf{C}_{0i})\right\}\mathbf{R}_i,$$

where $\mathbf{C}_{0i}$ and $\mathbf{C}_{1i}$ are of order t_i and are similar to $\mathbf{C}_0$ and $\mathbf{C}_1$, respectively. It turns out even in this case, the best unbiased estimating equation is same as the ML equation and it is given by

$$\sum_{i=1}^{n} \operatorname{tr}\left\{\left(1+\alpha^2\right)\mathbf{C}_{1i} - 2\alpha\left(\mathbf{I}_i + \mathbf{C}_{1i}\right)\right\}\mathbf{z}_i\mathbf{z}_i' + 2\alpha\left(1-\alpha^2\right)\left\{\sum_{i=1}^{n} t_i - n\right\} = 0.$$

We will now focus our attention on the subclass of (3.2.1) with added restriction $\operatorname{tr}(\mathbf{W}_i) = 0$ for all i. Explicitly, the subclass takes the form

$$\sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_i\mathbf{R}_i^{-1}\mathbf{z}_i\mathbf{z}_i'\right\} = 0 \quad \text{subject to} \quad \operatorname{tr}(\mathbf{W}_i) = 0. \tag{3.2.5}$$

The asymptotic variance of the estimate from this subclass of equations is same as before except for the constraint $\operatorname{tr}(\mathbf{W}_i) = 0$ for all i. To get the optimal weights we need to minimize the asymptotic variance subject to the restrictions on the weights. Introducing Lagrange multipliers (λ_i) and equating to zero the derivative of (3.2.2) with respect to $\mathbf{W}_i$, we get

$$\begin{aligned}&\left\{\sum_{i=1}^{n} \operatorname{tr}\left(\mathbf{W}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}\right)\right\}\left\{\mathbf{R}_i\mathbf{W}_i\mathbf{R}_i^{-1} + \mathbf{W}_i'\right\}\\ &- \left\{\sum_{i=1}^{n} \operatorname{tr}(\mathbf{W}_i\mathbf{R}_i^{-1}\mathbf{W}_i'\mathbf{R}_i + \mathbf{W}_i^2)\right\}\left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\right\} - \frac{\lambda_i}{2}\mathbf{I}_i\left\{\operatorname{tr}\left(\mathbf{W}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right)\right\}^3 = 0\end{aligned}$$

which can be further simplified as

$$\left\{\mathbf{R}_i\mathbf{W}_i\mathbf{R}_i^{-1} + \mathbf{W}_i'\right\} + c_1\left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\right\} - c_{2i}\mathbf{I}_i = 0, \tag{3.2.6}$$

where c_1 and c_{2i} are constants. Equation (3.2.6) is equivalent to

$$\mathbf{R}_i\mathbf{W}_i + (\mathbf{R}_i\mathbf{W}_i)' = c_{2i}\mathbf{R}_i - c_1\left\{\frac{\partial \mathbf{R}_i}{\partial \alpha}\right\}. \tag{3.2.7}$$

Unfortunately, (3.2.7) does not have an explicit solution for the optimal weight $\mathbf{W}_i$ since the constant terms are complicated. However, when $\mathbf{R}_i$ has an exchangeable structure we have an explicit solution, which is

$$\begin{aligned}\mathbf{W}_i &= \mathbf{R}_i^{-1}\frac{\partial \mathbf{R}_i}{\partial \alpha} - \operatorname{diag}\left(\mathbf{R}_i^{-1}\frac{\partial \mathbf{R}_i}{\partial \alpha}\right)\\ &= \frac{1}{(1-\alpha)[1+(t_i-1)\alpha]}\left(\mathbf{J}_i - \mathbf{I}_i\right).\end{aligned}$$

The constant could be ignored when $t_i = t$ and thus

$$\mathbf{W}_i = \mathbf{J}_i - \mathbf{I}_i \quad \text{or} \quad \mathbf{W}_i = \mathbf{R}_i - \mathbf{I}_i . \tag{3.2.8}$$

The estimating equation with the above weights turns out to be

$$\sum_{i=1}^{n} \operatorname{tr}\left\{\left(\frac{1}{1+(t_i-1)\alpha}\mathbf{J}_i - \mathbf{I}_i\right)\mathbf{z}_i\mathbf{z}_i'\right\} = 0 . \tag{3.2.9}$$

This equation does not have a closed form solution except in the case $t_i = t$ for all i. But the asymptotic variance of the estimate is in a closed form

$$\sigma_{\hat{\alpha}}^2 = \frac{2\sum_{i=1}^{n} t_i(t_i-1)}{\left[\sum_{i=1}^{n} \frac{t_i(t_i-1)}{(1-\alpha)\left[1+(t_i-1)\alpha\right]}\right]^2} .$$

Now suppose that $\mathbf{R}_i$ has an AR(1) structure. In this case the optimal weights are difficult to obtain. But one possibility is to choose $\mathbf{W}_i = \mathbf{R}_i - \mathbf{I}_i$. Substituting these weights in the estimating equation (3.2.1) we get

$$\sum_{i=1}^{n} \operatorname{tr}\left\{\left(\alpha(\mathbf{I}_i + \mathbf{C}_{i0}) - \mathbf{C}_{i1}\right)\mathbf{z}_i\mathbf{z}_i'\right\} = 0 , \tag{3.2.10}$$

where $\mathbf{C}_{i0}$ is matrix $\mathbf{C}_0$ of order t_i and $\mathbf{C}_{i1}$ is matrix $\mathbf{C}_1$ with order t_i as defined in Section II.1.2. Equation (3.2.10) has a closed form solution which is given by

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} \operatorname{tr}\left(\mathbf{C}_{i1}\mathbf{z}_i\mathbf{z}_i'\right)}{\sum_{i=1}^{n} \operatorname{tr}\left\{(\mathbf{I}_i + \mathbf{C}_{i0})\mathbf{z}_i\mathbf{z}_i'\right\}} .$$

The asymptotic variance of $\hat{\alpha}$ is

$$\frac{\sum_{i=1}^{n} \left(t_i - 1 - t_i\alpha^2 + \alpha^{2t_i}\right)}{\left[\sum_{i=1}^{n}(t_i-1)\right]^2} .$$

The unbiased estimating equations based on Cholesky decomposition of $\mathbf{R}_i^{-1}$ can be easily extended to the unbalanced case. They are given by

$$
\begin{aligned}
U_1: \quad & \sum_{i=1}^{n} \operatorname{tr}\left\{\left(\frac{\partial \mathbf{B}_{li}}{\partial \alpha}\boldsymbol{\Lambda}_{li}\mathbf{B}_{li}' + \frac{\partial \mathbf{B}_{ui}}{\partial \alpha}\boldsymbol{\Lambda}_{ui}\mathbf{B}_{ui}'\right)\mathbf{z}_i\mathbf{z}_i'\right\} = 0 \\
U_2: \quad & \sum_{i=1}^{n} \operatorname{tr}\left\{\left(\frac{\partial \mathbf{B}_{li}}{\partial \alpha}\mathbf{R}^{-1} + \frac{\partial \mathbf{B}_{ui}}{\partial \alpha}\mathbf{R}^{-1}\right)\mathbf{z}_i\mathbf{z}_i'\right\} = 0 \\
U_4: \quad & \sum_{i=1}^{n} \operatorname{tr}\left\{\left(\frac{\partial \mathbf{B}_{li}}{\partial \alpha}\frac{\partial \boldsymbol{\Lambda}_{li}}{\partial \alpha}\boldsymbol{\Lambda}_{li}\mathbf{B}_{li}' + \frac{\partial \mathbf{B}_{ui}}{\partial \alpha}\frac{\partial \boldsymbol{\Lambda}_{ui}}{\partial \alpha}\boldsymbol{\Lambda}_{ui}\mathbf{B}_{ui}'\right)\mathbf{z}_i\mathbf{z}_i'\right\} = 0.
\end{aligned}
$$

Clearly, they fall into the subclass of unbiased estimating equations with different weights. Unfortunately, none of them are optimal when $\mathbf{R}_i$ has an exchangeable or an AR(1) structure.

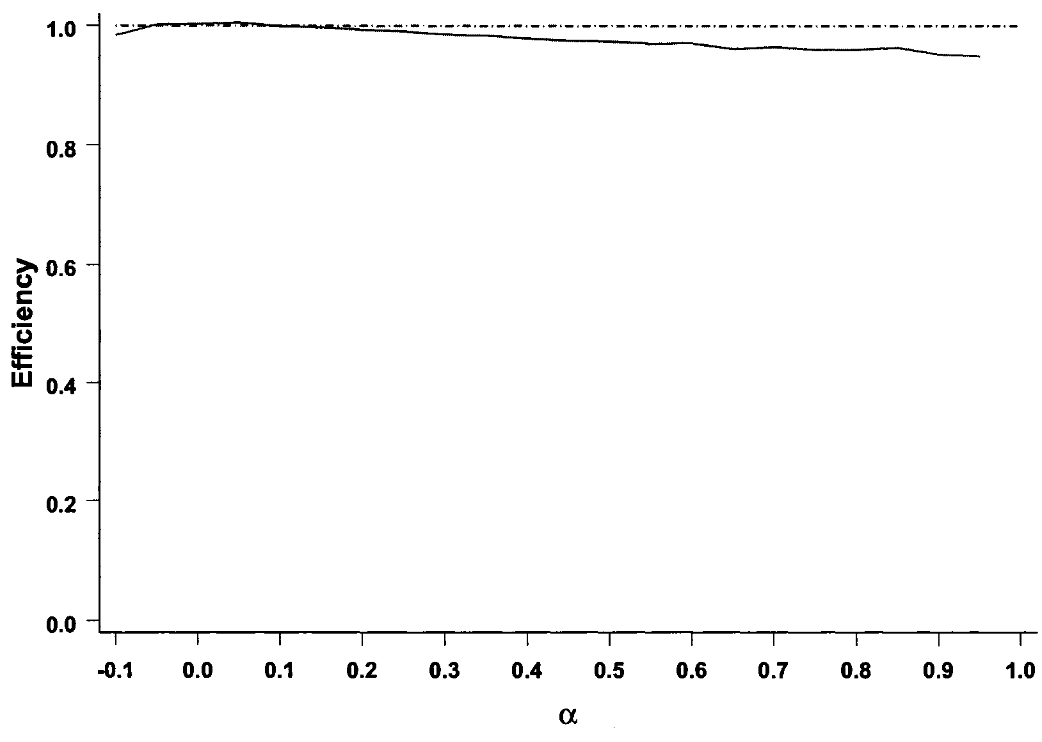


Figure 3.1: EXCH: Efficiency of optimal $\hat{\alpha}$ (n=30).

III.3 Relative Efficiency

In this section we study the small sample relative efficiencies of the estimates obtained using the unbiased estimating equations in the subclass (3.2.5) with different weights using simulations. First, we look at the efficiencies for multivariate normal data and then we explore the case when there is a departure from normality. Since correlation is scale invariant in our simulations we fixed the variance to be 1. The simulation steps are as follows.

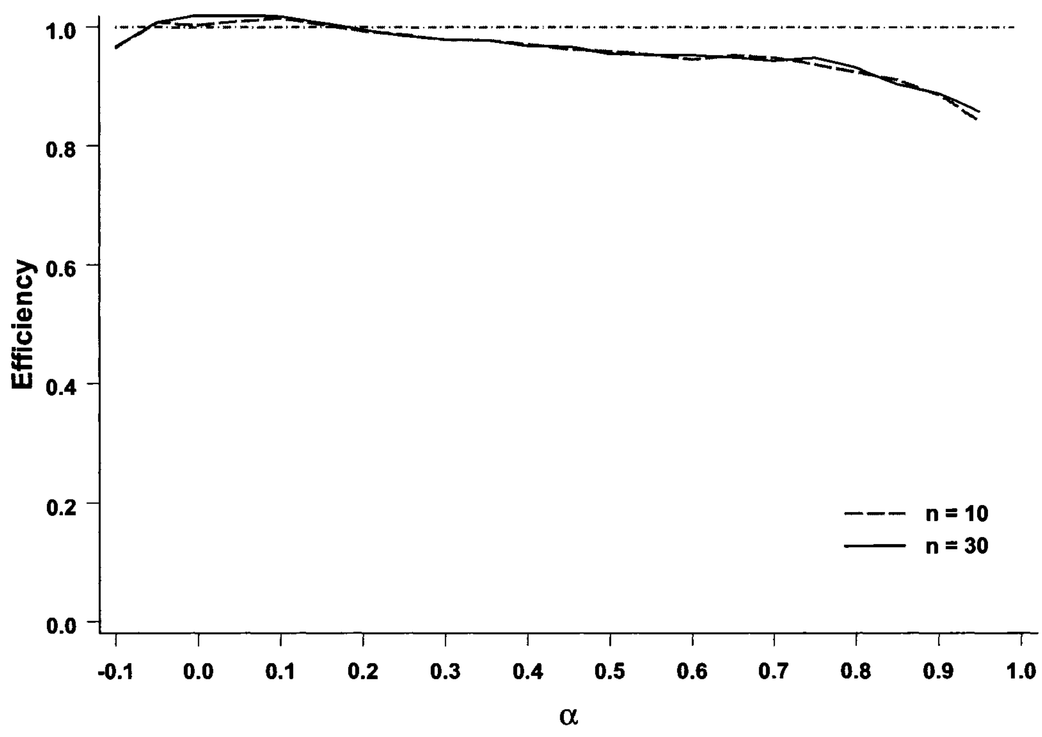


Figure 3.2: EXCH: Efficiencies of $\hat{\alpha}$ for small and large sample size.

(1) Generate n integers $\{t_i, 1 \leq i \leq n\}$ for the number of repeated measurements (or sizes of the clusters) ranging from 2 to 6 from the discrete uniform distribution.

(2) For $i = 1, 2, \ldots, n$, generate t_i univariate standard normal random numbers and stack them into a column vector $\mathbf{z}_i$.

(3) For a fixed value of α, construct the correlation matrix $\mathbf{R}_i$ (exchangeable or AR(1)).

(4) Let $\mathbf{y}_i = \mathbf{R}_i^{\frac{1}{2}}\mathbf{z}_i$, where $\mathbf{R}_i^{\frac{1}{2}}$ is the square root of $\mathbf{R}_i$. Clearly $\mathrm{E}(\mathbf{y}_i) = 0$ and $\mathrm{Cov}(\mathbf{y}_i) = \mathbf{R}_i$.

(5) Solve equation (3.2.5) with optimal weight $\mathbf{W}_i = \dfrac{1}{1+(t_i-1)\alpha}(\mathbf{J}_i - \mathbf{I}_i)$ and simplified weight $\mathbf{W}_i = \mathbf{R}_i - \mathbf{I}_i$ (or equivalently, solve equation (3.2.9)) to get estimate $\hat{\alpha}_S$ of α for exchangeable correlation structure. Similarly, use the weight $\mathbf{W}_i = \mathbf{R}_i - \mathbf{I}_i$ in (3.2.5) (or equivalently, solve (3.2.10)) to get $\hat{\alpha}_S$ for the AR(1) structure.

(6) Solve (3.1.1) to get the ML estimate $\hat{\alpha}_L$ of α.

(7) Calculate error squares $\tau_S^2 = (\hat{\alpha}_S - \alpha)^2$ and $\tau_L^2 = (\hat{\alpha}_L - \alpha)^2$.

(8) Repeat steps (1) – (7) a large number of times (r), say $r = 10,000$ times, and calculate $\sum \tau_S^2$ and $\sum \tau_L^2$, the sum of error squares for the estimates.

Then the relative efficiency of $\hat{\alpha}_S$ to $\hat{\alpha}_L$ is given by

$$\mathrm{Eff}(\hat{\alpha}_S, \hat{\alpha}_L) = \frac{\sum \tau_L^2}{\sum \tau_S^2}.$$

Table 3.1: Efficiency of α for various sample sizes

α	$t=4$	$t=6$	$t=9$	$t=15$	$t=20$
0.0	1.0103	1.0032	1.0021	1.0010	1.0003
0.1	1.0079	1.0033	1.0008	1.0006	0.9997
0.2	1.0037	0.9991	0.9968	0.9985	0.9981
0.3	0.9940	0.9918	0.9922	0.9947	0.9959
0.4	0.9820	0.9800	0.9858	0.9890	0.9905
0.5	0.9614	0.9654	0.9743	0.9829	0.9873
0.6	0.9533	0.9572	0.9558	0.9750	0.9813
0.7	0.9269	0.9270	0.9376	0.9448	0.9601
0.8	0.9236	0.9073	0.9129	0.9248	0.9312
0.9	0.9387	0.9073	0.8895	0.8947	0.8885

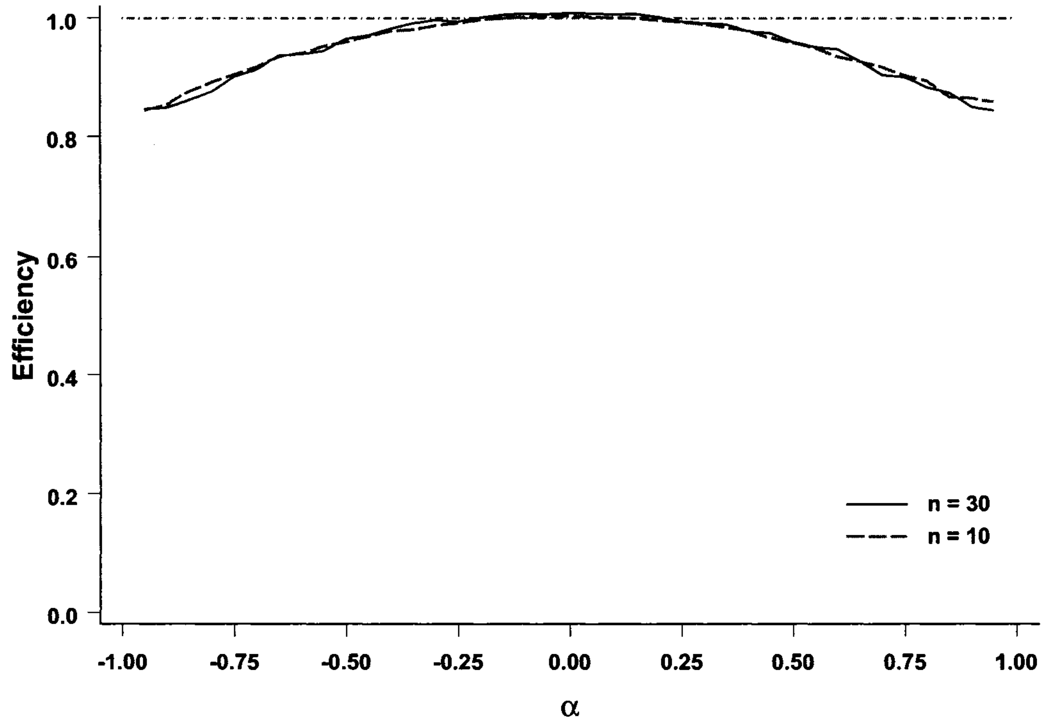


Figure 3.3: AR(1): Efficiencies of $\hat{\alpha}$ for small and large sample size.

Figure 3.1 shows the relative efficiency of $\hat{\alpha}_S$ obtained using weight $\mathbf{W}_i = \frac{1}{1+(t_i-1)\alpha}(\mathbf{J}_i - \mathbf{I}_i)$ with respect to the ML estimate, when the correlation structure is exchangeable, for $n = 30$. It is clear that the unbiased estimating equation approach with optimal weights is nearly as good as the ML estimate.

Figure 3.2 shows the relative efficiencies of the estimate $\hat{\alpha}_S$ using $\mathbf{W}_i = \mathbf{R}_i - \mathbf{I}_i$ and an exchangeable correlation matrix, for n=10, 30. The plot of the relative efficiencies when the correlation structure is AR(1) is in Figure 3.3. It is clear from these plots the unbiased estimating approach yields highly efficient estimates for normal data.

For balanced data, the estimates obtained using the unbiased equations from the subclass is identical to ML estimate for exchangeable structure and is also as good as the ML estimate for AR(1) structure. The plot of the efficiency for AR(1) structure is in Figure 3.4. Some numerical results are also presented in Table 3.1.

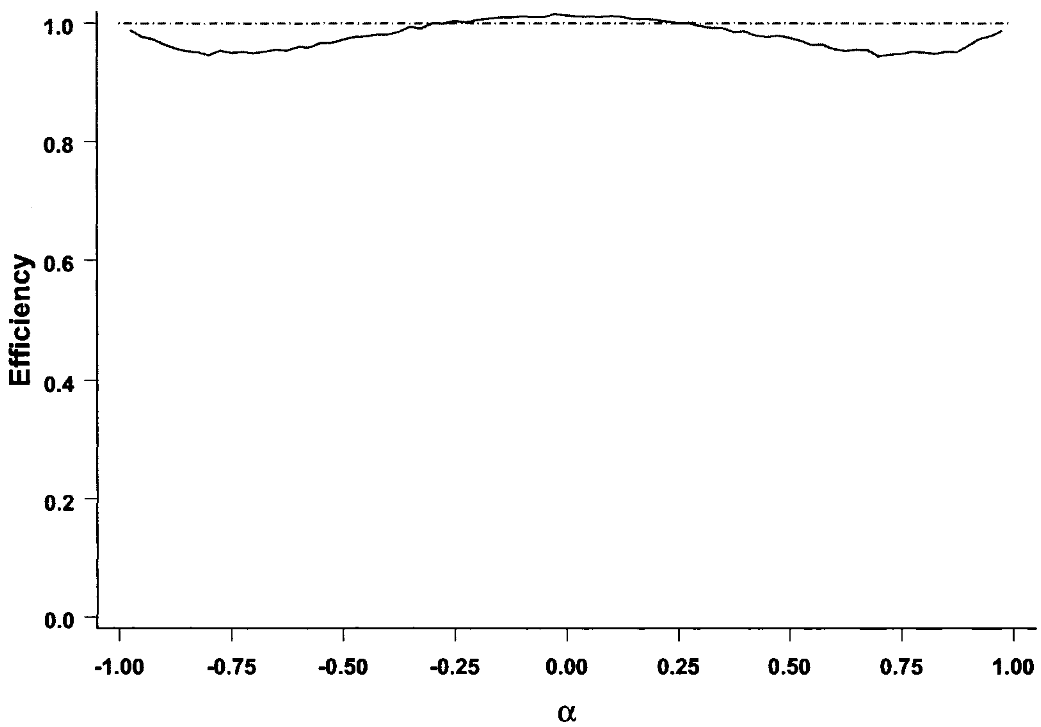


Figure 3.4: AR(1): ARE of $\hat{\alpha}$ for balanced data ($t = 3$).

To study the robustness of the estimates, we simulated data from a multivariate t distribution (Johnson et al., 1972), which was used by many authors as a model to study the departure from normality. To simulate random numbers from the multivariate t distribution, we first generated random normal variables $\mathbf{z}_i$ according to the procedure described above. Then we generated random numbers (λ_i) from a Chi-square (χ^2) distribution with 5 degrees of freedom. The ratios $\frac{\mathbf{z}_i}{\sqrt{\lambda_i/5}}$ are the desired multivariate t random variables. Figures 3.5 and 3.6 show the asymptotic relative efficiencies of the estimates $\widehat{\alpha}_S$, when the correlation matrix is exchangeable and AR(1), respectively. It is clear from the plots the estimates remain efficient when there is a departure from normality.

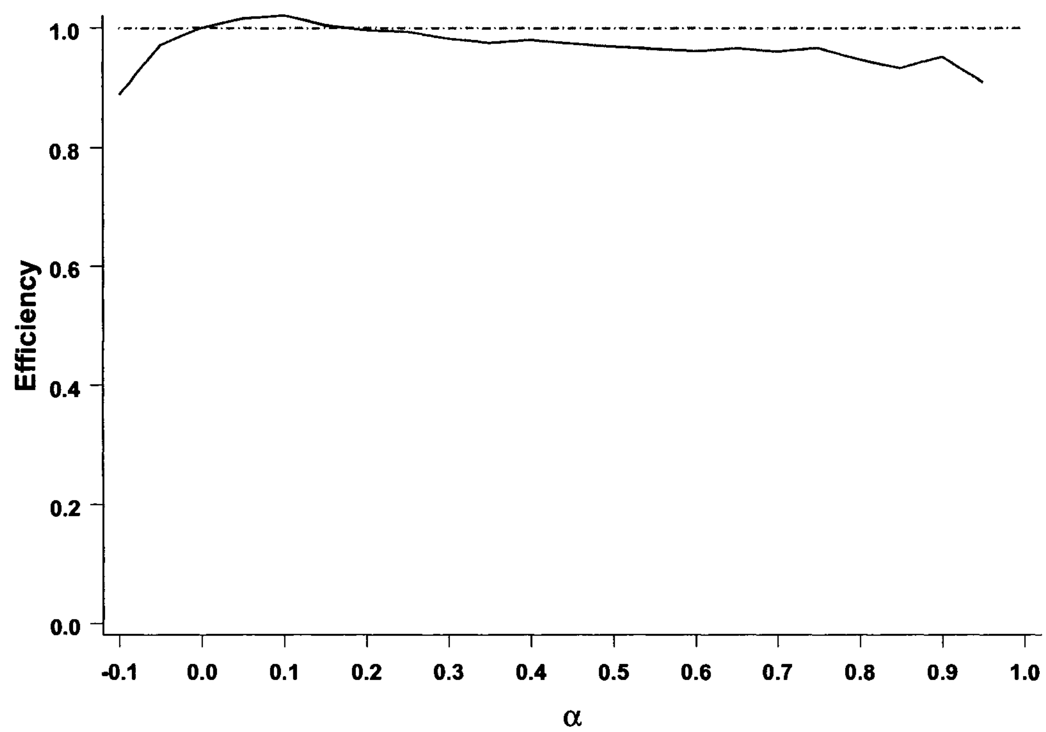


Figure 3.5: EXCH: Efficiency of $\hat{\alpha}$ for non-normal responses.

III.4 An Illustrative Example

Here we will illustrate our new methods of estimation with a real life example. The data is from the AIDS Clinical Trial Group (ACTG) Study 193A (Fitzmaurice et al., 2004). This is a randomized, double-blind, study of AIDS patients with advanced immune suppression (CD4 counts of less than or equal to 50 cells/mm^3). The patients in this study were assigned to dual or triple combinations of HIV-1 reverse transcriptase inhibitors. Specifically, patients were randomized to receive one of four daily regimens containing 600mg of zidovudine: zidovudine alternating monthly with 400mg didanosine; zidovudine plus 2.25mg of zalcitabine; zidovudine plus 400mg of didanosine; or zidovudine plus 400mg of didanosine plus 400mg of nevirapine (triple therapy). Measurements of CD4 counts were collected at baseline and at 8-week intervals during follow-up. However, the CD4 count data are unbalanced due to mistimed measurements and missing data that resulted from skipped visits and dropouts. The

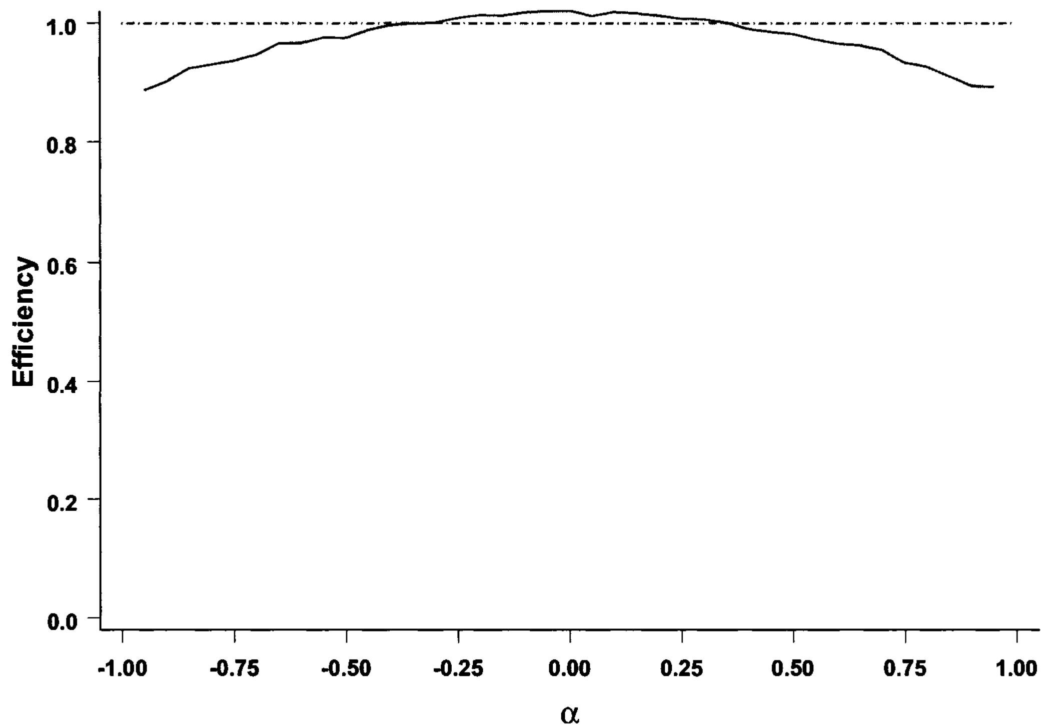


Figure 3.6: AR(1): Efficiency of $\hat{\alpha}$ for non-normal responses.

number of measurements of CD4 counts during the first 40 weeks of follow-up varied from 1 to 9, with a median of 4. The response variable is the log transformed CD4 counts, log(CD4 counts + 1), available on 1309 patients. The categorical variable Treatment is coded as 1 = zidovudine alternating monthly with 400mg didanosine, 2 = zidovudine plus 2.25mg of zalcitabine, 3 = zidovudine plus 400mg of didanosine, and 4 = zidovudine plus 400mg of didanosine plus 400mg of nevirapine. The variable week represents time since baseline (in weeks). Table 3.2 shows an abbreviated version of the data set.

The regression and the correlation parameter estimates obtained using the unbiased estimating equation U_S and the ML estimates are in Table 3.3 for the exchangeable correlation structure. The estimates and the standard errors are very similar.

Parallel results for the AR(1) correlation structure are in Table 3.4. Once again,

Table 3.2: Partial list of AIDS data

Subject	Treatment	Age (years)	Gender (1 = M, 0 = F)	Week	log(CD4 + 1)
1	2	36.43	1	0.00	3.135
1	2	36.43	1	7.57	3.045
1	2	36.43	1	15.57	2.773
1	2	36.43	1	23.57	2.833
1	2	36.43	1	32.57	3.219
1	2	36.43	1	40.00	3.045
2	4	47.85	1	0.00	3.068
2	4	47.85	1	8.00	3.892
2	4	47.85	1	16.00	3.970
2	4	47.85	1	23.00	3.611
2	4	47.85	1	30.71	3.332
2	4	47.85	1	39.00	3.091
3	1	60.29	1	0.00	3.738
4	3	36.60	1	0.00	4.119
4	3	36.60	1	7.14	4.111
4	3	36.60	1	16.14	4.710
4	3	36.60	1	32.43	2.833
⋮	⋮	⋮	⋮	⋮	⋮
1313	1	15.84	0	0.00	4.984
1313	1	15.84	0	7.29	4.159
1313	1	15.84	0	20.00	4.407
1313	1	15.84	0	27.00	3.556
1313	1	15.84	0	35.00	3.466

NOTES: Source: Fitzmaurice et al. (2004), *Applied longitudinal analysis.*

the estimates and the standard errors are in agreement. This example shows that the unbiased estimating approach, which is simpler to implement, is a great alternative to the ML estimation approach.

Table 3.3: Parameter estimates with EXCH correlation structure

	U_S (S.E.)		MLE (S.E.)	
Intercept	2.3326	(0.1469)	2.3325	(0.1453)
Treatment	0.0760	(0.0228)	0.0761	(0.0229)
Age	0.0116	(0.0031)	0.0116	(0.0031)
Gender	-0.1175	(0.0765)	-0.1175	(0.0788)
Scale	1.0689	—	1.0665	—
$\hat{\alpha}$	0.6412	(0.01299)	0.6388	(0.01230)

Table 3.4: Parameter estimates with AR(1) correlation structure

	U_S (S.E.)		MLE (S.E.)	
Intercept	2.3737	(0.1437)	2.3721	(0.1351)
Treatment	0.0620	(0.0223)	0.0628	(0.0212)
Age	0.0109	(0.00306)	0.0109	(0.00291)
Gender	-0.1330	(0.0742)	-0.1328	(0.0738)
Scale	1.0702	—	1.0568	—
$\hat{\alpha}$	0.7031	(0.01002)	0.6906	(0.00976)

CHAPTER IV

ANALYSIS OF FAMILIAL STRUCTURE

The correlation structures that we have studied in the previous chapters contain only a single parameter. In this chapter we consider structures that are characterized by more than one parameter. More specifically we focus our attention on a structure that has been widely used to study the inter-relationships in familial data, that is, data collected on families including parents and children. Numerous authors have studied the analysis of familial data and proposed several methods of estimation, which are either moment based or likelihood. For example, Donner and Koval (1980) studied likelihood estimation of intra-class correlation. Srivastava (1984) discussed likelihood estimation of inter-class correlation using transformation. Eliasziw and Donner (1990) compared different methods for inter-class correlation estimation. Srivastava et al. (1988) extended the work in Srivastava (1984) to the simultaneous estimation of intra- and inter-relationships, and Konishi et al. (1991) addressed the inferences on the correlations between different family members. However, there is no literature on parameter estimation through estimating equations approach and there is almost no discussion about the optimality of the correlation estimates.

The organization of this chapter is as follows. We first study properties of the familial correlation structure in Section IV.1. In Section IV.2 we discuss maximum likelihood estimation of the familial correlations for normal data. We derive the asymptotic covariance matrix of the ML estimates as well. As an alternative approach to estimation of the familial correlations, we present a general class of unbiased estimating equations and a useful subclass in Section IV.3. We study asymptotic properties of the estimates obtained solving those unbiased estimating equations. Since the optimal weights which minimize the asymptotic variances, are not in a simple form, we suggest some simpler weights that are nearly optimal. Expressions for the asymptotic covariance matrices for the near optimal weights are also given in Section IV.3. Simulation results to compare relative efficiencies under the normality assumption are presented in Section IV.4. Finally, results from a real life data analysis are given in Section IV.5.

IV.1 Familial Correlation Structure

A correlation model that has been widely used to model associations within families is the structure

$$\mathbf{R}_i = \begin{bmatrix} 1 & \rho \mathbf{1}_i' \\ \rho \mathbf{1}_i & (1-\alpha)\mathbf{I}_i + \alpha \mathbf{J}_i \end{bmatrix} = \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \alpha & \dots & \alpha \\ \rho & \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \alpha & \alpha & \dots & 1 \end{bmatrix}. \qquad (4.1.1)$$

Here ρ is the correlation between the mother (or parent) and her children, and α is the common correlation among t_i children.

To find the necessary and sufficient conditions for $\mathbf{R}_i$ to be positive definite, let us consider the Helmert matrix

$$\mathbf{M}_i = \begin{bmatrix} \frac{1}{\sqrt{t_i}} & \frac{1}{\sqrt{t_i}} & \frac{1}{\sqrt{t_i}} & \dots & \frac{1}{\sqrt{t_i}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \dots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{t_i(t_i-1)}} & \frac{1}{\sqrt{t_i(t_i-1)}} & \frac{1}{\sqrt{t_i(t_i-1)}} & \dots & \frac{-(t_i-1)}{\sqrt{t_i(t_i-1)}} \end{bmatrix}$$

of dimension t_i. It is easy to verify that $\mathbf{M}_i$ is an orthogonal matrix, that is, $\mathbf{M}_i\mathbf{M}_i' = \mathbf{I}_i$. Let

$$\begin{aligned} \boldsymbol{\Omega}_i &= \begin{bmatrix} 1 & \mathbf{0}_i' \\ \mathbf{0}_i & \mathbf{M}_i \end{bmatrix} \mathbf{R}_i \begin{bmatrix} 1 & \mathbf{0}_i' \\ \mathbf{0}_i & \mathbf{M}_i' \end{bmatrix} \\ &= \begin{bmatrix} 1 & \sqrt{t_i}\rho & 0 & \dots & 0 \\ \sqrt{t_i}\rho & 1+(t_i-1)\alpha & 0 & \dots & 0 \\ 0 & 0 & 1-\alpha & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1-\alpha \end{bmatrix}. \end{aligned}$$

Since $\mathbf{R}_i$ is positive definite if and only if $\boldsymbol{\Omega}_i$ is positive definite, the necessary and sufficient conditions for positive definiteness are

(1) $-\frac{1}{t_i-1} < \alpha < 1$,

(2) $t_i\rho^2 < 1 + (t_i - 1)\alpha$.

For example, when $t_i = 2$, the conditions become

$$(1)\ -1 < \alpha < 1 \qquad (2)\ -\sqrt{\frac{1+\alpha}{2}} < \rho < \sqrt{\frac{1+\alpha}{2}}\,, \tag{4.1.2}$$

and when $t_i = 3$, the conditions reduce to

$$(1)\ -\frac{1}{2} < \alpha < 1 \qquad (2)\ -\sqrt{\frac{1+2\alpha}{3}} < \rho < \sqrt{\frac{1+2\alpha}{3}}\,. \tag{4.1.3}$$

These ranges are shown in Figure 4.1. The area enclosed by the outer curve is the feasible region for $t = 2$ and the inner curve encloses the feasible region when $t = 3$. In general, the feasible range becomes narrower as t increases.

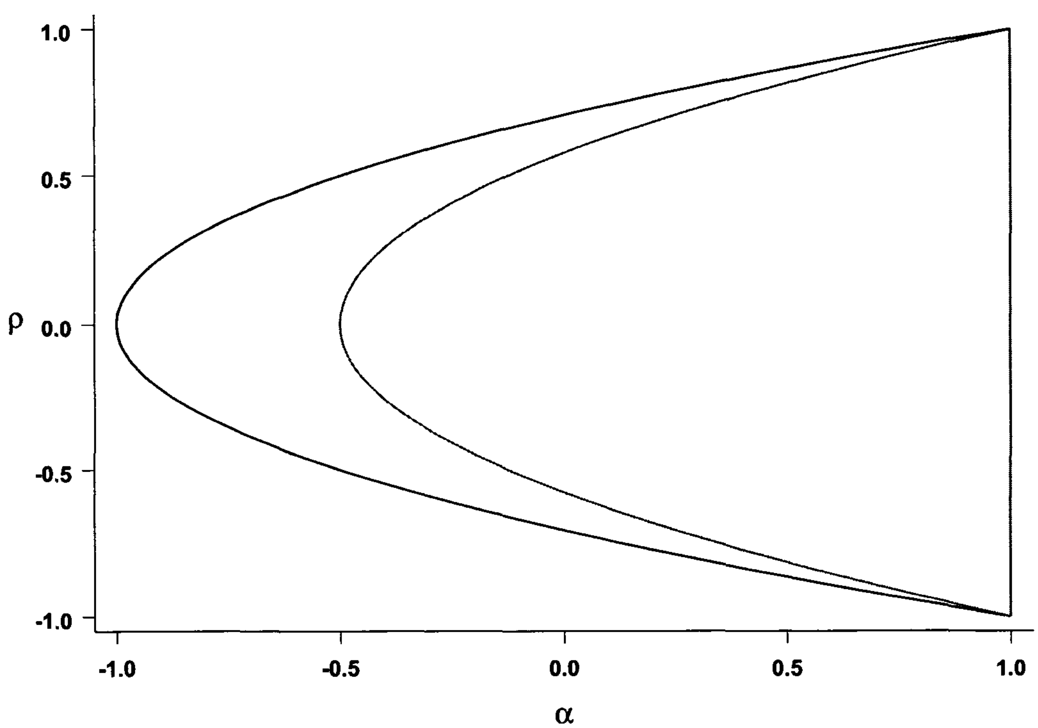


Figure 4.1: Range of (ρ, α) when $t = 2$ and $t = 3$.

The inverse of the familial correlation matrix is given by

$$\mathbf{R}_i^{-1} = \begin{bmatrix} \dfrac{1+(t_i-1)\alpha}{1+(t_i-1)\alpha-t_i\rho^2} & \dfrac{-\rho}{1+(t_i-1)\alpha-t_i\rho^2}\mathbf{1}_i' \\ \dfrac{-\rho}{1+(t_i-1)\alpha-t_i\rho^2}\mathbf{1}_i & \dfrac{1}{1-\alpha}\mathbf{I}_i - \dfrac{\alpha-\rho^2}{(1-\alpha)[1+(t_i-1)\alpha-t_i\rho^2]}\mathbf{J}_i \end{bmatrix}.$$

Appendix A.3 contains details of the Cholesky decompositions of $\mathbf{R}_i$ and $\mathbf{R}_i^{-1}$. Let σ_1^2 be the marginal variance of observation on the parent and σ_0^2 be the common variance of observations on the children. If $\mathbf{y}_i$ is a vector consisting of observations on the mother and her children, then the covariance matrix of $\mathbf{y}_i$ is

$$\begin{aligned} \mathbf{V}_i &= \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_0\rho & \dots & \sigma_1\sigma_0\rho \\ \sigma_1\sigma_0\rho & \sigma_0^2 & \dots & \sigma_1\sigma_0\alpha \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_0\rho & \sigma_1\sigma_0\alpha & \dots & \sigma_0^2 \end{bmatrix} \\ &= \mathbf{A}_i^{\frac{1}{2}}\mathbf{R}_i\mathbf{A}_i^{\frac{1}{2}} \end{aligned}$$

where $\mathbf{A}_i = \mathrm{diag}(\sigma_1^2, \sigma_0^2, \dots, \sigma_0^2)$.

IV.2 Maximum Likelihood Estimate

Recall that the ML estimates of the correlation parameters can be obtained by solving equation (1.3.6), which is same as

$$\sum_{i=1}^{n} \mathrm{tr}\left\{\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\alpha}}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' - \mathbf{V}_i)\right\} = \mathbf{0}.$$

Using the identity

$$\frac{\partial \mathbf{V}_i^{-1}}{\partial \boldsymbol{\alpha}} = -\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\alpha}}\mathbf{V}_i^{-1},$$

we can rewrite the ML equation for $\boldsymbol{\alpha}$ as

$$\sum_{i=1}^{n} \mathrm{tr}\left\{\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\alpha}}\mathbf{V}_i^{-1}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' - \mathbf{V}_i)\right\} = \mathbf{0}.$$

For the familial covariance matrix we have $\boldsymbol{\alpha} = (\rho, \alpha)'$ and the above is equivalent to

$$\sum_{i=1}^{n} \mathrm{tr}\left\{\mathbf{R}_i^{-1}\frac{\partial \mathbf{R}_i}{\partial \boldsymbol{\alpha}}\mathbf{R}_i^{-1}(\mathbf{z}_i\mathbf{z}_i' - \mathbf{R}_i)\right\} = \mathbf{0}, \qquad (4.2.1)$$

where $\mathbf{z}_i$ is the Pearson residual and $\mathbf{R}_i$ is the familial correlation matrix. The ML equation for $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_0^2)'$ given in (1.3.4) can be simplified as

$$\sum_{i=1}^{n} \mathrm{tr}\left\{\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\sigma}^2}\mathbf{V}_i^{-1}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' - \mathbf{V}_i\right)\right\} = \mathbf{0}. \qquad (4.2.2)$$

The elements of the Fisher information matrix for the parameters $\boldsymbol{\alpha} = (\rho, \alpha)'$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_0^2)'$ are

$$\begin{aligned}
-\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial\left(\boldsymbol{\sigma}^2\right)^2}\right\} &= \sum_{i=1}^{n} \mathrm{tr}\left\{\frac{\partial^2 \mathbf{A}_i^{-\frac{1}{2}}}{\partial\left(\boldsymbol{\sigma}^2\right)^2}\mathbf{A}_i^{\frac{1}{2}} + \frac{\partial \mathbf{A}_i^{-\frac{1}{2}}}{\partial \boldsymbol{\sigma}^2}\mathbf{R}_i^{-1}\frac{\partial \mathbf{A}_i^{-\frac{1}{2}}}{\partial \boldsymbol{\sigma}^2}\mathbf{A}_i^{\frac{1}{2}}\mathbf{R}_i\mathbf{A}^{\frac{1}{2}}\right\} \\
&\quad + \frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2 \log\left(\sigma_1^2\sigma_0^{2t_i}\right)}{\partial \boldsymbol{\sigma}^2} \\
-\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\sigma}^2 \partial \boldsymbol{\alpha}}\right\} &= \sum_{i=1}^{n} \mathrm{tr}\left\{\frac{\partial \mathbf{A}_i^{-\frac{1}{2}}}{\partial \boldsymbol{\sigma}^2}\frac{\partial \mathbf{R}_i^{-1}}{\partial \boldsymbol{\alpha}}\mathbf{R}_i\mathbf{A}^{\frac{1}{2}}\right\} \\
-\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\alpha}^2}\right\} &= \frac{1}{2}\sum_{i=1}^{n}\left\{\mathrm{tr}\left(\frac{\partial^2 \mathbf{R}_i^{-1}}{\partial \boldsymbol{\alpha}^2}\mathbf{R}_i\right) + \frac{\partial^2 \log|\mathbf{R}_i|}{\partial \boldsymbol{\alpha}^2}\right\}.
\end{aligned}$$

The asymptotic covariance matrix of the estimates, obtained by solving (4.2.2) and (4.2.1), is the inverse of

$$\begin{bmatrix}
-\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial\left(\boldsymbol{\sigma}^2\right)^2}\right\} & -\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\sigma}^2 \partial \boldsymbol{\alpha}}\right\} \\
-\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\sigma}^2 \partial \boldsymbol{\alpha}}\right\} & -\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\alpha}^2}\right\}
\end{bmatrix}.$$

Notice that equation (4.2.1) can also be expressed as

$$\sum_{i=1}^{n}\left\{\frac{\partial \mathrm{vech}(\mathbf{R}_i)}{\partial \boldsymbol{\alpha}}\right\}' \mathbf{G}_i'(\mathbf{R}_i^{-1} \otimes \mathbf{R}_i^{-1})\mathbf{G}_i\left\{\mathrm{vech}(\mathbf{z}_i\mathbf{z}_i') - \mathrm{vech}(\mathbf{R}_i)\right\} = 0,$$

and a similar argument as in Section III.1 proves that ML equation for $\boldsymbol{\alpha}$ is Godambe optimal.

IV.3 Classes of Unbiased Estimating Equations

Suppose the correlation parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)$ is multidimensional. In this situation we could consider a more general class of estimating equations

$$\sum_{i=1}^{n} \mathrm{tr}\left\{\mathbf{W}_{ji}\mathbf{R}_i^{-1}\left(\mathbf{z}_i\mathbf{z}_i' - \mathbf{R}_i\right)\right\} = 0, \qquad (4.3.1)$$

where $\mathbf{W}_{ji}$ is some weighting matrix in estimating the jth component α_j of $\boldsymbol{\alpha}$. For the familial correlation, we have $\alpha_1 = \rho$ and $\alpha_2 = \alpha$. The two equations in (4.3.1) are

$$\begin{aligned}
\sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\left(\mathbf{z}_i\mathbf{z}_i' - \mathbf{R}_i\right)\right\} &= 0, \\
\sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\left(\mathbf{z}_i\mathbf{z}_i' - \mathbf{R}_i\right)\right\} &= 0,
\end{aligned}$$

where $\mathbf{W}_{\rho i} \equiv \mathbf{W}_{1i}$ and $\mathbf{W}_{\alpha i} \equiv \mathbf{W}_{2i}$ are the weighting matrices for estimating ρ and α, respectively. Suppose $\widehat{\boldsymbol{\alpha}} = (\widehat{\rho}, \widehat{\alpha})'$ is the solution of the above equations. From Lemma 2.1 and Theorem 2.1, it follows that the asymptotic covariance matrix of $\widehat{\boldsymbol{\alpha}}$ is $\mathbf{I}_0^{-1}\mathbf{I}_1\mathbf{I}_0'^{-1}$ where

$$\begin{aligned}
\mathbf{I}_0 &= \begin{bmatrix}
\sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\} & \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\} \\
\sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\} & \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\}
\end{bmatrix} \\
\mathbf{I}_1 &= \begin{bmatrix}
\sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i + \mathbf{W}_{\rho i}^2\right\} & \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i + \mathbf{W}_{\rho i}\mathbf{W}_{\alpha i}\right\} \\
\sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i + \mathbf{W}_{\alpha i}\mathbf{W}_{\rho i}\right\} & \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i + \mathbf{W}_{\alpha i}^2\right\}
\end{bmatrix}.
\end{aligned}$$

To find the optimal weights, we may need to minimize the covariance matrix in some sense. One criteria is to use the determinant of the matrix

$$\left|\mathbf{I}_0^{-1}\mathbf{I}_1\mathbf{I}_0'^{-1}\right| = \left|\mathbf{I}_0^{-1}\right| \cdot \left|\mathbf{I}_1\right| \cdot \left|\mathbf{I}_0'^{-1}\right| = \frac{|\mathbf{I}_1|}{|\mathbf{I}_0|^2} = \frac{\xi_1}{\xi_0^2}$$

where

$$\begin{aligned}
\xi_1 &= \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i + \mathbf{W}_{\rho i}^2\right\} \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i + \mathbf{W}_{\alpha i}^2\right\} \\
&\quad - \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i + \mathbf{W}_{\alpha i}\mathbf{W}_{\rho i}\right\} \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i + \mathbf{W}_{\rho i}\mathbf{W}_{\alpha i}\right\} \\
\xi_0 &= \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\} \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\} \\
&\quad - \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\} \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\}.
\end{aligned}$$

Equating the derivative of the determinant with respect to $\mathbf{W}_{\rho i}$ to zero, and using the identities (Magnus and Neudecker, 1999)

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{W}_{\rho i}}\left\{\operatorname{tr}\left(\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i\right)\right\} &= 2\,\mathbf{R}_i\mathbf{W}_{\rho i}\mathbf{R}_i^{-1} \\
\frac{\partial}{\partial \mathbf{W}_{\rho i}}\left\{\operatorname{tr}\left(\mathbf{W}_{\rho i}^2\right)\right\} &= 2\,\mathbf{W}_{\rho i}' \\
\frac{\partial}{\partial \mathbf{W}_{\rho i}}\left\{\operatorname{tr}\left(\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i\right)\right\} &= \mathbf{R}_i\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1} \\
\frac{\partial}{\partial \mathbf{W}_{\rho i}}\left\{\operatorname{tr}\left(\mathbf{W}_{\rho i}\mathbf{W}_{\alpha i}\right)\right\} &= \mathbf{W}_{\alpha i}' \\
\frac{\partial}{\partial \mathbf{W}_{\rho i}}\left\{\operatorname{tr}\left(\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right)\right\} &= \mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho} \\
\frac{\partial}{\partial \mathbf{W}_{\rho i}}\left\{\operatorname{tr}\left(\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right)\right\} &= \mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i\right\} &= \operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i\right\} \\
\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{W}_{\rho i}\right\} &= \operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{W}_{\alpha i}\right\},
\end{aligned}$$

we have for $i = 1, 2, \ldots, n$,

$$\begin{aligned}
&\left\{\mathbf{R}_i\mathbf{W}_{\rho i}\mathbf{R}_i^{-1} + \mathbf{W}_{\rho i}'\right\}\xi_0\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i + \mathbf{W}_{\alpha i}^2\right\} \\
&\quad - \left\{\mathbf{R}_i\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1} + \mathbf{W}_{\alpha i}'\right\}\xi_0\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i + \mathbf{W}_{\rho i}\mathbf{W}_{\alpha i}\right\} \\
&\quad - \left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\right\}\xi_1\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\} \\
&\quad + \left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\right\}\xi_1\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\} = 0. \qquad (4.3.2)
\end{aligned}$$

Similarly, differentiate the determinant with respect to $\mathbf{W}_{\alpha i}$ and equating to zero, we have

$$\begin{aligned}
&\left\{\mathbf{R}_i\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1} + \mathbf{W}_{\alpha i}'\right\}\xi_0\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i + \mathbf{W}_{\rho i}^2\right\} \\
&\quad - \left\{\mathbf{R}_i\mathbf{W}_{\rho i}\mathbf{R}_i^{-1} + \mathbf{W}_{\rho i}'\right\}\xi_0\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i + \mathbf{W}_{\rho i}\mathbf{W}_{\alpha i}\right\} \\
&\quad - \left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\right\}\xi_1\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\}
\end{aligned}$$

$$+\left\{\mathbf{R}_i\frac{\partial\mathbf{R}_i^{-1}}{\partial\rho}\right\}\xi_1\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right\} = 0. \tag{4.3.3}$$

Here ξ_0, ξ_1 and all the traces are scalars. Looking at the component matrices in (4.3.2) and (4.3.3), we can see that a solution is given by

$$\mathbf{W}_{\rho i} = \mathbf{R}_i^{-1}\frac{\partial\mathbf{R}_i}{\partial\rho} \quad\text{and}\quad \mathbf{W}_{\alpha i} = \mathbf{R}_i^{-1}\frac{\partial\mathbf{R}_i}{\partial\alpha}.$$

Thus the optimal weights are identical to those of ML equations in this case.

Another criterion that we could use to determine the optimal weights is to minimize the trace of the covariance matrix. The trace of $\mathbf{I}_0^{-1}\mathbf{I}_1\mathbf{I}_0'^{-1}$ is given by ξ_2/ξ_0^2, where

$$\begin{aligned}
\xi_2 = &\left(\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right\}\right)^2\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i+\mathbf{W}_{\rho i}^2\right\}\\
&+\left(\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\rho}\mathbf{R}_i\right\}\right)^2\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i+\mathbf{W}_{\rho i}^2\right\}\\
&+\left(\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right\}\right)^2\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i+\mathbf{W}_{\alpha i}^2\right\}\\
&+\left(\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\rho}\mathbf{R}_i\right\}\right)^2\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i+\mathbf{W}_{\alpha i}^2\right\}\\
&-2\left(\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right\}\right)\left(\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right\}\right)\sum_{i=1}^{n}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i+\mathbf{W}_{\alpha i}\mathbf{W}_{\rho i}\right\}\\
&-2\left(\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\rho}\mathbf{R}_i\right\}\right)\left(\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\rho}\mathbf{R}_i\right\}\right)\sum_{i=1}^{n}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i+\mathbf{W}_{\alpha i}\mathbf{W}_{\rho i}\right\}
\end{aligned}$$

and ξ_0 is defined earlier before. Once again equating the partial derivative with respect to $\mathbf{W}_{\rho i}$ to zero, we get

$$\begin{aligned}
&\left\{\mathbf{R}_i\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}+\mathbf{W}_{\rho i}'\right\}c_1+\left\{\mathbf{R}_i\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}+\mathbf{W}_{\alpha i}'\right\}c_2\\
&\qquad+\left\{\mathbf{R}_i\frac{\partial\mathbf{R}_i^{-1}}{\partial\rho}\right\}c_3+\left\{\mathbf{R}_i\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\right\}c_4 = 0,
\end{aligned} \tag{4.3.4}$$

where the constants are

$$c_1 = \left(\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right\}\right)^2+\left(\sum_{i=1}^{n}\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\rho}\mathbf{R}_i\right\}\right)^2$$

$$\begin{aligned}
c_2 &= -\left(\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\}\right)\left(\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\}\right) \\
&\quad -\left(\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\}\right)\left(\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\}\right) \\
c_3 &= \left(\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\}\right)\left(\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}'_{\alpha i}\mathbf{R}_i + \mathbf{W}^2_{\alpha i}\right\}\right) \\
&\quad -\left(\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\}\right)\left(\sum_{i=1}^{n} \left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}'_{\rho i}\mathbf{R}_i + \mathbf{W}_{\alpha i}\mathbf{W}_{\rho i}\right\}\right) \\
c_4 &= \left(\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\}\right)\left(\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}'_{\alpha i}\mathbf{R}_i + \mathbf{W}^2_{\alpha i}\right\}\right) \\
&\quad -\left(\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\}\right)\left(\sum_{i=1}^{n} \left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}'_{\rho i}\mathbf{R}_i + \mathbf{W}_{\alpha i}\mathbf{W}_{\rho i}\right\}\right).
\end{aligned}$$

A close examination of the above expressions reveals that

$$\begin{aligned}
\mathbf{R}_i\mathbf{W}_{\rho i}\mathbf{R}_i^{-1} + \mathbf{W}'_{\rho i} &= c^*\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho} \\
\mathbf{R}_i\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1} + \mathbf{W}'_{\alpha i} &= c^\circ\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}
\end{aligned}$$

and consequently the optimal weights are

$$\mathbf{W}_{\rho i} = \mathbf{R}_i^{-1}\frac{\partial \mathbf{R}_i}{\partial \rho} \quad \text{and} \quad \mathbf{W}_{\alpha i} = \mathbf{R}_i^{-1}\frac{\partial \mathbf{R}_i}{\partial \alpha}.$$

Interestingly, these optimal weights are the same as those obtained by minimizing the determinant of the covariance matrix of $\boldsymbol{\alpha}$. Therefore, the optimal weights are same whether we minimize the determinant or the trace.

We can also construct a subclass of estimating equations by adding the constraints $tr(\mathbf{W}_{ij}) = 0$ to (4.3.1). The unbiased estimating equations in this subclass are

$$\begin{aligned}
\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}(\mathbf{z}_i\mathbf{z}'_i - \mathbf{R}_i)\right\} &= 0 \qquad \text{subject to} \quad \text{tr}(\mathbf{W}_{\rho i}) = 0 \\
\sum_{i=1}^{n} \text{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}(\mathbf{z}_i\mathbf{z}'_i - \mathbf{R}_i)\right\} &= 0 \qquad \text{subject to} \quad \text{tr}(\mathbf{W}_{\alpha i}) = 0.
\end{aligned}$$

The asymptotic covariance matrix of the parameter estimates for this subclass is the same as that of the general class, which is given by $\mathbf{I}_0^{-1}\mathbf{I}_1\mathbf{I}_0'^{-1}$, except that we will have additional constraints on the weights.

Now to find the optimal weights, we need to first differentiate the determinant $|\mathbf{I}_0^{-1}\mathbf{I}_1\mathbf{I}_0'^{-1}|$ with respect to $\mathbf{W}_{\rho i}$ and $\mathbf{W}_{\alpha i}$ using Lagrange multipliers λ_i. Thus to get the optimal weights, we need to solve

$$\begin{aligned}
&\left\{\mathbf{R}_i\mathbf{W}_{\rho i}\mathbf{R}_i^{-1} + \mathbf{W}'_{\rho i}\right\} \xi_0 \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}'_{\alpha i}\mathbf{R}_i + \mathbf{W}^2_{\alpha i}\right\} \\
&\quad - \left\{\mathbf{R}_i\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1} + \mathbf{W}'_{\alpha i}\right\} \xi_0 \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}'_{\alpha i}\mathbf{R}_i + \mathbf{W}_{\rho i}\mathbf{W}_{\alpha i}\right\} \\
&\quad - \left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\right\} \xi_1 \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\} \\
&\quad + \left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\right\} \xi_1 \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\} - \frac{\lambda_i}{2}\mathbf{I}_i\xi_0^3 = 0, \qquad (4.3.5)
\end{aligned}$$

and

$$\begin{aligned}
&\left\{\mathbf{R}_i\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1} + \mathbf{W}'_{\alpha i}\right\} \xi_0 \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}'_{\rho i}\mathbf{R}_i + \mathbf{W}^2_{\rho i}\right\} \\
&\quad - \left\{\mathbf{R}_i\mathbf{W}_{\rho i}\mathbf{R}_i^{-1} + \mathbf{W}'_{\rho i}\right\} \xi_0 \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}'_{\alpha i}\mathbf{R}_i + \mathbf{W}_{\rho i}\mathbf{W}_{\alpha i}\right\} \\
&\quad - \left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\right\} \xi_1 \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\} \\
&\quad + \left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\right\} \xi_1 \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\} - \frac{\lambda_i}{2}\mathbf{I}_i\xi_0^3 = 0, \qquad (4.3.6)
\end{aligned}$$

where ξ_0 and ξ_1 are the same as before. Observing the patterns in equations (4.3.5) and (4.3.6) and keeping in mind that all the traces are scalars, and using the restrictions $\operatorname{tr}(\mathbf{W}_{\rho i}) = 0$ and $\operatorname{tr}(\mathbf{W}_{\alpha i}) = 0$, we can see that (4.3.5) and (4.3.6) can be reduced to similar expressions to (3.2.7) with different subscripts. This means that the optimal weights within the subclass cannot be obtained explicitly for the familial correlation structure, either.

Now differentiating the trace of $\mathbf{I}_0^{-1}\mathbf{I}_1\mathbf{I}_0'^{-1}$ with respect to $\mathbf{W}_{\rho i}$ and using Lagrange multipliers λ_i, we get

$$\begin{aligned}
&\left\{\mathbf{R}_i\mathbf{W}_{\rho i}\mathbf{R}_i^{-1} + \mathbf{W}'_{\rho i}\right\} \xi_0\, c_1 + \left\{\mathbf{R}_i\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1} + \mathbf{W}'_{\alpha i}\right\} \xi_0\, c_2 \\
&\quad + \left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\right\} \left(\xi_0\, c_3 + \xi_2 \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\mathbf{R}_i\right\}\right) \\
&\quad + \left\{\mathbf{R}_i\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha}\right\} \left(\xi_0\, c_4 + \xi_2 \sum_{i=1}^{n} \operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial \mathbf{R}_i^{-1}}{\partial \rho}\mathbf{R}_i\right\}\right) + \frac{\lambda_i}{2}\mathbf{I}_i\xi_0^3 = 0.
\end{aligned}$$

We get similar equation for $\mathbf{W}_{\alpha i}$. And both equations do not have a closed form solution. Thus the optimal weights cannot be obtained explicitly. Computation of the optimal weights is also problematic. We suggest using the simpler weights

$$\mathbf{W}_{\rho i} = \begin{bmatrix} 0 & \mathbf{1}_i' \\ \mathbf{1}_i & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{W}_{\alpha i} = \begin{bmatrix} 0 & \mathbf{0}_i' \\ \mathbf{0} & \mathbf{J}_i - \mathbf{I}_i \end{bmatrix}$$

where $\mathbf{0}$ is a square matrix with all 0 elements and appropriate dimension. The estimating equations for ρ and α, with these simpler weights are

$$\sum_{i=1}^{n} \operatorname{tr}\left\{\begin{bmatrix} \dfrac{-t_i\rho}{1+(t_i-1)\alpha - t_i\rho^2} & \dfrac{1}{1+(t_i-1)\alpha - t_i\rho^2}\mathbf{1}_i' \\ \dfrac{1+(t_i-1)\alpha}{1+(t_i-1)\alpha - t_i\rho^2}\mathbf{1}_i & \dfrac{-\rho}{1+(t_i-1)\alpha - t_i\rho^2}\mathbf{J}_i \end{bmatrix}\mathbf{z}_i\mathbf{z}_i'\right\} = 0,$$

$$\sum_{i=1}^{n} \operatorname{tr}\left\{\begin{bmatrix} 0 & \mathbf{0}_i' \\ \dfrac{-(t_i-1)\rho}{1+(t_i-1)\alpha - t_i\rho^2}\mathbf{1}_i & \dfrac{1}{1-\alpha}\left\{\dfrac{1-\rho^2}{1+(t_i-1)\alpha - t_i\rho^2}\mathbf{J}_i - \mathbf{I}_i\right\} \end{bmatrix}\mathbf{z}_i\mathbf{z}_i'\right\} = 0.$$

(4.3.7)

Using the identities

$$\begin{aligned}
\operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\rho i}'\mathbf{R}_i\right\} &= \frac{t_i[2+2(t_i-1)\alpha+(t_i-1)^2\alpha^2-2t_i\rho^2]}{1+(t_i-1)\alpha-t_i\rho^2} \\
\operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i\right\} &= \frac{-t_i(t_i-1)^2\alpha\rho}{1+(t_i-1)\alpha-t_i\rho^2} \\
\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\mathbf{R}_i^{-1}\mathbf{W}_{\alpha i}'\mathbf{R}_i\right\} &= \frac{t_i(t_i-1)[1+(t_i-1)\alpha-\rho^2]}{1+(t_i-1)\alpha-t_i\rho^2} \\
\operatorname{tr}\left\{\mathbf{W}_{\rho i}^2\right\} &= 2t_i \\
\operatorname{tr}\left\{\mathbf{W}_{\rho i}\mathbf{W}_{\alpha i}\right\} &= 0 \\
\operatorname{tr}\left\{\mathbf{W}_{\alpha i}^2\right\} &= t_i(t_i-1) \\
\operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\rho}\mathbf{R}_i\right\} &= \frac{-t_i[2+(t_i-1)\alpha]}{1+(t_i-1)\alpha-t_i\rho^2} \\
\operatorname{tr}\left\{\mathbf{W}_{\rho i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right\} &= \frac{t_i(t_i-1)\rho}{1+(t_i-1)\alpha-t_i\rho^2} \\
\operatorname{tr}\left\{\mathbf{W}_{\alpha i}\frac{\partial\mathbf{R}_i^{-1}}{\partial\alpha}\mathbf{R}_i\right\} &= \frac{-t_i(t_i-1)(1-\rho^2)}{(1-\alpha)[1+(t_i-1)\alpha-t_i\rho^2]},
\end{aligned}$$

we can check that the matrices $\mathbf{I}_0$ and $\mathbf{I}_1$ reduce to

$$\mathbf{I}_0 = \begin{bmatrix} \sum_{i=1}^{n} \frac{-t_i[2+(t_i-1)\alpha]}{1+(t_i-1)\alpha - t_i\rho^2} & \sum_{i=1}^{n} \frac{t_i(t_i-1)\rho}{1+(t_i-1)\alpha - t_i\rho^2} \\ \sum_{i=1}^{n} \frac{t_i(t_i-1)\rho}{1+(t_i-1)\alpha - t_i\rho^2} & \sum_{i=1}^{n} \frac{-t_i(t_i-1)(1-\rho^2)}{(1-\alpha)[1+(t_i-1)\alpha - t_i\rho^2]} \end{bmatrix}$$

$$\mathbf{I}_1 = \begin{bmatrix} \sum_{i=1}^{n} \frac{[2+(t_i-1)\alpha]^2 - 4t_i\rho^2}{1+(t_i-1)\alpha - t_i\rho^2} & \sum_{i=1}^{n} \frac{-t_i(t_i-1)^2\alpha\rho}{1+(t_i-1)\alpha - t_i\rho^2} \\ \sum_{i=1}^{n} \frac{-t_i(t_i-1)^2\alpha\rho}{1+(t_i-1)\alpha - t_i\rho^2} & \sum_{i=1}^{n} \frac{t_i(t_i-1)[2+2(t_i-1)\alpha - (t_i+1)\rho^2]}{1+(t_i-1)\alpha - t_i\rho^2} \end{bmatrix}.$$

IV.4 Relative Efficiency

In this section we will study the relative efficiencies of the various estimates discussed in the previous sections for the familial correlations using simulations. The simulation steps are similar to that in Chapter III and are described below.

(1) Generate n integers $\{t_i, 1 < i < n\}$ for the number of children in each family ranging from 2 to 4 with the following probability distribution. The reason for using this distribution is that the number of children in the present day society are often less or equal to 4 and we need at least 2 children in the family for sibling-sibling correlation to exist.

t_i	2	3	4
$P(t_i)$	0.4	0.4	0.2

(2) For $i = 1, 2, \ldots, n$, generate t_i univariate standard normal random numbers and stack them into a column vector $\mathbf{z}_i$.

(3) For fixed value of ρ and α, construct the familial correlation matrix $\mathbf{R}_i$.

(4) Let $\mathbf{y}_i = \mathbf{R}_i^{\frac{1}{2}}\mathbf{z}_i$, where $\mathbf{R}_i^{\frac{1}{2}}$ is the square root of $\mathbf{R}_i$.

(5) Solve equations in (4.3.7) to get estimates $\hat{\rho}_S$ and $\hat{\alpha}_S$.

(6) Solve equation (4.2.1) to get ML estimates $\hat{\rho}_L$ and $\hat{\alpha}_L$.

(7) Calculate errors $\tau_S = (\hat{\rho}_S - \rho, \hat{\alpha}_S - \alpha)$ and $\tau_L = (\hat{\rho}_L - \rho, \hat{\alpha}_L - \alpha)$.

(8) Repeat steps (1) – (7) a large number of times (r), say $r = 10,000$, and stack errors into two $r \times 2$ vectors $\mathbf{e}_S$ and $\mathbf{e}_L$.

Then calculate the relative efficiencies based on trace and determinant criteria as follows:

$$\mathrm{Eff_t}(\widehat{\boldsymbol{\alpha}}_S, \widehat{\boldsymbol{\alpha}}_L) = \frac{\mathrm{tr}(\mathbf{e}_L'\mathbf{e}_L)}{\mathrm{tr}(\mathbf{e}_S'\mathbf{e}_S)} \quad \text{and} \quad \mathrm{Eff_d}(\widehat{\boldsymbol{\alpha}}_S, \widehat{\boldsymbol{\alpha}}_L) = \frac{|\mathbf{e}_L'\mathbf{e}_L|}{|\mathbf{e}_S'\mathbf{e}_S|}.$$

Figures 4.2 and 4.3 shows different perspectives of the surface of efficiencies $\mathrm{Eff_t}(\widehat{\boldsymbol{\alpha}}_S, \widehat{\boldsymbol{\alpha}}_L)$ based on trace criteria, for different values of ρ and α. Corresponding efficiencies $\mathrm{Eff_d}(\widehat{\boldsymbol{\alpha}}_S, \widehat{\boldsymbol{\alpha}}_L)$ based on the determinant criteria are shown in Figures 4.4 and 4.5. Table 4.1 and 4.2 contain some numerical values of these efficiencies for different values of α and positive ρ. The efficiencies can be obtained for negative values of ρ by symmetry. An examination of these Figures and Tables clearly show that the unbiased estimating approach yields highly efficient estimates over a wide range of the parameter space. Efficiencies for the case where the families are of equal size with three children are given in Figure 4.6 and 4.7. Notice that the efficiencies at the boundary of feasible regions of ρ and α are very high.

Table 4.1: Numerical values of efficiency based on trace criteria

	α									
ρ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.0	0.947	1.003	1.009	1.002	0.996	0.989	0.982	0.978	0.985	0.978
0.1	0.946	1.000	1.007	1.005	0.990	0.983	0.977	0.978	0.974	0.980
0.2	0.938	0.989	0.998	0.990	0.981	0.975	0.964	0.962	0.973	0.965
0.3	0.939	0.996	0.993	0.980	0.966	0.956	0.946	0.950	0.957	0.957
0.4	1.111	1.043	0.998	0.970	0.942	0.929	0.929	0.926	0.933	0.936
0.5		1.284	1.073	0.984	0.932	0.914	0.896	0.895	0.898	0.914
0.6			1.430	1.083	0.930	0.884	0.865	0.857	0.863	0.879
0.7					1.158	0.899	0.799	0.801	0.808	0.830
0.8							0.974	0.708	0.723	0.752
0.9									0.881	0.600

Table 4.2: Numerical values of efficiency based on determinant criteria

	α									
ρ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.0	0.898	1.009	1.019	1.002	0.984	0.958	0.934	0.906	0.886	0.865
0.1	0.898	1.000	1.013	1.006	0.971	0.948	0.921	0.901	0.887	0.853
0.2	0.899	0.975	0.987	0.968	0.944	0.925	0.889	0.860	0.846	0.814
0.3	0.931	0.979	0.962	0.928	0.898	0.873	0.842	0.824	0.808	0.773
0.4	1.227	1.036	0.941	0.883	0.830	0.802	0.787	0.763	0.727	0.698
0.5		1.440	1.037	0.882	0.788	0.745	0.698	0.676	0.647	0.629
0.6			1.634	1.061	0.802	0.689	0.629	0.587	0.557	0.525
0.7					1.302	0.732	0.569	0.496	0.462	0.417
0.8							0.924	0.498	0.363	0.305
0.9									0.325	0.242

IV.5 An Illustrative Example

In this section we apply the estimation methods on a real life example. Dern and Wiorkowski (1969) have discussed an interesting familial data. The data consists of pre- and post-storage measurements of erythrocyte Adenosine Triphosphate (ATP) levels from healthy Caucasian family members from 22 families. The pre-storage measurements were taken, in most cases, immediately after phlebotomy or after arrival of the sample in the laboratory. The post-storage measurements, in all cases, were taken after 21 days of storage in the refrigerator at 4 ± 1°C. All ATP levels are expressed as μ-moles per grams of hemoglobin. In addition to the ATP levels, ages of family members are also available for the analysis. Table 4.3 lists part of the originally data. The data is incomplete for three families in the sense of mothers' presence. In our analysis we dropped the three families, and used the complete data of 80 observations on the mothers and children in the remaining 19 families. Table 4.4 contains the results of our analysis using the unbiased estimating approach and the ML approach. The estimates are similar, but however, the standard errors obtained using the unbiased estimating equations, are lower and hence preferable.

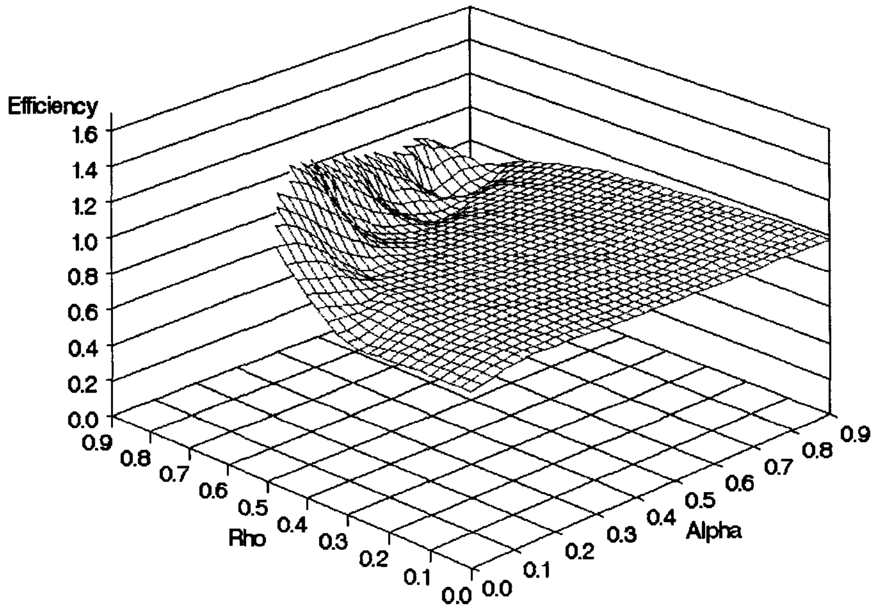


Figure 4.2: Efficiency based on trace criteria (45° view).

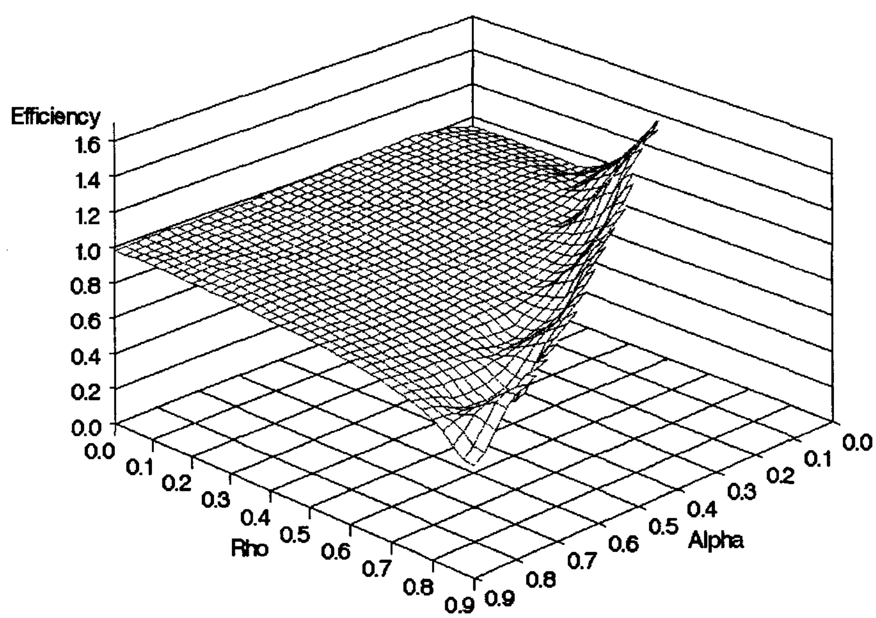


Figure 4.3: Efficiency based on trace criteria (225° view).

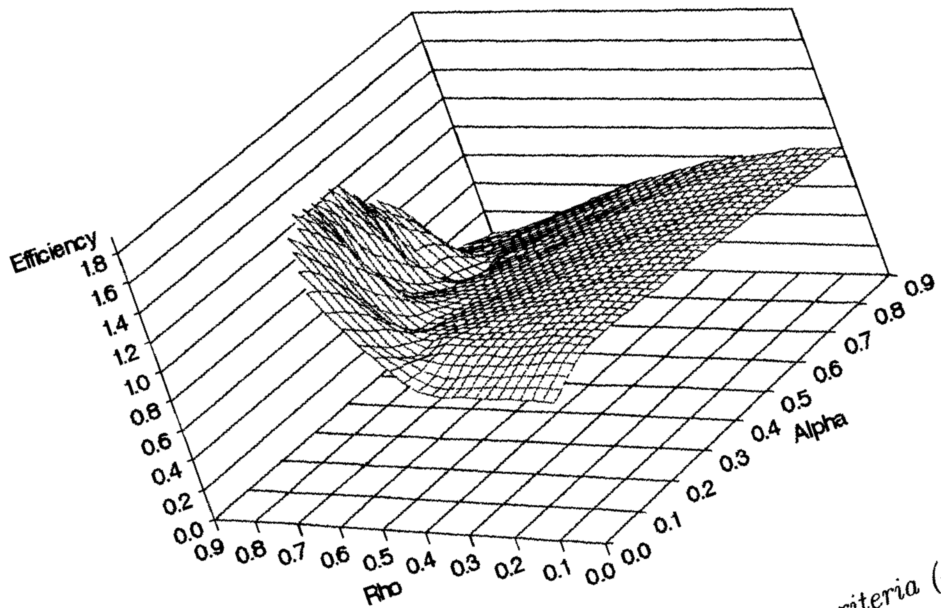


Figure 4.4: Efficiency based on determinant criteria (45° view).

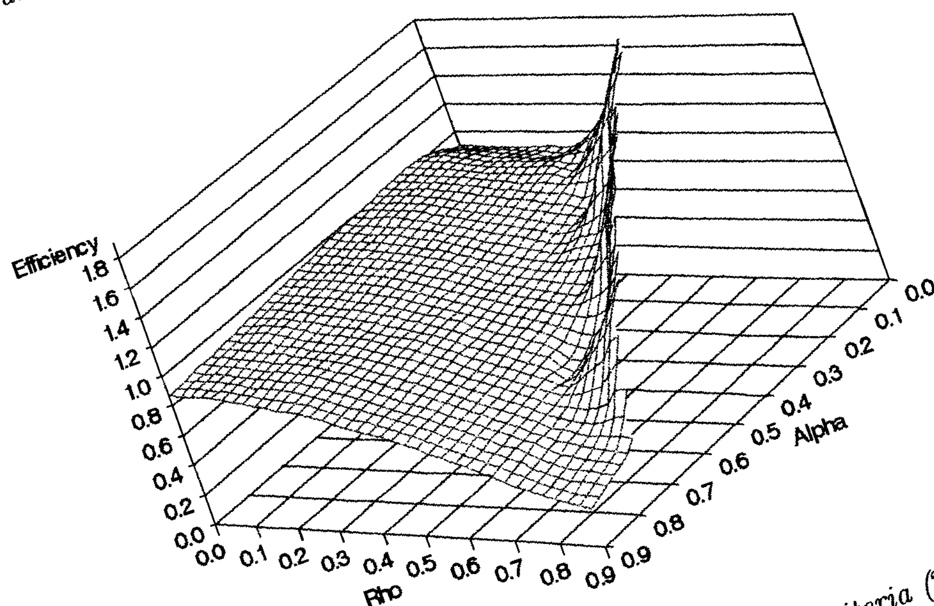


Figure 4.5: Efficiency based on determinant criteria (225° view).

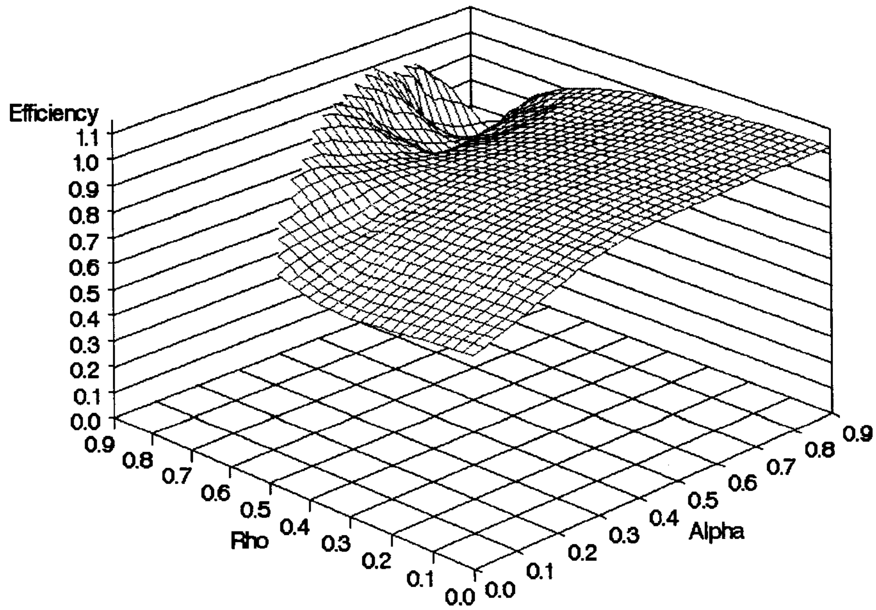


Figure 4.6: Efficiency based on trace criteria: balanced case (45° view).

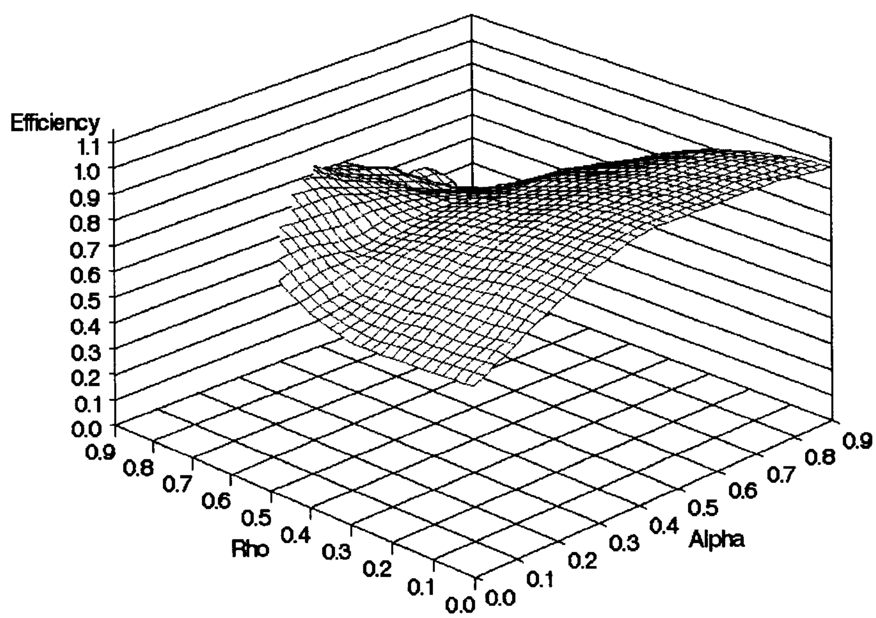


Figure 4.7: Efficiency based on determinant criteria: balanced case (45° view).

Table 4.3: Partial list of member ATP levels in 22 families

Family	Member	Age	Pre-ATP	Post-ATP
1	Mother	50	—	3.30
	Father	54	—	2.40
	Son	24	—	3.76
	Daughter	30	—	2.14
	Daughter	26	—	2.55
2	Mother	62	4.43	2.49
	Father	62	3.72	1.79
	Son	24	4.18	1.49
	Son	41	4.81	2.84
	Daughter	31	4.42	2.04
	Daughter	38	3.65	1.17
3	Mother	50	3.79	1.28
	Father	45	4.54	3.07
	Son	7	4.72	1.19
4	Mother	55	5.42	3.65
	Father	56	4.10	2.65
	Son	23	5.30	2.16
	Son	27	4.48	2.40
	Son	19	4.85	3.28
	Son	24	—	2.20
5	Mother	57	4.71	2.23
	Father	76	—	2.15
	Son	32	4.19	1.33
	Son	28	3.43	1.85
⋮	⋮	⋮	⋮	⋮
22	Mother	45	5.29	3.27
	Father	—	—	—
	Son	24	5.30	4.10
	Son	20	5.25	3.67

Table 4.4: Parameter estimates

	U_S (S.E.)		MLE (S.E.)	
Intercept	-0.7994	(0.3867)	-0.7281	(0.4354)
Gender	0.1624	(0.0771)	0.1804	(0.0930)
ATP (pre-storage)	0.7260	(0.0763)	0.7118	(0.0898)
Scale (mother)	0.5577	—	0.5569	—
Scale (siblings)	0.4617	—	0.5034	—
$\hat{\rho}$	0.3404	(0.0236)	0.3864	—
$\hat{\alpha}$	0.4192	(0.0163)	0.5845	—

CHAPTER V

ESTIMATION OF NUCLEAR FAMILIAL STRUCTURE

In Chapter IV we have discussed the analysis of data collected on single parent families. Here we extend those results for data taken on nuclear families, that is, data taken on two parent families. The correlation structure to model intra-family associations will have additional parameters, for example we need to account for the correlation between the parents and the correlation between the father and children. These additional parameters will require additional estimating equations, and the analysis poses challenging computational problems. We briefly sketch the generalizations, since the details are similar to the results in Chapter IV.

The organization of this chapter is as follows. We first study properties of the general familial correlation structure in Section V.1. ML estimation of the parameters for normal data is presented in Section V.2. Next, a general class and a subclass of unbiased estimating equations for parameter estimation are presented in Section V.3.

V.1 Nuclear Familial Correlation Structure

A correlation model that is appropriate to model associations within a nuclear family is the structure

$$\mathbf{R}_i = \begin{bmatrix} 1 & \gamma & \rho_1 & \rho_1 & \rho_1 & \dots & \rho_1 \\ \gamma & 1 & \rho_2 & \rho_2 & \rho_2 & \dots & \rho_2 \\ \rho_1 & \rho_2 & 1 & \alpha & \alpha & \dots & \alpha \\ \rho_1 & \rho_2 & \alpha & 1 & \alpha & \dots & \alpha \\ \rho_1 & \rho_2 & \alpha & \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_1 & \rho_2 & \alpha & \alpha & \alpha & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & \gamma & \rho_1 \mathbf{1}_i' \\ \gamma & 1 & \rho_2 \mathbf{1}_i' \\ \rho_1 \mathbf{1}_i & \rho_2 \mathbf{1}_i & (1-\alpha)\mathbf{I}_i + \alpha \mathbf{J}_i \end{bmatrix} \quad (5.1.1)$$

where γ is the correlation between the two parents, ρ_1 is the father-children correlation, ρ_2 is the mother-children correlation, and as before α is the common correlation among the t_i children. Pre- and Post-multiplying $\mathbf{R}_i$ by the matrix incorporating the Helmert matrix $\mathbf{M}_i$ which we defined in Chapter IV, we get

$$\mathbf{\Omega}_i = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \mathbf{0}_i' \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & \mathbf{0}_i' \\ \mathbf{0}_i & \mathbf{0}_i & \mathbf{M}_i \end{bmatrix} \mathbf{R}_i \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \mathbf{0}_i' \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & \mathbf{0}_i' \\ \mathbf{0}_i & \mathbf{0}_i & \mathbf{M}_i' \end{bmatrix}$$

$$= \begin{bmatrix} 1+\gamma & 0 & \sqrt{\frac{t_i}{2}}(\rho_1+\rho_2) & 0 & \dots & 0 \\ 0 & 1-\gamma & \sqrt{\frac{t_i}{2}}(\rho_1-\rho_2) & 0 & \dots & 0 \\ \sqrt{\frac{t_i}{2}}(\rho_1+\rho_2) & \sqrt{\frac{t_i}{2}}(\rho_1-\rho_2) & 1+(t_i-1)\alpha & 0 & \dots & 0 \\ 0 & 0 & 0 & 1-\alpha & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1-\alpha \end{bmatrix}.$$

Clearly, the necessary and sufficient conditions for $\mathbf{\Omega}_i$, or equivalently $\mathbf{R}_i$, to be positive definite are

(1) $-1 < \gamma < 1$

(2) $-\frac{1}{t_i-1} < \alpha < 1$

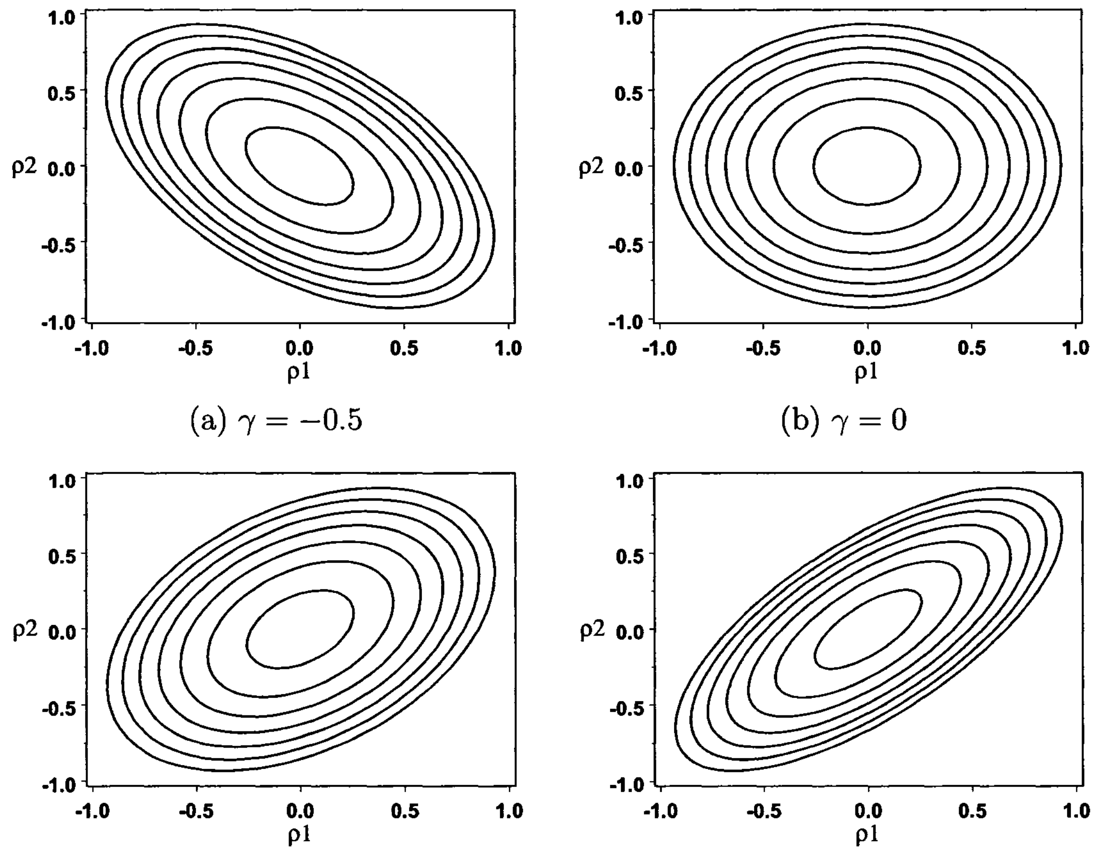
(3) $t_i(\rho_1^2 + \rho_2^2 - 2\gamma\rho_1\rho_2) < (1-\gamma^2)[1+(t_i-1)\alpha]$.

These conditions ensure that the determinants of all its principle minors of $\mathbf{R}_i$ are positive. Condition (3) can be further written as

$$\frac{(\rho_1 - \gamma\rho_2)^2}{1-\gamma^2} + \rho_2^2 < \frac{1+(t_i-1)\alpha}{t_i},$$

which means that for fixed γ and α satisfying conditions (1) and (2), the graph of ρ_1 versus ρ_2 is an ellipse. Figure 5.1 shows contour plots of α versus (ρ_1, ρ_2) for given values of γ when $t_i = 3$. The inverse of the familial correlation matrix is given by

$$\mathbf{R}_i^{-1} = \begin{bmatrix} \frac{1-\phi_2}{\zeta} & \frac{\phi_0-\gamma}{\zeta} & \zeta_1\mathbf{1}_i' \\ \frac{\phi_0-\gamma}{\zeta} & \frac{1-\phi_1}{\zeta} & \zeta_2\mathbf{1}_i' \\ \zeta_1\mathbf{1}_i & \zeta_2\mathbf{1}_i & \frac{1}{1-\alpha}\mathbf{I}_i - \frac{(1-\alpha)(\rho_1\zeta_1+\rho_2\zeta_2)+\alpha}{(1-\alpha)[1+(t-1)\alpha]}\mathbf{J}_i \end{bmatrix}$$



(a) $\gamma = -0.5$ (b) $\gamma = 0$

(c) $\gamma = 0.4$ (d) $\gamma = 0.7$

Figure 5.1: Contour plots of α *vs.* (ρ_1, ρ_2).

where

$$
\begin{aligned}
\phi_0 &= \frac{t\rho_1\rho_2}{1+(t-1)\alpha} \\
\phi_1 &= \frac{t\rho_1^2}{1+(t-1)\alpha} \\
\phi_2 &= \frac{t\rho_2^2}{1+(t-1)\alpha} \\
\zeta &= (1-\phi_1)(1-\phi_2)-(\gamma-\phi_0)^2 \\
\zeta_1 &= \frac{(1-\phi_2)(\gamma\rho_2-\rho_1)}{[1+(t-1)\alpha-t\rho_2^2][(1-\phi_1)(1-\phi_2)-(\gamma-\phi_0)^2]} \\
\zeta_2 &= \frac{-\rho_2}{[1+(t-1)\alpha](1-\phi_2)}-\frac{(\gamma\rho_2-\rho_1)(\gamma-\phi_0)}{[1+(t-1)\alpha-t\rho_2^2][(1-\phi_1)(1-\phi_2)-(\gamma-\phi_0)^2]}.
\end{aligned}
$$

Appendix A.4 contains the Cholesky decomposition matrices of $\mathbf{R}_i$ and $\mathbf{R}_i^{-1}$.

As for the covariance matrix for the variables in a nuclear family, it is reasonable to assume the variances for parents are different from those of the children. Let σ_1^2

be the variance of the measurement on the father, σ_2^2 be the variance for the mother and σ_0^2 be the common variance among the children. The covariance matrix is

$$\begin{aligned}\mathbf{V}_i &= \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\gamma & \sigma_1\sigma_0\rho_1 & \sigma_1\sigma_0\rho_1 & \cdots & \sigma_1\sigma_0\rho_1 \\ \sigma_1\sigma_2\gamma & \sigma_2^2 & \sigma_2\sigma_0\rho_2 & \sigma_2\sigma_0\rho_2 & \cdots & \sigma_2\sigma_0\rho_2 \\ \sigma_1\sigma_0\rho_1 & \sigma_2\sigma_0\rho_2 & \sigma_0^2 & \sigma_0^2\alpha & \cdots & \sigma_0^2\alpha \\ \sigma_1\sigma_0\rho_1 & \sigma_2\sigma_0\rho_2 & \sigma_0^2\alpha & \sigma_0^2 & \cdots & \sigma_0^2\alpha \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_0\rho_1 & \sigma_2\sigma_0\rho_2 & \sigma_0^2\alpha & \sigma_0^2\alpha & \cdots & \sigma_0^2 \end{bmatrix} \\ &= \mathbf{A}_i^{\frac{1}{2}}\mathbf{R}_i\mathbf{A}_i^{\frac{1}{2}}\end{aligned}$$

where $\mathbf{A}_i^{\frac{1}{2}} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \sigma_0^2, \ldots, \sigma_0^2)$.

V.2 Maximum Likelihood Estimate

In this section we discuss ML estimation for normal data collected on nuclear families. The results are extensions of the results that we have in Section IV.2. We present main equations omitting some details. The ML estimating equation for correlation parameters is

$$\sum_{i=1}^{n} \mathrm{tr}\left\{\mathbf{R}_i^{-1}\frac{\partial \mathbf{R}_i}{\partial \boldsymbol{\alpha}}\mathbf{R}_i^{-1}\left(\mathbf{z}_i\mathbf{z}_i' - \mathbf{R}_i\right)\right\} = \mathbf{0}, \tag{5.2.1}$$

where $\mathbf{z}_i$ is the Pearson residual. The ML equation for estimating the variances is

$$\sum_{i=1}^{n} \mathrm{tr}\left\{\mathbf{V}_i^{-1}\frac{\partial \mathbf{V}_i}{\partial \boldsymbol{\sigma}^2}\mathbf{V}_i^{-1}\left(\varepsilon_i\varepsilon_i' - \mathbf{V}_i\right)\right\} = \mathbf{0}, \tag{5.2.2}$$

where $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \sigma_0^2)'$ and correlation parameter $\boldsymbol{\alpha} = (\gamma, \rho_1, \rho_2, \alpha)'$. Expressions for the second order partial derivatives are

$$\begin{aligned}-\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial\left(\boldsymbol{\sigma}^2\right)^2}\right\} &= \sum_{i=1}^{n}\mathrm{tr}\left(\frac{\partial^2\mathbf{A}_i^{-\frac{1}{2}}}{\partial\left(\boldsymbol{\sigma}^2\right)^2}\mathbf{A}_i^{\frac{1}{2}} + \frac{\partial \mathbf{A}_i^{-\frac{1}{2}}}{\partial \boldsymbol{\sigma}^2}\mathbf{R}_i^{-1}\frac{\partial \mathbf{A}_i^{-\frac{1}{2}}}{\partial \boldsymbol{\sigma}^2}\mathbf{A}_i^{\frac{1}{2}}\mathbf{R}_i\mathbf{A}^{\frac{1}{2}}\right) \\ &\quad + \frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2 \log\left(\sigma_1^2\sigma_2^2\sigma_0^{2t_i}\right)}{\partial \boldsymbol{\sigma}^2} \\ -\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\sigma}^2 \partial \boldsymbol{\alpha}}\right\} &= \sum_{i=1}^{n}\mathrm{tr}\left\{\frac{\partial \mathbf{A}_i^{-\frac{1}{2}}}{\partial \boldsymbol{\sigma}^2}\frac{\partial \mathbf{R}_i^{-1}}{\partial \boldsymbol{\alpha}}\mathbf{R}_i\mathbf{A}^{\frac{1}{2}}\right\} \\ -\mathrm{E}\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\alpha}^2}\right\} &= \frac{1}{2}\sum_{i=1}^{n}\left\{\mathrm{tr}\left(\frac{\partial^2 \mathbf{R}_i^{-1}}{\partial \boldsymbol{\alpha}^2}\mathbf{R}_i\right) + \frac{\partial^2 \log|\mathbf{R}_i|}{\partial \boldsymbol{\alpha}^2}\right\}.\end{aligned}$$

Finally, the Fisher information is

$$\begin{bmatrix} -\mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial\left(\boldsymbol{\sigma}^2\right)^2}\right\} & -\mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \boldsymbol{\sigma}^2 \partial \boldsymbol{\alpha}}\right\} \\ -\mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \boldsymbol{\sigma}^2 \partial \boldsymbol{\alpha}}\right\} & -\mathrm{E}\left\{\dfrac{\partial^2 \ell}{\partial \boldsymbol{\alpha}^2}\right\} \end{bmatrix}.$$

V.3 Classes of Unbiased Estimating Equations

To estimate the correlation parameters in a nuclear family, we could consider a general class of estimating equations given by

$$\sum_{i=1}^{n} \mathrm{tr}\left\{\mathbf{W}_{ji}\mathbf{R}_i^{-1}\left(\mathbf{z}_i\mathbf{z}_i' - \mathbf{R}_i\right)\right\} \;=\; 0\,, \tag{5.3.1}$$

where $\mathbf{W}_{ji}$ are weights matrices. The range of j equals $1, 2, 3, 4$ corresponding to the four correlation parameters γ, ρ_1, ρ_2 and α, respectively. Using Lemma 2.1 and Theorem 2.1, we can get the asymptotic covariance matrix of the unbiased estimating equation estimate $\hat{\boldsymbol{\alpha}} = (\hat{\gamma}, \hat{\rho}_1, \hat{\rho}_2, \hat{\alpha})'$ as $\mathbf{I}_0^{-1}\mathbf{I}_1\mathbf{I}_0'^{-1}$, where $\mathbf{I}_0 = (u_{rs})$ and $\mathbf{I}_1 = (v_{rs})$ and

$$\begin{aligned} u_{rs} &= \sum_{i=1}^{n} \mathrm{tr}\left\{\mathbf{W}_{ri}\frac{\partial \mathbf{R}_i^{-1}}{\partial \alpha_{[s]}}\mathbf{R}_i\right\} \\ v_{rs} &= \sum_{i=1}^{n} \mathrm{tr}\left\{\mathbf{W}_{ri}\mathbf{R}_i^{-1}\mathbf{W}_{si}'\mathbf{R}_i + \mathbf{W}_{ri}\mathbf{W}_{is}\right\} \end{aligned}$$

with $\alpha_{[j]}$ is the jth element of $\boldsymbol{\alpha}$, that is, $\alpha_{[1]} \equiv \gamma$, $\alpha_{[2]} \equiv \rho_1$, $\alpha_{[3]} \equiv \rho_2$ and $\alpha_{[4]} \equiv \alpha$.

To find the optimal weights, we could minimize the determinant or the trace of this asymptotic covariance matrix. The resulting optimal weights coincide with our results in Chapter IV, and they are

$$\mathbf{W}_{ji} = \mathbf{R}_i^{-1}\frac{\partial \mathbf{R}_i}{\partial \alpha_{[j]}} \qquad \text{for } j = 1, 2, 3, 4.$$

We could also consider a subclass of estimating equations by adding the constraint $\mathrm{tr}(\mathbf{W}_{ji}) = 0$ to the equation (5.3.1). The asymptotic covariance matrix of the estimates derived from this subclass is same as that of the general class. And as in Chapter IV, we do not have closed form solutions for the optimal weights, either.

Once again, to facilitate the construction of the weights and to avoid convergence problem that may occur, we suggest using the simpler weights that are close to being optimal:

$$
\mathbf{W}_{1i} = \begin{bmatrix} 0 & 1 & \mathbf{0}_i' \\ 1 & 0 & \mathbf{0}_i' \\ \mathbf{0}_i & \mathbf{0}_i & \mathbf{0} \end{bmatrix} \qquad \mathbf{W}_{2i} = \begin{bmatrix} 0 & 0 & \mathbf{1}_i' \\ 0 & 0 & \mathbf{0}_i' \\ \mathbf{1}_i & \mathbf{0}_i & \mathbf{0} \end{bmatrix}
$$

$$
\mathbf{W}_{3i} = \begin{bmatrix} 0 & 0 & \mathbf{0}_i' \\ 0 & 0 & \mathbf{1}_i' \\ \mathbf{0}_i & \mathbf{1}_i & \mathbf{0} \end{bmatrix} \qquad \mathbf{W}_{4i} = \begin{bmatrix} 0 & 0 & \mathbf{0}_i' \\ 0 & 0 & \mathbf{0}_i' \\ \mathbf{0}_i & \mathbf{0}_i & \mathbf{J}_i - \mathbf{I}_i \end{bmatrix},
$$

corresponding to the four correlation parameters, γ, ρ_1, ρ_2 and α respectively. With these weights, the estimating equations can be simplified as

$$
\gamma: \quad \sum_{i=1}^{n} \operatorname{tr}\left\{ \begin{bmatrix} \frac{\phi_0 - \gamma}{\zeta} & \frac{1-\phi_1}{\zeta} & \zeta_2 \mathbf{1}_i' \\ \frac{1-\phi_2}{\zeta} & \frac{\phi_0 - \gamma}{\zeta} & \zeta_1 \mathbf{1}_i' \\ \mathbf{0}_i & \mathbf{0}_i & \mathbf{0}_I \end{bmatrix} \mathbf{z}_i \mathbf{z}_i' \right\} = 0
$$

$$
\rho_1: \quad \sum_{i=1}^{n} \operatorname{tr}\left\{ \begin{bmatrix} t_i\zeta_1 & t_i\zeta_2 & \frac{1 - t_i(\rho_1\zeta_1 + \rho_2\zeta_2)}{1+(t_i-1)\alpha}\mathbf{1}_i' \\ 0 & 0 & \mathbf{0}_i' \\ \frac{1-\phi_2}{\zeta}\mathbf{1}_i & \frac{\phi_0-\gamma}{\zeta}\mathbf{1}_i & \zeta_1 \mathbf{J}_i \end{bmatrix} \mathbf{z}_i \mathbf{z}_i' \right\} = 0
$$

$$
\rho_2: \quad \sum_{i=1}^{n} \operatorname{tr}\left\{ \begin{bmatrix} 0 & 0 & \mathbf{0}_i' \\ t_i\zeta_1 & t_i\zeta_2 & \frac{1 - t_i(\rho_1\zeta_1 + \rho_2\zeta_2)}{1+(t_i-1)\alpha}\mathbf{1}_i' \\ \frac{\phi_0-\gamma}{\zeta}\mathbf{1}_i & \frac{1-\phi_1}{\zeta}\mathbf{1}_i & \zeta_2 \mathbf{J}_i \end{bmatrix} \mathbf{z}_i \mathbf{z}_i' \right\} = 0
$$

$$
\alpha: \quad \sum_{i=1}^{n} \operatorname{tr}\left\{ \begin{bmatrix} 0 & 0 & \mathbf{0}_i' \\ 0 & 0 & \mathbf{0}_i' \\ (t_i-1)\zeta_1\mathbf{1}_i & (t_i-1)\zeta_2\mathbf{1}_i & \zeta_3\mathbf{J}_i - \frac{1}{1-\alpha}\mathbf{I}_i \end{bmatrix} \mathbf{z}_i \mathbf{z}_i' \right\} = 0
$$

where ζ, ζ_1 and ζ_2 are as defined before, and

$$
\zeta_3 = \frac{1-(t_i-1)(1-\alpha)(\rho_1\zeta_1+\rho_2\zeta_2)}{(1-\alpha)[1+(t_i-1)\alpha]}
$$

The elements of the 4×4 matrix $\mathbf{I}_0$ are

$$
\begin{aligned}
u_{11} &= \sum_{i=1}^{n} \frac{2-\phi_1-\phi_2}{\zeta} \\
u_{12} &= \sum_{i=1}^{n} t_i\zeta_2 \\
u_{13} &= \sum_{i=1}^{n} t_i\zeta_1 \\
u_{14} &= 0 \\
u_{22} &= \sum_{i=1}^{n} t_i \left\{ \frac{1-t_i(\rho_1\zeta_1+\rho_2\zeta_2)}{1+(t_i-1)\alpha} + \frac{1-\phi_2}{\zeta} \right\} \\
u_{23} &= \sum_{i=1}^{n} \frac{t_i(\phi_0-\gamma)}{\zeta} \\
u_{24} &= \sum_{i=1}^{n} t_i(t_i-1)\zeta_1 \\
u_{33} &= \sum_{i=1}^{n} t_i \left\{ \frac{1-t_i(\rho_1\zeta_1+\rho_2\zeta_2)}{1+(t_i-1)\alpha} + \frac{1-\phi_1}{\zeta} \right\} \\
u_{34} &= \sum_{i=1}^{n} t_i(t_i-1)\zeta_2 \\
u_{44} &= \sum_{i=1}^{n} \frac{t_i(t_i-1)\left\{1-(t_i-1)(1-\alpha)(\rho_1\zeta_1+\rho_2\zeta_2)\right\}}{(1-\alpha)[1+(t_i-1)\alpha]}
\end{aligned}
$$

and the elements of the 4×4 matrix $\mathbf{I}_1$ are

$$
\begin{aligned}
v_{11} &= \sum_{i=1}^{n} \frac{2\zeta+2-\phi_1-\phi_2+2\gamma(\phi_0-\gamma)}{\zeta} \\
v_{12} &= \sum_{i=1}^{n} t_i \left\{ \zeta_2+\gamma\zeta_1+\frac{\rho_1(\phi_0-\gamma)+\rho_2(1-\phi_2)}{\zeta} \right\} \\
v_{13} &= \sum_{i=1}^{n} t_i \left\{ \zeta_1+\gamma\zeta_2+\frac{\rho_2(\phi_0-\gamma)+\rho_1(1-\phi_1)}{\zeta} \right\} \\
v_{14} &= \sum_{i=1}^{n} t_i(t_i-1)(\rho_1\zeta_2+\rho_2\zeta_1) \\
v_{22} &= \sum_{i=1}^{n} t_i \left\{ 2+2t_i\rho_1\zeta_1+\frac{1-t_i(\rho_1\zeta_1+\rho_2\zeta_2)}{1+(t_i-1)\alpha} + \frac{(1-\phi_2)\{1+(t_i-1)\alpha\}}{\zeta} \right\} \\
v_{23} &= \sum_{i=1}^{n} t_i \left\{ t_i\rho_1\zeta_2+t_i\rho_2\zeta_1+\frac{\gamma\{1-t_i(\rho_1\zeta_1+\rho_2\zeta_2)\}}{1+(t_i-1)\alpha} + \frac{(\phi_0-\gamma)\{1+(t_i-1)\alpha\}}{\zeta} \right\} \\
v_{24} &= \sum_{i=1}^{n} t_i(t_i-1) \left\{ \zeta_1\{1+(t_i-1)\alpha\}+\frac{\rho_1\{1-t_i(\rho_1\zeta_1+\rho_2\zeta_2)\}}{1+(t_i-1)\alpha} \right\}
\end{aligned}
$$

$$
\begin{aligned}
v_{33} &= \sum_{i=1}^{n} t_i \left\{ 2 + 2t_i\rho_2\zeta_2 + \frac{1 - t_i(\rho_1\zeta_1 + \rho_2\zeta_2)}{1 + (t_i - 1)\alpha} + \frac{(1-\phi_1)\{1 + (t_i - 1)\alpha\}}{\zeta} \right\} \\
v_{34} &= \sum_{i=1}^{n} t_i(t_i - 1) \left\{ \zeta_2\{1 + (t_i - 1)\alpha\} + \frac{\rho_2\{1 - t_i(\rho_1\zeta_1 + \rho_2\zeta_2)\}}{1 + (t_i - 1)\alpha} \right\} \\
v_{44} &= \sum_{i=1}^{n} t_i(t_i - 1)\{2 - (t_i - 1)(\rho_1\zeta_1 + \rho_2\zeta_2)\}.
\end{aligned}
$$

These are useful to calculate the asymptotic covariance of the estimates of the correlation parameters.

CHAPTER VI

ANALYSIS OF FAMILIAL BINARY OUTCOMES

In previous chapters, we have studied modelling and estimation of correlation parameters for continuous responses. However, correlated discrete data, in particular binary, are common in many scientific studies including medical, social and biological research. In this chapter we focus our attention on the analysis of familial binary data, that is, binary data collected on families. Unlike continuous data, the ranges of correlations between binary variables are constrained by the marginal means. The organization of this chapter is as follows. We first study feasible ranges of the correlations, and more generally ranges of different measures of associations for familial binary variables. These probabilistic results are important for developing theoretically sound methods of estimation for the association measures. In Section VI.2, we study latent variable models for familial binary variables. In particular, we investigate stochastic representations for the multivariate probit model (Ashford and Sowden, 1970). Finally, we present a binary data analysis example to illustrate the estimation procedures.

VI.1 Ranges of Measures of Associations

In a recent paper Chaganty and Joe (2006) studied ranges of correlation parameters between three binary random variables for the unstructured and common structured matrices, for example, exchangeable and AR(1). Here we extend their results to the familial correlation structure. We also study the ranges of other measures of association such as odds ratios, kappa statistics, and relative risks for familial binary variables.

VI.1.1 Ranges of Correlation Coefficients

To begin with, let y_i, y_j be two binary variables with marginal means p_i, p_j and correlation ρ_{ij}. It is well known (Chaganty and Joe, 2006) that a necessary and sufficient condition for the bivariate binary distribution of (y_i, y_j) to exist is

$$L(p_i,\, p_j) \le \rho_{ij} \le U(p_i,\, p_j), \tag{6.1.1}$$

where

$$\begin{aligned} L(p_i, p_j) &= \max\left\{-\sqrt{\frac{p_i p_j}{q_i q_j}}, -\sqrt{\frac{q_i q_j}{p_i p_j}}\right\} \\ U(p_i, p_j) &= \min\left\{\sqrt{\frac{p_i q_j}{q_i p_j}}, \sqrt{\frac{q_i p_j}{p_i q_j}}\right\} \end{aligned} \tag{6.1.2}$$

for $p_i, p_j \in (0, 1)$ and $q_i = 1 - p_i$, $q_j = 1 - p_j$. Inequalities (6.1.1) can be obtained from the Fréchet bounds applied to bivariate binary distributions.

Familial Correlation

Suppose now y_1, y_2, y_3 are three binary variables with means p_1, p_2, p_3 and familial correlation structure (4.1.1) with $t_i = 2$. We are interested in finding the range of the correlations ρ and α as a function of p_1, p_2, p_3. Let $p_{ij} = \Pr(y_i = 1, y_j = 1) = \mathrm{E}(y_i\, y_j)$, $1 \le i < j \le 3$, and $p_{123} = \Pr(y_1 = y_2 = y_3 = 1) = \mathrm{E}(y_1\, y_2\, y_3)$. With this notation, the eight trivariate probabilities can be written as in Table 6.1. We have the following result for the range of familial correlations for binary random variables. Let $\sigma_i = \sqrt{p_i\, q_i}$, where $q_i = 1 - p_i$.

Table 6.1: Trivariate probability mass function

$y_1\, y_2\, y_3$	Probability
1 1 1	p_{123}
0 1 1	$p_{23} - p_{123}$
1 0 1	$p_{13} - p_{123}$
1 1 0	$p_{12} - p_{123}$
0 0 1	$p_3 - p_{13} - p_{23} + p_{123}$
0 1 0	$p_2 - p_{12} - p_{23} + p_{123}$
1 0 0	$p_1 - p_{12} - p_{13} + p_{123}$
0 0 0	$1 - p_1 - p_2 - p_3 + p_{12} + p_{13} + p_{23} - p_{123}$

Theorem 6.1 *Consider three binary random variables y_1, y_2 and y_3 with means p_1, p_2, p_3 in the interval $(0,\, 1)$ and structured correlation matrix $\mathbf{R}$ given by (4.1.1) with parameters ρ and α and $t_i = 2$. Then a joint distribution for the three binary variables exists if and only if the following two conditions are satisfied:*

(1) $L(p_2, p_3) \leq \alpha \leq U(p_2, p_3)$

(2) $\max\{L(p_1,p_2), L(p_1,p_3), L_1(\alpha)\} \leq \rho \leq \min\{U(p_1,p_2), U(p_1,p_3), U_1(\alpha)\}$

where

$$\begin{aligned} L_1(\alpha) &= -\frac{\alpha\,\sigma_2\sigma_3 + p_1p_2p_3 + q_1q_2q_3}{\sigma_1(\sigma_2+\sigma_3)} \\ U_1(\alpha) &= \frac{\alpha\,\sigma_2\sigma_3 + p_1q_2q_3 + q_1p_2p_3}{\sigma_1(\sigma_2+\sigma_3)}. \end{aligned} \tag{6.1.3}$$

Proof: A necessary and sufficient condition for the existence of the joint distribution is that the eight trivariate probabilities given in Table 6.1 are non-negative. This leads to the condition

$$\begin{aligned} &p_{123L} = \max\{0, p_{12}+p_{13}-p_1,\ p_{12}+p_{23}-p_2,\ p_{13}+p_{23}-p_3\} \\ &\leq p_{123} \leq \min\{p_{12}, p_{13}, p_{23},\ 1-p_1-p_2-p_3+p_{12}+p_{13}+p_{23}\} = p_{123U} \end{aligned}$$

or equivalently

$$\begin{aligned} &p_{123L} = \max\{0, p_{12}+p_{13}-p_1,\ p_{12}+p_{23}-p_2,\ p_{13}+p_{23}-p_3\} \\ &\leq \min\{p_{12}, p_{13}, p_{23},\ 1-p_1-p_2-p_3+p_{12}+p_{13}+p_{23}\} = p_{123U}. \end{aligned} \tag{6.1.4}$$

There are sixteen pairwise inequalities in (6.1.4) given by

$$\begin{array}{lll} (1) & p_{12} \geq 0 & \\ (2) & p_1 \geq p_{12} & \Longleftrightarrow \quad p_{13} \geq p_{12}+p_{13}-p_1 \\ (3) & p_2 \geq p_{12} & \Longleftrightarrow \quad p_{23} \geq p_{12}+p_{23}-p_2 \\ (4) & 1-p_1-p_2+p_{12} \geq 0 & \\ & \Longleftrightarrow \quad 1-p_1-p_2-p_3+p_{12}+p_{13}+p_{23} \geq p_{13}+p_{23}-p_3 & \\ (5) & p_{13} \geq 0 & \\ (6) & p_1 \geq p_{13} & \Longleftrightarrow \quad p_{12} \geq p_{12}+p_{13}-p_1 \\ (7) & p_3 \geq p_{13} & \Longleftrightarrow \quad p_{23} \geq p_{13}+p_{23}-p_3 \\ (8) & 1-p_1-p_3+p_{13} \geq 0 & \\ & \Longleftrightarrow \quad 1-p_1-p_2-p_3+p_{12}+p_{13}+p_{23} \geq p_{12}+p_{23}-p_2 & \\ (9) & p_{23} \geq 0 & \\ (10) & p_2 \geq p_{23} & \Longleftrightarrow \quad p_{12} \geq p_{12}+p_{23}-p_2 \\ (11) & p_3 \geq p_{23} & \Longleftrightarrow \quad p_{13} \geq p_{13}+p_{23}-p_3 \\ (12) & 1-p_2-p_3+p_{23} \geq 0 & \\ & \Longleftrightarrow \quad 1-p_1-p_2-p_3+p_{12}+p_{13}+p_{23} \geq p_{12}+p_{13}-p_1 & \end{array} \tag{6.1.5}$$

$$\begin{array}{ll} (13) & p_{12} \geq p_{13} + p_{23} - p_3 \\ (14) & p_{13} \geq p_{12} + p_{23} - p_2 \\ (15) & p_{23} \geq p_{12} + p_{13} - p_1 \\ (16) & 1 - p_1 - p_2 - p_3 + p_{12} + p_{13} + p_{23} \geq 0\,. \end{array} \tag{6.1.6}$$

Note that $p_{12} = p_1p_2 + \rho\,\sigma_1\sigma_2$, $p_{13} = p_1p_3 + \rho\,\sigma_1\sigma_3$, and $p_{23} = p_2p_3 + \alpha\,\sigma_2\sigma_3$. The first set of inequalities (1), (2), (3) and (4) in (6.1.5) hold if and only if $L(p_1, p_2) \leq \rho \leq U(p_1, p_2)$; the second set of inequalities (5), (6), (7) and (8) in (6.1.5) hold if and only if $L(p_1, p_3) \leq \rho \leq U(p_1, p_3)$; the third set of inequalities (9), (10), (11) and (12) in (6.1.5) hold if and only if $L(p_2, p_3) \leq \alpha \leq U(p_2, p_3)$. Inequalities (15) and (16) in (6.1.6) hold if and only if $L_1(\alpha) \leq \rho \leq U_1(\alpha)$. As for inequalities (13) and (14) in (6.1.6), if we define

$$\begin{aligned} L_2(\alpha) &= \frac{\alpha\,\sigma_2\sigma_3 - p_1p_2q_3 - q_1q_2p_3}{\sigma_1(\sigma_2 - \sigma_3)} \\ U_2(\alpha) &= \frac{-\alpha\,\sigma_2\sigma_3 + p_1q_2p_3 + q_1p_2q_3}{\sigma_1(\sigma_2 - \sigma_3)} \\ V_1 &= \frac{p_1p_2q_3 + q_1q_2p_3}{\sigma_2\sigma_3} \\ V_2 &= \frac{p_1q_2p_3 + q_1p_2q_3}{\sigma_2\sigma_3} \end{aligned}$$

then there exist three possibilities: (a) if $\sigma_2 > \sigma_3$, then inequalities (13) and (14) hold if and only if $L_2(\alpha) \leq \rho \leq U_2(\alpha)$; (b) if $\sigma_2 < \sigma_3$, then inequalities (13) and (14) hold if and only if $U_2(\alpha) \leq \rho \leq L_2(\alpha)$; (c) if $\sigma_2 = \sigma_3$ then inequalities (13) and (14) hold if and only if $\alpha \leq \min\{V_1, V_2\}$.

Now we will show that for $L(p_2, p_3) \leq \alpha \leq U(p_2, p_3)$, the conditions given by inequalities (13) and (14) are redundant under each case. Without loss of generality, we first assume (a) $\sigma_2 > \sigma_3$, we will show that $U_2(\alpha)$ always lies above $\min\{U(p_1, p_2), U(p_1, p_3)\}$ and $L_2(\alpha)$ always lies below $\max\{L(p_1, p_2), L(p_1, p_3)\}$. Note that $U_2(\alpha)$ and $L_2(\alpha)$ are linearly decreasing and increasing in α, respectively. We only need to show that at the point $\alpha^* = U(p_2, p_3)$, $U_2(\alpha^*)$ is greater than any of the bounds $U(p_1, p_2)$ and $U(p_1, p_3)$ and $L_2(\alpha^*)$ is less than any of the bounds $L(p_1, p_2)$ and $L(p_1, p_3)$. In Table 6.2 and Table 6.3, we give the upper and lower bounds for ρ from conditions $U(p_1, p_2)$, $U(p_1, p_3)$ and $L(p_1, p_2)$, $L(p_1, p_3)$ under different situations, which will help us in comparing the different quantities for the upper or lower bounds of ρ.

Table 6.2: $\min\{U(p_1,p_2),U(p_1,p_3)\}$ *for* ρ *(*$\sigma_2 > \sigma_3$*)*

	$p_1 \le q_1$	$p_1 > q_1$
$p_2 < p_3$	$\sqrt{\frac{p_1q_3}{q_1p_3}}$	$\min\left\{\sqrt{\frac{q_1p_2}{p_1q_2}},\sqrt{\frac{p_1q_3}{q_1p_3}}\right\}$
$p_2 > p_3$	$\min\left\{\sqrt{\frac{p_1q_2}{q_1p_2}},\sqrt{\frac{q_1p_3}{p_1q_3}}\right\}$	$\sqrt{\frac{q_1p_3}{p_1q_3}}$

Table 6.3: $\max\{L(p_1,p_2),L(p_1,p_3)\}$ *for* ρ *(*$\sigma_2 > \sigma_3$*)*

	$p_1 \le q_1$	$p_1 > q_1$
$p_2 < p_3$	$\max\left\{-\sqrt{\frac{p_1p_2}{q_1q_2}},-\sqrt{\frac{q_1q_3}{p_1p_3}}\right\}$	$-\sqrt{\frac{q_1q_3}{p_1p_3}}$
$p_2 > p_3$	$-\sqrt{\frac{p_1p_3}{q_1q_3}}$	$\max\left\{-\sqrt{\frac{q_1q_2}{p_1p_2}},-\sqrt{\frac{p_1p_3}{q_1q_3}}\right\}$

When $p_2 < p_3$ or equivalently $p_2q_3 < p_3q_2$, we have $\alpha^* = U(p_2,p_3) = \sqrt{\frac{p_2q_3}{q_2p_3}}$, then

$$\begin{aligned}
U_2(\alpha^*) - \sqrt{\frac{p_1q_3}{q_1p_3}} &= \frac{-p_2q_3 + p_1p_3q_2 + q_1q_3p_2}{\sigma_1(\sigma_2-\sigma_3)} - \sqrt{\frac{p_1q_3}{q_1p_3}} \\
&= \frac{p_1(p_3q_2 - p_2q_3) - p_1\left(\sqrt{p_2q_2q_3/p_3} - q_3\right)}{\sigma_1(\sigma_2-\sigma_3)} \\
&= \frac{p_1q_2 - p_1\sqrt{p_2q_2q_3/p_3}}{\sigma_1(\sigma_2-\sigma_3)} \\
&= \frac{p_1\sqrt{q_2}\left(\sqrt{p_3q_2} - \sqrt{p_2q_3}\right)}{\sigma_1(\sigma_2-\sigma_3)\sqrt{p_3}} > 0\,.
\end{aligned}$$

This indicates that $U_2(\alpha)$ is at least greater than one of the upper bounds $U(p_1,p_2)$ and $U(p_1,p_3)$, which further indicates $U_2(\alpha) > \min\{U(p_1,p_2),U(p_1,p_3)\}$ for $L(p_2,p_3) \le \alpha \le U(p_2,p_3)$.

When $p_2 > p_3$ or equivalently $p_2q_3 > p_3q_2$, we have $\alpha^* = U(p_2,p_3) = \sqrt{\frac{p_3q_2}{q_3p_2}}$, then

$$U_2(\alpha^*) - \sqrt{\frac{q_1p_3}{p_1q_3}} = \frac{-p_3q_2 + p_1p_3q_2 + q_1q_3p_2}{\sigma_1(\sigma_2-\sigma_3)} - \sqrt{\frac{q_1p_3}{p_1q_3}}$$

$$
\begin{aligned}
&= \frac{q_1(p_2q_3 - p_3q_2) - q_1\left(\sqrt{p_2q_2p_3/q_3} - p_3\right)}{\sigma_1(\sigma_2-\sigma_3)} \\
&= \frac{q_1p_2 - q_1\sqrt{p_2q_2p_3/q_3}}{\sigma_1(\sigma_2-\sigma_3)} \\
&= \frac{p_1\sqrt{p_2}\left(\sqrt{p_2q_3} - \sqrt{p_3q_2}\right)}{\sigma_1(\sigma_2-\sigma_3)\sqrt{q_3}} > 0\,,
\end{aligned}
$$

which indicates that $U_2(\alpha)$ is at least greater than one of the upper bounds $U(p_1, p_2)$ and $U(p_1, p_3)$, which further indicates $U_2(\alpha) > \min\{U(p_1, p_2), U(p_1, p_3)\}$ for $L(p_2, p_3) \le \alpha \le U(p_2, p_3)$.

Similar argument shows that $L_2(\alpha) < \max\{L(p_1, p_2), L(p_1, p_3)\}$. Here we only give the main steps. When $p_2 > p_3$ or equivalently $p_2q_3 > p_3q_2$, we have $\alpha^* = U(p_2, p_3) = \sqrt{\dfrac{q_3p_2}{p_3q_2}}$, and

$$
\begin{aligned}
L_2(\alpha^*) - \left\{-\sqrt{\frac{q_1q_3}{p_1p_3}}\right\} &= \frac{p_2q_3 - p_1p_2q_3 - q_1q_2p_3}{\sigma_1(\sigma_2-\sigma_3)} + \sqrt{\frac{q_1q_3}{p_1p_3}} \\
&= \frac{q_1(p_2q_3 - p_3q_2) + q_1\left(\sqrt{p_2q_2q_3/p_3} - q_3\right)}{\sigma_1(\sigma_2-\sigma_3)} \\
&= \frac{-q_1\sqrt{q_2}\left(\sqrt{p_3q_2} - \sqrt{p_2q_3}\right)}{\sigma_1(\sigma_2-\sigma_3)\sqrt{p_3}} < 0\,.
\end{aligned}
$$

When $p_2 < p_3$ or equivalently $p_2q_3 < p_3q_2$, we have $\alpha^* = U(p_2, p_3) = \sqrt{\dfrac{p_3q_2}{q_3p_2}}$. Thus

$$
\begin{aligned}
L_2(\alpha^*) - \left\{-\sqrt{\frac{p_1p_3}{q_1q_3}}\right\} &= \frac{p_3q_2 - p_1p_2q_3 - q_1q_2p_3}{\sigma_1(\sigma_2-\sigma_3)} + \sqrt{\frac{p_1p_3}{q_1q_3}} \\
&= \frac{p_1(p_3q_2 - p_2q_3) + p_1\left(\sqrt{p_2q_2p_3/q_3} - p_3\right)}{\sigma_1(\sigma_2-\sigma_3)} \\
&= \frac{-p_1\sqrt{p_2}\left(\sqrt{p_2q_3} - \sqrt{p_3q_2}\right)}{\sigma_1(\sigma_2-\sigma_3)\sqrt{q_3}} < 0\,,
\end{aligned}
$$

and therefore $L_2(\alpha) \le \rho \le U_2(\alpha)$ always yields redundant constrains when $\sigma_2 > \sigma_3$. The case when $\sigma_2 < \sigma_3$ follows by symmetry. Now let us assume (c) $\sigma_2 = \sigma_3 = \sigma$, the conditions are $\alpha \le \min\{V_1, V_2\}$, which does not constrain ρ. It is easy to verify that if (i) $p_2 < 1/2$ and $p_2 \le p_3$, then $U(p_2, p_3) = \sqrt{\dfrac{p_2q_3}{q_2p_3}}$ and

$$
\begin{aligned}
V_1 - U(p_2, p_3) &= \frac{q_1(p_3q_2 - p_2q_3)}{\sigma^2} > 0 \\
V_2 - U(p_2, p_3) &= \frac{p_1(p_3q_2 - p_2q_3)}{\sigma^2} > 0
\end{aligned}
$$

and (ii) $p_2 > 1/2$ and $p_2 \geq p_3$, then $U(p_2, p_3) = \sqrt{\dfrac{q_2 p_3}{p_2 q_3}}$ and

$$\begin{aligned} V_1 - U(p_2, p_3) &= \frac{p_1(p_2 q_3 - p_3 q_2)}{\sigma^2} > 0 \\ V_2 - U(p_2, p_3) &= \frac{q_1(p_2 q_3 - p_3 q_2)}{\sigma^2} > 0 . \end{aligned}$$

This means that V_1 and V_2 always yield redundant constraints on α. Therefore, we can conclude that the necessary and sufficient conditions for the three binary random variables to exist are as stated in Theorem 6.1. This completes the proof of the theorem $\diamond$

It is interesting to note that when $p_2 = p_3 = p$, $\sigma = \sqrt{pq}$ and $q = 1 - p$, the necessary and sufficient conditions in Theorem 6.1 further reduce to

(1) $\max\{-p/q, -q/p\} \leq \alpha \leq 1$

(2) $\max\left\{L(p_1, p), \dfrac{-(\alpha\sigma^2 + p_1 p^2 + q_1 q^2)}{2\sigma_1\sigma}\right\} \leq \rho \leq \min\left\{U(p_1, p), \dfrac{\alpha\sigma^2 + p_1 q^2 + q_1 p^2}{2\sigma_1\sigma}\right\}$

and when $p_2 = 1 - p_3 = p$, we have

(1) $-1 \leq \alpha \leq \min\{p/q,\, q/p\}$

(2) $\max\left\{L(p_1, q), \dfrac{-\sigma(1+\alpha)}{2\sigma_1}\right\} \leq \rho \leq \min\left\{U(p_1, q), \dfrac{\sigma(1+\alpha)}{2\sigma_1}\right\}$.

Further, when $p_1 = p_2 = p_3 = p$, a trivariate binary distribution for $\mathbf{y}$ with correlation structure (4.1.1) exists if and only if

(1) $\max\{-p/q,\, -q/p\} \leq \alpha \leq 1$

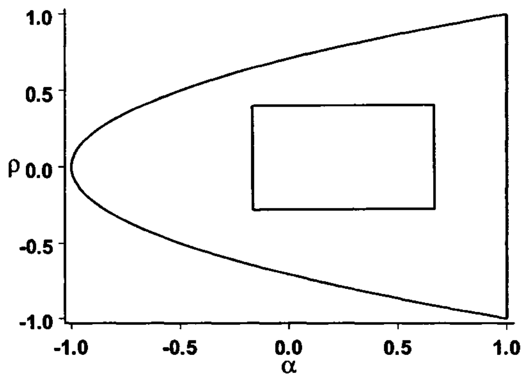
(2) $\max\left\{-p/q, -q/p, \dfrac{-(\alpha pq + p^3 + q^3)}{2pq}\right\} \leq \rho \leq \dfrac{1+\alpha}{2}$.

More specially, if $p = 1/2$, then the necessary and sufficient conditions become

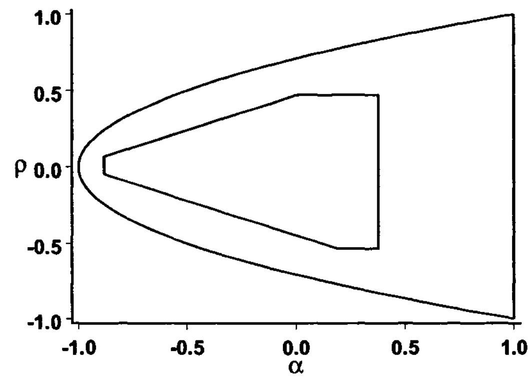
(1) $-1 \leq \alpha \leq 1$ (2) $\dfrac{-(1+\alpha)}{2} \leq \rho \leq \dfrac{1+\alpha}{2}$.

These ranges form a proper subset of the constraints (4.1.2) given in Section IV.1, which are also the bounds for ρ and α for Gaussian random variables. Note that the two dimensional region of the ranges of ρ and α given by (4.1.2) contain the region given by (1) and (2) in Theorem 6.1 for any $\mathbf{p} = (p_1, p_2, p_3)$.

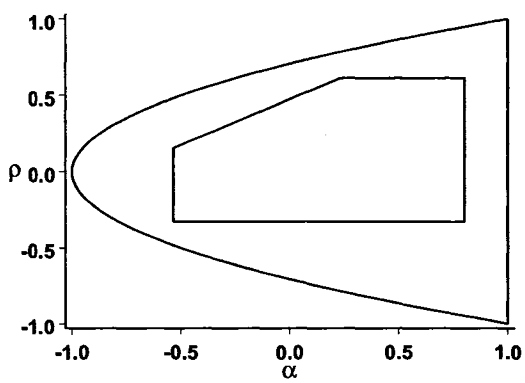
Figure 6.1 and 6.2 show the feasible regions of $(\alpha,\ \rho)$ for different $\mathbf{p}$ and special cases when $\sigma_2 = \sigma_3$. For Gaussian variables, feasible region is the area enclosed by the parabola, whereas the embedded figure within the parabola is the feasible region for the binary variables.



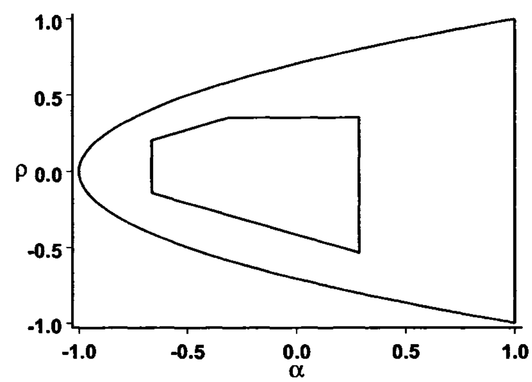
(a) $p_1 = 0.4$, $p_2 = 0.2$, $p_3 = 0.1$



(b) $p_1 = 0.4$, $p_2 = 0.3$, $p_3 = 0.75$



(c) $p_1 = 0.2$, $p_2 = 0.4$, $p_3 = 0.3$



(d) $p_1 = 0.4$, $p_2 = 0.3$, $p_3 = 0.84$

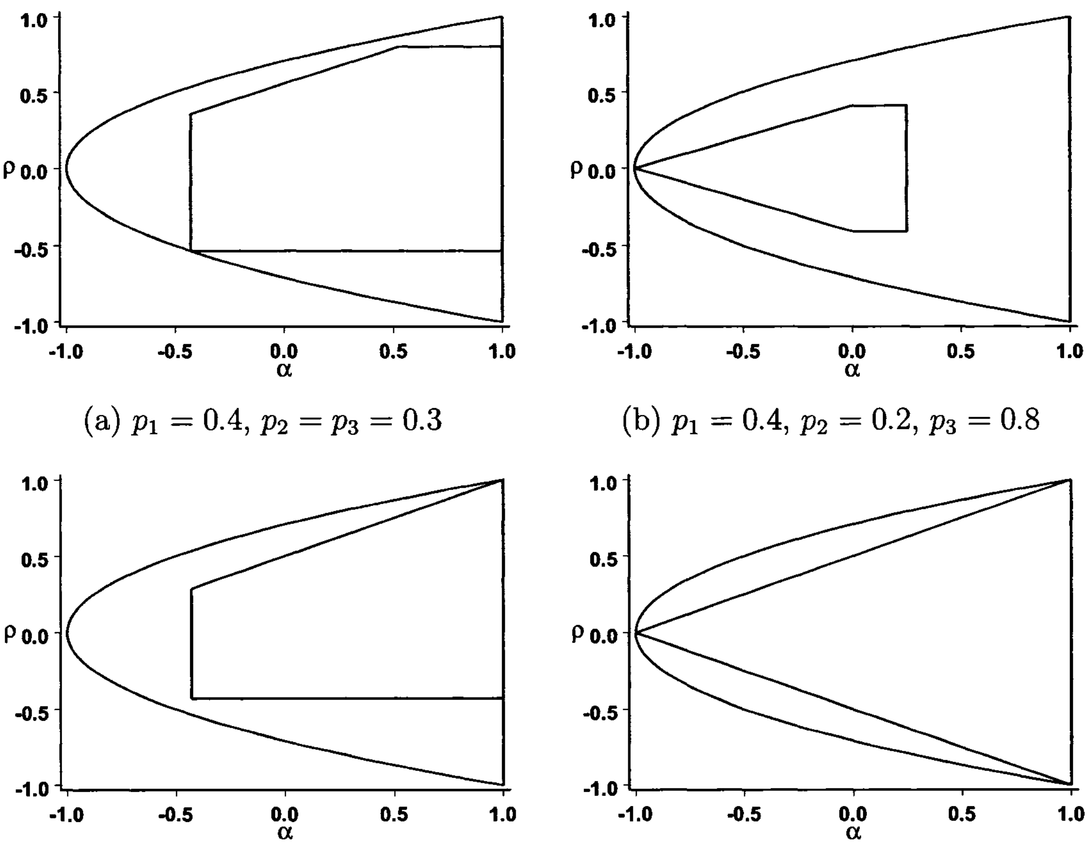
Figure 6.1: Region of (α, ρ) for familial structure.

For a given α, the unattainable range of ρ for binary variables is

$$-\sqrt{(1+\alpha)/2} < \rho < \min_{\mathbf{p}\in\mathcal{A}} \max\{L(p_1,p_2), L(p_1,p_3), L_1(\alpha)\}$$

or

$$\max_{\mathbf{p}\in\mathcal{A}} \min\{U(p_1,p_2), U(p_1,p_3), U_1(\alpha)\} < \rho < \sqrt{(1+\alpha)/2}$$



(a) $p_1 = 0.4$, $p_2 = p_3 = 0.3$

(b) $p_1 = 0.4$, $p_2 = 0.2$, $p_3 = 0.8$

(c) $p_1 = p_2 = p_3 = 0.3$

(d) $p_1 = p_2 = p_3 = 0.5$

Figure 6.2: Region of (α, ρ) for familial structure: special cases.

where $\mathcal{A} = [\mathbf{p} \in (0,1)^3 : L(p_2, p_3) \leq \alpha \leq U(p_2, p_3)]$. Table 6.4 contains the range of ρ, computed numerically, that is unattainable by familial binary variables for given values of α.

A Variant of Familial Correlation

Suppose that the parent-sibling correlation has an auto-regressive pattern. In this case a reasonable model for the familial correlations between a parent and the first and second child is the structure

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \alpha \\ \rho^2 & \alpha & 1 \end{bmatrix}. \tag{6.1.7}$$

For Gaussian variables, the ranges of parameters ρ and α for the structure (6.1.7)

Table 6.4: Unattainable range of ρ for given α

α	Unattainable range of ρ
-0.9	(-0.2236, -0.2205) ∪ (0.2212, 0.2236)
-0.6	(-0.4472, -0.4448) ∪ (0.4460, 0.4472)
-0.3	(-0.5916, -0.5892) ∪ (0.5904, 0.5916)
0.0	(-0.7071, -0.6988) ∪ (0.7013, 0.7071)
0.1	(-0.7416, -0.7202) ∪ (0.7201, 0.7416)
0.2	(-0.7746, -0.7416) ∪ (0.7418, 0.7746)
0.3	(-0.8062, -0.7616) ∪ (0.7613, 0.8062)
0.4	(-0.8367, -0.7877) ∪ (0.7871, 0.8367)
0.5	(-0.8660, -0.8158) ∪ (0.8143, 0.8660)
0.6	(-0.8944, -0.8428) ∪ (0.8433, 0.8944)
0.7	(-0.9220, -0.8762) ∪ (0.8748, 0.9220)
0.8	(-0.9487, -0.9119) ∪ (0.9123, 0.9487)
0.9	(-0.9747, -0.9533) ∪ (0.9531, 0.9747)

are given by

$$-1 < \rho < 1 \text{ and } \rho^3 - (1-\rho^2)\sqrt{1+\rho^2} < \alpha < \rho^3 + (1-\rho^2)\sqrt{1+\rho^2}\,,$$

or equivalently,

$$\begin{array}{lll} & -\frac{5\sqrt{3}}{9} < \alpha < \frac{5\sqrt{3}}{9} & \text{and} \quad T_{1S}(\alpha) < \rho < T_{2L}(\alpha)\,, \\ \text{or} & -1 < \alpha \le -\frac{5\sqrt{3}}{9} & \text{and} \quad T_{1S}(\alpha) < \rho < T_{2S}(\alpha);\ T_{2M}(\alpha) < \rho < T_{2L}(\alpha)\,, \\ \text{or} & \frac{5\sqrt{3}}{9} \le \alpha < 1 & \text{and} \quad T_{1S}(\alpha) < \rho < T_{1M}(\alpha);\ T_{1L}(\alpha) < \rho < T_{2L}(\alpha)\,, \end{array}$$

where T_{1S}, T_{1M}, T_{1L} and T_{2S}, T_{2M}, T_{2L} are the roots of the equations $\rho^3 + \sqrt{\rho^6 - \rho^4 - \rho^2 + 1} = \alpha$ and $\rho^3 - \sqrt{\rho^6 - \rho^4 - \rho^2 + 1} = \alpha$ for fixed value of α in the range of $(-1, 1)$, respectively. Note that $T_{1S} < T_{1M} < T_{1L}$ and $T_{2S} < T_{2M} < T_{2L}$.

We have the following result for the ranges of ρ and α for binary variables.

Theorem 6.2 *Let* $\mathbf{y} = (y_1, y_2, y_3)$ *be a binary vector with mean* $\mathbf{p} = (p_1, p_2, p_3)$, $0 < p_i < 1$, *and correlation matrix* $\mathbf{R}$ *given by (6.1.7) with parameters* ρ *and* α. *Then a joint binary distribution for* $\mathbf{y}$ *exists if and only if the following two conditions are satisfied:*

(1) $\max\{L(p_1, p_2), -\sqrt{U(p_1, p_3)}\} \le \rho \le \min\{U(p_1, p_2), \sqrt{U(p_1, p_3)}\}$

(2) $\max\{L(p_2,p_3), L_1(\rho), L_2(\rho)\} \leq \alpha \leq \min\{U(p_2,p_3), U_1(\rho), U_2(\rho)\}$,

where

$$\begin{aligned}
L_1(\rho) &= (\sigma_1\sigma_3\rho^2 + \sigma_1\sigma_2\rho - p_1q_2q_3 - q_1p_2p_3)/\sigma_2\sigma_3 \\
L_2(\rho) &= (-\sigma_1\sigma_3\rho^2 - \sigma_1\sigma_2\rho - p_1p_2p_3 - q_1q_2q_3)/\sigma_2\sigma_3 \\
U_1(\rho) &= (-\sigma_1\sigma_3\rho^2 + \sigma_1\sigma_2\rho + p_1p_2q_3 + q_1q_2p_3)/\sigma_2\sigma_3 \\
U_2(\rho) &= (\sigma_1\sigma_3\rho^2 - \sigma_1\sigma_2\rho + p_1q_2p_3 + q_1p_2q_3)/\sigma_2\sigma_3\,,
\end{aligned}$$

or alternatively,

(i) $L(p_2,p_3) \leq \alpha \leq U(p_2,p_3)$

(ii) $\max\{L(p_1,p_2), -\sqrt{U(p_1,p_3)}\,, L_1'(\alpha)\} \leq \rho \leq \min\{U(p_1,p_2), \sqrt{U(p_1,p_3}\}, U_1'(\alpha)\}$ *and* $\rho \leq L_2'(\alpha)$, $\rho \geq U_2'(\alpha)$; $\rho \leq L_3'(\alpha)$, $\rho \geq U_3'(\alpha)$,

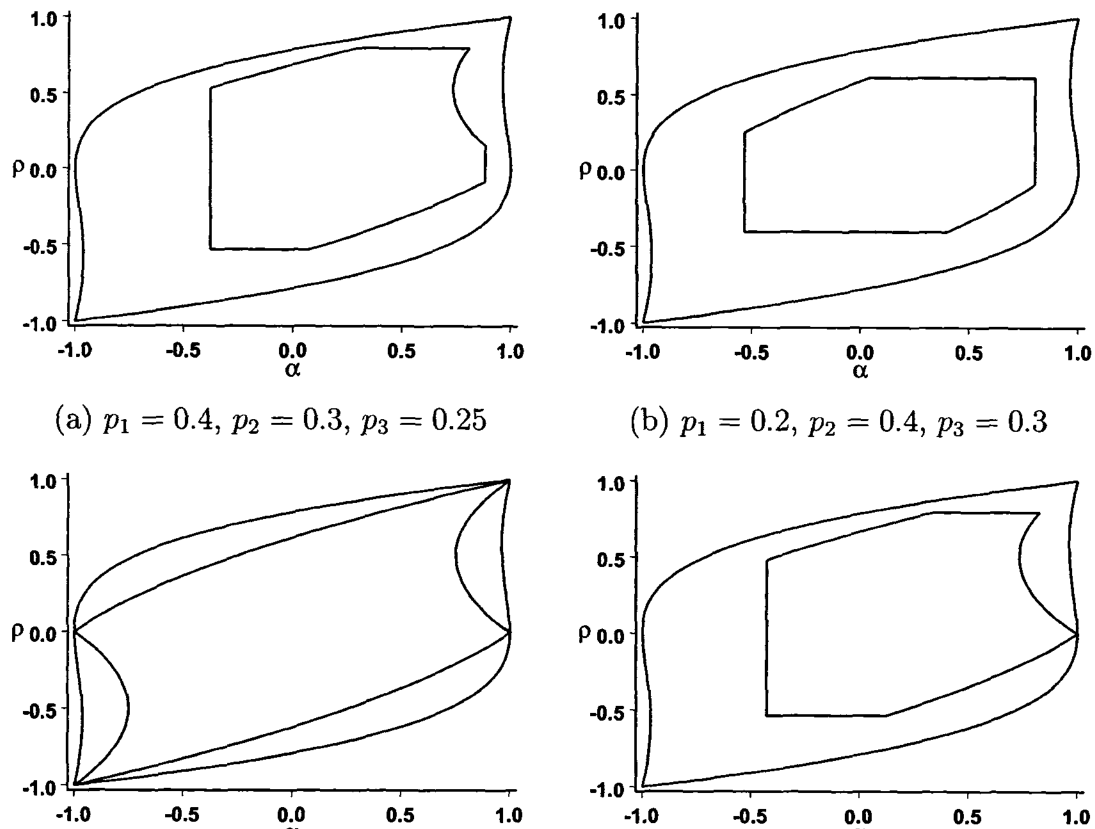
where

$$\begin{aligned}
L_1'(\alpha) &= \frac{\sigma_1\sigma_2 - \sqrt{\sigma_1^2\sigma_2^2 - 4\sigma_1\sigma_3(\sigma_2\sigma_3\alpha - p_1p_2q_3 - q_1q_2p_3)}}{2\sigma_1\sigma_3} \\
L_2'(\alpha) &= \frac{\sigma_1\sigma_2 - \sqrt{\sigma_1^2\sigma_2^2 + 4\sigma_1\sigma_3(\sigma_2\sigma_3\alpha - p_1q_2p_3 - q_1p_2q_3)}}{2\sigma_1\sigma_3} \\
L_3'(\alpha) &= \frac{-\sigma_1\sigma_2 - \sqrt{\sigma_1^2\sigma_2^2 - 4\sigma_1\sigma_3(\sigma_2\sigma_3\alpha + p_1p_2p_3 + q_1q_2q_3)}}{2\sigma_1\sigma_3} \\
U_1'(\alpha) &= \frac{-\sigma_1\sigma_2 + \sqrt{\sigma_1^2\sigma_2^2 + 4\sigma_1\sigma_3(\sigma_2\sigma_3\alpha + p_2p_3q_1 - q_2q_3p_1)}}{2\sigma_1\sigma_3} \\
U_2'(\alpha) &= \frac{\sigma_1\sigma_2 + \sqrt{\sigma_1^2\sigma_2^2 + 4\sigma_1\sigma_3(\sigma_2\sigma_3\alpha - p_1q_2p_3 - q_1p_2q_3)}}{2\sigma_1\sigma_3} \\
U_3'(\alpha) &= \frac{-\sigma_1\sigma_2 + \sqrt{\sigma_1^2\sigma_2^2 - 4\sigma_1\sigma_3(\sigma_2\sigma_3\alpha + p_1p_2p_3 + q_1q_2q_3)}}{2\sigma_1\sigma_3}\,.
\end{aligned}$$

The proof of this theorem parallels the proof of Theorem 6.1. A necessary and sufficient condition for the existence of the joint distribution is that the eight trivariate probabilities given in Table 6.1 are non-negative. Note that $p_{12} = p_1\,p_2 + \rho\,\sigma_1\sigma_2$, $p_{13} = p_1\,p_3 + \rho^2\,\sigma_1\,\sigma_3$, and $p_{23} = p_2\,p_3 + \alpha\,\sigma_2\,\sigma_3$. Using the notation of 16 pairwise inequalities in (6.1.5) and (6.1.6), we have $L(p_1,p_2) \leq \rho \leq U(p_1,p_2)$ from (1), (2), (3) and (4); $L(p_1,p_3) \leq \rho^2 \leq U(p_1,p_3)$ or $\sqrt{U(p_1,p_3)} \leq \rho^2 \leq \sqrt{U(p_1,p_3)}$ from (5), (6), (7) and (8); $L(p_2,p_3) \leq \alpha \leq U(p_2,p_3)$ from (9), (10), (11) and (12);

$\max\{L_1(\alpha), L_2(\alpha)\} \leq \rho \leq \min\{U_1(\alpha), U_2(\alpha)\}$ from (13), (14), (15) and (16). A rearrangement of these conditions results in the other expression. This completes the proof of the theorem. $\diamond$

Figure 6.3 shows the permissible range of (α, ρ) for different values of $\mathbf{p}$. The feasible region for Gaussian variables is the area enclosed by the outer curve, and the embedded curve contains the feasible region for binary variables.



(a) $p_1 = 0.4$, $p_2 = 0.3$, $p_3 = 0.25$

(b) $p_1 = 0.2$, $p_2 = 0.4$, $p_3 = 0.3$

(c) $p_1 = p_2 = p_3 = 0.5$

(d) $p_1 = 0.4$, $p_2 = p_3 = 0.3$

Figure 6.3: Region of (α, ρ) for variant familial structure.

Theorem 6.2 shows that for a fixed $\rho \in (-1, 1)$, the range of α unattainable by binary distributions with correlation structure (6.1.7) is given by

$$\rho^3 - (1-\rho^2)\sqrt{1+\rho^2} < \alpha < \min_{\mathbf{p} \in \mathcal{B}} \max\{L(p_2, p_3), L_1(\rho), L_2(\rho)\}$$

or

$$\max_{\mathbf{p} \in \mathcal{B}} \min\{U(p_2, p_3), U_1(\rho), U_2(\rho)\} < \alpha < \rho^3 + (1-\rho^2)\sqrt{1+\rho^2}$$

where
$\mathcal{B} = \left[\mathbf{p} \in (0,1)^3 : \max\{L(p_1,p_2), -\sqrt{U(p_1,p_3)}\} \leq \rho \leq \min\{U(p_1,p_2), \sqrt{U(p_1,p_3)}\}\right]$.

Table 6.5 contains the unattainable range of α calculated numerically for some given values of ρ. The unattainable range of α when $\rho < 0$ has a similar pattern.

Table 6.5: Unattainable range of α for given ρ

ρ	Unattainable range of α
0.1	(-0.9939, -0.9923) ∪ (0.9934, 0.9959)
0.2	(-0.9710, -0.9689) ∪ (0.9807, 0.9870)
0.3	(-0.9231, -0.9210) ∪ (0.9635, 0.9771)
0.4	(-0.8407, -0.8364) ∪ (0.9438, 0.9687)
0.5	(-0.7135, -0.7101) ∪ (0.9258, 0.9635)
0.6	(-0.5303, -0.5271) ∪ (0.9056, 0.9624)
0.7	(-0.2795, -0.2722) ∪ (0.8996, 0.9655)
0.8	(0.0510, 0.1421) ∪ (0.9024, 0.9730)
0.9	(0.4734, 0.6460) ∪ (0.9301, 0.9846)

Extension to Nuclear Familial Correlation

A natural extension would be to consider a nuclear family with two parents and more than two children. For example, we could consider correlation structure of the form (5.1.1) given in Section V.1. Recall that the necessary and sufficient conditions for (5.1.1) to be positive definite are

(1) $-1 < \gamma < 1$

(2) $-\frac{1}{t_i - 1} < \alpha < 1$

(3) $t_i(\rho_1^2 + \rho_2^2 - 2\gamma\rho_1\rho_2) < (1-\gamma^2)[1 + (t_i - 1)\alpha]$.

The conditions for the existence of a joint distribution for binary random variables with correlation structure (5.1.1) are complicated. However, note that the results in Section V.1 are necessary conditions for the existence of the three dimensional marginal binary distributions, corresponding to the sub-correlation matrices.

VI.1.2 Ranges of Odds Ratios

An alternative measure of association for binary variables is the odds ratio, which is less constrained than the correlations. The odds ratio for a pair of binary random variables y_i and y_j is defined as (Agresti, 2002)

$$\psi_{ij} = \frac{P(y_i = 1, y_j = 1)P(y_i = 0, y_j = 0)}{P(y_i = 1, y_j = 0)P(y_i = 0, y_j = 1)} = \frac{p_{ij}(1 - p_i - p_j + p_{ij})}{(p_i - p_{ij})(p_j - pij)}. \qquad (6.1.8)$$

Note that $E(y_i y_j) = p_{ij} = p_{ij}(\psi_{ij}) = C(p_i, p_j, \psi_{ij})$ where for fixed ψ, the function $C(u, v, \psi)$ is the Plackett copula (Joe, 1997) given by

$$C(u, v, \psi) = \begin{cases} \dfrac{1 + (u+v)(\psi - 1) - \sqrt{[1 + (u+v)(\psi - 1)]^2 - 4\psi(\psi - 1)uv}}{2(\psi - 1)} & \text{if } \psi \neq 1 \\ uv & \text{if } \psi = 1 . \end{cases}$$

For a fixed ψ, $C(u, v, \psi)$ is simply a bivariate distribution function with uniform $(0, 1)$ margins. The next theorem gives feasible ranges for the familial odds ratios.

Theorem 6.3 *Let y_1 and (y_2, y_3) be binary outcomes on a parent and two children, respectively. Suppose ψ is the common odds ratio between the parent and the two children, and let ψ_0 be the odds ratio between the children. A trivariate binary distribution for $\mathbf{y} = (y_1, y_2, y_3)$ with mean $\mathbf{p} = (p_1, p_2, p_3)$, $0 < p_i < 1$, exists if and only if*

$$0 \leq \psi_0 < \infty \text{ and } \psi_L(\mathbf{p}, \psi_0) \leq \psi \leq \psi_U(\mathbf{p}, \psi_0), \qquad (6.1.9)$$

where the lower bound $\psi_L(\mathbf{p}, \psi_0) = 0$ if $p_1 + p_2 + p_3 \leq 1 + p_{23}(\psi_0)$ or $p_1 + p_{23}(\psi_0) \geq 1$. Otherwise, $\psi_L(\mathbf{p}, \psi_0)$ is the positive root of the equation $1 - p_1 - p_2 - p_3 + p_{12}(x) + p_{13}(x) + p_{23}(\psi_0) = 0$. The upper bound $\psi_U(\mathbf{p}, \psi_0) = \infty$ if $p_1 \leq p_{23}(\psi_0)$ or $p_2 + p_3 \leq p_1 + p_{23}(\psi_0)$. Otherwise $\psi_U(\mathbf{p}, \psi_0)$ is the positive root of the equation $p_{23}(\psi_0) - p_{12}(x) - p_{13}(x) + p_1 = 0$.

Proof. It is well known that the range of ψ_{ij} as a function of p_{ij} is $[0, \infty)$, clearly, the range for ψ_0 is $[0, \infty)$. Notice that $p_{ij}(\psi) = C(p_i, p_j, \psi)$ is increasing in its first two arguments and is also increasing in ψ for $\psi \in [0, \infty)$, and $\lim_{\psi \to \infty} p_{ij}(\psi) = \min(p_i, p_j)$. We will show that inequalities (13) and (14) in (6.1.5) always hold for $0 \leq \psi < \infty$. Let $g(\psi) = p_{12} - p_{13} + p_3 - p_{23}$, if $p_2 > p_3$, then $C(p_1, p_2, \psi) > C(p_1, p_3, \psi)$ and $p_3 > C(p_2, p_3, \psi_0)$ and therefore $g(\psi) > 0$; if $p_1 > p_3$, then $C(p_1, p_2, \psi) > C(p_3, p_2, \psi)$

and $p_3 > C(p_1, p_3, \psi_0)$ and therefore $g(\psi) > 0$. Now we only need to prove $g(\psi) > 0$ when $p_3 > \max\{p_1, p_2\}$. Imitating the proof in Chaganty and Joe (2006), let $(U, V) \sim C(u, v, \psi)$, then $\partial C(u, v, \psi)/\partial u = \Pr(V \leq v | U = u)$ is increasing in v. Consider $g = g(p_1)$ as a function of p_1 varying in $[0, p_3)$ for fixed values of p_2, p_3 and ψ, we have $g(0) = p_3 - p_{23} \geq 0$ and $g(p_3) = p_3 - C(p_2, p_3, \psi) \geq 0$. Since $p_3 > p_2$, we also have

$$
\begin{aligned}
\frac{\partial g(p_1)}{\partial p_1} &= \frac{\partial p_{12}}{\partial p_1} - \frac{\partial p_{13}}{\partial p_1} \\
&= \Pr(V \leq p_2 | U = p_1) - \Pr(V \leq p_3 | U = p_1) \leq 0\,.
\end{aligned}
$$

Thus g is nonnegative for all $p_3 \geq \max\{p_1, p_2\}$ and ψ. This proves (13). By symmetry, inequality (14) also holds for $0 \leq \psi < \infty$.

Also notice that for (15) in (6.1.5), $k(\psi) = p_{23} - p_{12} - p_{13} + p_1$ is decreasing in ψ and for (16) in (6.1.5), $l(\psi) = 1 - p_1 - p_2 - p_3 + p_{12} + p_{13} + p_{23}$ is increasing in ψ for fixed value of ψ_0. In order for $k(\psi)$ and $l(\psi)$ to be nonnegative for all $\psi \in [0, \infty)$, we must have $k(\infty) \geq 0$ and $l(0) \geq 0$. Therefore,

$$
k(\infty) \geq 0 \iff \begin{cases} p_{23} \geq p_1 \\ p_1 + p_{23} \geq p_2 + p_3 \end{cases}
$$

$$
l(0) \geq 0 \iff \begin{cases} p_1 + p_{23} \geq 1 \\ p_1 + p_2 + p_3 \leq 1 + p_{23}\,. \end{cases}
$$

In addition, if $k(\psi)$ and $l(\psi)$ are not always nonnegative for $\psi \in [0, \infty)$, then (15) and (16) in (6.1.5) hold if and only if $\psi_L(p_1, p_2, p_3) \leq \psi < \psi_U(p_1, p_2, p_3)$. The other twelve pairwise inequalities in (6.1.5) hold trivially for $0 \leq \psi_0 < \infty$ and $0 \leq \psi < \infty$. This completes the proof of the theorem. $\diamond$

Table 6.6 contains the range of ψ given p_1, p_2, p_3, and ψ_0 for the different cases that can occur.

VI.1.3 Ranges of Kappa Statistics

Another measure of association between two binary variables is the kappa statistic. Let y_i and y_j be two binary random variables with marginal means p_i and p_j. The kappa statistic κ_{ij} is defined as

$$
\kappa_{ij} = \frac{2(p_{ij} - p_i p_j)}{p_i + p_j - 2p_i p_j} = \frac{2(p_{ij} - p_i p_j)}{p_i q_j + p_j q_i}\,.
$$

Table 6.6: Bounds for the common odds ratio

p_1	p_2	p_3	ψ_0	p_{23}	ψ_L	ψ_U
0.3	0.1	0.2	2	0.0315	0	∞
0.1	0.3	0.4	1.5	0.1407	0	∞
0.4	0.2	0.3	0.5	0.0390	0	24.5390
0.5	0.8	0.75	1.5	0.6128	0	∞
0.85	0.6	0.65	2	0.4281	0	∞
0.65	0.8	0.75	0.5	0.5816	0	22.0627
0.3	0.5	0.7	2	0.3859	0.0537	∞
0.6	0.3	0.35	4	0.1727	0.0265	∞
0.5	0.3	0.4	0.5	0.0866	0.0866	11.5492

See Agresti (2002). Since $\max\{0, p_i + p_j - 1\} \leq p_{ij} \leq \min\{p_i, p_j\}$, we have

$$K_l(p_1, p_2) \leq \kappa \leq K_u(p_1, p_2), \tag{6.1.10}$$

where

$$\begin{aligned} K_l(a,b) &= \max\left\{\frac{-2ab}{a(1-b)+(1-a)b}, \frac{-2(1-a)(1-b)}{a(1-b)+(1-a)b}\right\}, \\ K_u(a,b) &= \min\left\{\frac{2a(1-b)}{a(1-b)+(1-a)b}, \frac{2(1-a)b}{a(1-b)+(1-a)b}\right\}. \end{aligned}$$

In practice $\kappa_{ij} = 1$ indicates a perfect agreement and $\kappa_{ij} = 0$ indicates a completely random agreement between the binary variables y_i and y_j. Note that

$$p_{ij} = p_i p_j + \kappa_{ij}\left(\frac{p_i + p_j}{2} - p_i p_j\right) = p_i p_j + \kappa_{ij}\, d_{ij} \tag{6.1.11}$$

where $d_{ij} = (p_i q_j + q_i p_j)/2$. Equation (6.1.11) resembles the relation between p_{ij} and the correlation ρ_{ij}. The equations are similar except that $\sigma_i \sigma_j$, the geometric mean of $p_i q_j$ and $p_j q_i$, is replaced by d_{ij}, which is the arithmetic mean of $p_i q_j$ and $p_j q_i$. Therefore, if κ and κ_0 denote the parent-sibling and common sibling-sibling kappa statistics, the feasible ranges of these two kappa's can be deduced from Theorem 6.1 as

(1) $K_l(p_2, p_3) \leq \kappa_0 \leq K_u(p_2, p_3)$

(2) $\max\{K_l(p_1, p_2), K_l(p_1, p_3), K_1(\kappa_0)\} \leq \kappa \leq \min\{K_u(p_1, p_2), K_u(p_1, p_3), K_2(\kappa_0)\}$,

where

$$\begin{aligned} K_1(\kappa_0) &= \frac{-(\kappa_0\, d_{23} + p_1p_2p_3 + q_1q_2q_3)}{d_{12} + d_{13}} \\ K_2(\kappa_0) &= \frac{\kappa_0\, d_{23} + p_1q_2q_3 + q_1p_2p_3}{d_{12} + d_{13}}. \end{aligned}$$

VI.1.4 Ranges of Relative Risks

Relative risk is another important measure of association for binary variables. The relative risk of y_j with respect to y_i is defined as the ratio of the conditional probability that $y_j = 1$ given y_i, or mathematically (Agresti, 2002)

$$\theta_{j|i} = \frac{P(y_j = 1|y_i = 1)}{P(y_j = 1|y_i = 0)} = \frac{p_{ij}(1 - p_i)}{p_i(p_j - p_{ij})}. \tag{6.1.12}$$

Equation (6.1.12) can be rewritten as

$$p_{ij} = \frac{\theta_{j|i} p_i p_j}{1 + (\theta_{j|i} - 1)p_i}.$$

For the familial binary case, it may be reasonable to assume that the relative risk of the children given their mother's status is same, that is, $\theta_{2|1} = \theta_{3|1} = \theta$. Suppose that α is the sibling-sibling correlation. Then we have

$$p_{12} = \frac{\theta\, p_1p_2}{1 + (\theta - 1)\, p_1}, \quad p_{13} = \frac{\theta\, p_1p_3}{1 + (\theta - 1)\, p_1}, \quad p_{23} = p_2\, p_3 + \alpha\sigma_2\, \sigma_3\,.$$

A trivariate binary distribution for $\mathbf{y}$ exists if and only if the following three conditions hold:

(a) $\max\{0, p_1 + p_2 - 1\} \le p_{12} \le \min\{p_1, p_2\}$

(b) $\max\{0, p_2 + p_3 - 1\} \le p_{23} \le \min\{p_2, p_3\}$

(c) $\max\{0, p_{12} + p_{23} - p_2\} + \max\{0, p_1 + p_2 + p_3 - p_{12} - p_{23} - 1\}$

$\le p_{13} \le \min\{p_{12}, p_{23}\} + \min\{p_1 - p_{12}, p_3 - p_{23}\}\,.$

Note that conditions (a), (b) and (c) are equivalent to (6.1.4) (see Chaganty and Joe, 2006). It is easy to check that (a) holds if and only if

$$\max\{0, 1 - q_2/p_1\} \le \theta \le \{q_1/(\max(p_1, p_2) - p_1)\}$$

and (b) holds if and only if $L(p_2, p_3) \le \alpha \le U(p_2, p_3)$. However, simplification of (c) is cumbersome and does not yield neat expressions for the joint range of θ and α.

VI.2 Multivariate Probit Model and Parameter Estimation

The classical model for analyzing multivariate binary response variables $\mathbf{y}_i$ is the multivariate probit model (Ashford and Sowden, 1970). The mass function of $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{it_i})$ is given by

$$\Pr(\mathbf{y}_i) = \int_{\mathcal{C}_{t_i}} \cdots \int_{\mathcal{C}_1} \frac{1}{(2\pi)^{\frac{t_i}{2}} |\mathbf{R}_i|^{\frac{1}{2}}} \exp\left\{-\frac{\mathbf{z}_i'\mathbf{R}_i^{-1}\mathbf{z}_i}{2}\right\} d\mathbf{z}_i \tag{6.2.1}$$

where

$$\mathcal{C}_j = \begin{cases} (-\infty, \mu_j) & \text{if } y_j = 1 \\ (\mu_j, \infty) & \text{if } y_j = 0 \end{cases}$$

Note that $\Pr(\mathbf{y}_i = \mathbf{1}) = \Phi_{t_i}(\boldsymbol{\mu}_i)$, where $\boldsymbol{\mu}_i = (\mu_1, \ldots, \mu_{t_i})$ and $\Phi_{t_i}(\cdot)$ is the cumulative multivariate normal probability function of dimension t_i. Accurate computation of the joint probability (6.2.1) is a challenging problem, and many evaluation methods and approximations were proposed (Henery, 1981; Genz, 1992 and Joe, 1995). However, we could reduce the multiple integral to a one-dimensional integral for some structured correlation matrices using stochastic representations. For example, suppose that $\mathbf{R}_i$ is an exchangeable structure with parameter α. Let $U_0, U_1, \ldots, U_{t_i}$ be independent identically distributed as standard normal. Consider the stochastic representation (Kotz et al., 2000)

$$Z_j = \sqrt{\alpha}\, U_0 + \sqrt{1-\alpha}\, U_j \qquad \text{for } j = 1, 2, \ldots, t_i. \tag{6.2.2}$$

It is easy to verify that

$$\begin{aligned} \mathrm{Var}(Z_j) &= \alpha + (1-\alpha) = 1 \\ \mathrm{Cov}(Z_j, Z_k) &= \mathrm{Cov}(\sqrt{\alpha}\, U_0, \sqrt{\alpha}\, U_0) \\ &= \alpha\, \mathrm{Cov}(U_0, U_0) \\ &= \alpha\,, \end{aligned}$$

and therefore the correlation between Z_j and Z_k $(j \neq k)$ is given by

$$\mathrm{Corr}\{Z_j, Z_k\} = \frac{\mathrm{Cov}\{Z_j, Z_k\}}{\sqrt{\mathrm{Var}(Z_j)\mathrm{Var}(Z_k)}} = \alpha\,.$$

Using the above stochastic representation, we can see that the multiple integral in (6.2.1) can be reduced to

$$\Pr(\mathbf{y}_i) = \int_{-\infty}^{\infty} \phi(u_0) \prod_{j=1}^{t_i} p_j^{y_{ij}} (1-p_j)^{(1-y_{ij})}\, du_0\,.$$

where $p_j = \Phi\left(\frac{\mu_{ij} - \sqrt{\alpha}\, u_0}{\sqrt{1-\alpha}}\right)$. Suppose that the correlation matrix $\mathbf{R}_i$ has a familiar structure with parameters ρ and α. In this case we can introduce another standard normal random variable U_M, which is independent of U_j's. Consider the stochastic representation

$$\begin{aligned} Z_j &= \sqrt{\alpha}\, U_0 + \sqrt{1-\alpha}\, U_j \\ Z_M &= \frac{\rho}{\sqrt{\alpha}} U_0 + \sqrt{1 - \frac{\rho^2}{\alpha}}\, U_M . \end{aligned} \tag{6.2.3}$$

Clearly, $\mathrm{Corr}(Z_j, Z_k) = \alpha$ for $j \neq k$, and

$$\begin{aligned} \mathrm{Var}(Z_M) &= \frac{\rho^2}{\alpha} + \left\{1 - \frac{\rho^2}{\alpha}\right\} &= 1 \\ \mathrm{Corr}\left\{Z_j, Z_M\right\} &= \frac{\mathrm{Cov}\{Z_j, Z_M\}}{\sqrt{\mathrm{Var}(Z_j)\mathrm{Var}(Z_M)}} &= \rho . \end{aligned}$$

In this case we have

$$\Pr(\mathbf{y}_i = \mathbf{1}) = \int_{-\infty}^{\infty} \phi(u_0)\, \Phi\left(\frac{\mu_{i1} - \frac{\rho}{\sqrt{\alpha}} u_0}{\sqrt{1 - \frac{\rho^2}{\alpha}}}\right) \prod_{j=2}^{t_i} \Phi\left(\frac{\mu_{ij} - \sqrt{\alpha}\, u_0}{\sqrt{1-\alpha}}\right) du_0 .$$

Suppose that $\mathbf{R}_i$ corresponds to a family structure that includes an additional parameter γ representing the correlation between the parents. In this case we can introduce independent standard normal variables U_F and U_P, which are also independent of U_j and U_M. Consider the stochastic representation

$$\begin{aligned} Z_j &= \sqrt{\alpha}\, U_0 + \sqrt{1-\alpha}\, U_j \\ Z_M &= \frac{\rho_2}{\sqrt{\alpha}} U_0 + \sqrt{1 - \gamma + \frac{\rho_1\rho_2 - \rho_2^2}{\alpha}}\, U_M + \sqrt{\gamma - \frac{\rho_1\rho_2}{\alpha}}\, U_P \\ Z_F &= \frac{\rho_1}{\sqrt{\alpha}} U_0 + \sqrt{1 - \gamma + \frac{\rho_1\rho_2 - \rho_1^2}{\alpha}}\, U_F + \sqrt{\gamma - \frac{\rho_1\rho_2}{\alpha}}\, U_P . \end{aligned} \tag{6.2.4}$$

Clearly, $\mathrm{Corr}(Z_j, Z_k) = \alpha$. We can check that

$$\begin{aligned} \mathrm{Var}(Z_M) &= \frac{\rho_2^2}{\alpha} + \left\{1 - \gamma + \frac{\rho_1\rho_2 - \rho_2^2}{\alpha}\right\} + \left\{\gamma - \frac{\rho_1\rho_2}{\alpha}\right\} &= 1 \\ \mathrm{Var}(Z_F) &= \frac{\rho_1^2}{\alpha} + \left\{1 - \gamma + \frac{\rho_1\rho_2 - \rho_1^2}{\alpha}\right\} + \left\{\gamma - \frac{\rho_1\rho_2}{\alpha}\right\} &= 1 \end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}\{Z_j, Z_M\} &= \mathrm{Cov}\left\{\sqrt{\alpha}\,U_0, \frac{\rho_2}{\sqrt{\alpha}}U_0\right\} \\
&= \rho_2 \\
\mathrm{Cov}\{Z_j, Z_F\} &= \mathrm{Cov}\left\{\sqrt{\alpha}\,U_0, \frac{\rho_1}{\sqrt{\alpha}}U_0\right\} \\
&= \rho_1 \\
\mathrm{Cov}\{Z_M, Z_F\} &= \mathrm{Cov}\left\{\frac{\rho_2}{\sqrt{\alpha}}U_0 + \sqrt{\gamma - \frac{\rho_1\rho_2}{\alpha}}\,U_P, \frac{\rho_1}{\sqrt{\alpha}}U_0 + \sqrt{\gamma - \frac{\rho_1\rho_2}{\alpha}}\,U_P\right\} \\
&= \frac{\rho_1\rho_2}{\alpha} + \left\{\gamma - \frac{\rho_1\rho_2}{\alpha}\right\} \\
&= \gamma\,.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Corr}\{Z_j, Z_M\} &= \frac{\mathrm{Cov}\{Z_j, Z_M\}}{\sqrt{\mathrm{Var}(Z_j)\mathrm{Var}(Z_M)}} &= \rho_2 \\
\mathrm{Corr}\{Z_j, Z_F\} &= \frac{\mathrm{Cov}\{Z_j, Z_F\}}{\sqrt{\mathrm{Var}(Z_j)\mathrm{Var}(Z_F)}} &= \rho_1 \\
\mathrm{Corr}\{Z_M, Z_F\} &= \frac{\mathrm{Cov}\{Z_M, Z_F\}}{\sqrt{\mathrm{Var}(Z_M)\mathrm{Var}(Z_F)}} &= \gamma\,.
\end{aligned}$$

In this case the multiple integral in the expression for $\Pr(\mathbf{y}_i = \mathbf{1})$, reduces to double integral

$$\int_{-\infty}^{\infty} \phi(u_p) \int_{-\infty}^{\infty} \phi(u_0)\Phi\left(\psi_{i1}\right)\Phi\left(\psi_{i2}\right)\prod_{j=3}^{t_i}\Phi\left(\frac{\mu_{ij} - \sqrt{\alpha}\,u_0}{\sqrt{1-\alpha}}\right) du_0\, du_p$$

or

$$\int_{-\infty}^{\infty} \phi(u_0)\prod_{j=3}^{t_i}\Phi\left(\frac{\mu_{ij} - \sqrt{\alpha}\,u_0}{\sqrt{1-\alpha}}\right)\int_{-\infty}^{\infty} \phi(u_p)\Phi\left(\psi_{i1}\right)\Phi\left(\psi_{i2}\right) du_p\, du_0$$

with

$$\begin{aligned}
\psi_{i1} &= \frac{\mu_{i1} - \frac{\rho_1}{\sqrt{\alpha}}u_0 - \sqrt{\gamma - \frac{\rho_1\rho_2}{\alpha}}\,u_p}{\sqrt{1 - \gamma + \frac{\rho_1\rho_2 - \rho_1^2}{\alpha}}} \\
\psi_{i2} &= \frac{\mu_{i2} - \frac{\rho_2}{\sqrt{\alpha}}u_0 - \sqrt{\gamma - \frac{\rho_1\rho_2}{\alpha}}\,u_p}{\sqrt{1 - \gamma + \frac{\rho_1\rho_2 - \rho_2^2}{\alpha}}}\,.
\end{aligned}$$

Note that while the stochastic representations solve computational difficulties, they however, add additional constraints to the parameter space. For example, the feasible range for α in an exchangeable correlation structure is $[0, 1)$. For familial structure, we have the range of the correlation parameters as

$$0 \leq \alpha < 1 \quad \text{and} \quad \rho^2 < \alpha$$

which is also contained in the region given in Section IV.1.

The maximum likelihood estimates of the latent correlations can be obtained by maximizing the log-likelihood

$$\ell = \text{constant} + \sum_{i=1}^{n} \log\{\Pr(\mathbf{y}_i)\} \tag{6.2.5}$$

or solving the likelihood equations

$$\begin{aligned} \sum_{i=1}^{n} \frac{1}{\Pr(\mathbf{y}_i)} \frac{\partial \Pr(\mathbf{y}_i)}{\partial \boldsymbol{\beta}} &= \mathbf{0} \\ \sum_{i=1}^{n} \frac{1}{\Pr(\mathbf{y}_i)} \frac{\partial \Pr(\mathbf{y}_i)}{\partial \boldsymbol{\alpha}} &= \mathbf{0}. \end{aligned}$$

Computation of ML estimates and the asymptotic standard errors of the estimates is extremely time consuming. An alternative method is to solve the unbiased estimating equation

$$\sum_{i=1}^{n} \frac{\partial \widetilde{\mathbf{R}}_i}{\partial \boldsymbol{\alpha}} \mathbf{V}_i^{-1} \left\{ \widetilde{\mathbf{z}_i \mathbf{z}_i'} - \widetilde{\mathbf{R}}_i \right\} = \mathbf{0}$$

where $\widetilde{\mathbf{R}}_i = \text{vech}(\mathbf{R}_i)$, $\widetilde{\mathbf{z}_i\mathbf{z}_i'} = \text{vech}(\mathbf{z}_i\mathbf{z}_i')$ and $\mathbf{V}_i$ is the covariance matrix of $\widetilde{\mathbf{z}_i\mathbf{z}_i'}$. We could also add the additional restriction $\text{tr}(\mathbf{W}_i) = 0$ so that a subclass of estimating equations can be expressed as

$$\sum_{i=1}^{n} \text{tr}\left\{ \mathbf{W}_i \mathbf{R}_i^{-1} \mathbf{z}_i \mathbf{z}_i' \right\} = 0.$$

VI.3 An Illustrative Example

To illustrate the analysis of binary data, we modify the familial data set used in Section IV.5. We have generated binary data on the mother and her children by

Table 6.7: Familial data set with binary outcomes

Family	Member	Age	Pre-ATP	Post-ATP
2	Mother	62	4.43	1 (2.49)
	Son	24	4.18	1 (1.49)
	Son	41	4.81	0 (2.84)
	Daughter	31	4.42	1 (2.04)
	Daughter	38	3.65	1 (1.17)
3	Mother	50	3.79	1 (1.28)
	Son	7	4.72	1 (1.19)
4	Mother	55	5.42	0 (3.65)
	Son	23	5.30	1 (2.16)
	Son	27	4.48	1 (2.40)
	Son	19	4.85	0 (3.28)
5	Mother	57	4.71	1 (2.23)
	Son	32	4.19	1 (1.33)
	Son	28	3.43	1 (1.85)
⋮	⋮	⋮	⋮	⋮
22	Mother	45	5.29	0 (3.27)
	Son	24	5.30	0 (4.10)
	Son	20	5.25	0 (3.67)

truncating the ATP levels using the medians as the cutoff values. In another word, if the post ATP level of mother is greater than or equal to the median 3.05 of the sample data on mothers' post ATP levels, it is coded as "0", indicating high level; otherwise, it is coded as "1", indicating low level. Similarly, if the post ATP level of a child is greater than or equal to the median 2.84 of the sample data on children's post ATP levels, it is coded as "0"; otherwise, it is coded as "1". Therefore, we have binary outcomes, representing low-high post ATP levels for 19 families along with the covariates pre-ATP levels, age and gender. Table 6.7 contains a partial list of the modified data set.

For this binary data, we have computed the ML estimates of the regression and the latent correlations using the probit model, as well as using the unbiased estimating equation approach. The parameter estimates, which are in agreement, are presented in Table 6.8.

Table 6.8: Parameter estimates for familial binary outcomes

	UEE	*MLE*
Intercept	-7.8620	-7.4949
Gender	0.4345	0.4154
Pre-ATP	1.5842	1.5067
ρ	0.5129	0.6080
α	0.4884	0.5763

CHAPTER VII

SUMMARY

In this thesis, we have studied alternative approaches to maximum likelihood for estimating parameters in structured correlation matrices that are usually employed in analyzing longitudinal and clustered or more generally correlated data. These alternative approaches are based on constructing general classes of weighted unbiased estimating equations using Cholesky decompositions of the inverse of the correlation matrix. When the response variables are distributed as multivariate normal, we have proved that the Godambe's optimal unbiased estimating equation coincides with the likelihood equation. For a general class of weighted unbiased estimation equations, we have obtained optimal weights by minimizing the asymptotic variances. However, unbiased equations employing these optimal weights are difficult to solve for some structures, for example the familial correlation structure. Therefore, we have introduced an additional constraint on the weights and studied properties of the subclass of unbiased estimating equations. We have also suggested, for common correlation structures including the familial structure, weights in a closed form that are close to being optimal. Using simulations we have shown that these approximate weights yield highly efficient and robust estimates, which are easy to compute and do not run into computational problems.

When the response variables are binary, it is well known that the ranges of common measures of associations are restricted by the marginal means. Understanding these restrictions is the key for developing efficient methods of estimation for the associations. In this thesis we have studied ranges of association measures including correlations, odds ratios, kappa statistics and relative risks for familial binary variables. We have generalized the classical multivariate probit model to estimate familial correlations. Computing the maximum likelihood estimates was facilitated by the use of a stochastic representation of the latent familial variables. We have also studied the use of weighted unbiased estimation equations. The results were comparable with the maximum likelihood estimates.

REFERENCES

Agresti, A. (2002), *Categorical Data Analysis*, New York: John Wiley & Sons, Inc., 2nd ed.

Ashford, J. R., and Sowden, R. R. (1970), "Multi-variate Probit Analysis," *Biometrics*, 26, 535–546

Chaganty, N. R. (2003), "Analysis of Growth Curves with Patterned Correlation Matrices Using Quasi-least Squares," *Journal of Statistical Planning and Inference*, 117, 123–139.

Chaganty, N. R., and Shi, G. (2004), "A Note on the Estimation of Autocorrelation in Repeated Measurements," *Communications in Statistics*, 33, 1157–1170.

Chaganty, N. R., and Joe, H. (2006), "Range of Correlation Matrices for Dependent Bernoulli Random Variables," *Biometrika*, 93, 197–206.

Dern, R. J., and Wiorkowski, J. J. (1969), "Studies on the Preservation of Human Blood. IV. the Hereditary Component of pre- and poststorage Erythrocyte Adenosine Triphosphate Levels," *Journal of Laboratory & Clinical Medicine*, 73, 1019–1029.

Diggle, P. J., Heagerty, P., Liang, K.-Y., and Zeger, S. L. (2002), *Analysis of Longitudianal Data*, Oxford: Oxford University Press Inc., 2nd ed.

Donner, A., and Koval, J. J. (1980), "The Estimation of Intraclass Correlation in the Analysis of Family Data," *Biometrics*, 36, 19–25.

Eliasziw, M., and Donner, A. (1990), "Comparison of Recent Estimators of Interclass Correlation from Familial Data," *Biometrics*, 46, 391–398.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2004), *Applied Longitudinal Analysis*, New York: John Wiley & Sons, Inc.

Genz, A. (1992), "Numerical Computation of the Multivariate Normal Probabilities," *Journal of Compuational and Graphical Statistics*, 1, 141–150.

Godambe, V. P. (1960), "An Optimum Property of Regular Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 1208–1212.

Hardin, J. W., and Hilbe, J. M. (2003), *Generalized Estimating Equations*, New York: Chapman & Hall.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, New York: Springer-Verlag.

Henery, R. J. (1981), "An Approximation to Certain Multivariate Normal Probabilities," *Journal of Royal Statistical Society B*, 43, 81–85.

Joe, H. (1995), "Approximations to Multivariate Normal Rectangle Probabilities Based on Conditional Expectations," *Journal of the American Statistical Association*, 90, 957–964.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, New York: Chapman and Hall.

Johnson, N. L., and Kotz, S. (1972) *Distributions in Statistics: Continuous Multivariate Distributions*, New York: John Wiley & Sons, Inc.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000), *Continuous Multivariate Distributions*, New York: John Wiley & Sons, Inc., 2nd ed.

Konish, S., Khatri, C. G., and Rao, C. R. (1991), "Inferences on Multivariate Measures of Interclass and Intraclass Correlations in Familial Data," *Journal of Royal Statistical Society B*, 53, 649–659.

Lee, J. C. (1988), "Prediction and Estimation of Growth Curves with Special Covariance Structures," *Journal of the American Statistical Association*, 83, 432–440.

Liang, K.-Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Magnus, J. R., and Neudecker, H. (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: John Wiley & Sons, Inc.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, New York: Chapman and Hall, 2nd ed.

Myers, R. H., Montgomery, D. C., and Vining, G. G. (2002), *Generalized Linear Models*, New York: John Wiley & Sons, Inc.

Nelder, J. A., and Wedderburn, R. (1972), "Generalized Linear Models," *Journal of Royal Statistical Society A*, 135, 370–384.

Prentice, R. L., and Zhao, L. P. (1991), "Estimating Equations for Parameters in the Means and Covariances of Multivariate Discrete and Continuous Responses," *Biometrika*, 47, 825–839.

Qu, A., Lindsay, B. G., and Li, B. (2000), "Improving Generalised Estimating Equations Using Quadratic Inference Functions," *Biometrika*, 87, 823–836.

Rao, C. R., and Rao, M. B. (1998), *Matrix Algebra and Its Applications to Statistics and Economics*, Singapore: World Scientific.

Srivastava, M. S. (1984), "Estimation of Interclass Correlations in Familial Data," *Biometrika*, 71, 177–185.

Srivastava, M. S., Keen, K. J., and Katapa, R. S. (1988), "Estimation of Interclass and Intraclass Correlations in Multivariate Familial Data," *Biometrics*, 44, 141–150.

Vonesh, E. F., and Chinchilli, V. M. (1997), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, New York: Marcel Deckker.

Wang, Y.-G., and Carey, V. J. (2004), "Unbiased Estimating Equations from Working Correlation Models for Irregularly Timed Repeated Measures," *Journal of the American Statistical Association*, 99, 845–853.

Wu, C.-T., Gumpertz, M. L., and Boos, D. D. (2001), "Comparison of GEE, MINQUE, ML, and REML Estimating Equations for Normally Distributed Data," *the American Statistician*, 55, 125–130.

Zeger, S. L., and Liang, K. (1992), "An Overview of Methods for the Analysis of Longitudinal Data," *Statistics in Medicine*, 11, 1825–1839.

APPENDIX

CHOLESKY DECOMPOSITIONS

We want to decompose structured matrix $\mathbf{R}$ into $\mathbf{P\Gamma P'}$ or $\mathbf{R}^{-1}$ into $\mathbf{B\Lambda B'}$ where $\mathbf{P}$ and $\mathbf{B}$ are upper (or lower) triangular matrices with unit leading elements and $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ are diagonal matrices. The structures that $\mathbf{R}$ assumes are: (1) component symmetry (CS) (or exchangeable); (2) first order autoregressive (AR(1)); (3) familial (single parent); (4) nuclear familial (two parents).

A.1 Exchangeable Correlation Matrix

The exchangeable correlation matrix is defined as $\mathbf{R} = (1-\alpha)\mathbf{I} + \alpha\mathbf{J}$ of order t. The Cholesky decomposition matrices of $\mathbf{R}$ are:

$$\mathbf{P}_l = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ \alpha & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha & \frac{\alpha}{1+\alpha} & \dots & 1 & 0 \\ \alpha & \frac{\alpha}{1+\alpha} & \dots & \frac{\alpha}{1+(t-2)\alpha} & 1 \end{bmatrix}_{t\times t}$$

$$\mathbf{\Gamma}_l = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1-\alpha^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{(1-\alpha)[1+(t-2)\alpha]}{1+(t-3)\alpha} & 0 \\ 0 & 0 & \dots & 0 & \frac{(1-\alpha)[1+(t-1)\alpha]}{1+(t-2)\alpha} \end{bmatrix}_{t\times t}$$

or

$$\mathbf{P}_u = \begin{bmatrix} 1 & \frac{\alpha}{1+(t-2)\alpha} & \dots & \frac{\alpha}{1+\alpha} & \alpha \\ 0 & 1 & \dots & \frac{\alpha}{1+\alpha} & \alpha \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \alpha \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{t\times t}$$

$$\mathbf{\Gamma}_u = \begin{bmatrix} \frac{(1-\alpha)[1+(t-1)\alpha]}{1+(t-2)\alpha} & 0 & \cdots & 0 & 0 \\ 0 & \frac{(1-\alpha)[1+(t-2)\alpha]}{1+(t-3)\alpha} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1-\alpha^2 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}_{t\times t}$$

The inverse of $\mathbf{R}$ is well known as $\frac{1}{1-\alpha}\mathbf{I} - \frac{\alpha}{(1-\alpha)[1+(t-1)\alpha]}\mathbf{J}$, and the decomposition matrices of $\mathbf{R}^{-1}$ are:

$$\mathbf{B}_l = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ \frac{-\alpha}{1+(t-2)\alpha} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{-\alpha}{1+(t-2)\alpha} & \frac{-\alpha}{1+(t-3)\alpha} & \cdots & 1 & 0 \\ \frac{-\alpha}{1+(t-2)\alpha} & \frac{-\alpha}{1+(t-3)\alpha} & \cdots & -\alpha & 1 \end{bmatrix}_{t\times t}$$

$$\mathbf{\Lambda}_l = \begin{bmatrix} \frac{1+(t-2)\alpha}{(1-\alpha)[1+(t-1)\alpha]} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1+(t-3)\alpha}{(1-\alpha)[1+(t-2)\alpha]} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{1-\alpha^2} & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}_{t\times t}$$

or

$$\mathbf{B}_u = \begin{bmatrix} 1 & -\alpha & \cdots & \frac{-\alpha}{1+(t-3)\alpha} & \frac{-\alpha}{1+(t-2)\alpha} \\ 0 & 1 & \cdots & \frac{-\alpha}{1+(t-3)\alpha} & \frac{-\alpha}{1+(t-2)\alpha} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \frac{-\alpha}{1+(t-2)\alpha} \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}_{t\times t}$$

$$\mathbf{\Lambda}_u = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{1-\alpha^2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1+(t-3)\alpha}{(1-\alpha)[1+(t-2)\alpha]} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1+(t-2)\alpha}{(1-\alpha)[1+(t-1)\alpha]} \end{bmatrix}_{t\times t}$$

A.2 First Order Autoregressive Correlation Matrix

The first order autoregressive correlation matrix is defined as

$$\mathbf{R} = \begin{bmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{t-1} \\ \alpha & 1 & \alpha & \dots & \alpha^{t-2} \\ \alpha^2 & \alpha & 1 & \dots & \alpha^{t-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{t-1} & \alpha^{t-2} & \alpha^{t-3} & \dots & 1 \end{bmatrix}_{t\times t}$$

then the decomposition matrices are:

$$\mathbf{P}_l = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ \alpha & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha^{t-2} & \alpha^{t-3} & \dots & 1 & 0 \\ \alpha^{t-1} & \alpha^{t-2} & \dots & \alpha & 1 \end{bmatrix}_{t\times t} \quad \mathbf{\Gamma}_l = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1-\alpha^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1-\alpha^2 & 0 \\ 0 & 0 & \dots & 0 & 1-\alpha^2 \end{bmatrix}_{t\times t}$$

or

$$\mathbf{P}_u = \begin{bmatrix} 1 & \alpha & \dots & \alpha^{t-2} & \alpha^{t-1} \\ 0 & 1 & \dots & \alpha^{t-3} & \alpha^{t-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \alpha \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{t\times t} \quad \mathbf{\Gamma}_u = \begin{bmatrix} 1-\alpha^2 & 0 & \dots & 0 & 0 \\ 0 & 1-\alpha^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1-\alpha^2 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{t\times t}$$

The inverse of $\mathbf{R}$ is known as $\frac{1}{1-\alpha^2}\left(\mathbf{I} + \alpha^2\mathbf{C}_0 - \alpha\mathbf{C}_1\right)$ where diagonal matrix $\mathbf{C}_0 = \text{diag}(0, 1, 1, \dots, 1, 0)$ and $\mathbf{C}_1$ is tridiagonal matrix with 0 on the main diagonal and 1 on the upper and lower diagonals. The decomposition matrices of $\mathbf{R}^{-1}$ are:

$$\mathbf{B}_l = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -\alpha & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & -\alpha & 1 \end{bmatrix}_{t\times t} \quad \mathbf{\Lambda}_l = \begin{bmatrix} \frac{1}{1-\alpha^2} & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{1-\alpha^2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1}{1-\alpha^2} & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{t\times t}$$

or

$$\mathbf{B}_u = \begin{bmatrix} 1 & -\alpha & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -\alpha \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{t\times t} \quad \mathbf{\Lambda}_u = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{1-\alpha^2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1}{1-\alpha^2} & 0 \\ 0 & 0 & \dots & 0 & \frac{1}{1-\alpha^2} \end{bmatrix}_{t\times t}$$

A.3 Familial Correlation Matrix

The familial correlation matrix with one parent is defined as

$$\mathbf{R} = \begin{bmatrix} 1 & \rho\mathbf{1}'_{1\times t} \\ \rho\mathbf{1}_{t\times 1} & (1-\alpha)\mathbf{I}_{t\times t} + \alpha\mathbf{J}_{t\times t} \end{bmatrix}_{(t+1)\times(t+1)}$$

then the decomposition matrices are:

$$\mathbf{P}_l = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & 0 \\ \rho & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \rho & \frac{\alpha-\rho^2}{1-\rho^2} & \dots & 1 & 0 & 0 \\ \rho & \frac{\alpha-\rho^2}{1-\rho^2} & \dots & \frac{\alpha-\rho^2}{1-\rho^2+(t-3)(\alpha-\rho^2)} & 1 & 0 \\ \rho & \frac{\alpha-\rho^2}{1-\rho^2} & \dots & \frac{\alpha-\rho^2}{1-\rho^2+(t-3)(\alpha-\rho^2)} & \frac{\alpha-\rho^2}{1-\rho^2+(t-2)(\alpha-\rho^2)} & 1 \end{bmatrix}_{(t+1)\times(t+1)}$$

$$\mathbf{\Gamma}_l = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1-\rho^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{(1-\alpha)[1-\rho^2+(t-2)(\alpha-\rho^2)]}{1-\rho^2+(t-3)(\alpha-\rho^2)} & 0 \\ 0 & 0 & \dots & 0 & \frac{(1-\alpha)[1-\rho^2+(t-1)(\alpha-\rho^2)]}{1-\rho^2+(t-2)(\alpha-\rho^2)} \end{bmatrix}_{(t+1)\times(t+1)}$$

or

$$\mathbf{P}_u = \begin{bmatrix} 1 & \frac{\rho}{1+(t-1)\alpha} & \frac{\rho}{1+(t-2)\alpha} & \dots & \frac{\rho}{1+\alpha} & \rho \\ 0 & 1 & \frac{\alpha}{1+(t-2)\alpha} & \dots & \frac{\alpha}{1+\alpha} & \alpha \\ 0 & 0 & 1 & \dots & \frac{\alpha}{1+\alpha} & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & \alpha \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{(t+1)\times(t+1)}$$

$$\mathbf{\Gamma}_u = \begin{bmatrix} 1-\frac{t\rho^2}{1+(t-1)\alpha} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{(1-\alpha)[1+(t-1)\alpha]}{1+(t-2)\alpha} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{(1-\alpha)(1+2\alpha)}{1+\alpha} & 0 & 0 \\ 0 & 0 & \dots & 0 & 1-\alpha^2 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}_{(t+1)\times(t+1)}$$

The decomposition matrices of $\mathbf{R}^{-1}$ are:

$$\mathbf{B}_l = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ \frac{-\rho}{1+(t-1)\alpha} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{-\rho}{1+(t-1)\alpha} & \frac{-\alpha}{1+(t-2)\alpha} & \dots & 1 & 0 \\ \frac{-\rho}{1+(t-1)\alpha} & \frac{-\alpha}{1+(t-2)\alpha} & \dots & -\alpha & 1 \end{bmatrix}_{(t+1)\times(t+1)}$$

$$\mathbf{\Lambda}_l = \begin{bmatrix} \frac{1+(t-1)\alpha}{1+(t-1)\alpha-t\rho^2} & 0 & \dots & 0 & 0 \\ 0 & \frac{1+(t-2)\alpha}{(1-\alpha)[1+(t-1)\alpha]} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1}{1-\alpha^2} & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{(t+1)\times(t+1)}$$

or

$$\mathbf{B}_u = \begin{bmatrix} 1 & -\rho & \dots & \frac{-\rho(1-\alpha)}{1-\rho^2+(t-3)(\alpha-\rho^2)} & \frac{-\rho(1-\alpha)}{1-\rho^2+(t-2)(\alpha-\rho^2)} \\ 0 & 1 & \dots & \frac{\rho^2-\alpha}{1-\rho^2+(t-3)(\alpha-\rho^2)} & \frac{\rho^2-\alpha}{1-\rho^2+(t-2)(\alpha-\rho^2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \frac{\rho^2-\alpha}{1-\rho^2+(t-2)(\alpha-\rho^2)} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{(t+1)\times(t+1)}$$

$$\mathbf{\Lambda}_u = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{1-\rho^2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1-\rho^2+(t-3)(\alpha-\rho^2)}{(1-\alpha)[1-\rho^2+(t-2)(\alpha-\rho^2)]} & 0 \\ 0 & 0 & \dots & 0 & \frac{1-\rho^2+(t-2)(\alpha-\rho^2)}{(1-\alpha)[1-\rho^2+(t-1)(\alpha-\rho^2)]} \end{bmatrix}_{(t+1)\times(t+1)}$$

A.4 Nuclear Familial Correlation Matrix

The complete (two parents) familial correlation structure is defined as

$$\mathbf{R} = \begin{bmatrix} 1 & \gamma & \rho_1 \mathbf{1}'_{1\times t} \\ \gamma & 1 & \rho_2 \mathbf{1}'_{1\times t} \\ \rho_1 \mathbf{1}_{t\times 1} & \rho_2 \mathbf{1}_{t\times 1} & (1-\alpha)\mathbf{I}_{t\times t} + \alpha \mathbf{J}_{t\times t} \end{bmatrix}_{(t+2)\times(t+2)}$$

if we define $\psi = 1 - \frac{\rho_1^2+\rho_2^2-2\gamma\rho_1\rho_2}{1-\gamma^2}$, $\psi_\alpha = \alpha - \frac{\rho_1^2+\rho_2^2-2\gamma\rho_1\rho_2}{1-\gamma^2}$; $\phi_0 = \frac{t\rho_1\rho_2}{1+(t-1)\alpha}$, $\phi_1 = \frac{t\rho_1^2}{1+(t-1)\alpha}$, and $\phi_2 = \frac{t\rho_2^2}{1+(t-1)\alpha}$, then the decomposition matrices are:

$$\mathbf{P}_l = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \gamma & 1 & 0 & \cdots & 0 & 0 & 0 \\ \rho_1 & \frac{\rho_2-\gamma\rho_1}{1-\gamma^2} & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \rho_1 & \frac{\rho_2-\gamma\rho_1}{1-\gamma^2} & \frac{\psi_\alpha}{\psi} & \cdots & 1 & 0 & 0 \\ \rho_1 & \frac{\rho_2-\gamma\rho_1}{1-\gamma^2} & \frac{\psi_\alpha}{\psi} & \cdots & \frac{\psi_\alpha}{\psi+(t-3)\psi_\alpha} & 1 & 0 \\ \rho_1 & \frac{\rho_2-\gamma\rho_1}{1-\gamma^2} & \frac{\psi_\alpha}{\psi} & \cdots & \frac{\psi_\alpha}{\psi+(t-3)\psi_\alpha} & \frac{\psi_\alpha}{\psi+(t-2)\psi_\alpha} & 1 \end{bmatrix}_{(t+2)\times(t+2)}$$

$$\mathbf{\Gamma}_l = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1-\gamma^2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \psi & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{(1-\rho)[\psi+(t-3)\psi_\alpha]}{\psi+(t-4)\psi_\alpha} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{(1-\rho)[\psi+(t-2)\psi_\alpha]}{\psi+(t-3)\psi_\alpha} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \frac{(1-\rho)[\psi+(t-1)\psi_\alpha]}{\psi+(t-2)\psi_\alpha} \end{bmatrix}_{(t+2)\times(t+2)}$$

or

$$\mathbf{P}_u = \begin{bmatrix} 1 & \frac{\gamma-\phi_0}{1-\phi_2} & \frac{\rho_1}{1+(t-1)\alpha} & \cdots & \frac{\rho_1}{1+2\alpha} & \frac{\rho_1}{1+\alpha} & \rho_1 \\ 0 & 1 & \frac{\rho_2}{1+(t-1)\alpha} & \cdots & \frac{\rho_2}{1+2\alpha} & \frac{\rho_2}{1+\alpha} & \rho_2 \\ 0 & 0 & 1 & \cdots & \frac{\alpha}{1+2\alpha} & \frac{\alpha}{1+\alpha} & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \frac{\alpha}{1+\alpha} & \alpha \\ 0 & 0 & 0 & \cdots & 0 & 1 & \alpha \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}_{(t+2)\times(t+2)}$$

$$\mathbf{\Gamma}_u = \begin{bmatrix} 1-\phi_1-\frac{(\gamma-\phi_0)^2}{1-\phi_2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1-\phi_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{(1-\alpha)[1+(t-1)\alpha]}{1+(t-2)\alpha} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-\alpha^2 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}_{(t+2)\times(t+2)}$$

The decomposition matrices of $\mathbf{R}^{-1}$ are:

$$\mathbf{B}_l = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\frac{\gamma-\phi_0}{1-\phi_2} & 1 & 0 & \cdots & 0 & 0 & 0 \\ \frac{\gamma\rho_2-\rho_1}{1+(t-1)\alpha-t\rho_2^2} & \frac{-\rho_2}{1+(t-1)\alpha} & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\gamma\rho_2-\rho_1}{1+(t-1)\alpha-t\rho_2^2} & \frac{-\rho_2}{1+(t-1)\alpha} & \frac{-\alpha}{1+(t-2)\alpha} & \cdots & 1 & 0 & 0 \\ \frac{\gamma\rho_2-\rho_1}{1+(t-1)\alpha-t\rho_2^2} & \frac{-\rho_2}{1+(t-1)\alpha} & \frac{-\alpha}{1+(t-2)\alpha} & \cdots & \frac{-\alpha}{1+\alpha} & 1 & 0 \\ \frac{\gamma\rho_2-\rho_1}{1+(t-1)\alpha-t\rho_2^2} & \frac{-\rho_2}{1+(t-1)\alpha} & \frac{-\alpha}{1+(t-2)\alpha} & \cdots & \frac{-\alpha}{1+\alpha} & -\alpha & 1 \end{bmatrix}_{(t+2)\times(t+2)}$$

$$\mathbf{\Lambda}_l = \begin{bmatrix} \frac{1}{1-\phi_1-\frac{(\gamma-\phi_0)^2}{1-\phi_2}} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{1-\phi_2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1+(t-2)\alpha}{(1-\alpha)[1+(t-1)\alpha]} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{1-\alpha^2} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}_{(t+2)\times(t+2)}$$

or

$$\mathbf{B}_u = \begin{bmatrix} 1 & -\gamma & \frac{-(\rho_1-\gamma\rho_2)}{1-\gamma^2} & \frac{-(1-\alpha)(\rho_1-\gamma\rho_2)}{\psi(1-\gamma^2)} & \cdots & \frac{-(1-\alpha)(\rho_1-\gamma\rho_2)}{[\psi+(t-2)\psi_\alpha](1-\gamma^2)} \\ 0 & 1 & \frac{-(\rho_2-\gamma\rho_1)}{1-\gamma^2} & \frac{-(1-\alpha)(\rho_2-\gamma\rho_1)}{\psi(1-\gamma^2)} & \cdots & \frac{-(1-\alpha)(\rho_2-\gamma\rho_1)}{[\psi+(t-2)\psi_\alpha](1-\gamma^2)} \\ 0 & 0 & 1 & \frac{-\psi_\alpha}{\psi} & \cdots & \frac{-\psi_\alpha}{\psi+(t-2)\psi_\alpha} \\ 0 & 0 & 0 & 1 & \cdots & \frac{-\psi_\alpha}{\psi+(t-2)\psi_\alpha} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{(t+2)\times(t+2)}$$

$$\mathbf{\Lambda}_u = \begin{bmatrix} \dfrac{1}{1-\phi_1-\frac{(\gamma-\phi_0)^2}{1-\phi_2}} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{1-\phi_2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1+(t-2)\alpha}{(1-\alpha)[1+(t-1)\alpha]} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{1-\alpha^2} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}_{(t+2)\times(t+2)}$$

VITA

Yihao Deng
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529

Education

PhD	Old Dominion University, Norfolk, VA, U.S.A. (August 2006) Major: Computational and Applied Mathematics (Statistics)
MS	Old Dominion University, Norfolk, VA, U.S.A. (May 2004) Major: Computational and Applied Mathematics (Statistics)
MS	Dalian University of Technology, Dalian, P.R.China. (July 2001) Major: Management Engineering
BS	Dongbei University of Finance & Economics Dalian, P.R.China. (July 1998) Major: Insurance

Experience

Graduate Teaching Assistant (08/2001 - 05/2006)
Old Dominion University, Norfolk, VA

Statistical Consultant (08/2003 - 05/2006)
Old Dominion University, Norfolk, VA

Publications

Chaganty, N. R. and **Deng, Y.** (2007), *"Ranges of measures of associations for familial binary variables."* To appear in Communications in Statistics.
Deng, Y. and Chaganty, N. R. (2006), *"Optimal unbiased estimating equations for correlation parameter estimation in longitudinal studies."* Under preparation.

Typeset using LaTeX.