THE ANALYSIS OF INCOMPLETE LONGITUDINAL DATA
WITH MODELED COVARIANCE STRUCTRUES

by

Lisa Morrissey LaVange

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1449

August 1983

THE ANALYSIS OF INCOMPLETE LONGITUDINAL DATA WITH MODELED COVARIANCE STRUCTURES

by

Lisa Morrissey LaVange

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill

1983

Approved by:

Ronald W. Helms
Adviser

Gary G. Koch
Reader

James D. Hosking
Reader

ABSTRACT

LISA MORRISSEY LAVANGE. The Analysis of Incomplete Longitudinal Data with Modeled Covariance Structures. (Under the direction of RONALD HELMS).

A general linear model formulation is presented that is applicable to the analysis of incomplete longitudinal data. Other models that have appeared in the literature in the context of incomplete multivariate data analysis are shown to be special cases of this general model with appropriate constraints imposed on the model parameters. Problems of estimation due to particularly sparse data are reviewed and the concept of imposing constraints on the model parameters as an alternative is discussed. Methods of estimation and hypothesis testing assuming covariance models are proposed.

Two covariance models associated with time series data are proposed for use with a general incomplete model first introduced by Kleinbaum. Consistent estimation of the variance-covariance matrix is derived for the incomplete data case assuming an autoregressive error process of order one and a finite moving average error process.

A model is proposed for incomplete multivariate data analysis assuming a time series model for each dependent variable. Estimation is developed for the regression coefficients and the covariance parameters associated with each time series model as well as between series. The autoregressive model of order one and the moving average model of order one are considered for each dependent variable. The methods are illustrated with an analysis of a longitudinal study of spirometry in children.

ACKNOWLEDGEMENTS

I gratefully acknowledge my advisor, Dr. Ronald Helms, who introduced me to the field of linear models and multivariate analysis, and I thank him for his continued guidance and support throughout this research effort. I also thank the members of my committee, Prof. Gary Koch, Dr. James Hosking, Prof. Michael Symons, and Prof. Berton Kaplan, for providing helpful comments along the way.

To Martha Clegg I express my appreciation of her meticulous typing of this manuscript and her patience during the many revisions.

I would also like to acknowledge my father, Mr. Joe Morrissey, who continually encouraged and supported my educational endeavors. Finally, I am indebted to my husband, Mr. Ben LaVange, for the assistance and encouragement he has provided and for his patience during my pursuit of the doctoral degree.

Table of Contents

Table of Contents
(continued)

List of Tables

List of Tables (continued)

List of Figures

NOTATION AND ABBREVIATIONS

$\underset{\sim}{y}$ denotes a column vector of specified dimension.

$\underset{\sim}{y}'$ denotes the transpose of $\underset{\sim}{y}$.

$\underset{\sim}{A} = ((a_{ij}))$ denotes a matrix of specified dimensions with a_{ij} as the element in the i^{th} row and j^{th} column.

$E(\underset{\sim}{y})$ denotes the expectation of a vector of random variables $\underset{\sim}{y}$.

$Var(\underset{\sim}{y})$ denotes the variance-covariance matrix of $\underset{\sim}{y}$.

$Cov(\underset{\sim}{x}, \underset{\sim}{y})$ denotes the variance-covariance matrix for $\underset{\sim}{x}$ and $\underset{\sim}{y}$.

$tr(\underset{\sim}{A})$ denotes the trace of $\underset{\sim}{A}$.

$\underset{\sim}{A} \otimes \underset{\sim}{B}$ denotes the Kronecher product of $\underset{\sim}{A}$ and $\underset{\sim}{B}$ given by $((a_{jj}\ B))$.

$\underset{\sim}{I}_p$ denotes the identity matrix of order p.

$\underset{\sim}{y} \sim N(\underset{\sim}{u}, \underset{\sim}{\Sigma})$ means that $\underset{\sim}{y}$ is distributed according to the multivariate normal distribution with mean $\underset{\sim}{\mu}$ and variance-covariance matrix $\underset{\sim}{\Sigma}$.

$\underset{\sim}{y}_n \overset{L}{\rightarrow} \underset{\sim}{y}$ means that the random vector $\underset{\sim}{y}_n$ converges in distribution (or law) to the random vector $\underset{\sim}{y}$

$a_n \rightarrow a$ means that the limit of the series of real numbers a_n is equal to a.

$\theta_n \overset{L}{\rightarrow} \theta$ means that θ_n converges in probability to θ, or that θ_n is consistent for θ.

$\underset{\sim}{y} \sim X^2_w$ means that $\underset{\sim}{y}$ is distributed according to the central chi-square distribution with w degrees of freedom.

$\sigma_{ij(k)}$ denotes the element in the i^{th} row and j^{th} column of the k^{th} submatrix $\underset{\sim}{\Sigma}_k$ where $\underset{\sim}{\Sigma}_k = ((\sigma_{ij}))_k$.

INTRODUCTION

This research is concerned with the multivariate analysis of incomplete longitudinal data assuming a modeled covariance structure. Estimation is proposed for two classes of covariance models commonly associated with time series data. The asymptotic properties of estimators of linear model parameters are discussed and appropriate tests of hypotheses are presented. The methods are illustrated with the analysis of a longitudinal dataset from an ongoing study.

Chapter 1 contains a brief review of the literature in two parts. In the first part, the existing methods of analysis of multivariate data are reviewed. In the second part, covariance models frequently associated with longitudinal data are summarized. A brief review of the analysis of an econometric model with which a variety of time series covariance models are often associated is provided in this section.

In Chapter 2, linear models that are applicable to incomplete longitudinal data are reviewed. A general linear model formulation (hereafter referred to as the LM) is introduced. It is shown that several models already reviewed are special cases of the LM when suitable constraints on the model parameters are imposed. Best asymptotically normal (BAN) estimation of model parameters is established for a special case of the LM that is particularly suited for incomplete observation vectors in a repeated measurements design. The use of Wald statistics to test hypotheses about these model parameters assuming covariance constraints is discussed.

In Chapter 3, attention is focused on the general incomplete model (GIM) introduced by Kleinbaum (1970, 1973). Estimation and hypothesis testing for the model assuming a constrained covariance matrix is presented. Cases for which the GIM analysis suggested by Kleinbaum is not feasible unless a covariance model is assumed are discussed, and an example of such a case is presented for illustration.

With motivation established for assuming a covariance model in the GIM, two covariance structures appropriate for the analysis of incomplete longitudinal data are presented. In Chapter 4 consistent estimation is derived for the covariance parameters associated with an autoregressive process of order one (AR(1)) with missing data. Estimation and hypothesis testing for the GIM assuming that each subject's residuals are generated by an AR(1) process are presented. An example is provided to illustrate the techniques.

In Chapter 5 the assumption is made that a subject's errors about the population regression line are generated by a finite moving average process of order M (MA(M)). Consistent estimation of the variance and covariance parameters associated with this process is derived and incorporated into the estimation procedure for the GIM parameters. An illustrative example is provided.

In Chapter 6 a model formulation is presented for multiple response variables measured longitudinally on each subject. Estimation of model parameters is presented assuming the errors associated with each response variable are generated by individual AR(1) processes and by individual MA(1) processes. This estimation includes the covariances between measurements of a single response variable taken at different times as well as covariances between response variables measured across time.

In Chapter 7, the methods proposed are illustrated with the analysis of a longitudinal study of spirometry in children. Data on three spirometric variables are analyzed separately assuming both an AR(1) and an

MA(4) error process for each subject. The data are then combined for an analysis assuming multiple AR(1) error processes. Practical aspects of the covariance model selection process are discussed. Recommendations for future research are made in Chapter 8.

CHAPTER 1

REVIEW OF THE LITERATURE

1.1 Introduction

The literature review for the analysis of incomplete longitudinal data is presented in two parts. The first section is concerned with estimators and hypothesis testing for general multivariate models with missing data. Covariance models appropriate for the analysis of longitudinal data are reviewed in the second section. The notation established here for the general linear multivariate model (GLMM) will be referred to throughout this review.

The GLMM usually arises from experiments in which p measurements are taken on a sample of n subjects. The p measurements are univariate random variables that might represent p distinct characteristics or one characteristic measured at p time points. Unless otherwise noted, $\underset{\sim}{Y}$ will refer to the (n x p) data matrix where y_{ij} represents the j^{th} response for the i^{th} subject. The GLMM is as follows:

$$E(\underset{\sim}{Y}) = \underset{\sim}{X}\underset{\sim}{B}$$

$$\mathrm{Var}(i^{th} \text{ row of } \underset{\sim}{Y}) = \underset{\sim}{\Sigma}$$

where

$\underset{\sim}{X}$(n x q) is a known design matrix of rank $\leq$ q.

$\underset{\sim}{B}$(q x p) is a matrix of unknown model parameters.

$\underset{\sim}{\Sigma}$(p x p) is a positive definite symmetric matrix.

If it is further assumed that the rows of $\underset{\sim}{Y}$ follow a normal distribution, then the model is said to be a GLMM with normality.

1.2 Analysis of Incomplete Multivariate Data

Early work in multivariate data analysis with missing values focused on maximum likelihood estimation of the population parameters. Under the assumption of a bivariate normal distribution, Wilks (1932) derived the likelihood equations for samples with missing values in both variates. Due to the complexity of finding solutions, he also suggested estimates based upon a "complete" sample in which sample means were substituted for the missing values.

Anderson (1957) proposed a method for obtaining the maximum likelihood estimates (MLE's) of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ in a p-variate normal sample with the missing data structured hierarchically. Specifically it was assumed that the likelihood for the data observed on the i^{th} variate was conditional upon the data for the preceding (i-1) variates. An example of such a structure to which this method applies is as follows:

$$x_1, \ldots, x_n, \; x_{n+1}, \ldots, x_N$$
$$y_1, \ldots, y_n, \; y_{n+1}, \ldots, y_N$$
$$z_1, \ldots, z_n$$

Assuming that x, y, and z are jointly distributed according to a trivariate normal distribution, Anderson's method can be used to produce MLE's of the population parameters.

Nicholson (1957) applied Wilks' likelihood equations to the multivariate normal (MVN) sample with missing values for one variate only, giving MLE's for the population mean vector and covariance matrix.

None of the above papers address the issue of observations missing at random as opposed to missing by design. Trawinski and Bargmann (1964) derived the likelihood equations for the general multivariate normal case in which observations are missing by design. An iterative solution of the likelihood equations was presented.

Several early papers approached the missing data problem via techniques other than maximum likelihood, thus avoiding any stringent distributional assumptions as well as problems of computational feasibility. Dear (1959) proposed a method for estimating randomly missing values in the design matrix of a linear model. This technique estimates the unknown values with the first principal component for the known portion of the matrix.

Buck (1960) suggested replacing missing values with regression estimates, which may be described as follows. In the simplest case, only one missing value for each experimental unit, multiple regression equations are solved using only those observational units with complete data. Appropriate predicted values are then substituted for missing data when needed. The extension of this method to cases where more than one variable may be missing for some units requires computation of multiple regression estimates for all possible combinations of missing data. The authors also gave the resulting correction for bias needed in the estimation of covariance parameters.

Glasser (1965) used an intuitive approach to estimate linear regression parameters for samples in which at least two subjects have complete data for each pair of independent variables. His method, often called "pairwise deletion," uses information from all subjects with non-missing values for the i^{th} and j^{th} variates to compute the respective covariance. These estimates combined with sample means and variances yield regression estimates that are unbiased and consistent.

Timm (1970) published the results of a comparison study of several of the methods already discussed, namely those of Dear, Buck, and substitution of means due to Wilks. He also included estimation based only on the complete data vectors, commonly referred to as listwise deletion. Only estimation of the population variance-covariance and correlation matrices was evaluated. After varying the number of variates, proportion of missing data, sample size, and average correlations, the author concluded that in general the methods of Buck and Dear were preferred, with Dear's method being computationally more desirable.

Gleason and Staelin (1975) extended Timm's study to include a new procedure in which missing data are estimated via a transformation of the largest principal components of the data matrix. This procedure was compared to three others for estimating missing data values: methods due to Wilks, Dear, and an extension of the method due to Buck. In addition, these four techniques plus Glasser's pairwise deletion were compared with respect to estimation of the population correlation matrix. In both studies, Gleason and Staelin's principal components method fared at least as well if not better than the others under most conditions. The root-mean-square standardized residual was the criteria used to measure the accuracy of the various methods.

Two different methods have been proposed recently that yield iterative solutions for maximum likelihood estimates in the general multivariate normal case, i.e., no assumptions on the structure of the missing data. The first of these is due to Hocking and Smith (1968) and Hartley and Hocking (1971). Hocking and Smith dealt with estimation of $\underset{\sim}{\Sigma}$ only. They suggested subsetting the observations into groups with identical patterns of missing values. Then the MLE's based on the group with complete data are refined by adjoining linear functions of MLE's from each group separately until all data has been used. At each stage an upper bound on the gain in precision from incorporating the additional information is available. While unable to verify that the final estimates were maximum likelihood in cases with more than two groups, the authors did show that the estimates attained large sample properties similar to those of MLE's.

Hartley and Hocking (1971) summarized the various aspects of the incomplete data problem. They also defined a procedure which generalized the method due to Hocking and Smith and illustrated that their estimates were indeed maximum likelihood. The assumptions are that N observations sampled from a p-variate normal distribution with mean vector $\underset{\sim}{\mu}$ and covariance matrix $\underset{\sim}{\Sigma}$ are divided into T groups, group t having n_t observations, each observation within group t having the same missing variates. The likelihood equations are developed by noting that the total likelihood is just the product of the group likelihoods, resulting in equations that are in terms of the t groups of MLE's. The authors then proposed an iterative solution which is shown to converge quickly in some examples. The large sample covariance matrices for the estimates are available at each step of the iteration. The difficulty with this method is that the parameters for a group with small n_t may be unestimable. The likelihood equations of Hartley and Hocking can be solved analytically for the special cases of nested groups considered by Anderson (1957). The authors also verified that their procedure yields the same likelihood equations as Trawinski and Bargmann for cases in which the mean vector depends on other unknown parameters.

The second method of maximum likelihood for the general incomplete data problem was derived by Woodbury and Hasselblad (1970) and Orchard and Woodbury (1972) and is based on Woodbury's "missing information principle." According to this principle, the missing values are random variables and the likelihood for the sample involves the conditional distribution of the complete data given the observed data. The resulting likelihood equations are often more readily solvable than those based only on the observed data.

In this paper, Orchard and Woodbury applied their principle to several examples, including a multivariate normal sample. First note that for the k^{th} $(p \times 1)$ observation vector of a GLMM, we could estimate complete date as

$$\hat{\underset{\sim}{Y}}_k = \underset{\sim}{Y}_{k,0} + \hat{\underset{\sim}{Y}}_{k,m}$$

where $\underset{\sim}{Y}_{k,0}$ contains zeros in place of missing values and observed values elsewhere, while $\hat{\underset{\sim}{Y}}_{k,m}$ contains estimates in place of missing values and zeroes elsewhere. The estimate to be used for $\hat{\underset{\sim}{Y}}_{k,m}$ is the expectation of the MVN distribution conditional upon the observed data; i.e.,

$$\hat{\underset{\sim}{Y}}_{k,m} = \hat{\underset{\sim}{\mu}}_m + \underset{\sim}{\Sigma}_{m,0} \; \underset{\sim}{\Sigma}_{0,0}^{-1} \; (\underset{\sim}{Y}_{k,0} - \hat{\underset{\sim}{\mu}}_0).$$

Initial estimates for $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ are based on the group of complete data vectors. Note that the parameters of the conditional expectation above will be the same for all observations with equivalent missing data patterns. The estimated complete data matrix can then be used to refine estimates of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ for the next iteration.

Dempster, Laird, and Rubin (1976) argued that the algorithm proposed by Orchard and Woodbury (1972) is a speical case of their two-step iterative EM algorithm. During the estimation (E) step of a particular iteration, sufficient statistics of the hypothetical complete data matrix $\underset{\sim}{X}$ are estimated conditional upon the observed data $\underset{\sim}{Y}$. The maximization step (M) then consists of maximum likelihood estimation of parameters based on X. If ϕ is the parameter set and $g(\underset{\sim}{Y}|\underset{\sim}{\phi})$ the sampling distribution for Y, the authors proved that when $g(\underset{\sim}{Y}|\underset{\sim}{\phi})$ is a member of the regular exponential family of distributions the EM algorithm converges to $\underset{\sim}{\phi}^*$ which maximizes $[\log g(\underset{\sim}{Y}|\underset{\sim}{\phi})]$, i.e., $\underset{\sim}{\phi}^*$ is MLE. The convergence properties for $\underset{\sim}{\phi}_n$ are also given in the paper. If the EM algorithm is applied to a MVN $(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$ sample, the sufficient statistics to be computed at the E step are the column sums of $\underset{\sim}{X}$ and the sums of squares and cross-products for the columns of $\underset{\sim}{X}$. The missing components of these statistics are replaced with the conditional expectations given the observed data for each row of $\underset{\sim}{X}$. The M step is straightforward using the sample moments. It is obvious that this algorithm produces the same estimates as that of Orchard and Woodbury.

Dempster, Rubin, and Tsutakawa (1981) applied the EM algorithm to a variance components estimation problem with missing data. Laird and Ware (1983) illustrated the use of the EM algorithm with a random effects model. They proposed a two-stage model for analyzing unbalanced longitudinal data. Iterative estimation of the fixed and random effects was presented assuming equal correlations between observations on a subject.

Other recent papers approach the missing data problem without relying on the likelihood of the sample. Kleinbaum (1970, 1973) investigated estimation and hypothesis testing from a "More General Linear Model" (MGLM) framework that includes samples with missing data as well as different design matrices for different dependent variables.

Two unbiased, consistent estimators of $\underset{\sim}{\Sigma}$ were derived based on pairwise deletion which can be used to produce nonlinear BAN estimators for functions of the model parameters. Two iterative procedures were proposed to refine these initial estimates. In addition, Wald statistics were constructed from the BAN estimators which can be used to test linear hypotheses about functions of the treatment parameters. The author illustrated these methods for several examples. Kleinbaum also extended the growth curve model proposed by Potthoff and Roy (1964) to include missing data and developed BAN estimators of the model parameters.

Hosking (1980) examined a special case of the incomplete data problem. Randomly missing values were assumed to occur in the dependent variables only of a GLMM with normality. Four techniques for estimation of the model parameters were considered; listwise deletion and methods due to

(i) Hocking and Smith (1968) and Hartley and Hocking (1972), ("HHS");

(ii) Woodbury and Hasselblad (1970) and Orchard and Woodbury (1972), ("WOH"); and

(iii) Kleinbaum (1970, 1973) ("KLN").

Hosking first generalized the methods of HHS and WOH to include maximum likelihood estimation of $\underset{\sim}{\beta}$ where $E(\underset{\sim}{Y}) = \underset{\sim}{\mu} = \underset{\sim}{X}\underset{\sim}{\beta}$. He also presented a derivation of MLE's for $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$ based on the work of Trawinski and Bargmann (1964). Although the maximum likelihood techniques were shown to be BAN, no relationships were found between those estimators and the KLN estimators.

The Monte Carlo study included the four techniques above, pairwise deletion, and estimation of the complete data matrix by standard GLM methods. Sample size, proportion of missing values, and the average intercorrelation of the dependent variables were varied. For the latter factor, two correlation structures were used for "low" and "high" respectively. Four matrix-valued measures and three summary (scalar) measures of accuracy were used. One major result was that the HHS algorithm failed to work for the general model $E(\underset{\sim}{Y}) = \underset{\sim}{X}\underset{\sim}{\beta}$. (Hartley and Hocking used $\underset{\sim}{X} = 1$, $\underset{\sim}{\beta} = \underset{\sim}{\mu}(1 \times p)$). In some cases the subgroup parameters were inestimable due to small subgroup sizes while in other cases, the algorithm simply did not converge, or converged to unacceptable estimates. The WOH and KLN algorithms converged in all cases and were superior to listwise deletion with respect to all measures of accuracy. In general, the KLN technique was preferred slightly for the estimation of $\underset{\sim}{\Sigma}$ while the WOH technique was preferred for estimation of $\underset{\sim}{\beta}$.

Longitudinal studies designed to contain missing data have also appeared in the literature. Rao and Rao (1966) recommended a "linked cross-sectional" study to determine growth norms and rates for Indian school-age boys. The height and weight of each child aged 5 to 16 years was measured yearly for no more than 3 years. The following data patterns could therefore occur: measurements at one age only, at two consecutive ages, at three ages, and at ages one year apart. The subjects were grouped according to place of residence and regional differences were studied.

Rao and Rao used univariate weighted least squares theory to produce estimates of the mean height (weight) at each age $(\mu_5, \ldots, \mu_{16})$ by assuming the following:

(i) Equal variances at each age for each subject;

(ii) Equal correlations between observations.

Pooled estimates of σ^2, ρ_1, and ρ_2 over all subgroups were used in the analysis where ρ_1 signifies the correlation of any two observations i years apart. The results of the analysis showed linked cross-sectional designs to be "efficient" in the sense that the standard errors of the growth norms were almost half the magnitude of those resulting from the standard cross-sectional analysis. The savings of such studies are compared to longitudinal studies with respect to time and cost were also discussed.

Woolson, Leeper, and Clarke (1978) and Woolson and Leeper (1980) suggested analyzing the data arising from a linked cross-sectional study in the framework of Kleinbaum's MGLM. It was assumed that the covariance matrix for each subject's responses had the form $\sigma^2 \underset{\sim}{\Gamma}_j$ for a known, symmetric, positive definite matrix $\underset{\sim}{\Gamma}_j$ and unknown σ^2. The problem then reduced to one in which univariate least squares theory applies. For the special case where $\underset{\sim}{X}\underset{\sim}{\beta} = \underset{\sim}{\mu}$, the resulting equations are equivalent to the least squares equations solved by Rao and Rao.

Woolson, et al. applied their technique to a linked cross-sectional study, here referred to as a "mixed longitudinal design." Growth data were collected from all 12 grades of a school system for five consecutive years, covering ages from six to 18 years. It was of interest to estimate growth norms and rates at each age. Thus some measurements are missing due to the design of the study and some due to attrition. Following the above model, two separate analyses were performed. For the first analysis, $\underset{\sim}{\Gamma}_j$ was assumed to have a fixed, known structure and for the second, pooled estimates of correlations across ages were used. In addition, a standard cross-sectional analysis was carried out. Little differences in the parameter estimates was observed for the two estimates of $\underset{\sim}{\Gamma}_j$, however, the mixed longitudinal analysis yielded considerable reduction in standard errors of the estimates when compared to the cross sectional analysis. In conclusion the authors indicated that strong justification is needed for the choice of $\underset{\sim}{\Gamma}_j$ in the analysis.

Schwertman and Allen (1979) examined the problem of estimation with Kleinbaum's MGLM when the estimate $\hat{\Sigma}$ is non-positive definite. The authors proposed an iterative smoothing procedure that produces a positive semi-definite estimate that is "closest" to $\hat{\Sigma}$. A simulation study was carried out to investigate the reliability of the Wald statistics computed with this smoothing procedure. The results indicated that smoothing was frequently needed with small sample sizes and particularly sparse data, and that the smoothing process seemed to stablize the Wald statistics to reasonable values.

Leeper and Woolson (1982) performed a simulation study to determine the small sample properties of the Wald statistics proposed by Kleinbaum (1970) using three estimators of Σ. The pairwise deletion estimator originally proposed by Kleinbaum, the smoothed estimator of Σ proposed by Schwertman and Allen, and a second smoothed estimator developed by the authors were compared. The smoothed estimators eliminated the problem of negative test statistics but greatly affected the distribution of the test statistics.

1.3 Covariance Models Associated with Longitudinal Data

Modeling of the population covariance matrix is often employed in conjunction with the analysis of longitudinal data. Potthoff and Roy (1964) suggested specifying a structure for Σ in the analysis of growth curve data. In their example, four measurements on a subject were presumed to be serially correlated and the following model for Σ was applied:

$$\underset{\sim}{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

where ρ is estimated from a separate but similar experiment.

Machin (1975) assumed this same general structure for $\underset{\sim}{\Sigma}$ in a comparison of two longitudinal designs. Instead of measuring n subjects on the same p occasions, he proposed taking fewer measurements on each subject and compensating by following more subjects such that the total number of observations remained unchanged. Not all measurements in the second study were taken on the same occasions. Assuming a GLMM with normality and known $\underset{\sim}{\Sigma}$, expressions for the variances of the model parameters, $\underset{\sim}{\beta}$, were derived. The relative efficiency of the two designs with respect to $\underset{\sim}{\beta}$ was examined for a specified structure of $\underset{\sim}{\Sigma}$ (the serial correlation matrix described above). The second study proved more efficient whenever ρ was positive.

Morrison (1972) considered the analysis of repeated measurements in the presence of constraints on the covariance matrix. His primary consideration was testing the equality of the elements of the mean vector, $\underset{\sim}{\mu}$, from a multinormal sample. The model assumes p responses not necessarily ordered in time, as in the GCM. The hypothesis of interest is then

$$H_0: \ \mu_1 = \ldots = \mu_p .$$

For a general, positive definite $\underset{\sim}{\Sigma}$, H_0 can be tested with Hotelling's T^2 statistic. If the "symmetric variance-covariance" model is assumed for $\underset{\sim}{\Sigma}$, i.e., $cov(x_i, x_j) = \rho\sigma^2$ and $var(x_i) = \sigma^2$ the generalized likelihood ratio test due to Wilks (1946) applies. Morrison developed two different test statistics suitable for a class of "reducible" covariance matrices. This class includes matrices such that $\underset{\sim}{C}\underset{\sim}{\Sigma}\underset{\sim}{C}'$ is diagonal for some $\underset{\sim}{C}$, (p - 1)x p, of rank (p - 1) whose elements are independent of $\underset{\sim}{\Sigma}$. One example is the covariance matrix associated with a Wiener stochastic process:

$$\underset{\sim}{\Sigma} = \sigma^2 \cdot \begin{bmatrix} t_1 & t_1 \cdots t_1 \\ t_1 & t_2 \cdots t_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ t_1 & t_2 \cdots t_p \end{bmatrix} .$$

Another class of matrices considered are those reducible to a "compound symmetry pattern." The successive serial correlation matrix is an example:

$$\underset{\sim}{\Sigma} = \sigma^2 \cdot \begin{bmatrix} 1 & \rho & 0 \ldots 0 \\ \rho & 1 & \rho \ldots 0 \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ 0 & 0 & 0 \ldots 1 \end{bmatrix} .$$

The information statistic due to Geisser (1963) can be used to test the repeated measures hypothesis for this class. Morrison proposed a likelihood ratio test developed by Bhargava (1962) for longitudinal studies with hierarchically missing data. The gains in sensitivity of the tests that follow from assuming a structure of $\underset{\sim}{\Sigma}$ are illustrated by comparing the average squared lengths of the simultaneous confidence intervals.

Modeling of covariance matrices also appears extensively in the analysis of econometric data. Longitudinal data consisting of responses measured at T time points on A cross-sectional units are often analyzed in the field of econometrics with a cross-sectional time series (CSTS) model. The general definition of the CSTS model is as follows:

$$\underset{\sim}{y}_i = \underset{\sim}{x}_i \underset{\sim}{B} + \underset{\sim}{u}_i \qquad , \; i=1,\ldots,A,$$

where

$$\underset{\sim}{y}_i = (y_{i1}, y_{i2},\ldots,y_{iT})'$$

is a vector of responses for the i^{th} cross-sectional unit taken at T distinct time points, $\underset{\sim}{x}_i$ is a known (T x q) design matrix, $\underset{\sim}{B}$ is the (q x 1) vector of unknown model parameters, invariant over time, and $\underset{\sim}{u}_i$ is the (T x 1) vector of residuals. Several methods of estimation for this model have been proposed in the literature and are briefly reviewed here. Each method differs with respect to the assumptions made about the behavior of the $\underset{\sim}{u}_i$. For a more comprehensive review of the literature on pooled cross-sectional time series data the reader is referred to Dielman (1983).

Early work in this area relied heavily on the method of "seemingly unrelated regressions" due to Zellner (1972). To apply this method, the following assumptions for the CSTS model must be made:

$$\text{(i)}\; E(\underset{\sim}{u}_i) = \underset{\sim}{0},$$

$$\text{(ii)}\; E(\underset{\sim}{u}_i \underset{\sim}{u}_i') = \sigma_{ij} \underset{\sim}{I}_T, \quad i,j=1,\ldots,A.$$

Zellner suggested estimating the $\{\sigma_{ij}\}$ via ordinary least squares (OLS) regression, i.e.,

$$\hat{\sigma}_{ij} = \frac{1}{T-q} (\underset{\sim}{y}_i - \underset{\sim}{X}_i \hat{\underset{\sim}{B}}_0)' (\underset{\sim}{y}_j - \underset{\sim}{X}_j \hat{\underset{\sim}{B}}_0)$$

where $\hat{\underset{\sim}{B}}_0$ is the OLS estimator of $\underset{\sim}{B}$. The weighted least squares (WLS) estimator of $\underset{\sim}{B}$ computed using the $\hat{\sigma}_{ij}$ is shown to be unbiased and to follow the same asymptotic distribution as $T\to\infty$ as the WLS estimator of $\underset{\sim}{B}$ when the σ_{ij} are known.

Zellner's method does not allow for correlations among measurements on a given cross-sectional unit, which is often unreasonable in econometric analyses. Several methods have been proposed that incorporate various correlation structures with respect to time in the CSTS model. Parks (1967) developed a three-stage estimation procedure assuming that the $\underset{\sim}{y}_i$ follow an autoregressive process of order one, AR(1), namely

$$u_{it} = \rho_i u_{i,t-1} + e_{it} \quad ,t=2,\ldots,T$$

where

$$E(e_{it}) = 0$$

$$E(e_{it} e_{js}) = \begin{cases} \sigma_{ij} & \text{if } t=s \\ 0 & \text{otherwise} \end{cases}$$

assuming the usual initial conditions it can be shown that

$$E(u_{it} u_{js}) = \sigma_{ij}\rho_i^{|t-s|}/(1 - \rho_i\rho_j).$$

The ρ_i are estimated in the first stage, the σ_{ij} are estimated in the second stage, and the third stage consists of computing the WLS estimator of $\underset{\sim}{B}$ with these estimates. Let $\underset{\sim}{X} = (\underset{\sim}{x}_1, \underset{\sim}{x}_2,\ldots\underset{\sim}{x}_A)'$. Then assuming $\underset{\sim}{X}'\underset{\sim}{X}/T$ approaches a positive definite matrix as $T\to\infty$ and $E(\underset{\sim}{B})$ exists, Parks' estimator of $\underset{\sim}{B}$ is unbiased and has the same asymptotic distribution as the WLS estimator when all variance-covariance parameters are known.

Kmenta (1971) simplified Parks' assumptions and thereby proposed a two-stage estimator of $\underset{\sim}{B}$. By assuming $\sigma_{ij} = 0$ if $i \neq j$ the model reduces to one in which the A cross-sectional units are heteroscedastic but autocorrelations are present only within each subject with respect to time, i.e.,

$$E(u_{it} u_{js}) = \begin{cases} \sigma_i^2 \rho_i^{|t-s|}/(1 - \rho_i^2) & \text{if } i = j \\ 0 & \text{if } i \neq j . \end{cases}$$

Kmenta's estimator enjoys the same asymptotic properties as Parks' estimator but can be computed in two steps.

Fuller and Battese (1974) considered estimation for the CSTS model with a "crossed-error" covariance structure instead of specifying a time series model. The residuals are assumed to have the form

$$u_{ij} = a_i + b_j + e_{ij}$$

where a_i is the cross-sectional unit effect, b_j is the time effect, and e_{ij} is the residual effect. It is assumed that a_i, b_j, and e_{ij} are independently distributed with mean zero and variances σ_a^2, σ_b^2, and σ_e^2, all positive.

Fuller and Battese suggest estimating the variance components with the "fitting-of-constants" method (e.g., Searle, 1971). The WLS estimator of $\underset{\sim}{B}$ computed with the estimated variance components is then shown to be unbiased and asymptotically distributed as the WLS estimator of B when all variance components are known. The authors describe a computational scheme involving OLS regressions of transformed data with minimal matrix manipulations.

Da Silva (1975) introduced a model more general than those described above. He claims that a major limitation of the methods due to Zellner, Parks, and Kmenta is the assumption of a fixed number of cross-sectional units. The resulting asymptotic theory holds only as the number of time points increases to infinity. It is often desirable in econometric analysis to assume that the cross-sectional units are a sample from a population of interest. To solve this short-coming, da Silva's model incorporates a random cross-sectional unit effect, a random time effect, and a finite moving average model for residuals. He argues that while most econometric analysis is carried out under the assumption of an AR(1) residual model, an infinite moving average method is the most general model for stationary processes and can be closely approximated with a finite moving average model, MA(q).

It is assumed that

$$u_{it} = a_i + b_t + e_{it} \qquad ,i=1,\ldots,A,\ t=1,\ldots,T,$$

where a_i and b_t are as specified above in the crossed-error model due to Fuller and

$$e_{it} = \alpha_0\, \varepsilon_t + \alpha_1\, \varepsilon_{t-1} + \ldots + \alpha_M\, \varepsilon_{t-M}, \quad M \le T-1.$$

The $\{\varepsilon_t\}$ are iid random variables with $E(\varepsilon_t^2) = \sigma_\varepsilon^2 > 0$. We have

$$\mathrm{Cov}(e_{it}, e_{is}) = \begin{cases} \gamma(|t-s|) & , \quad |t-s| \le M \\ 0 & , \quad |t-s| > M. \end{cases}$$

Da Silva formulates the variance of $\underset{\sim}{y} = (\underset{\sim}{y}_1,\ldots,\underset{\sim}{y}_A)'$, $\underset{\sim}{V}$, in terms of the M + 3 unknown variance components. He then applies a method due to Seely (1969) for estimating the components and computes $\hat{\underset{\sim}{V}}$. A method due to Hannan (1963) can be used to transform $\hat{\underset{\sim}{V}}$ into a matrix, $\underset{\sim}{V}^*$, with a known spectral decomposition. The WLS estimate of $\underset{\sim}{B}$ computed using $\underset{\sim}{V}^*$ is then shown to be unbiased and asymptotically normal and efficient due to the properties of these Hannan estimators.

Andersen (1981) applied time series covariance models to a two-way analysis of variance design. Under the assumptions of an autoregressive error process and a moving average error process, both of order one, the authors derived an approximate F test for testing row and column effects. Their work relies heavily on that due to Box (1954, 1958) and Anderson (1971).

CHAPTER 2

MODEL FORMULATIONS AND RESULTS

2.1 Model Definitions

In this section the various multivariate linear models in the literature are reviewed for application to incomplete longitudinal datasets. Each model that is introduced is shown to be a special case of a general linear model, defined below, with suitable constraints imposed.

Consider the following, very general definition of a linear model (LM):

(2.1.1) $$E[\underset{\sim}{y}] = \underset{\sim}{Z}\,\underset{\sim}{\gamma}$$

$$\mathrm{Var}[\underset{\sim}{y}] = \underset{\sim}{\Omega}$$

where $\underset{\sim}{y}$ is an (N x 1) vector of observations,

$\underset{\sim}{\gamma}$ is a (q x 1) vector of unknown model parameters,

$\underset{\sim}{Z}$ is an (N x q) design matrix corresponding to $\underset{\sim}{\gamma}$ with full column rank (= q),

$\underset{\sim}{\Omega}$ is an (N x N) positive definite symmetric covariance matrix.

One can add the normality assumption (optional):

$$\underset{\sim}{y} \sim N(\underset{\sim}{Z}\,\underset{\sim}{\gamma}, \underset{\sim}{\Omega}).$$

In this most general case $\underset{\sim}{\Omega}$ consists of N x (N + 1)/2 unknown elements. By imposing constraints on these elements, the model reduces to other well-known models. For example, suppose

$$\underset{\sim}{\Omega} = \sigma^2\, \underset{\sim}{I}_N$$

where σ^2 is an unknown variance parameter. Then the observations in $\underset{\sim}{y}$

are independent, assuming a normal distribution, and homoscedastic. The Gauss-Markov Theorem guarantees that best linear unbiased estimates of $\underset{\sim}{\gamma}$ exist and are given by

$$\hat{\underset{\sim}{\gamma}} = (\underset{\sim}{Z}' \underset{\sim}{Z})^{-1} \underset{\sim}{Z}' \underset{\sim}{y} .$$

The General Linear Multivariate Model (GLMM) defined in Chapter 1 is also a special case of the LM (2.1.1). Given a data matrix $\underset{\sim}{Y}$ arising from an experiment in which p responses were measured on n subjects, suppose $\underset{\sim}{Y} \sim$ GLMM ($\underset{\sim}{X}$ $\underset{\sim}{B}$, $\underset{\sim}{\Sigma}$) where $\underset{\sim}{X}$, $\underset{\sim}{B}$, and $\underset{\sim}{\Sigma}$ are as defined in (1.1.1). Let

$$\underset{np \times 1}{\underset{\sim}{y}} = [\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_n]'$$

$$\underset{np \times pq}{\underset{\sim}{Z}} = \begin{bmatrix} \underset{\sim}{I}_p \otimes \underset{\sim}{x}_1' \\ \vdots \\ \underset{\sim}{I}_p \otimes \underset{\sim}{x}_n' \end{bmatrix}$$

$$\underset{pq \times 1}{\underset{\sim}{\gamma}} = \begin{bmatrix} \underset{\sim}{b}_1 \\ \vdots \\ \underset{\sim}{b}_p \end{bmatrix}$$

$$\underset{np \times np}{\underset{\sim}{\Omega}} = \underset{\sim}{I}_n \otimes \underset{\sim}{\Sigma}$$

where

$\underset{\sim}{y}_i$ denotes the i^{th} row of $\underset{\sim}{Y}$, $i = 1,\ldots,n$,

$\underset{\sim}{x}_i$ denotes the i^{th} row of $\underset{\sim}{X}$, $i = 1,\ldots,n$,

$\underset{\sim}{b}_j$ denotes the j^{th} column of $\underset{\sim}{B}$, $j = 1,\ldots,p$.

With these definitions and constraints on the structure of $\underset{\sim}{Z}$ and $\underset{\sim}{\Omega}$, the linear model given in (2.1.1) reduces to the GLMM with parameters $\underset{\sim}{B}$ and $\underset{\sim}{\Sigma}$.

Kleinbaum (1970, 1973) expanded the GLMM into a more general framework, which he called the More General Linear Model (MGLM), in order to allow for missing data as well as different design matrices for different dependent variables. In the MGLM it is assumed that each of p response variates was measured on some subset of n experimental units. For the s^{th} response variate, s = 1,...,p let

$\underset{\sim}{y}_s$ denote the $(N_s \times 1)$ vector of observations,

$\underset{\sim}{b}_s$ denote the $(q_s \times 1)$ vector of model parameters,

$\underset{\sim}{A}_s$ denote the $(N_s \times q_s)$ design matrix.

The "vector-version" of the MGLM given by Kleinbaum is equivalent to the LM with

(2.1.2) $$\underset{N\times 1}{\underset{\sim}{y}} = \begin{bmatrix} \underset{\sim}{y}_1 \\ \vdots \\ \underset{\sim}{y}_p \end{bmatrix}, \quad \underset{Q\times 1}{\underset{\sim}{\gamma}} = \begin{bmatrix} \underset{\sim}{b}_1 \\ \vdots \\ \underset{\sim}{b}_p \end{bmatrix},$$

$$\underset{N\times Q}{\underset{\sim}{Z}} = \begin{bmatrix} \underset{\sim}{A}_1 & & \underset{\sim}{0} \\ & \ddots & \\ \underset{\sim}{0} & & \underset{\sim}{A}_p \end{bmatrix},$$

$$\underset{N\times N}{\underset{\sim}{\Omega}} = \begin{bmatrix} \sigma_{11}\,\underset{\sim}{I}_{N_1} & \cdots & \sigma_{1p}\,\underset{\sim}{U}_{1p} \\ \vdots & & \\ \sigma_{1p}\,\underset{\sim}{U}'_{1p} & \cdots & \sigma_{pp}\,\underset{\sim}{I}_{N_p} \end{bmatrix},$$

$$N = \sum_{s=1}^{p} N_s, \quad Q = \sum_{s=1}^{p} q_s, \quad \underset{\sim}{\Sigma} = ((\sigma_{ij})), \quad \underset{\sim}{U}_{rs} = ((u_{ij}))_{rs},$$

and $$u_{ij(rs)} = \begin{cases} 1 & \text{if the } i^{th} \text{ observation on the } r^{th} \text{ variate and the } j^{th} \text{ observation on the } s^{th} \text{ variate are on the same subject} \\ 0 & \text{otherwise.} \end{cases}$$

Thus the LM formulation of (2.1.1) encompasses the MGLM when the above definitions and constraints are applied. The MGLM allows a great deal of freedom in model assumptions. The number of observations on a response variate (N_s) is allowed to vary as well as the number of model parameters (q_s) used to predict each response.

The General Incomplete Model, GIM, was introduced by Kleinbaum (1970) as a special case of the MGLM in which the vector of model parameters, $\underset{\sim}{b}_s$, is of dimension (q x 1) for all response variates, $s = 1,\ldots,p$. Missing observations are allowed, however, and the subjects are grouped according to patterns of missing data. For the j^{th} missing data group, $j = 1,\ldots, u$, we have the model:

$$E[\underset{\sim}{Y}_j] = \underset{\sim}{D}_j \underset{\sim}{B} \underset{\sim}{K}_j$$

$$\text{Var}[i^{th} \text{ row of } \underset{\sim}{Y}_j] = \underset{\sim}{K}_j' \underset{\sim}{\Sigma} \underset{\sim}{K}_j$$

where

$\underset{\sim}{Y}_j$ is the $(n_j \times m_j)$ matrix of observations,

$\underset{\sim}{B}$ is the $(q \times p)$ matrix of model parameters,

$\underset{\sim}{D}_j$ is the $(n_j \times q)$ design matrix,

$\underset{\sim}{K}_j$ is the $(p \times n_j)$ incidence matrix defining the variates observed in the j^{th} group, i.e., $\underset{\sim}{K}_j = ((k_{\ell i}))_j$ where

$$k_{\ell i_{(j)}} = \begin{cases} 1 & \text{if the } i^{th} \text{ variate measured on the } j^{th} \text{ group corresponds to the } \ell^{th} \text{ variate in the complete data matrix, } i = 1,\ldots,m_j,\ \ell = 1,\ldots,p, \\ 0 & \text{otherwise.} \end{cases}$$

To see that the GIM is a special case of the LM, consider the "vector-version" of this model. Let $\underset{\sim}{y}_j$ denote the $(n_j m_j \times 1)$ vector of columns of $\underset{\sim}{Y}_j$ (i.e., $\underset{\sim}{Y}_j$ "rolled out by columns"). Then the GIM is equivalent to the LM with

$$\underset{N\times 1}{\tilde{y}} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_u \end{bmatrix}, \quad \underset{pq \times 1}{\tilde{\gamma}} = \begin{bmatrix} \tilde{b}_1 \\ \vdots \\ \tilde{b}_p \end{bmatrix},$$

$$\underset{N \times pq}{\tilde{\tilde{Z}}} = \begin{bmatrix} \tilde{K}_1' \otimes \tilde{D}_1 \\ \vdots \\ \tilde{K}_u' \otimes \tilde{D}_u \end{bmatrix},$$

$$\underset{N\times N}{\tilde{\tilde{\Omega}}} = \begin{bmatrix} \tilde{K}_1' \tilde{\tilde{\Sigma}} \tilde{K}_1 \otimes \tilde{I}_{n_1} & & \tilde{0} \\ & \ddots & \\ \tilde{0} & & \tilde{K}_u' \tilde{\tilde{\Sigma}} \tilde{K}_u \otimes \tilde{I}_{n_u} \end{bmatrix},$$

$$\tilde{N} = \sum_{j=1}^{u} n_j m_j .$$

The GLMM, MGLM, and GIM include the assumption that responses measured on a subject are a subset of a well-defined set of p response variates. It is often the case in longitudinal data that responses for an individual are measured at arbitrary or even unique time intervals. The following definitions will adapt the LM to such a situation.

For the j^{th} individual, $j = 1, \ldots, n$, let

> $\tilde{y}_j$ denote the $(m_j \times 1)$ vector of measurements or responses,
>
> $\tilde{\gamma}$ denote the $(q \times 1)$ vector of model parameters,
>
> $\tilde{Z}_j$ denote the $(m_j \times q)$ design matrix.

The following individual model can be defined:

(2.1.4)
$$E[\tilde{y}_j] = \tilde{Z}_j \tilde{\gamma}$$

$$Var[\tilde{y}_j] = \tilde{V}_j .$$

where it is assumed that $\underset{\sim}{y}_j$ is independent of $\underset{\sim}{y}_{j'}$ for $j \neq j'$. The n individual models can be combined into the following version of the LM:

(2.1.5)
$$E[\underset{\sim}{y}] = \underset{\approx}{Z}\,\underset{\sim}{\gamma}\,,$$

$$\mathrm{Var}[\underset{\sim}{y}] = \underset{\approx}{\Omega}\,,$$

where

$$\underset{N\times 1}{\underset{\sim}{y}} = \begin{bmatrix} \underset{\sim}{y}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{y}_n \end{bmatrix}, \qquad \underset{N\times q}{\underset{\approx}{Z}} = \begin{bmatrix} \underset{\sim}{Z}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{Z}_n \end{bmatrix},$$

$$\underset{N\times N}{\underset{\approx}{\Omega}} = \begin{bmatrix} \underset{\sim}{V}_1 & & & \underset{\sim}{0} \\ & \cdot & & \\ & & \cdot & \\ \underset{\sim}{0} & & & \underset{\sim}{V}_n \end{bmatrix}, \qquad N = \sum_{j=1}^{n} m_j\,.$$

This model will be referred to as the linear model for individual observation times, LMI.

Laird and Ware (1983) considered the use of a two-stage random effects model for the analysis of longitudinal data with arbitrary observation times for each subject. Their model is a special case of the LMI in which each covariance matrix, $\underset{\sim}{V}_j$, is assumed to have a particular form. At the first stage, the j^{th} subject's responses are modeled as

(2.1.6)
$$\underset{\sim}{y}_j = \underset{\sim}{Z}_j\,\underset{\sim}{\gamma} + \underset{\sim}{X}_j\,\underset{\sim}{\alpha}_j + \underset{\sim}{e}_j$$

where
$$\mathrm{Var}(\underset{\sim}{e}_j) = \underset{\sim}{V}_j$$

and
$$\mathrm{Cov}(\underset{\sim}{e}_j, \underset{\sim}{e}_{j'}) = 0,\ j \neq j'\,.$$

Here $\underset{\sim}{\gamma}$ and $\underset{\sim}{\alpha}_j$ are assumed to be fixed effects. At the second stage, the distribution of $\underset{\sim}{\alpha}_j$ across individuals is introduced:

$$E(\underset{\sim}{\alpha}_j) = 0,$$

$$\mathrm{Var}(\underset{\sim}{\alpha}_j) = \underset{\sim}{U},$$

$$\mathrm{Cov}(\underset{\sim}{\alpha}_j, \underset{\sim}{\alpha}_{j'}) = \underset{\sim}{0} \text{ for } j \neq j' ,$$

$$\mathrm{Cov}(\underset{\sim}{\alpha}_j, \underset{\sim}{e}_j) = \underset{\sim}{0} ,$$

where $\underset{\sim}{U}$ is a positive definite, symmetric matrix. Assuming a normal distribution for $\underset{\sim}{y}_j$ and $\underset{\sim}{\alpha}_j$, we have

(2.1.7)
$$E(\underset{\sim}{y}_j) = \underset{\sim}{Z}_j \underset{\sim}{\gamma}$$

$$\mathrm{Var}(\underset{\sim}{y}_j) = \underset{\sim}{V}_j + \underset{\sim}{X}_j \underset{\sim}{U} \underset{\sim}{X}_j' .$$

This is just the model formulation for the LMI with a particular structure imposed on the individual variance matrices.

A simple example of the model given in (2.1.7) often used for longitudinal data is the variance components model given by

$$\underset{\sim}{V}_j = \sigma_e^2 \underset{\sim}{I}_{m_j}$$

and

$$\underset{\sim}{U} = \sigma_a^2 \cdot \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \vdots & & & \\ \rho & \cdot & \cdots & 1 \end{bmatrix} .$$

Here σ_e^2 is called the error variance component and σ_a^2 the individual variance component.

It should be noted that the GIM is a special case of the LMI in which $\underset{\sim}{y}_j$ consists of the rows of $\underset{\sim}{Y}_j$ and $\underset{\sim}{V}_j = \underset{\sim}{K}_j' \underset{\sim}{\Sigma} \underset{\sim}{K}_j$ for each subject in the j^{th} data group. Since multiple design matrices for each subject are not being considered here, the LMI and GIM afford most of the generality needed for this research.

2.2 Estimation for the LM

In this section the results on estimation for the models defined in 2.1 are reviewed for the case in which the covariance parameters are known. In addition, questions of estimability of the model parameters are discussed.

In the analysis of data which satisfy the GLMM assumptions, notation, and structure, it is usually of interest to estimate the parameter matrices $\underset{\sim}{B}$ and $\underset{\sim}{\Sigma}$ as well as functions of the form $\underset{\sim}{C}\ \underset{\sim}{B}\ \underset{\sim}{U}$ for constant matrices $\underset{\sim}{C}$ and $\underset{\sim}{U}$ of conforming dimensions. As is well-known, best linear unbiased estimation (BLUE) is possible in this setting (e.g., Morrison, 1976). The BLUE estimates for $\underset{\sim}{B}$ are the ordinary least squares estimates given by

$$\hat{\underset{\sim}{B}} = (\underset{\sim}{X}'\ \underset{\sim}{X})^{-1}\ \underset{\sim}{X}'\ \underset{\sim}{Y}\ .$$

For both the MGLM and the GIM, estimation of the parameter vector $\underset{\sim}{b}$ is of interest, as well as estimation of functions of the form $\underset{\sim}{H}\ \underset{\sim}{b}$ for known $\underset{\sim}{H}$. In MGLM notation these functions can be written as

$$\underset{\sim}{H}\ \underset{\sim}{b} = \sum_{s=1}^{p} \underset{\sim}{c}_s\ \underset{\sim}{b}_s$$

for constant vectors $\underset{\sim}{c}_s$. Kleinbaum (1970) illustrated that linear functions of the form $\underset{\sim}{C}\ \underset{\sim}{B}\ \underset{\sim}{U}$ are special cases of (2.2.1) where $\underset{\sim}{b}$ is the vector-version of $\underset{\sim}{B}$ (i.e., $\underset{\sim}{B}$ "rolled-out by columns"). In particular, when $\underset{\sim}{H} = \underset{\sim}{U} \times \underset{\sim}{C}'$ the two expressions are identical. Similarly, in the LM setting it is of primary concern to estimate $\underset{\sim}{\gamma}$ and linear functions of the form $\underset{\sim}{H}\ \underset{\sim}{\gamma}$.

Under the LM assumptions with normality, $\underset{\sim}{y} \sim N(\underset{\sim}{Z}\ \underset{\sim}{\gamma},\ \underset{\sim}{\Omega})$ where $\underset{\sim}{Z}$ is of full rank q. Therefore all linear functions $\underset{\sim}{\theta} = \underset{\sim}{H}\ \underset{\sim}{\gamma}$ are estimable in the sense that an unbiased linear estimate exists for $\underset{\sim}{\theta}$, viz., (2.2.2) $$\hat{\underset{\sim}{\theta}} = \underset{\sim}{H} (\underset{\sim}{Z}' \underset{\sim}{Z})^{-1} \underset{\sim}{Z}' \underset{\sim}{y}$$

and $E(\hat{\underset{\sim}{\theta}}) = \underset{\sim}{\theta}$. However these estimates may not be best in that they may not achieve minimum variance among all linear estimates of $\underset{\sim}{\theta}$. Of $\underset{\sim}{\Omega}$ is known, then Gauss-Markov theory can be applied to produce weighted least squares (WLS) estimates of $\underset{\sim}{\theta}$ which are BLUE (e.g., Rao, 1973)

(2.2.3) $$\hat{\underset{\sim}{\theta}} = \underset{\sim}{H} \hat{\underset{\sim}{\gamma}} = \underset{\sim}{H} (\underset{\sim}{Z}' \underset{\sim}{\Omega}^{-1} \underset{\sim}{Z})^{-1} \underset{\sim}{Z}' \underset{\sim}{\Omega}^{-1} \underset{\sim}{y}$$

with

(2.2.4) $$\text{Var}(\hat{\underset{\sim}{\theta}}) = \underset{\sim}{H} (\underset{\sim}{Z}' \underset{\sim}{\Omega}^{-1} \underset{\sim}{Z})^{-1} \underset{\sim}{H}' .$$

In practice it is usually the case that $\underset{\sim}{\Omega}$ is not known. While it is not possible to obtain BLUE estimates in this case, one can compute best asymptotically normal (BAN) estimates of $\underset{\sim}{H} \underset{\sim}{\gamma}$ under additional assumptions. Kleinbaum (1970) considered BAN estimation for the MGLM and GIM. The discussion that follows presents BAN estimation procedures for the LMI.

2.3 BAN Estimation for the LMI

In this section, the asymptotic properties of weighted least squares (WLS) estimates are presented for the model parameters, $\underset{\sim}{\gamma}$, in the LMI when the covariance matrix, $\underset{\sim}{\Omega}$, is unknown. Since it has been shown that the GIM is a special case of the LMI in which subjects can be grouped according to like missing data patterns, the results presented for the LMI parameters apply to the GIM parameters also.

For the LMI given in (2.1.5) with normality, $\underset{\sim}{y} \sim N(\underset{\sim}{Z} \underset{\sim}{\gamma}, \underset{\sim}{\Omega})$ with log likelihood function given by

(2.3.1) $$\log L_n = -\frac{N}{2} \log 2\Pi - \frac{1}{2} \log |\Omega| - \frac{1}{2} (\underset{\sim}{y} - \underset{\sim}{Z} \underset{\sim}{\gamma})' \underset{\sim}{\Omega}^{-1} (\underset{\sim}{y} - \underset{\sim}{Z} \underset{\sim}{\gamma}) .$$

The Fisher information matrix with respect to $\underset{\sim}{\gamma}$, $\underset{\sim}{F}_n(\underset{\sim}{\gamma})$, is given by

$$(2.3.2) \quad \underset{\sim}{F}_n(\gamma) = E_{\gamma}\left[-\frac{\partial^2 \log L_n}{\partial \gamma \, \partial \gamma'}\right]$$

$$= \underset{\sim}{Z}' \underset{\sim}{\Omega}^{-1} \underset{\sim}{Z}$$

after applying several well-known rules of matrix differentiation. The Cramer-Rao lower bound for the variance of an unbiased estimator of γ is given by

$$(2.3.3) \quad \underset{\sim}{F}_n^{-1} = (\underset{\sim}{Z}' \underset{\sim}{\Omega}^{-1} \underset{\sim}{Z})^{-1}$$

provided $\underset{\sim}{F}_n$ is positive definite. A best asymptotically normal (BAN) estimator for γ satisfies

$$(2.3.4) \quad \underset{\sim}{F}_n^{\frac{1}{2}} (\hat{\gamma}_n - \gamma) \xrightarrow{L} N(\underset{\sim}{0}, \underset{\sim}{I}).$$

Thus $\hat{\gamma}_n$ is "best" in the sense that the lower bound for the variance matrix is achieved asymptotically. Clearly, any estimator that is a BLUE is also a BAN estimator.

Before continuing, it is necessary to define a consistent matrix estimator. A matrix estimator is said to be consistent for a parameter matrix if each element is a consistent estimator of the corresponding element in the parameter matrix. For example, $\hat{\underset{\sim}{\Sigma}}$ is consistent for $\underset{\sim}{\Sigma}$ if we have

$$\hat{\sigma}_{ij} \xrightarrow{p} \sigma_{ij}$$

where $\underset{\sim}{\Sigma} = ((\sigma_{ij}))$.

A Lemma concerning continuous functions of consistent estimators is presented below. Although this result is well-known, a proof is provided for completeness.

Lemma 2.3.1. Given a random sample $y_1,\ldots,y_n$ with $y_i \sim G(\underset{\sim}{\theta})$, let $\underset{\sim}{T}_n(y_1,\ldots,y_n)$ be a consistent estimator of $\underset{\sim}{\theta}$, $\underset{\sim}{\theta} \in \Theta$, the parameter space.

Let $g(\underset{\sim}{\theta})$ be a continuous function defined on Θ. Then $g(\underset{\sim}{T}_n)$ is a consistent estimator of $g(\underset{\sim}{\theta})$.

Proof. Since $\underset{\sim}{T}_n$ is consistent for $\underset{\sim}{\theta}$ we have

$$\Pr\{|\underset{\sim}{T}_n - \underset{\sim}{\theta}| > \varepsilon\} \to 0 \text{ as } n \to \infty$$

or

$$\Pr\{|\underset{\sim}{T}_n - \underset{\sim}{\theta}| < \varepsilon\} \to 1 \text{ as } n \to \infty.$$

For any $\varepsilon > 0$, the continuity of g guarantees that there exists an $\eta > 0$ such that $\{|\underset{\sim}{T}_n - \underset{\sim}{\theta}| < \eta\}$ implies $\{|g(\underset{\sim}{T}_n) - g(\theta)| < \varepsilon\}$. Therefore

$$1 \geq \Pr\{\ |g(\underset{\sim}{T}_n) - g(\underset{\sim}{\theta})| < \varepsilon\ \} \geq \Pr\{\ |\underset{\sim}{T}_n - \underset{\sim}{\theta}| < \eta\} \to 1$$

as $n \to \infty$.

$$\therefore\ \Pr\{\ |g(\underset{\sim}{T}_n) - g(\underset{\sim}{\theta})| > \varepsilon\ \} \to 0 \text{ as } n \to \infty.$$

q.e.d.

For the model formulation given in (2.1.5) the following assumptions are made:

(2.3.5) (i) The vector $\underset{\sim}{y}$ consists of n independent random vectors, $\underset{\sim}{y}_i$, each of dimension (m_i x 1) with $m_i \leq p$ where p is a fixed positive constant, and $\underset{\sim}{y}_i \sim N(\underset{\sim}{Z}_i\underset{\sim}{\gamma}, \underset{\sim}{V}_i)$.

(ii) The elements of $\underset{\sim}{V}_i$ are continuous functions of a vector of parameters, $\underset{\sim}{\phi}$, that does not depend on i for i = 1,...,n; i.e. the parameters are common across the vectors of observations, $\underset{\sim}{y}_i$.

(iii) If N_k denotes the total number of vectors contributing to the estimation of γ_k (i.e., vectors with $z_{ik} \neq 0$) then as $n \to \infty$, $N_k \to \infty$ also for $k = 1,\ldots,q$.

Assumption (ii) above is made to ensure the estimability of $\underset{\sim}{\Omega}$ whenever $\underset{\sim}{\phi}$ is estimable. Assumption (iii) is made to ensure the estimability of $\underset{\sim}{\gamma}$. In a repeated measurements design, this assumption is equivalent to requiring an adequate number of measurements at each of the response times while allowing for missing data.

The following Theorem establishes a BAN estimation procedure for $\boldsymbol{\gamma}$ under the assumptions specified above. The sample of n random vectors is divided into two subsequences of equal length. Consistent estimators of $\boldsymbol{\phi}$ are produced using the first subsequence, and consistent estimators of $\boldsymbol{\gamma}$ are computed using the second subsequence. Without loss of generality it is assumed that n/2 is an integer. An additional assumption must first be made.

(2.3.6) (iv) There exists consistent estimates of $\boldsymbol{\phi}$, and hence of $\boldsymbol{\Omega}$, based on the subsequence of n/2 odd-numbered vectors, $\{\mathbf{y}_{2i-1}\}$, $i = 1,\ldots,n$.

Under assumptions (i) - (iv) above, the following result can be established.

Theorem 2.3.2. For the model given in (2.1.5) and assumptions (i) - (iv) in (2.3.5) and (2.3.6), let $\hat{\boldsymbol{\Omega}} = \text{Diag}\,(\hat{\mathbf{V}}_1,\ldots,\hat{\mathbf{V}}_n)$ denote the consistent estimate of $\boldsymbol{\Omega}$ based on the subsequence of odd-numbered random vectors, $\mathbf{y}_1, \mathbf{y}_3,\ldots,\mathbf{y}_{n-1}$. Let

$$\text{(2.3.7)} \qquad \hat{\boldsymbol{\gamma}}_n = \sum_{i=1}^{n/2} \mathbf{Z}'_{2i}\, \hat{\mathbf{V}}^{-1}_{2i}\, \mathbf{Z}_{2i} \;{}^{-1} \sum_{i=1}^{n/2} \mathbf{Z}'_{2i}\, \hat{\mathbf{V}}^{-1}_{2i}\, \mathbf{y}_{2i}$$

denote the WLS estimate of $\boldsymbol{\gamma}$ based on the independent subsequence of even-numbered vectors, $\mathbf{y}_2, \mathbf{y}_4,\ldots,\mathbf{y}_n$. If, as $n \to \infty$,

$$\text{(2.3.8)} \qquad (n/2)^{-1} \sum_{i=1}^{n/2} \mathbf{Z}'_{2i}\, \mathbf{V}^{-1}_{2i}\, \mathbf{Z}_{2i} \to \mathbf{C}$$

where $\mathbf{C}$ is a constant positive definite symmetric matrix, then $\hat{\boldsymbol{\gamma}}_n$ is a BAN estimator of $\boldsymbol{\gamma}$ based on a sample of n/2 random vectors.

Proof. For simplicity of notation, let all summations range over $i = 1,\ldots,n/2$. Let $\boldsymbol{\gamma}^*_n$ denote the BLUE of $\boldsymbol{\gamma}$ based on the even-numbered subsequence when $\boldsymbol{\Omega}$ is known; namely

$$\boldsymbol{\gamma}^*_n = [\,\Sigma\, \mathbf{Z}'_{2i}\, \mathbf{V}^{-1}_{2i}\, \mathbf{Z}_{2i}]^{-1}\; \Sigma\, \mathbf{Z}'_{2i}\, \mathbf{V}^{-1}_{2i}\, \mathbf{y}_{2i}\ .$$

Then

$$\sqrt{n/2}\quad (\underset{\sim}{\gamma}_n^* - \underset{\sim}{\gamma}) = \sqrt{n/2}\quad [\Sigma\, \underset{\sim}{Z}_{2i}'\, \underset{\sim}{V}_{2i}^{-1}\, \underset{\sim}{Z}_{2i}]^{-1}\quad \Sigma\, \underset{\sim}{Z}_{2i}'\, \underset{\sim}{V}_{2i}^{-1}\, \underset{\sim}{u}_{2i}$$

where $\underset{\sim}{y}_i = \underset{\sim}{Z}_i\, \underset{\sim}{\gamma} + \underset{\sim}{u}_i$. By assumption (i), $\underset{\sim}{u}_i \sim N(0, \underset{\sim}{V}_i)$ for $i = 1,\ldots,n$, therefore

$$\sqrt{n/2}\quad (\underset{\sim}{\gamma}_n^* - \underset{\sim}{\gamma}) \sim N(\underset{\sim}{0}, (n/2)\ [\Sigma\, \underset{\sim}{Z}_{2i}'\, \underset{\sim}{V}_{2i}^{-1}\, \underset{\sim}{Z}_{2i}]^{-1})$$

$$\overset{L}{\rightarrow}\ N(\underset{\sim}{0}, \underset{\sim}{C}^{-1})\ .$$

Consider

$$\sqrt{n/2}\ (\underset{\sim}{\gamma}_n^* - \hat{\underset{\sim}{\gamma}}_n) = \sqrt{n/2}\ \{[\Sigma\, \underset{\sim}{Z}_{2i}'\, \underset{\sim}{V}_{2i}^{-1}\, \underset{\sim}{Z}_{2i}]^{-1}\ \Sigma\, \underset{\sim}{Z}_{2i}'\, \underset{\sim}{V}_{2i}^{-1}\, \underset{\sim}{u}_{2i}$$

$$-\ [\Sigma\, \underset{\sim}{Z}_{2i}'\, \hat{\underset{\sim}{V}}_{2i}^{-1}\, \underset{\sim}{Z}_{2i}]^{-1}\ \Sigma\, \underset{\sim}{Z}_{2i}'\, \hat{\underset{\sim}{V}}_{2i}^{-1}\, \underset{\sim}{u}_{2i}\}$$

$$= \sqrt{n/2}\ [\Sigma\, \underset{\sim}{Z}_{2i}'\, \underset{\sim}{V}_{2i}^{-1}\, \underset{\sim}{Z}_{2i}]^{-1}\ \{\Sigma\, \underset{\sim}{Z}_{2i}'\underset{\sim}{V}_{2i}^{-1}\underset{\sim}{u}_{2i}$$

$$-\ [\Sigma\, \underset{\sim}{Z}_{2i}'\, \underset{\sim}{V}_{2i}^{-1}\, \underset{\sim}{Z}_{2i}]\ [\Sigma\, \underset{\sim}{Z}_{2i}'\, \hat{\underset{\sim}{V}}_{2i}^{-1}\, Z_{2i}]^{-1}\ \Sigma\, \underset{\sim}{Z}_{2i}'\, \hat{\underset{\sim}{V}}_{2i}^{-1}\, \underset{\sim}{u}_{2i}\}$$

By assumption, $\hat{\underset{\sim}{V}}_i \overset{p}{\rightarrow} \underset{\sim}{V}_i$; therefore $\hat{\underset{\sim}{V}}_i^{-1} \overset{p}{\rightarrow} \underset{\sim}{V}_i^{-1}$ and

$(\underset{\sim}{Z}_i'\ \hat{\underset{\sim}{V}}_i^{-1}\ \underset{\sim}{Z}_i)^{-1} \overset{p}{\rightarrow} (\underset{\sim}{Z}_i'\ \underset{\sim}{V}_i^{-1}\ \underset{\sim}{Z}_i)^{-1}$.

We have

$$\sqrt{n/2}\quad (\underset{\sim}{\gamma}_n^* - \hat{\underset{\sim}{\gamma}}_n) = \sqrt{n/2}\quad [\Sigma\, \underset{\sim}{Z}_{2i}'\, \underset{\sim}{V}_{2i}^{-1}\, \underset{\sim}{Z}_{2i}]^{-1}\ \Sigma\, \underset{\sim}{Z}_{2i}'\ (\underset{\sim}{V}_{2i}^{-1} - \hat{\underset{\sim}{V}}_{2i}^{-1})\ \underset{\sim}{u}_{2i}\ .$$

Let

$$\underset{\sim}{A}_n = (n/2)\ [\Sigma\, \underset{\sim}{Z}_{2i}'\, \underset{\sim}{V}_{2i}^{-1}\, \underset{\sim}{Z}_{2i}]^{-1}$$

and

$$\underset{\sim}{B}_n = (n/2)^{-\frac{1}{2}}\ \Sigma\, \underset{\sim}{Z}_{2i}'\ (\underset{\sim}{V}_{2i}^{-1} - \hat{\underset{\sim}{V}}_{2i}^{-1})\ \underset{\sim}{u}_{2i}\ .$$

Then $\underset{\sim}{A}_n \to \underset{\sim}{C}^{-1}$. Since $\hat{\phi}$ was computed from an independent subsequence, $\hat{\underset{\sim}{V}}_{2i}$ is independent of $\underset{\sim}{u}_{2i}$, $i = 1,\ldots,n/2$. Therefore

$$E[\underset{\sim}{B}_n] = (n/2)^{-\frac{1}{2}} \; \Sigma \; \underset{\sim}{Z}'_{2i} \; (\underset{\sim}{V}_{2i}^{-1} - E[\hat{\underset{\sim}{V}}_{2i}^{-1}]) \; E(\underset{\sim}{u}_{2i})$$

$$= \underset{\sim}{0}$$

and

$$\mathrm{Var}[\underset{\sim}{B}_n] = E_{\hat{\underset{\sim}{\phi}}} \; [\mathrm{Var}(\underset{\sim}{B}_n \mid \hat{\underset{\sim}{\phi}})] + \mathrm{Var}_{\hat{\underset{\sim}{\phi}}} \; [E \; (\underset{\sim}{B}_n \mid \hat{\underset{\sim}{\phi}})]$$

$$= E_{\hat{\underset{\sim}{\phi}}} \; [(n/2)^{-1} \; \Sigma \; \mathrm{Var}[\underset{\sim}{Z}'_{2i} \; (\underset{\sim}{V}_{2i}^{-1} - \hat{\underset{\sim}{V}}_{2i}^{-1}) \; \underset{\sim}{u}_{2i} \mid \hat{\underset{\sim}{\phi}}]$$

$$= E_{\hat{\phi}} \; [(n/2)^{-1} \; \Sigma \; \underset{\sim}{Z}'_{2i} \; (\underset{\sim}{V}_{2i}^{-1} - \hat{\underset{\sim}{V}}_{2i}^{-1}) \; \underset{\sim}{V}_{2i} \; (\underset{\sim}{V}_{2i}^{-1} - \hat{\underset{\sim}{V}}_{2i}^{-1}) \; \underset{\sim}{Z}_{2i} \; .$$

Since $\hat{\underset{\sim}{V}}_i^{-1} \overset{p}{\to} V_{\sim i}^{-1}$ $\mathrm{Var}(\underset{\sim}{B}_n) \to \underset{\sim}{0}$ as $n \to \infty$. This implies that $\underset{\sim}{B}_n \overset{p}{\to} E(\underset{\sim}{B}_n) = \underset{\sim}{0}$ and since $\underset{\sim}{A}_n \to \underset{\sim}{C}^{-1}$ and $\underset{\sim}{B}_n \overset{p}{\to} \underset{\sim}{0}$, then $\underset{\sim}{A}_n \; \underset{\sim}{B}_n \overset{p}{\to} \underset{\sim}{0}$.

Thus $\sqrt{n/2} \; (\underset{\sim}{\gamma}_n^* - \hat{\underset{\sim}{\gamma}}) \overset{p}{\to} \underset{\sim}{0}$.

In order to establish $\hat{\underset{\sim}{\gamma}}_n$ as a BAN estimator of $\underset{\sim}{\gamma}$ it remains to be shown that

$$[\; \Sigma \; \underset{\sim}{Z}'_{2i} \; \underset{\sim}{V}_{2i}^{-1} \; \underset{\sim}{Z}_{2i}]^{\frac{1}{2}} \quad (\hat{\underset{\sim}{\gamma}}_n - \underset{\sim}{\gamma}) \overset{L}{\to} N(\underset{\sim}{0}, \underset{\sim}{I}).$$

Let

$$\underset{\sim}{A}_n^{-\frac{1}{2}} = (n/2)^{-\frac{1}{2}} \; [\; \Sigma \; \underset{\sim}{Z}'_{2i} \; \underset{\sim}{V}_{2i}^{-1} \; \underset{\sim}{Z}_{2i}]^{\frac{1}{2}}.$$

Then $\underset{\sim}{A}_n^{-\frac{1}{2}} \to \underset{\sim}{C}^{\frac{1}{2}}$. We have seen that

$$\sqrt{n/2} \; (\underset{\sim}{\gamma}_n^* - \underset{\sim}{\gamma}) \overset{L}{\to} N(\underset{\sim}{0}, \; \underset{\sim}{C}^{-1}) \; .$$

By a Theorem due to Sverdrup (Puri and Sen, 1971), if $\sqrt{n/2} \; (\underset{\sim}{\gamma}_n^* - \hat{\underset{\sim}{\gamma}}_n) \to \underset{\sim}{0}$, then $\sqrt{n/2} \; (\hat{\underset{\sim}{\gamma}}_n - \underset{\sim}{\gamma})$ has the same asymptotic distribution as $\sqrt{n/2} \; (\underset{\sim}{\gamma}_n^* - \underset{\sim}{\gamma})$;

namely $\sqrt{n/2}\,(\hat{\gamma}_n - \gamma) \xrightarrow{L} N(\mathbf{0}, \mathbf{C}^{-1})$.

Therefore $\hat{\gamma}_n$ is a BAN estimator of γ based on a sample of n/2 random vectors as $n \to \infty$. q.e.d.

It should be noted that although the procedure just described relies on independent estimation of the variance-covariance parameters, ϕ, and the mean parameters, γ, it is arguable that the use of the entire sample for the estimation of both sets of parameters is also a good procedure. However, the asymptotic properties of the resulting estimators have not been readily attainable.

One requirement for BAN estimation of model parameters in the LMI with normality is a consistent estimate of ϕ. For the GIM, this requirement is equivalent to a consistent estimate of Σ. Kleinbaum (1970) suggested using an estimate of Σ computed via pairwise deletion. In this procedure, each covariance parameter, σ_{rs}, is estimated using only those subjects with measurements for both the r^{th} and s^{th} response variates. This estimate, $\hat{\sigma}_{rs}$, is equivalent to the usual covariance estimate for the GLMM with two dependent variables. Similarly, the variance estimates, $\hat{\sigma}_{rr}$, are computed as the usual estimate of variance in a univariate model by using all measurements on the r^{th} response variate. Each parameter estimate must be divided by the appropriate degrees of freedom. Sufficient conditions for the existence of this pairwise deletion estimator, $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$, are given by

(2.3.9) $$N_{rs} \geq 2, \qquad r,s = 1,\ldots,p$$

where N_{rs} is the number of subjects with measurements on the r^{th} and s^{th} response variates, and Σ is of dimension (p x p). Furthermore, $\hat{\Sigma}$ is unbiased and therefore consistent for Σ in the GIM provided

(2.3.10) $\lim_{n\to\infty} N_{rs}/n$ exists and $\neq 0$, $r,s = 1,\ldots,p$

where n is the total number of subjects (ref. Kleinbaum, 1970, Theorem 2.4.1 and Corollary 2.4.3).

In analyzing data that satisfy the GLMM assumptions, it may not be possible to compute covariance estimates with pairwise deletion techniques due to missing values (i.e., conditions (2.3.9) and (2.3.10) may not be satisfied). One approach to computing reasonable covariance estimates is to impose constraints on the $p(p + 1)/2$ unknown parameters of $\underset{\sim}{\Sigma}$. For example, assume $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}(\underset{\sim}{\phi})$ where $\underset{\sim}{\phi}$ is a $(t \times 1)$ vector of unknown parameters with $t < p(p + 1)/2$. Then it may be possible to consistently estimate $\underset{\sim}{\Sigma}$ based on a suitable estimate of $\underset{\sim}{\phi}$.

This method can be illustrated with a simple example. Let

$$\underset{\sim}{Y} = \begin{bmatrix} y_{11} & \cdot & y_{13} & \cdot \\ y_{21} & \cdot & y_{23} & \cdot \\ \cdot & y_{32} & \cdot & y_{34} \\ \cdot & y_{42} & \cdot & y_{44} \end{bmatrix}$$

where "$\cdot$" denotes a missing value, and assume $\underset{\sim}{Y} \sim \text{GLMM}(\underset{\sim}{X}\,\underset{\sim}{B}, \underset{\sim}{\Sigma})$. Clearly if $\underset{\sim}{\Sigma} = ((\sigma_{ij}))$, σ_{12} and σ_{34} are not reasonably estimable from this data. However, suppose the following model is assumed for $\underset{\sim}{\Sigma}$:

$$\underset{\sim}{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \rho & \cdot & \cdot & \cdot \end{bmatrix} .$$

Then $\underset{\sim}{\phi} = (\sigma^2, \rho)$ can be estimated by considering the four observations as independent bivariate normal random vectors and computing the sample variance and correlation. Assuming a model for the covariance structure enables an estimate of $\underset{\sim}{\Sigma}$ to be computed.

Clearly the result of Theorem 2.3.2 applies when a model is assumed for $\underset{\sim}{\Sigma}$ in the GIM provided that the elements of $\underset{\sim}{\Sigma}$ are continuous functions of $\underset{\sim}{\phi}$, and that consistent estimates of $\underset{\sim}{\phi}$ are available from the appropriate subsequence. The topic will be considered in detail in Chapters 3 through 6.

2.4 Testing Hypotheses in the LMI

In addition to deriving WLS estimators for the MGLM parameters, Kleinbaum also suggested creating Wald statistics based on these estimators for use in testing hypotheses about the parameters. A brief review of the theory of Wald statistics is given in this section (ref. Wald, 1943).

Given a random sample of size n, $\underset{\sim}{y}_1,\ldots,\underset{\sim}{y}_n$, whose distribution depends on a parameter vector $\underset{\sim}{\xi}$, suppose $\underset{\sim}{g}(\underset{\sim}{\xi})$ is a (w x 1) continuous vector function of $\underset{\sim}{\xi}$ with continuous first partial derivatives. Let $\underset{\sim}{F}_n(\underset{\sim}{\xi})$ denote the Fisher Information matrix for $\underset{\sim}{\xi}$ and define

$$(2.4.1) \qquad C_n = \left[\frac{\partial\, \underset{\sim}{g}(\underset{\sim}{\xi})}{\partial\, \underset{\sim}{\xi}}\right]' \underset{\sim}{F}_n^{-1}(\underset{\sim}{\xi}) \left[\frac{\partial\, \underset{\sim}{g}(\underset{\sim}{\xi})}{\partial\, \underset{\sim}{\xi}}\right].$$

If $\hat{\underset{\sim}{\xi}}$ is a consistent estimate of $\underset{\sim}{\xi}$ and $\underset{\sim}{g}(\hat{\underset{\sim}{\xi}})$ is BAN for $\underset{\sim}{g}(\underset{\sim}{\xi})$ then

$$(2.4.2) \qquad [\underset{\sim}{g}(\hat{\underset{\sim}{\xi}}) - \underset{\sim}{g}(\underset{\sim}{\xi})]' \; \hat{\underset{\sim}{C}}_n^{-1} \; [\underset{\sim}{g}(\hat{\underset{\sim}{\xi}}) - \underset{\sim}{g}(\underset{\sim}{\xi})]$$

has asymptotically a central chi-square distribution with w degrees of freedom, X_w^2, where $\hat{\underset{\sim}{C}}_n = \underset{\sim}{C}_n(\hat{\underset{\sim}{\xi}})$. Based on this result, the statistic suggested by Wald (1943) to test the hypothesis

$$(2.4.3) \qquad H_0: \; \underset{\sim}{g}(\underset{\sim}{\xi}) = \underset{\sim}{0}$$

is given by

(2.4.4) $$W_n = [g(\hat{\xi})]' \hat{C}_n^{-1} [g(\hat{\xi})].$$

When H_0 is true,

(2.4.5) $$W_n \xrightarrow{L} \chi_w^2 \quad \text{as} \quad n \to \infty.$$

In the LMI it is of interest to test hypotheses of the form

(2.4.6) $$H_0: \ H \gamma = 0$$

where H is a (w x q) constant matrix. The following theorem provides a test for H_0.

Theorem 2.4.1. For the LMI with normality, assume that the conditions of Theorem 2.3.2 are met. Let $\hat{\Omega} = \Omega(\hat{\phi})$ denote the consistent estimator of Ω based on the odd-numbered subsequence, and let $\hat{\gamma}_n$ be as defined in (2.3.7). Then if H in (2.4.6) is of full column rank, w, and if (2.4.6) is true,

(2.4.7) $$W_n = (H \hat{\gamma}_n)' [H (Z' \hat{\Omega}^{-1} Z)^{-1} H']^{-1} (H \hat{\gamma}_n)$$

has a limiting central chi-square distribution, as $n \to \infty$, with w degrees of freedom.

Proof: From Theorem 2.3.2, $H \hat{\gamma}$ is a BAN estimator of $H \gamma$. The Fisher information matrix for γ is

$$F_n(\gamma) = Z' \Omega^{-1} Z ,$$

and $$\frac{\partial H\gamma}{\partial \gamma} = H .$$

Thus W_n is a Wald statistic with the desired limiting distribution.

q.e.d.

Here again, the technique of estimating ϕ and γ from independent subsequences has been employed to guarantee that W_n follows an asymptotic chi-square distribution. Although Wald test statistics can be computed using the entire sample for estimation of both ϕ and γ, the test statistics may not follow an asymptotic chi-square distribution.

CHAPTER 3

THE GIM WITH COVARIANCE CONSTRAINTS

3.1 Introduction

In this chapter the analysis of incomplete longitudinal data with covaraince constraints is considered for cases in which the dataset satisfies the GIM assumptions. A numerical example is given to illustrate the methods of analysis.

The data are assumed to arise from an experiment in which p responses were measured on n subjects but some of the responses may be missing. Let $\{v_1,\ldots,v_p\}$ denote the p response variates. The i^{th} subject's response vector consists of $m_i \leq p$ measurements where $\{v_1,\ldots v_{m_i}\} \varepsilon \{v_1,\ldots,v_p\}$. It was shown in section 2.1 that Kleinbaum's GIM is a suitable model for this data configuration. Recall that with this model the subjects are grouped according to similar patterns of observed variates. Let j=1,...,u index these data groups. Then, repeating the notation of section 2.1 for convenience, we have the following model for the j^{th} group:

(3.1.1) $$E(\underset{\sim}{Y}_j) = \underset{\sim}{D}_j \underset{\sim}{B} \underset{\sim}{K}_j,$$

$$\text{Var}(i^{th} \text{ row of } \underset{\sim}{Y}_j) = \underset{\sim}{K}_j' \underset{\sim}{\Sigma} \underset{\sim}{K}_j \qquad ,i = 1,\ldots,n_j,$$

$$\text{Cov}(i^{th} \text{ row of } \underset{\sim}{Y}_j,\ \ell^{th} \text{ row of } \underset{\sim}{Y}_j) = 0, \qquad i \neq \ell,$$

where $\underset{\sim}{Y}_j$, $\underset{\sim}{D}_j$ $\underset{\sim}{B}$, $\underset{\sim}{K}_j$, and $\underset{\sim}{\Sigma}$ are as defined in (2.1.3). It was shown that this model is a special case of the LM with the vector of model parameters, $\underset{\sim}{\gamma}$, equal to the vector-version of $\underset{\sim}{B}$ ("rolled out by columns"). Kleinbaum's estimator of $\underset{\sim}{\gamma}$, and hence of $\underset{\sim}{B}$, can be computed using any consistent estimator of $\underset{\sim}{\Sigma}$. This estimator is expressed in terms of the u groups' parameters as follows:

$$(3.1.2) \qquad \hat{\underset{\sim}{\gamma}} = (\underset{\sim}{Z}' \hat{\underset{\sim}{\Omega}}^{-1} \underset{\sim}{Z})^{-1} \underset{\sim}{Z}' \hat{\underset{\sim}{\Omega}}^{-1} \underset{\sim}{y}$$

$$= \left[\sum_{j=1}^{u} (K_j \otimes D_j') (\underset{\sim}{K}_j' \hat{\underset{\sim}{\Sigma}} \underset{\sim}{K}_j \otimes \underset{\sim}{I}_{n_j})^{-1} (\underset{\sim}{K}_j' \otimes \underset{\sim}{D}_j) \right]^{-1} \cdot$$

$$\sum_{j=1}^{u} (K_j \otimes D_j') (\underset{\sim}{K}_j' \hat{\underset{\sim}{\Sigma}} \underset{\sim}{K}_j \otimes \underset{\sim}{I}_{n_j})^{-1} \underset{\sim}{y}_j$$

$$= \left[\sum_{j=1}^{u} K_j (\underset{\sim}{K}_j' \hat{\underset{\sim}{\Sigma}} \underset{\sim}{K}_j)^{-1} \underset{\sim}{K}_j' \otimes \underset{\sim}{D}_j' \underset{\sim}{D}_j \right]^{-1} \sum_{j=1}^{u} K_j (\underset{\sim}{K}_j' \underset{\sim}{\Sigma} \underset{\sim}{K}_j)^{-1} \otimes \underset{\sim}{D}_j' \underset{\sim}{y}_j$$

where $\underset{\sim}{y}_j$ is the vector-version of $\underset{\sim}{Y}_j$.

From a computational point of view, this expression is more desirable than the expression for the general LM estimator. For large datasets, since $\underset{\sim}{\Omega}$ is an (np x np) matrix, its inversion could require large amounts of computer time and storage. If (3.1.2) is used to compute estimates, the data can be processed one group at a time and the summation terms accumulated over groups. The dimension of $\underset{\sim}{K}_j' \underset{\sim}{\Sigma} \underset{\sim}{K}_j$ is $(m_j \times m_j)$ which is never larger than (p x p).

Expression (3.1.2) defines a BAN estimtor provided that $\hat{\underset{\sim}{\Sigma}}$ is consistent for $\underset{\sim}{\Sigma}$ and the conditions of Theorem 2.3.1 are satisfied. Kleinbaum introduced a pairwise deletion estimator for the MGLM that is unbiased and consistent for $\boldsymbol{\Sigma}$. The form of this estimator simplifies somewhat for the GIM. Let

$$(3.1.3) \qquad \mathbf{Q}_j = \mathbf{Y}_j' \, [\mathbf{I}_{n_j} - \mathbf{D}_j(\mathbf{D}_j' \, \mathbf{D}_j)^{-1} \, \mathbf{D}_j'] \, \mathbf{Y}_j$$

and

$$(3.1.4) \qquad \mathbf{S} = \sum_{j=1}^{u} \mathbf{K}_j \, \mathbf{Q}_j \, \mathbf{K}_j' \, .$$

Then the pairwise deletion estimator, $\hat{\boldsymbol{\Sigma}}_{PD} = ((\hat{\sigma}_{k\ell}))$ is given by

$$(3.1.5) \qquad \hat{\sigma}_{k\ell} = s_{k\ell}/(N_{k\ell}-q) \quad , \qquad k,\ell=1,\ldots,p,$$

where

$\mathbf{S} = ((s_{k\ell}))$,

$N_{k\ell}$ = the number of subjects on which both the k^{th} and ℓ^{th} response variates are observed,

q = rank $(\mathbf{D}_j' \, \mathbf{D}_j)$ for any $j=1,\ldots,u$.

This estimator is consistent for $\boldsymbol{\Sigma}$ in the GIM provided $N_{k\ell} \to \infty$ as $n \to \infty$ for $k,\ell=1,\ldots,p$ (Kleinbaum, 1970).

Kleinbaum also suggested an iterative refinement of $\hat{\boldsymbol{\gamma}}$. Note that the pairwise deletion estimator $\hat{\boldsymbol{\Sigma}}_{PD}$ is computed using ordinary least squares (OLS) residuals. Given an estimate of $\boldsymbol{\gamma}$, residuals could be computed and a new estimate of $\boldsymbol{\Sigma}$ produced by a method similar to that described above. Briefly, let $\hat{\boldsymbol{\gamma}}^{(1)}$ denote the estimate computed by substituting $\hat{\boldsymbol{\Sigma}}_{PD}$ in (3.1.2). Let

$$(3.1.6) \qquad \mathbf{R}_j = \mathbf{Y}_j - \mathbf{D}_j \, \hat{\mathbf{B}}^{(1)} \, \mathbf{K}_j$$

where $\hat{\boldsymbol{\gamma}}^{(1)}$ has been reshaped into the (q x p) matrix $\hat{\mathbf{B}}^{(1)}$, and let

$$(3.1.7) \qquad \mathbf{S} = \sum_{j=1}^{u} \mathbf{K}_j \, \mathbf{R}_j' \, \mathbf{R}_j \, \mathbf{K}_j' \quad .$$

Then a second estimate of $\underset{\sim}{\Sigma}$ can be computed as in (3.1.5) and, in turn, used to produce a refined estimate, $\hat{\underset{\sim}{\gamma}}^{(2)}$. The steps defined by (3.1.6), (3.1.7), (3.1.5), and (3.1.2) can be repeated until the estimates converge. Note that at each step, (3.1.6), (3.1.7), and (3.1.5) define another pairwise deletion estimation of $\underset{\sim}{\Sigma}$.

As will be illustrated in the next section, it may not always be possible to estimate $\underset{\sim}{\Sigma}$ with pairwise deletion techniques due to the sparseness of the data. In such cases, a constrained formulation of $\underset{\sim}{\Sigma}$ will be required. If this formulation defines $\underset{\sim}{\Sigma}$ as a function of parameters which can be consistently estimated from least squares residuals, then iterative estimation of the parameters as described here will still be possible.

Kleinbaum indicated satisfactory convergence properties for the iterative procedure defined by (3.1.2)-(3.1.7) for several example datasets. Provided that the formulation of $\underset{\sim}{\Sigma}$ as a function of $\underset{\sim}{\phi}$ is a valid assumption and that consistent estimation of $\underset{\sim}{\phi}$ is possible from the $\underset{\sim}{R}_j$, similar convergence properties would be expected. Hosking (1980) suggested using the maximum relative change in the elements of $\hat{\underset{\sim}{B}}$ and $\hat{\underset{\sim}{\Sigma}}$ as the convergence criterion. In the constrained case it is only necessary to examine the changes in $\hat{\underset{\sim}{B}}$ and $\hat{\underset{\sim}{\phi}}$.

3.2 An Example

In this section the analysis of a dataset satisfying the GIM assumptions is illustrated. It will be shown that due to sparseness of the data, Kleinbaum's proposed estimation procedure based on the pairwise deletion estimator was not possible. Several alternative models for $\underset{\sim}{\Sigma}$ were considered. Thus this example serves as additional motivation for pursuing estimation with a constrained formulation of $\underset{\sim}{\Sigma}$.

The covariance models used in the analysis are based on the assumption that $\underset{\sim}{\Sigma}$ is a function of two parameters, a common variance for each response and a common covariance between adjacent (in time) response variates. Let $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}(\underset{\sim}{\phi})$ and $\underset{\sim}{\phi} = (\sigma^2, \rho)$ where the elements of $\underset{\sim}{\Sigma}$ are continuous functions of the elements of $\underset{\sim}{\phi}$. By Lemma 2.3.2, if $\hat{\sigma}^2$ and $\hat{\rho}$ are consistent estimators of σ^2 and ρ respectively, then $\underset{\sim}{\Sigma}(\hat{\sigma}^2, \hat{\rho})$ is consistent for $\underset{\sim}{\Sigma}\ (\sigma^2, \rho)$.

One model of this form often used in analyzing repeated measurement data is the equi-correlation model:

(3.2.1) $$\underset{\sim}{\Sigma}_{EQ} = \sigma^2 \cdot \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \rho & \cdot & \cdots & 1 \end{bmatrix}.$$

Rao and Rao (1966) and Woolson, et. al. (1978) assumed this model in analyzing linked cross-sectional data. Consistent estimates of σ^2 and $\sigma^2\rho$ can be defined in terms of the elements of $\hat{\underset{\sim}{\Sigma}}_{PD}$. Let

(3.2.2) $$\hat{\sigma}^2 = \sum_{k=1}^{p} s_{kk} \Big/ \left[\sum_{k=1}^{p} N_{kk} - q\right]$$

and

$$\widehat{\sigma^2\rho} = \sum_{\substack{k=1 \\ k\neq \ell}}^{p} \sum_{\ell=1}^{p} s_{k\ell} \Big/ \left[\sum_{\substack{k=1 \\ k\neq \ell}}^{p} \sum_{\ell=1}^{p} N_{k\ell} - q\right]$$

where

$s_{k\ell}$ is defined in (3.1.4). The consistency of $\hat{\sigma}^2$ and $\widehat{\sigma^2\rho}$ is obvious since the average of consistent estimates is also consistent.

Another covariance pattern commonly associated with measurements collected over time is the tri-diagonal matrix

$$(3.2.4)\qquad \underset{\sim}{\Sigma}_{TR} = \sigma^2 \cdot \begin{bmatrix} 1 & \rho & 0 & \cdots & 0 \\ \rho & 1 & \rho & 0\cdots & 0 \\ 0 & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & \cdot & \cdot & \cdot & 1 \end{bmatrix} .$$

Under this model the assumption is made that correlation exists only between responses that are adjacent in time. Computing $\hat{\sigma}^2$ as in (3.2.2) and

$$(3.2.5)\qquad \widehat{\sigma^2\rho} = \sum_{k=1}^{p-1} S_{k,k+1} \Big/ \left[\sum_{k=1}^{p-1} N_{k,k+1} - q\right]$$

produces a consistent estimate for $\hat{\underset{\sim}{\Sigma}}_{TR}$. Note however that this is not a particularly good estimate of $\underset{\sim}{\Sigma}_{TR}$ since it makes no use of the fact that the remaining off-diagonal elements are equal to zero. Discussion of estimation under this model will be considered in detail in Chapter 5. It should be noted that iterative refinement of $\underset{\sim}{\Sigma}$ is possible under both of these models simply by computing $\underset{\sim}{S}$ as in (3.1.7) at each step.

A covariance model that appears frequently in growth curve analyses is the serial correlation model:

$$(3.2.6)\qquad \underset{\sim}{\Sigma}_{SE} = \sigma^2 \cdot \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-1} \\ \rho & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \rho^{p-1} & \cdot & \cdot & \cdot & 1 \end{bmatrix} .$$

Under this model, the correlation between measurements is assumed to decrease with time. Consistent estimation assuming this model is discussed in detail in Chapter 4. It will be assumed for the subsequent example analysis that consistent estimates are available for σ^2 and ρ.

To illustrate the use of the GIM with covariance constraints for analyzing incomplete data, 20% of the measurements from a longitudinal experiment were deleted at random. The original dataset, taken from

Danford, et. al. (1960) contained responses from 45 subjects suffering from cancerous lesions. The subjects were divided into a control group and three treatment groups. Varying amounts of whole-body x-radiation were administered to the treatment groups. The subjects were trained on a psychomotor device and a baseline score was recorded. Average daily scores were then taken on 10 consecutive days. For ease of matrix manipulaton only five post-treatment scores were used in this example. A listing of the dataset with observations deleted at random is given in Table 3.2.1. The observations have been rearranged into groups with the same missing data patterns.

The purpose of the analysis was to determine if there were any significant differences in test scores among the four treatment groups after adjusting for the pre-treatment measurement. In order to use the GIM for the analysis, subjects were grouped according to missing data patterns. There are 16 such data groups, with complete observation vectors for 16 subjects and from one to four subjects in the remaining groups.

Since measurements were deleted at random, it is possible to compare the results from the proposed incomplete data analysis with those from a standard linear model analysis of the complete data. Separate intercepts and a common slope were fit for the four treatment groups using the complete data. If y_{ijk} denotes the j^{th} response for the k^{th} subject in the i^{th} treatment group and x_{ijk} denotes the corresponding pre-treatment score, then the model parameters are:

(3.2.7) $$E[y_{ijk}] = b_{ij} + b_{5j}\, x_{ijk}, \qquad i=1,\ldots,4,\; j=1,\ldots,5,\; k=1,\ldots n_i$$

and

Table 3.2.1. Payne Data with 20% of Observations Randomly Missing

GROUP	PRE	Y2	Y4	Y6	Y8	Y10
1	155	191	219	237	252	245
1	85	201	224	246	255	281
1	15	24	38	46	62	74
2	33	50	44	45	50	52
2	179	206	221	224	246	229
2	30	37	48	64	83	90
2	51	131	181	195	158	215
2	137	172	168	190	211	221
3	94	169	182	188	181	152
3	69	67	43	55	73	76
3	69	137	95	129	133	91
3	51	76	72	107	128	133
4	86	75	71	157	173	156
4	131	183	206	197	226	240
4	172	263	276	267	283	298
4	224	248	257	260	299	300
3	190	224	249	293	295	•
3	188	235	265	263	285	•
4	183	217	241	229	233	•
1	64	81	92	126	•	140
2	16	45	37	51	•	45
3	148	202	184	207	•	163
4	71	107	101	78	•	71
1	106	214	265	282	•	•
2	121	188	224	230	•	•
2	84	97	47	•	110	187
1	191	242	266	•	286	•
3	109	102	135	•	•	171
3	127	149	•	207	220	219
4	201	229	•	217	244	246
4	113	159	•	162	167	195
4	246	269	•	306	295	311
3	156	198	•	217	219	•
3	181	199	•	232	•	250
4	115	168	•	•	194	212
2	53	•	105	122	93	132
2	92	•	148	146	148	169
3	178	•	255	251	254	275
2	108	•	29	57	47	•
3	99	•	130	153	144	•
2	114	•	155	•	208	173
2	188	•	•	260	286	296
3	140	•	•	228	245	262
2	205	•	•	282	298	•
3	207	•	•	279	307	•

$$\mathrm{Cov}(y_{ijk}, y_{i'j'k'}) = \begin{cases} \sigma_{ij'} & \text{if } i=i' \text{ and } k=k' \\ 0 & \text{otherwise.} \end{cases}$$

Estimates of $\underset{\sim}{B} = ((b_{ij}))$ and $\underset{\sim}{\Sigma} = ((\sigma_{ij}))$ computed from the complete data are given in Table 3.2.2. The $p(p + 1)/2 = 15$ covariance parameter estimates are given above the diagonal and the estimated correlations are given below the diagonal. The general linear hypothesis of no group differences on any of the five days was tested as follows:

$$H_1: \; \underset{\sim}{C}_1 \underset{\sim}{B} \underset{\sim}{U} = 0$$

where $\underset{\sim}{C}_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \end{bmatrix}$

and $\underset{\sim}{U}$ is the identity matrix. The multivariate test statistic is given in Table 3.2.2, along with univariate test statistics for no group differences on each of the five days separately (H_2-H_6). The hypothesis of no average differences between the groups was tested with $\underset{\sim}{C}_1$ defined above and

$$\underset{\sim}{U} = 1/5 \; [1 \; 1 \; 1 \; 1 \; 1]' \; .$$

The results for this test are also given in Table 3.2.2 (H_7).

Four separate analyses were run on the incomplete dataset corresponding to different assumptions about $\underset{\sim}{\Sigma}$. The first analysis was carried out via a linear models software package (LINMOD) that uses listwise deletion for incomplete datasets. Therefore only 16 subjects were included in the analysis. The parameter estimates and standard multivariate test statistics are given in Table 3.2.3.

The GIM with an unconstrained covariance structure was assumed for the second incomplete data analysis. For the j^{th} data group, j=1,...,16, we have the model:

Table 3.2.2 Analysis of Complete Dataset (n = 45)

Estimates

$\hat{\underset{\sim}{B}}$:

	Y_2	Y_4	Y_6	Y_8	Y_{10}
X_1	36.09	42.35	65.17	65.16	78.11
X_2	25.90	18.72	42.32	39.86	55.15
X_3	32.04	23.99	45.62	43.91	47.78
X_4	33.19	15.96	33.54	32.74	47.12
PRE	1.029	1.187	1.115	1.196	1.182

$\hat{\underset{\sim}{\Sigma}}/\hat{\underset{\sim}{R}}$:

	Y_2	Y_4	Y_6	Y_8	Y_{10}
Y_2	1069	1128	1003	964	953
Y_4	.864	1596	1286	1230	1331
Y_6	.821	.862	1394	1332	1337
Y_8	.756	.789	.915	1521	1435
Y_{10}	.677	.774	.831	.855	1855

Tests of Hypotheses:

	H_1^*	H_2	H_3	H_4	H_5	H_6	H_7
F Value	.672	.171	.627	.907	.901	.801	.670
Degrees of Freedom (NUM/DEN)	15/125	3/40	3/40	3/40	3/40	3/40	3/40
p-value	.807	.915	.601	.446	.449	.501	.576

*Approximate F for Wilk's Lambda.

Table 3.2.3 Analysis of Incomplere Data with Listwise Deletions (n = 16)

Estimates

$\hat{\underset{\sim}{B}}$:

	Y_2	Y_4	Y_6	Y_8	Y_{10}
X_1	40.30	59.27	87.11	92.42	112.24
X_2	19.68	30.14	53.33	51.21	72.61
X_3	30.38	13.88	45.49	47.80	39.96
X_4	14.91	20.28	59.39	69.91	90.28
PRE	1.157	1.189	1.050	1.144	1.032

$\hat{\underset{\sim}{\Sigma}}/\hat{\underset{\sim}{R}}$:

	Y_2	Y_4	Y_6	Y_8	Y_{10}
Y_2	1635	1966	1681	1241	1602
Y_4	.940	2676	2250	1620	2230
Y_6	.844	.883	2426	1800	2217
Y_8	.806	.822	.959	1451	1749
Y_{10}	.787	.856	.894	.912	2536

Tests of Hypotheses:

	H_1^*	H_2	H_3	H_4	H_5	H_6	H_7
F Value	.781	.244	.490	.447	.981	1.234	.585
Degrees of Freedom (NUM/DEN)	15/26	3/11	3/11	3/11	3/11	3/11	3/11
p-value	.686	.864	.696	.725	.437	.344	.637

*Approximate F for Wilk's Lambda.

$$E[\underset{\sim}{Y}_j] = \underset{\sim}{D}_j \underset{\sim}{B} \underset{\sim}{K}_j$$

$$\mathrm{Var}[\underset{\sim}{Y}_j] = I_{n_j} \otimes \underset{\sim}{K}_j' \underset{\sim}{\Sigma} \underset{\sim}{K}_j$$

where $\underset{\sim}{D}_j$ is the (n_j x 5) design matrix,

$\underset{\sim}{B}$ is defined above,

$\underset{\sim}{K}_j$ is the incidence matrix (5 x q_j).

For example, the last data group contains scores from days six and eight on two subjects. For this pattern we have

$$\underset{\sim}{K}_j = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

$$\underset{\sim}{D}_j = \begin{bmatrix} 0 & 1 & 0 & 0 & 285 \\ 0 & 0 & 1 & 0 & 287 \end{bmatrix}.$$

The 15 unknown elements of $\underset{\sim}{\Sigma}$ were estimated using pairwise deletion techniques. The resulting estimate, $\hat{\Sigma}_{PD}$ was non-positive definite, thus no tests of hypotheses were performed. The parameter estimates are given in Table 3.2.4. Note the estimated correlation coefficient equal to 1.039. The eigenvalues of $\hat{\Sigma}_{PD}$ are also presented in this table. Since this analysis proved unsatisfactory, the next two analyses incorporated constraints on $\underset{\sim}{\Sigma}$.

It is reasonable to assume for this experiment that the correlations among the dependent variables decrease with the time separation. For example, we would expect scores for days two and four to be more highly correlated than scores taken on days two and 10. The estimates of $\underset{\sim}{\Sigma}$ based on the complete data tend to support this assumption (Table 3.2.2). Therefore, the second analysis of the incomplete data was carried out assuming a serial correlation structure with the GIM, $\underset{\sim}{\Sigma}_{SE}$.

Table 3.2.4 Analysis of Incomplete Data with Pairwise Deletion

Estimates

$\hat{\underset{\sim}{B}}$:

	Y_2	Y_4	Y_6	Y_8	Y_{10}
X_1	35.99	43.22	64.55	73.26	81.38
X_2	26.68	18.03	42.16	37.36	53.24
X_3	32.28	20.26	50.20	48.52	43.58
X_4	33.07	24.91	36.13	36.49	45.29
PRE	1.029	1.180	1.103	1.189	1.178

$\hat{\underset{\sim}{\Sigma}}/\hat{\underset{\sim}{R}}$:

	Y_2	Y_4	Y_6	Y_8	Y_{10}
Y_2	884	1209	869	712	767
Y_4	.912	1986	1755	1594	1238
Y_6	.771	1.039	1438	1394	1386
Y_8	.652	.973	1.000	1351	1166
Y_{10}	.624	.672	0.884	.767	1710

Eigenvalues of $\underset{\sim}{\Sigma}$:

6501.87
690.849
370.141
-19.513
-174.901

Estimation of σ^2 and ρ was incorporated into the iterative algorithm described in section 3.1 and used to produce estimates of $\underset{\sim}{B}$ and $\underset{\sim}{\Sigma}_{SE}$. A relative change of less than 10^{-5} in the parameter estimates was the convergence criterion. The final estimates are given in Table 3.2.5 along with Wald test statistics for the seven hypotheses described earlier. This analysis appears to be more efficient than the analysis with listwise deletion comparing the two sets of variance-covariance estimates. This would be expected since that analysis was based on only 16 subjects.

The fourth incomplete data analyis was carried out assuming a tri-diagonal covariance structure, $\underset{\sim}{\Sigma}_{TR}$, with the GIM. Based on the estimate of $\underset{\sim}{\Sigma}$ from the complete data, this model appears unreasonable. Positive correlations exist between all measurements on a subject, and not just those adjacent in time. In fact the estimates of $\underset{\sim}{B}$ and $\underset{\sim}{\Sigma}_{TR}$ failed to converge.

This example is one in which an analysis of the incomplete data could not have been carried out with the GIM unless some constraints were imposed on the structure of $\underset{\sim}{\Sigma}$. Assuming a serial correlation structure appears to be a reasonable way to efficiently analyze the data. Estimation of the serial correlation parameters is the topic of the next chapter.

Table 3.2.5 Analysis of Incomplete Data with a Serial Correlation Structure (n = 45)

Estimates

$\hat{\underset{\sim}{B}}$:

	Y_2	Y_4	Y_6	Y_8	Y_{10}
X_1	35.86	42.18	65.81	69.44	82.60
X_2	26.25	18.36	41.18	36.83	56.41
X_3	33.56	19.08	50.25	46.46	44.17
X_4	32.90	18.52	35.15	33.69	47.28
PRE	1.030	1.188	1.105	1.199	1.166

$\hat{\underset{\sim}{\Sigma}}/\hat{\underset{\sim}{R}}$:

	Y_2	Y_4	Y_6	Y_8	Y_{10}
Y_2	1475	1345	1227	1119	1021
Y_4	.912	1475	1345	1227	1119
Y_6	.832	.912	1475	1345	1227
Y_8	.759	.832	.912	1475	1345
Y_{10}	.692	.759	.832	.912	1475

Tests of Hypotheses:

	H_1	H_2	H_3	H_4	H_5	H_6	H_7
Wald Test	22.40	.33	1.94	2.74	3.70	4.06	2.22
Degress of Freedom	15	3	3	3	3	3	3
p-value	.098	.954	.585	.433	.296	.255	.528

CHAPTER 4

GIM ANALYSIS WITH AUTOREGRESSIVE COVARIANCE MODELS

4.1 Introduction

The analysis of longitudinal data is typically concerned with characterizing population trends in one or more response variables. From these trends tests about subgroups in the population can be carried out as well as predictions of future growth or behavior. Associated with the responses of an individual is a set of residuals about the population trend or regression equation. When autocorrelations appear to be present in the subjects' residual values, a reasonable assumption that can be made is that the residuals arise from a stationary Gaussian process, and thus their behavior can be modeled with a suitable time series model. The advantages of such an assumption are that estimation of covariance parameters is relatively straightforward for some classes of time series, and asymptotic theory is often available for these estimates.

The use of time series models to explain residual behavior appears extensively in the field of econometrics. The most common model used by econometricians to analyze repeated measurements or longitudinal data is the cross-sectional time series model (CSTS). A brief review of the literature concerning estimation for the CSTS model is given in section 1.3. This model differs from the classical GLMM in that the vector of regression coefficients is assumed to be the same at each time point or for each response variable.

In this and the following chapters, methods of covariance parameter estimation often used in conjunction with CSTS models are adpated for use in the GLMM with missing data. Estimation is presented in this chapter for the GIM assuming that the residuals follow an autoregressive process of order one. An illustrative example of the methods is given.

4.2 Estimation of Model Parameters

In this section analysis of the GIM is developed under the assumption that the residuals for each subject are generated by an autoregressive process of order one, AR(1). This assumption is reasonable for some classes of longitudinal data since it implies that the correlation between two responses decreases with time.

Recall that for the complete data case we have the model

$$\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{B} + \underset{\sim}{U}$$

where $\underset{\sim}{U}$ is the (n x p) matrix of errors. The vector of errors for the i^{th} subject, $\underset{\sim}{u}_i = (u_{i1},\ldots,u_{ip})'$ is said to follow an AR(1) process if (Anderson, 1971)

(4.2.1) $$u_{it} = \rho\, u_{i,t-1} + e_{it} \quad , \quad \begin{matrix} t = 1,\ldots,\infty, \\ i = 1,\ldots,n, \end{matrix}$$

where

$$e_{it} \sim N(0, \sigma_e^2),$$

$$\mathrm{Cov}(e_{it}, e_{is}) = 0 \ ,$$

$$\mathrm{Cov}(u_{i,t-1}, e_{it}) = 0 \ .$$

By successive substitution, u_{it} can be expressed as a linear combination of all previous terms in the series, namely

(4.2.2) $$u_{it} = \rho^t\, u_{i0} + \rho^{t-1}\, e_{i1} + \ldots + \rho\, e_{i,t-1} + e_{it}.$$

Assuming the initial condition

$$u_{i0} \sim N(0, \sigma_e^2/(1 - \rho^2)) \quad , \quad i = 1,\ldots,n,$$

it can be shown that (Anderson, 1971)

$$\text{(4.2.3)} \qquad \operatorname{var}(u_{it}) \equiv \sigma_u^2 = \sigma_e^2/(1 - \rho^2)$$

and

$$\text{(4.2.4)} \qquad \operatorname{cov}(u_{it}, u_{is}) = \rho^{|t-s|}\, \sigma_e^2/(1 - \rho^2), \qquad \begin{array}{r} i = 1,\ldots,n, \\ t,s = 1,\ldots,\infty. \end{array}$$

Converting to matrix notation, the assumption of an AR(1) process yields a variation of the serial correlation matrix introduced in Chapter 3 as the variance-covariance model, i.e., we have $\underset{\sim}{Y} \sim \text{GLMM}\ (\underset{\sim}{X}\, \underset{\sim}{B},\ \underset{\sim}{\Sigma}_s)$ with

$$\text{(4.2.5)} \qquad \underset{\sim}{\Sigma}_s = \frac{\sigma_e^2}{(1-\rho^2)} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p-1} \\ \rho & & & & \\ \vdots & & & & \\ \rho^{p-1} & & \cdots & & 1 \end{bmatrix} .$$

Kmenta (1971) proposed a three-stage estimator for the CSTS model which is described in section 1.2. The following estimation process for the GIM assuming an AR(1) error process is an adaptation of the method proposed by Kmenta to the multivariate missing data case. It is important to note that, unlike the CSTS model, the vector of model parameters, $\underset{\sim}{b}_t$, t=1,...,p, varies with time in the GIM.

Recall that to apply the GIM the response vectors are first grouped according to like patterns of missing data, and each group is therefore "complete" in a subset of the p responses or time points. Repeating the notation for convenience, we have the following model for the j^{th} data group:

$$\underset{\sim}{Y}_j = \underset{\sim}{D}_j\, \underset{\sim}{B}\, \underset{\sim}{K}_j + \underset{\sim}{U}_j$$

where $\underset{\sim}{Y}_j$, $\underset{\sim}{D}_j$, $\underset{\sim}{B}$, and $\underset{\sim}{K}_j$ are defined in (2.1.3) and $\underset{\sim}{U}_j$ denotes the $(n_j \times m_j)$ matrix of errors. Assuming that the rows of $\underset{\sim}{U}_j$ were generated by independent and identical AR(1) processes, we have

$$\mathrm{Var}(\underset{\sim}{Y}_j) = \underset{\sim}{K}_j' \underset{\sim}{\Sigma}_s \underset{\sim}{K}_j$$

where $\underset{\sim}{\Sigma}_s$ is defined in (4.2.5).

As seen in Chapter 2, BAN estimation of $\underset{\sim}{B}$ in this model is possible provided a consistent estimate of $\underset{\sim}{\Sigma}_s$ is available and the conditions of Theorem 2.3.2 are met. Equation (3.1.2) gives the simplified form for this estimator in terms of the vector version of the model (i.e., $\underset{\sim}{Y}_j$ and $\underset{\sim}{B}$ are "rolled out" by columns). The following procedure defines a WLS estimator, $\hat{\underset{\sim}{B}}_s$, computed in two stages. At the first stage, ordinary least squares (OLS) estimates of $\underset{\sim}{B}$ are used to generate estimates of the covariance parameters. The weighted least squares (WLS) estimate of $\underset{\sim}{B}$ is then computed at the second stage.

First some additional notation must be established. Let $\underset{\sim}{R}_j$ denote the $(n_j \times m_j)$ matrix of residuals for the j^{th} data group computed using an estimate, $\hat{\underset{\sim}{B}}$, as follows:

(4.2.7) $$\underset{\sim}{R}_j = \underset{\sim}{Y}_j - \underset{\sim}{D}_j \hat{\underset{\sim}{B}} \underset{\sim}{K}_j \, .$$

Let $\underset{\sim}{r}_{t_{(j)}}$ denote the t^{th} column of $\underset{\sim}{R}_j$, $t=1,\ldots,m_j$. Let T_j denote the set of indices of the time points for which the j^{th} group has complete responses. Suppose that there are a_j pairs of consecutive time points in T_j. Let A_j denote the set of indices of time points for which the preceding time point is also in T_j. For example, if group j consists of responses measured at the first, second, third, fifth, and sixth time points, then $T_j = \{1,2,3,5,6\}$, $a_j = 3$, and $A_j = \{2,3,6\}$. Let N_t denote the the total number of responses at time t, $t \leq p$. Then $N = \sum_{t=1}^{p} N_t = \sum_{j=1}^{u} n_j m_j$. Finally, let N_a denote the total number of observations adjacent in time, i.e., $N_a = \sum_{j=1}^{u} n_j a_j$.

With these definitions consider the following estimation procedure.

(4.2.8) (i) Compute OLS estimates $\hat{\underset{\sim}{B}}^{(1)}$ from (3.1.2).

(ii) Compute residual matrices $\underset{\sim}{R}_j$ by substituting $\hat{\underset{\sim}{B}}^{(1)}$ in (4.2.7).

(iii) Compute $\underset{\sim}{S} = \sum_{t=1}^{p} \underset{\sim}{K}_j \underset{\sim}{R}_j' \underset{\sim}{R}_j \underset{\sim}{K}_j'$.

(iv) Compute $\hat{\sigma}_u^2 = \sum_{t=1}^{p} s_{tt}/(N - q)$ where

$\underset{\sim}{S} = ((s_{tw}))$ and $q = \text{rank}\ (\underset{\sim}{D}_j'\underset{\sim}{D}_j)$ for any $j=1,\ldots,u$.

(v) Compute $\hat{\rho}$ as follows:

$$\hat{\rho} = \frac{\sum_{j=1}^{u} \sum_{t\varepsilon A_j} \underset{\sim}{r}_{t(j)}' \ \underset{\sim}{r}_{t-1(j)}}{\sum_{j=1}^{u} \sum_{t\varepsilon A_j} \underset{\sim}{r}_{t-1(j)}' \ \underset{\sim}{r}_{t-1(j)}} .$$

(vi) Compute $\hat{\underset{\sim}{\Sigma}}_s$ from (4.2.5) with $\hat{\sigma}_u^2$ and $\hat{\rho}$.

(vii) Compute WLS estimates $\hat{\underset{\sim}{B}}_s$ from (3.1.2) with $\hat{\underset{\sim}{\Sigma}}_s$.

This estimator differs from that proposed by Kmenta (1971) due to the presence of missing values. In addition, the data transformation proposed by Kmenta for the estimation of σ_e^2 is not feasible with incomplete data. A consistent estimate of σ_e^2 can be computed using $\hat{\sigma}_u^2$ and $\hat{\rho}$ as follows:

$$\hat{\sigma}_e^2 = \hat{\sigma}_u^2(1 - \hat{\rho}^2).$$

The estimator $\hat{\underset{\sim}{B}}_s$ does not enjoy the small sample properties of Kmenta's estimator. Furthermore, because the number of model parameters in the GIM depends upon the number of time points, the asymptotic properties established for estimators of CSTS model prameters as the number of time points becomes infinite are not relevant here. The following theorem establishes $\hat{\underset{\sim}{\Sigma}}_s$ as a consistent estimator of $\underset{\sim}{\Sigma}_s$ as the number of subjects with consecutive responses increases for a fixed time period, $t=1,\ldots,p$. Sections of the proof are similar in spirit to a proof due to Theil (1971).

Theorem 4.2.1. The estimator $\hat{\underset{\sim}{\Sigma}}_s$ is a consistent estimator of $\underset{\sim}{\Sigma}_s$ in the GIM with normality when the errors associated with each subject's responses are assumed to follow independent and identical AR(1) processes.

Proof. For $t=1,\ldots,p$, $s_{tt}/(N_t - q)$ is the usual mean square estimate for the variance of the responses at time t resulting from a univariate OLS analysis. Assuming normality, this estimate is known to be unbiased and consistent as $N_t \to \infty$ (e.g., Anderson, 1971). Under the assumptions of the model, the responses are homoscedastic across time with common variance given by $\sigma_u^2 = \sigma_e^2/(1 - \rho^2)$. Therefore, as $N\to\infty$ such that N_t/N does not vanish, the estimate $\hat{\sigma}_u^2$ defined in step (iv) is consistent for σ_u^2.

To see that $\hat{\rho}$ is consistent for ρ, recall that the error terms follow an AR(1) process and thus satisfy

$$u_{it_{(j)}} = \rho\, u_{i,t-1_{(j)}} + e_{it_{(j)}}, \qquad i=1,\ldots,n_j,\ t\varepsilon A_j,\ j=1,\ldots,u, \tag{4.2.9}$$

where $\underset{\sim}{U}_j = ((u_{it}))_j$ and the $e_{it_{(j)}}$ are independent random variables distributed as $N(0, \sigma_e^2)$. Note that (4.2.9) specifies a set of precisely N_a regression equations. Letting $\underset{\sim}{R}_j = ((r_{it}))_j$, $\hat{\rho}$ can be expressed as

$$\hat{\rho} = \frac{\sum_{j=1}^{u} \sum_{t\varepsilon A_j} \sum_{i=1}^{n_j} r_{it_{(j)}}\, r_{i,t-1_{(j)}}}{\sum_{j=1}^{u} \sum_{t\varepsilon A_j} \sum_{i=1}^{n_j} r^2_{i,t-1_{(j)}}}\,. \tag{4.2.10}$$

By definition,

(4.2.11)
$$\underset{\sim}{R}_j = \underset{\sim}{Y}_j - \underset{\sim}{D}_j \, \hat{\underset{\sim}{B}}^{(1)} \, \underset{\sim}{K}_j$$
$$= \underset{\sim}{U}_j - \underset{\sim}{D}_j \, (\hat{\underset{\sim}{B}}^{(1)} - \underset{\sim}{B})K_j$$

where $\underset{\sim}{B}$ is the true (unknown) matrix of model parameters. $\hat{\underset{\sim}{B}}^{(1)}$ is the OLS estimator of $\underset{\sim}{B}$ and therefore $\hat{\underset{\sim}{B}}^{(1)} \overset{p}{\rightarrow} \underset{\sim}{B}$ as $n = \left[\sum_{j=1}^{u} n_j\right] \rightarrow \infty$ provided that the usual regularity conditions hold; viz., as $n \rightarrow \infty$, $\underset{\sim}{D}_j' \underset{\sim}{D}_j/n_j$ converges to a positive definite matrix for $j = 1,\ldots,u$. If we require that $n \rightarrow \infty$, then equation (4.2.11) implies that $r_{it_{(j)}}$ converges to $u_{it_{(j)}}$ in distribution for $i=1,\ldots,n_j$, $t\varepsilon A_j$, and $j=1,\ldots,u$.

Let

(4.2.12)
$$z = \frac{\sum_{j=1}^{u} \sum_{t\varepsilon A_j} \sum_{i=1}^{n_j} u_{it_{(j)}} \; u_{i,t-1_{(j)}}}{\sum_{j=1}^{u} \sum_{t\varepsilon A_j} \sum_{i=1}^{n_j} u^2_{i,t-1_{(j)}}} .$$

Then $\hat{\rho}$ converges in distribution to z as $n\rightarrow\infty$. Replacing $u_{it_{(j)}}$ in (4.2.12) with the right hand side of (4.2.9) and multiplying both the numerator and denominator by $1/N_a$ we get

(4.2.13)
$$z = \rho \cdot \left[\frac{\frac{1}{N_a} \sum_{j=1}^{u} \sum_{t\varepsilon A_j} \sum_{i=1}^{n_j} u^2_{i,t-1_{(j)}}}{\frac{1}{N_a} \sum_{j=1}^{u} \sum_{t\varepsilon A_j} \sum_{i=1}^{n_j} u^2_{i,t-1_{(j)}}}\right] + \left[\frac{\frac{1}{N_a} \sum_{j=1}^{u} \sum_{t\varepsilon A_j} \sum_{i=1}^{n_j} u_{i,t-1_{(j)}} \; e_{it_{(j)}}}{\frac{1}{N_a} \sum_{j=1}^{u} \sum_{t\varepsilon A_j} \sum_{i=1}^{n_j} u^2_{i,t-1(j)}}\right]$$
$$= \rho + z_1/z_2.$$

To show that $\hat{\rho} \xrightarrow{p} \rho$ it suffices to show that $z_1 \xrightarrow{p} 0$ and $z_2 \xrightarrow{p} c$, where c is any positive constant. Dropping the subscript limits for convenience, we have

$$E(z_1) = \frac{1}{N_a} \sum_j \sum_t \sum_i E(u_{i,t-1_{(j)}} e_{it_{(j)}}) .$$

From (4.2.2), $u_{i,t-1_{(j)}}$ can be written as a linear combination of $e_{i,t-1_{(j)}}$, $e_{i,t-2_{(j)}}$, etc. The independence of the $e_{it_{(j)}}$ implies that $E(z_1) = 0$.

By Chebyshev's Inequality (e.g., Cramer, 1946) if $Var(z_1) \to 0$, then $z_1 \to E(z_1)$. Due to the independence of subjects we have

$$(4.2.14) \quad Var(z_1) = \frac{1}{N_a^2} \sum_j \sum_i E(\sum_t u_{i,t-1_{(j)}} e_{it_{(j)}})^2$$

$$= \frac{1}{N_a^2} \sum_j \sum_i [E \sum_t u^2_{i,t-1_{(j)}} e^2_{it_{(j)}}$$

$$+ 2 E \sum_{\substack{t,s \varepsilon A_j, \\ t<s}} \sum u_{i,t-1_{(j)}} e_{it_{(j)}} u_{i,s-1_{(j)}} e_{is_{(j)}}]$$

$$= \frac{1}{N_a^2} \sum_j \sum_i [\sum_t E(u^2_{i,t-1_{(j)}}) E(e^2_{it_{(j)}})$$

$$+ 2 \sum_{t<s} \sum E(e_{is_{(j)}}) E(u_{i,t-1_{(j)}} u_{i,s-1_{(j)}} e_{it_{(j)}})].$$

The last equality in (4.2.14) holds because $e_{it_{(j)}}$ is independent of $u_{i,t-1_{(j)}}$, as noted above, and $e_{is_{(j)}}$ is independent of $u_{i,t-1_{(j)}}$, $u_{i,s-1_{(j)}}$, and $e_{it_{(j)}}$ for $t<s$. However $E(e_{is_{(j)}}) = 0$, therefore

$$Var(z_1) = \frac{1}{N_a^2} (N_a \; Var(e_{it_{(j)}}) \; Var(u_{i,t-1_{(j)}}))$$

$$= \frac{1}{N_a}\left[\frac{\sigma_e^4}{1-\rho^2}\right].$$

If $N_a \to \infty$, then $\mathrm{Var}(z_1) \to 0$ and thus $z_1 \xrightarrow{p} 0$.

It remains to be shown that z_2 converges to a constant as $n_j \to \infty$ for all j. We have

$$E(z_2) = \frac{1}{N_a} \sum_j \sum_i \sum_t E(u^2_{i,t-1_{(j)}}) = \frac{\sigma_e^2}{1-\rho^2}.$$

$$\mathrm{Var}(z_2) = \frac{1}{N_a^2} \sum_j \sum_i \mathrm{Var}(\sum_t u^2_{i,t-1_{(j)}}).$$

Letting $\sigma_u^2 = \sigma_e^2/(1-\rho^2)$, we have

(4.2.15) $$\mathrm{Var}(\sum_t u^2_{i,t-1_{(j)}}) = \sum_t E(u^2_{i,t-1_{(j)}} - \sigma_u^2)^2$$

$$+ \sum_{t<s}\sum E(u^2_{i,t-1_{(j)}} - \sigma_u^2)\, E(u^2_{i,s-1_{(j)}} - \sigma_u^2)$$

$$= \sum_t (E\, u^4_{i,t-1_{(j)}} - \sigma_u^4)$$

$$+ \sum_{t<s}\sum (E\, u^2_{i,t-1_{(j)}}\, u^2_{i,s-1_{(j)}} - \sigma_u^4)$$

To simplify the second term on the right-hand side of (4.2.15) note that $u_{i,s-1_{(j)}}$ can be expressed as a function of $u_{i,t-1_{(j)}}$ for $t<s\varepsilon A_j$. If $s-t = k$, then by successively substituting in (4.2.9) we get

$$u_{i,s-1_{(j)}} = \rho^k\, u_{i,t-1_{(j)}} + \sum_{\ell=1}^{k} \rho^{k-\ell}\, e_{i,t-1+\ell_{(j)}},$$

and

$$u^2_{i,t-1_{(j)}}\, u^2_{i,s-1_{(j)}} = \rho^{2k}\, u^4_{i,t-1_{(j)}} + \sum_{\ell=1}^{k} u^2_{i,t-1_{(j)}}\, \rho^{2(k-\ell)}\, e^2_{i,t-1+\ell_{(j)}} +$$

$$+ \sum_{\ell<m}\sum u^2_{i,t-1_{(j)}} \rho^{k-\ell} \rho^{k-m} e_{i,t-1+\ell_{(j)}} e_{i,t-1+m_{(j)}}$$

$$+ 2 \sum_{\ell=1} \rho^{2k-\ell} u^3_{i,t-1_{(j)}} e_{i,t-1+\ell_{(j)}} .$$

Taking the expectation, we get

$$E(u^2_{i,t-1_{(j)}} u^2_{i,s-1_{(j)}}) = \rho^{2k} E(u^4_{i,t-1_{(j)}}) + \sum_{\ell=1}^{k} \rho^{2(k-\ell)} \sigma^2_u \sigma^2_e$$

because $u_{i,t-1_{(j)}}$ is independent of $e_{i,t-1+\ell_{(j)}}$ for $\ell > 0$ and $E(e_{i,t-1+\ell_{(j)}}) = 0$. We have

$$\mathrm{Var}(z_2) = \frac{1}{N_a^2} \sum_j \sum_i \sum_t (E\, u^4_{i,t-1_{(j)}} - \sigma^4_u) + \sum_{t<s}\sum [\rho^{2k} (E\, u^4_{i,t-1_{(j)}} - \sigma^4_u) + \sum_{\ell=1}^{k} \rho^{2(k-\ell)} \sigma^2_u \sigma^2_e] .$$

The assumption of normality guarantees that the $u_{it_{(j)}}$ have a finite fourth moment; thus $\mathrm{Var}(z_2)$ is of the order $1/N_a$. If $N_a \to \infty$, $\mathrm{Var}(z_2) \to 0$, and therefore $z_2 \overset{p}{\to} E(z_2) = \sigma^2_u > 0$.

Finally, $z_1/z_2 \overset{p}{\to} 0$ and thus $\hat{\rho} \overset{p}{\to} \rho$.

Let $\hat{\Sigma}_s = \Sigma_s (\hat{\rho}, \hat{\sigma}^2_u)$. By Lemma 2.3.1, $\hat{\Sigma}_s$ is consistent for Σ_s.

q.e.d.

The estimate $\hat{\rho}$ defined in step (v) of (4.2.8) is computed using only complete pairs of responses at consecutive time points. Implicit in the use of this estimator is the assumption that the autocorrelation among the unobserved responses is identical to the observed responses. If the process that generated the missing data is suspected to be non-random, this estimator should be used with caution, for the properties of $\hat{\rho}$ would then depend on the missing data process. In particular the estimator may not be consistent for the population autocorrelation.

From Theorm 2.3.2, $\hat{\underset{\sim}{B}}_s$ will be a BAN estimator if independent subsequences of the sample are used to estimate $\underset{\sim}{\Sigma}_s$ and $\underset{\sim}{B}_s$. For small sample sizes it may be more practical to use the entire sample for estimation of both parameter matrices, although the asymptotic properties are not guaranteed. The WLS estimator can also be refined through iterative estimation of the covariance parameters. Let $\hat{\underset{\sim}{B}}_s^{(k)}$ denote the WLS estimate at the end of the k^{th} iteration. Then estimates for σ_u^2 and ρ can be computed using

$$\underset{\sim}{R}_j^{(k)} = Y_j - \underset{\sim}{D}_j \hat{\underset{\sim}{B}}^{(k)} \underset{\sim}{K}_j$$

in step (iii) of (4.2.8) for the (k + 1)st iteration.

The consistent estimate, $\hat{\underset{\sim}{\Sigma}}_s$, defined in (4.2.8) can be used to estimate the asymptotic variance-covariance matrix of $\hat{\underset{\sim}{B}}_s$ as discussed in section 2.4. If $\hat{\underset{\sim}{b}}_s$ denotes the vector-version of $\hat{\underset{\sim}{B}}_s$ then

$$(4.2.16)\quad \hat{\text{Var}}(\hat{\underset{\sim}{b}}_s) = \left[\sum_{j=1}^{u} \underset{\sim}{K}_j (\underset{\sim}{K}_j' \hat{\underset{\sim}{\Sigma}}_s \underset{\sim}{K}_j)^{-1} \underset{\sim}{K}_j' \otimes \underset{\sim}{D}_j' \underset{\sim}{D}_j\right]^{-1}$$

Note that if all subjects have responses at time t then

$$(4.2.17)\quad \hat{\text{Var}}(\hat{\underset{\sim}{b}}_{t_{(s)}}) = \hat{\sigma}_u^2 \sum_{j=1}^{u} (\underset{\sim}{D}_j' \underset{\sim}{D}_j)^{-1}$$

where $\hat{b}_{t_{(s)}}$ is the t^{th} column of $\hat{\underset{\sim}{B}}_s$. Thus estimates of model parameters corresponding to time points with complete data will have identical standard errors under these model assumptions. This is illustrated in the next section. The expression in (4.2.16) can also be used to generate Wald statistics for testing hypotheses about the model parameters, provided that the conditions of Theorem 2.4.1 are met.

4.3 An Example

To illustrate the estimation process described in the previous section, 150 observation vectors consisting of five responses each were generated for analysis. First error vectors were generated via an autoregressive process as follows:

$$u_{i0} = \sqrt{\sigma_e^2}\; e_{i0} ,$$

$$u_{it} = \rho\, u_{i,t-1} + e_{it} \quad , \qquad i = 1,\ldots,150, \; t = 2,\ldots,5$$

where

$$\sigma_e^2 = 0.25,$$

$$\rho = 0.70,$$

$$e_{it} = \sim N(0,0.25) .$$

The e_{it} are pseudo-random variables generated with the SAS function NORMAL (Barr, et. al, 1979). Descriptive statistics for the generated errors at the five time points are given in Table 4.3.1. The dependent variables were then generated according to the following model:

$$\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{B}$$

where $\underset{\sim}{X} = (\underset{\sim}{X}_1 \;\; \underset{\sim}{X}_2)$ is the design matrix, $\underset{\sim}{X}_1 = \underset{\sim}{1}$ and $\underset{\sim}{X}_2$ consists of elements ranging from 10 to 50, and

$$\underset{\sim}{B} = \begin{bmatrix} 1.0 & 1.5 & 2.0 & 2.5 & 3.0 \\ 0.2 & 0.6 & 0.8 & 0.9 & 1.0 \end{bmatrix} .$$

Thus the model parameters correspond to an intercept and a slope with respect to a covariate at each time point.

For computational simplicity, only two patterns of missing data were introduced. The responses at time t = 2 were deleted in 50 observations and the responses at time t = 3 were deleted in 50 observations. We have the following GIM parameters:

Table 4.3.1 Descriptive Statistics for Generated Errors (N = 150) at Times 1 - 5

	MEAN	S.D.
u_1	-0.0135	0.6834
u_2	0.0290	0.6879
u_3	0.0354	0.6964
u_4	-0.0565	0.7101
u_5	0.0288	0.6987

Correlation Matrix

	u_1	u_2	u_3	u_4	u_5
u_1	1	0.6883	0.5415	0.3998	0.3268
u_2	0.6883	1	0.7367	0.4564	0.3567
u_3	0.5415	0.7367	1	0.6479	0.4034
u_4	0.3998	0.4564	0.6479	1	0.6831
u_5	0.3268	0.3567	0.4034	0.6831	1

$$u = 3,$$

$$n_j = 50 \qquad \text{for } j=1,2,3,$$

$$\underset{\sim}{K}_1 = \underset{\sim}{I}_5,$$

$$\underset{\sim}{K}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \underset{\sim}{K}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$T_1 = \{1,2,3,4,5\}, \quad T_2 = \{1, 2, 3, 5\}, \quad T_3 = \{1,2,4,5\},$$

$$A_1 = \{2,3,5,5\}, \quad A_2 = \{2,3\}, \quad A_3 = \{2,3,4\} .$$

Initial estimates, $\hat{\underset{\sim}{B}}^{(1)}$, were produced with software developed for analyzing GIM data. Residuals about these OLS estimates were then pooled as described in steps (ii) - (iv) to produce the estimates

$$\hat{\rho} = 0.689$$

and

$$\hat{\sigma}_u^2 = 0.494 .$$

The estimate of ρ is based on 450 pairs of consecutive responses and the estimate of σ_u^2 is based on 650 responses.

Second stage estimates, $\hat{\underset{\sim}{B}}_s$, were computed by employing the GIM software with $\hat{\underset{\sim}{\Sigma}}_s = \underset{\sim}{\Sigma}_s(\hat{p}, \hat{\sigma}_u^2)$.

The estimated model parameters and the corresponding estimated standard errors are given in Table 4.3.2. Note that the standard errors corresponding to the first three time points are identical. This is due to the fact that responses at those times are complete and (4.2.17) holds.

For comparison purposes, results from a GLMM analysis of the complete dataset, prior to deleting responses at random, are given in Table 4.3.3. Confidence intervals constructed about the estimates in Table 4.3.2 contain both the population parameters and the complete data estimates. The standard errors in Table 4.3.2 compare favorably with those from the complete data analysis in Table 4.3.3. Although the entire sample was used to estimate $\underset{\sim}{B}_s$, this procedure yields reasonable results.

Table 4.3.2 Estimates and Standard Errors for $\underset{\sim}{B}_s$

$\hat{\underset{\sim}{B}}_s$:	Y_1	Y_2	Y_3	Y_4	Y_5
Intercept	1.09	1.81	2.26	2.39	2.93
Slope	0.20	0.59	0.79	0.90	1.00
SE ($\hat{\underset{\sim}{B}}_s$):					
Intercept	0.1597	0.1597	0.1597	0.1745	0.1833
Slope	0.0050	0.0050	0.0050	0.0054	0.0057

Table 4.3.3 GLMM Estimates from the Complete Dataset

$\hat{\underset{\sim}{B}}_s$:	Y_1	Y_2	Y_3	Y_4	Y_5
Intercept	1.09	1.81	2.26	2.40	3.08
Slope	0.20	0.59	0.79	0.90	1.00
SE ($\hat{\underset{\sim}{B}}_s$):					
Intercept	0.1557	0.1549	0.1576	0.1620	0.1593
Slope	0.0048	0.0048	0.0049	0.0050	0.0049

CHAPTER 5

GIM ANALYSIS WITH MOVING AVERAGE COVARIANCE MODELS

5.1 Introduction

The covariance structure associated with a moving average process is frequently used in the analysis of econometric data. This covariance model, which is not as restrictive as that associated with an autoregressive process, is useful in analyzing longitudinal data when it is known that the errors are correlated but the analyst is reluctant to specify the exact nature of the autocorrelation.

Estimation of the GIM parameters under the assumption of a moving average model for the errors is presented in section 5.2. A two-stage estimation for $\underset{\sim}{B}$ is proposed. Variance components are estimated in the first stage and then used to compute a WLS estimate in the second stage. An illustrative example is presented in section 5.3.

5.2 Estimation of Model Paramters

Let $\underset{\sim}{U}$ denote the (n x p) matrix of residuals for the hypothetical complete data matrix. The error vector $u_i = (u_{i1},\ldots,u_{ip})'$ for the i^{th} subject is said to follow a moving average process of order M, MA(M), (ref. Anderson, 1971) if

$$(5.2.1) \quad u_{it} = e_{it} + a_1 e_{i,t-1} + a_2 e_{i,t-2} + \cdots + a_M e_{i,t-M}, \qquad i=1,\ldots,n, \quad t \in (-\infty\ \infty), \quad M \le p-1$$

where

$$e_{it} \sim N(0,\sigma_e^2) \text{ for } t \in (-\infty, \infty)$$

and $$\mathrm{Cov}(e_{it}, e_{is}) = 0 \quad \text{for } t \neq s .$$

Thus we have

$$E(u_{it}) = 0,$$

$$\mathrm{Var}(u_{it}) = \sigma_e^2(1 + a_1^2 + \ldots + a_M^2),$$

$$\mathrm{Cov}(u_{it}, u_{i,t-s}) = \begin{cases} \sigma_e^2 \,(a_s + a_1\, a_{s+1} + \ldots + a_{M-s}\, a_M), & |s| \leq M \\ 0 & , \quad |s| > M. \end{cases}$$

Without loss of generality, the covariance structure of $\underset{\sim}{u}_i$ can be formulated as follows:

(5.2.3) $$\mathrm{Cov}(u_{it}, u_{i,t-s}) = \begin{cases} \gamma_{|s|} & , \quad |s| \leq M \\ 0 & , \quad |s| > M . \end{cases}$$

There are M+1 unknown variance-covariance parameters. The dispersion matrix for each $\underset{\sim}{u}_i$ is given by

(5.2.4) $$\underset{\sim}{\Gamma} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_M & 0 \cdots & 0 \\ \gamma_1 & \gamma_0 & \gamma_1 \cdots & \gamma_{M-1} & \gamma_M \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & & \cdot \quad \cdot & \cdot & & \gamma_0 \end{bmatrix} .$$

A decomposition suggested by da Silva (1974) reduces the problem of estimating $\underset{\sim}{\Gamma}$ to the estimation of M+1 variance components. Let

(5.2.5) $$\underset{\sim}{V}_0 = \underset{\sim}{I}_p$$

and $$\underset{\sim}{V}_k = ((v_{rs}))_k \quad , \quad k=1,\ldots,M,$$

where $$v_{rs_{(k)}} = \begin{cases} 1 & , \quad |r-s| = k \\ 0 & , \quad |r-s| \neq k . \end{cases}$$

Then

(5.2.6) $$\mathrm{Var}(\underset{\sim}{u}_i) = \sum_{k=0}^{M} \gamma_k \underset{\sim}{V}_k .$$

Since the $\underset{\sim}{V}_k$ are known matrices, several methods of variance components estimation are available in the literature (e.g., Searle, 1971) for estimating the γ_k. Da Silva (1974) proposed a method due to Seely (1969) for estimating variance components in the CSTS model. Since this method produces consistent estimates, it is adapted here for use in the GIM.

Seely's method can be summarized briefly as follows. Suppose $\underset{\sim}{z}$ is an (n x 1) vector such that $E(\underset{\sim}{z}) = \underset{\sim}{X}\,\underset{\sim}{\alpha}$ for some full-rank matrix $\underset{\sim}{X}$ and vector of unknown parameters $\underset{\sim}{\alpha}$. Assume that the variance of $\underset{\sim}{z}$ is given by (5.2.6). Let $\underset{\sim}{Q} = \underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$. Let $\underset{\sim}{g}$ be a solution to the equation

(5.2.7) $$\underset{\sim}{H}\,\underset{\sim}{g} = \underset{\sim}{c}$$

where $\underset{\sim}{H} = ((h_{ij}))$, $h_{ij} = \text{trace}(\underset{\sim}{Q}\,\underset{\sim}{V}_i\,\underset{\sim}{Q}\,\underset{\sim}{V}_j)$, $i,j = 0,\ldots,M$,

and $\underset{\sim}{c} = (c_j)$, $c_j = \underset{\sim}{z}'\,\underset{\sim}{Q}\,\underset{\sim}{V}_j\,\underset{\sim}{Q}\,\underset{\sim}{z}$, $j = 0,\ldots,M$.

Then $\underset{\sim}{g}$ is an unbiased estimate of $(\gamma_0, \gamma_1,\ldots,\gamma_M)'$. Of course a unique solution will exist if and only if $\underset{\sim}{H}$ is non-singular. The solution of (5.2.7) is also the minimum norm quadratic unbiased estimate (Rao, 1972), or MINQUE, when no a prior information about the γ_k is available. This estimation procedure has many desirable properties.

To apply this variance estimation procedure to the GIM, let $\underset{\sim}{y}_i$ denote the vector version of the incomplete $(n_j \times m_j)$ data matrix $\underset{\sim}{Y}_j$, j=1,...,u, in which $\underset{\sim}{Y}_j$ is rolled out by columns. We have

(5.2.8) $$E(\underset{\sim}{y}_j) = (\underset{\sim}{K}_j' \otimes \underset{\sim}{D}_j)\,\underset{\sim}{b}$$

and

$$\text{Var}(\underset{\sim}{y}_j) = (\underset{\sim}{K}_j'\,\underset{\sim}{\Sigma}\,\underset{\sim}{K}_j) \otimes \underset{\sim}{I}_{n_j} , \qquad j=1,\ldots,u,$$

where $\underset{\sim}{b}$ is the vector version of $\underset{\sim}{B}$ and $\underset{\sim}{K}_j$ is the matrix indicating complete response times for the j^{th} data group. Assuming a moving average process for the errors, $\underset{\sim}{\Sigma}$ has the form given in (5.2.6). The variance of $\underset{\sim}{y}_j$ can be simplified as follows:

(5.2.9)
$$\begin{aligned} Var(\underset{\sim}{y}_j) &= \underset{\sim}{K}_j' \underset{\sim}{\Sigma} \underset{\sim}{K}_j \otimes \underset{\sim}{I}_{n_j} \\ &= \underset{\sim}{K}_j' \left(\sum_{k=0}^{M} \gamma_k \underset{\sim}{V}_k\right) \underset{\sim}{K}_j \otimes \underset{\sim}{I}_{n_j} \\ &= \sum_{k=0}^{M} (\gamma_k \underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j) \otimes \underset{\sim}{I}_{n_j} \\ &= \sum_{k=0}^{M} \gamma_k (\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \otimes \underset{\sim}{I}_{n_j}). \end{aligned}$$

Let $\underset{\sim}{W}_{k_{(j)}} = \underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \otimes \underset{\sim}{I}_{n_j}$. Then the $\underset{\sim}{W}_{k_{(j)}}$ are known and Seely's method can be applied to estimate $\underset{\sim}{\gamma} = (\gamma_0\ \gamma, \ldots \gamma_M)'$.

Consider the model for a data group with at least one missing response, i.e., $\underset{\sim}{K}_j \neq \underset{\sim}{I}_p$. Recall that the m_j columns of $\underset{\sim}{K}_j$ are a subset of the columns of $\underset{\sim}{I}_p$. $\underset{\sim}{K}_j$ contains at least one row whose elements are all zero. The design matrix, $\underset{\sim}{K}_j' \otimes \underset{\sim}{D}_j$, is therefore less than full rank. Rather than employing a generalized inverse in the estimation of $\underset{\sim}{\gamma}$, an alternate formulation for $E(\underset{\sim}{y}_j)$ can be derived. Let

(5.2.10)
$$\underset{\sim}{B}_j = \underset{\sim}{B}\, \underset{\sim}{K}_j$$

and let $\underset{\sim}{b}_j$ denote the vector-version of $\underset{\sim}{B}_j$. Then it is easily seen that

$$E(\underset{\sim}{y}_j) = (\underset{\sim}{K}_j' \otimes \underset{\sim}{D}_j)\, \underset{\sim}{b} = (\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j)\, \underset{\sim}{b}_j \,.$$

Clearly $\underset{\sim}{I}_{m_j} \otimes \underset{\sim}{D}_j$ is of full rank when $\underset{\sim}{D}_j$ is.

Let

(5.2.11) $\underset{\sim}{Q}_j = \underset{\sim}{I}_{n_j m_j} - [(\underset{\sim}{I}_{m_j} \otimes \underset{\sim}{D}_j)((\underset{\sim}{I}_{m_j} \times \underset{\sim}{D}_j)' (\underset{\sim}{I}_{m_j} \otimes \underset{\sim}{D}_j))^{-1} (\underset{\sim}{I}_{m_j} \otimes \underset{\sim}{D}_j)']$

$$= \underset{\sim}{I}_{n_j m_j} - [\underset{\sim}{I}_{m_j} \otimes \underset{\sim}{D}_j (\underset{\sim}{D}_j' \underset{\sim}{D}_j)^{-1} D_j'] \ .$$

Then an estimate of the vector of unknown variance components is given by the solution to the following equation:

(5.2.12) $\underset{\sim}{H}_j \underset{\sim}{\gamma}_j = \underset{\sim}{c}_j$

where $\underset{\sim}{H}_j = ((h_{k\ell}))_j \quad , \quad h_{k\ell_{(j)}} = \text{trace}(\underset{\sim}{Q}_j \underset{\sim}{W}_{k_{(j)}} \underset{\sim}{Q}_j \underset{\sim}{W}_{\ell_{(j)}}),$

$\underset{\sim}{c}_j = (c_k)_j \quad , \quad c_{k_{(j)}} = \underset{\sim}{y}_j' \underset{\sim}{Q}_j \underset{\sim}{W}_{k_{(j)}} \underset{\sim}{Q}_j \underset{\sim}{y}_j \ ,$

and $\underset{\sim}{W}_{k_{(j)}}$ is defined above for $k = 0,\ldots,M$. Because of the patterns of the $W_{k_{(j)}}$, the off-diagonal elements of $\underset{\sim}{H}_j$ vanish. To see this first consider the terms $h_{0k_{(j)}}$, $k=1,\ldots,M$.

$$h_{0k_{(j)}} = \text{trace}\ (\underset{\sim}{Q}_j \underset{\sim}{W}_{0_{(j)}} \underset{\sim}{Q}_j \underset{\sim}{W}_{k_{(j)}})$$
$$= \text{trace}\ (\underset{\sim}{Q}_j \underset{\sim}{W}_{k_{(j)}})$$

since $\underset{\sim}{W}_{0_{(j)}} = \underset{\sim}{I}_{n_j m_j}$ for all j and $\underset{\sim}{Q}_j^2 = \underset{\sim}{Q}_j$. We have

$\text{trace}(\underset{\sim}{Q}_j \underset{\sim}{W}_{k_{(j)}}) = \text{trace}[\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \otimes I_{n_j} - \underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \otimes \underset{\sim}{D}_j (\underset{\sim}{D}_j' \underset{\sim}{D}_j)^{-1} \underset{\sim}{D}_j']$

$= \text{trace}(\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j)\ \text{trace}(\underset{\sim}{I}_{n_j}) - \text{trace}(\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j)\ \text{trace}(\underset{\sim}{D}_j (\underset{\sim}{D}_j' \underset{\sim}{D}_j)^{-1} \underset{\sim}{D}_j').$

By definition, (5.2.5), all diagonal elements of $\underset{\sim}{V}_k$ are zero. Furthermore, premultiplying by $\underset{\sim}{K}_j'$ and post-multiplying by $\underset{\sim}{K}_j$ is equivalent to deleting the i^{th} row and column of $\underset{\sim}{V}_k$ if the i^{th} response variate was not observed in the j^{th} group, for a total of $p - m_j$ such deletions. Thus the diagonal elements of $\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j$ are a subset of the diagonal elements of $\underset{\sim}{V}_k$ and are all equal to zero. This implies that the trace $(\underset{\sim}{Q}_j \underset{\sim}{W}_{k_{(j)}}) = 0$.

Now consider the terms $h_{k\ell_{(j)}}$, $k,\ell = 1,\ldots,M$ and $k \neq \ell$. We have

$$(5.2.13)\quad h_{k\ell_{(j)}} = \text{trace}\,(\underset{\sim}{Q}_j \underset{\sim}{W}_{k_{(j)}} \underset{\sim}{Q}_j W_{\ell_{(j)}})$$
$$= \text{trace}\,(\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_\ell \underset{\sim}{K}_j \otimes \underset{\sim}{I}_{n_j}$$
$$- \underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_\ell \underset{\sim}{K}_j \otimes \underset{\sim}{D}_j (\underset{\sim}{D}_j' \underset{\sim}{D}_j)^{-1} \underset{\sim}{D}_j')$$
$$= \text{trace}\,(\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_\ell \underset{\sim}{K}_j)\,(n_j - \text{trace}\,(D_j(D_j' D_j)^{-1} D_j')).$$

If $\underset{\sim}{v}_{i_{(k)}}$ denotes the i^{th} row (or column) of $\underset{\sim}{V}_k$ then from (5.2.5) it is clear that $\underset{\sim}{v}_{i_{(k)}}' \underset{\sim}{v}_{i_{(\ell)}} = 0$. Let $\underset{\sim}{v}_{i_{(k)(j)}}$ denote the i^{th} row of $\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j$. Then it is also clear that $\underset{\sim}{v}_{i_{(k)(j)}}' \underset{\sim}{v}_{i_{(\ell)(j)}} = 0$. Thus

$$(5.2.14)\quad \text{trace}\,(\underset{\sim}{V}_k \underset{\sim}{V}_\ell) = \text{trace}\,(\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_\ell \underset{\sim}{K}_j) = 0, \qquad h,\ell = 1,\ldots,M,\ k \neq \ell.$$

Similar arguments can be used to verify that the diagonal elements $h_{kk_{(j)}}$ are not necessarily zero. In particular, since $\underset{\sim}{v}_{i_{(k)}}' \underset{\sim}{v}_{i_{(k)}} \neq 0$, $h_{kk_{(j)}} \neq 0$ for the complete data group, i.e., $\underset{\sim}{K}_j = \underset{\sim}{I}_p$. However, it is possible that $\underset{\sim}{K}_j$ and $\underset{\sim}{V}_k$ are such that $\underset{\sim}{v}_{i_{(k)(j)}}' \underset{\sim}{v}_{i_{(k)(j)}} = 0$ for $i = 1,\ldots,m_j$. In this case, it will not be possible to estimate γ_k from the j^{th} data group. Let δ_{jk}^* equal 1 if γ_k can be estimated from group j and zero otherwise, i.e., if trace $(\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j) = 0$. Then if $\delta_{jk}^* = 1$ define

(5.2.15) $$\hat{\gamma}_{k_{(j)}} = c_{k_{(j)}} / h_{kk_{(j)}} , \qquad j = 1,\ldots,u, \quad k = 0,\ldots,M,$$

and

$$\hat{\gamma}_k = \frac{\sum_{j=1}^{u} \delta^*_{jk} \hat{\gamma}_{k_{(j)}}}{\sum_{j=1}^{u} \delta^*_{jk}} , \qquad k = 0,\ldots,M.$$

Because the $\hat{\gamma}_k$ were computed via Seely's method, they are unbiased for γ_k. The following theorem will establish the γ_k as consistent estimators also. Let

$$\hat{\Sigma}_M = \sum_{k=0}^{M} \hat{\gamma}_k V_k .$$

Then (3.1.2) can be used to produce a weighted least squares estimate, $\hat{B}_M$, for the GIM with a moving average error process by substituting $\hat{\Sigma}_M$ for Σ. As with the autoregressive model, the asymptotic theory for the number of time points increasing to infinity is not relevant in the context of the GIM. The following Theorem establishes $\hat{\Sigma}_M$ as a consistent estimator of Σ_M as the number of subjects in each data group, n_j, increases.

Theorem 5.2.1. The estimator $\hat{\Sigma}_M$ is a consistent estimator for Σ_M in the GIM with normality when the errors associated with each subject's responses are assumed to follow independent and identical MA(M) processes with $M \leq p - 1$.

Proof. Each $\hat{\gamma}_{k_{(j)}}$ is computed using Seely's method, thus

$$E(\hat{\gamma}_{k_{(j)}}) = \gamma_k \text{ provided } \delta^*_{jk} = 1 \text{ and } E(\hat{\gamma}_k) = \gamma_k, \qquad k = 0,\ldots,M.$$

It remains to be shown that $Var(\hat{\gamma}_k) \to 0$ as $n_j \to \infty$, $j = 1,\ldots,u$. We have, for $\delta^*_{jk} = 1$,

$$\hat{\gamma}_{k_{(j)}} = c_{k_{(j)}} / h_{kk_{(j)}}$$

where

(5.2.16) $$h_{kk_{(j)}} = \text{trace}\ (\underset{\sim}{Q}_j\ \underset{\sim}{W}_{k_{(j)}}\ \underset{\sim}{Q}_j\ \underset{\sim}{W}_{k_{(j)}})$$

$$= \text{trace}\ (\underset{\sim}{K}_j'\ \underset{\sim}{V}_k\ \underset{\sim}{K}_j \otimes \underset{\sim}{I}_{n_j} - \underset{\sim}{K}_j'\ \underset{\sim}{V}_k\ \underset{\sim}{K}_j \otimes \underset{\sim}{X}_j)^2$$

$$= \text{trace}\ [(\underset{\sim}{K}_j'\ \underset{\sim}{V}_k\ \underset{\sim}{K}_j\ \underset{\sim}{K}_j'\ \underset{\sim}{V}_k\ \underset{\sim}{K}_j \otimes \underset{\sim}{I}_{n_j}) - (\underset{\sim}{K}_j'\ \underset{\sim}{V}_k\ \underset{\sim}{K}_j\ \underset{\sim}{K}_j'\ \underset{\sim}{V}_k\ \underset{\sim}{K}_j \otimes \underset{\sim}{X}_j)]$$

$$= \text{trace}\ (\underset{\sim}{K}_j'\ \underset{\sim}{V}_k\ \underset{\sim}{K}_j\ \underset{\sim}{K}_j'\ \underset{\sim}{V}_k\ \underset{\sim}{K}_j)\ (n_j - \text{trace}(X_j))$$

and

$$\underset{\sim}{X}_j = \underset{\sim}{D}_j\ (\underset{\sim}{D}_j'\ \underset{\sim}{D}_j)^{-1}\ \underset{\sim}{D}_j' .$$

Note that $\underset{\sim}{X}_j$ is an indempotent matrix and therefore the trace (X_j) is equal to the rank (X_j), which is equal to q for all j = 1,...,u. Note also that the usual regularity conditions are assumed to hold for each data group, i.e., $\underset{\sim}{D}_j'\ \underset{\sim}{D}_j/n_j$ converges to a positive definite matrix as $n_j \to \infty$, j = 1,...,u. We have

$$h_{kk_{(j)}} = \text{trace}(\underset{\sim}{K}_j'\ \underset{\sim}{V}_k\ \underset{\sim}{K}_j\ \underset{\sim}{K}_j'\ \underset{\sim}{V}_k\ \underset{\sim}{K}_j)(n_j - q) .$$

An explicit form for the variance of $c_{k_{(j)}}$ is available since $c_{k_{(j)}}$ is a quadratic form in $\underset{\sim}{y}_j$ and by assumption, $\underset{\sim}{y}_j \sim N(\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j,\ \sum_{k=0} \gamma_k\ \underset{\sim}{W}_{k_{(j)}})$.

We have (e.g., Morrison, 1976)

(5.2.17) $$\text{Var}(c_{k_{(j)}}) = 2\ \text{trace}\ (\underset{\sim}{Q}_j\ W_{k_{(j)}}\ \underset{\sim}{Q}_j\ \sum_{i=0}^{M} \gamma_j\ W_{i_{(j)}})^2$$

$$+ 4(\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j)\ \underset{\sim}{B}_j)'\ \underset{\sim}{Q}_j\ \underset{\sim}{W}_{k_{(j)}}\ \underset{\sim}{Q}_j\ (\sum_{i=1}^{M} \gamma_i\ W_{i_{(j)}})\ \underset{\sim}{Q}_j\ \underset{\sim}{W}_{k_{(j)}}\ \underset{\sim}{Q}_j\ (\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j)\ \underset{\sim}{B}_j .$$

Noting that

$$\underset{\sim}{B}_j' (\underset{\sim}{I}_{n_h} \otimes \underset{\sim}{D}_j)' \underset{\sim}{Q}_j = \underset{\sim}{B}_j'(\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j)' \cdot$$

$$\{\underset{\sim}{I} - (\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j)[\underset{\sim}{I}_{n_j} \times \underset{\sim}{D}_j)'(\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j)]^{-1}(\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j)'\}$$

$$= B_j' [\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j)' - (\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j)']$$

$$= 0$$

we have

(5.2.18)

$$\mathrm{Var}(c_{k_{(j)}}) = 2 \text{ trace } [\sum_{i=0}^{M} \sum_{\ell=0}^{M} \gamma_i \gamma_\ell \underset{\sim}{Q}_j \underset{\sim}{W}_{k_{(j)}} \underset{\sim}{Q}_j \underset{\sim}{W}_{i_{(j)}} \underset{\sim}{Q}_j \underset{\sim}{W}_{k_{(j)}} \underset{\sim}{Q}_j \underset{\sim}{W}_{\ell_{(j)}}]$$

$$= 2 \text{ trace } \{\sum_{i=0}^{M} \sum_{\ell=0}^{M} \gamma_i \gamma_\ell [\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_i \underset{\sim}{K}_j \otimes (\underset{\sim}{I}_{n_j} - \underset{\sim}{X}_j)] \cdot$$

$$[\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_\ell K_j \otimes (\underset{\sim}{I}_{n_j} - \underset{\sim}{X}_j)]\}$$

$$= (n_j - q) \text{ trace } [\sum_{i=0}^{M} \sum_{\ell=0}^{M} \gamma_i \gamma_\ell \underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_i \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_\ell \underset{\sim}{K}_j] \cdot$$

The variance of $\hat{\gamma}_{k_{(j)}}$ is then given by

(5.2.19) $\mathrm{Var}(\hat{\gamma}_{k_{(j)}}) = \mathrm{Var}(c_{k_{(j)}}) / h^2_{kk_{(j)}}$

$$= \frac{(n_j \times q) \text{ trace } \sum_{i=0}^{M} \sum_{\ell=0}^{M} \gamma_i \gamma_\ell \underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_i \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_\ell \underset{\sim}{K}_j}{(n_j - q)^2 [\text{trace}(\underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j \underset{\sim}{K}_j' \underset{\sim}{V}_k \underset{\sim}{K}_j)]^2} .$$

For a fixed number of observation times in the j^{th} group, m_j, as $n_j \to \infty$, then clearly $\mathrm{Var}(\hat{\gamma}_{k_{(j)}}) \to 0$ and thus $\hat{\gamma}_{k_{(j)}} \overset{p}{\to} E(\hat{\gamma}_{k_{(j)}}) = \gamma_k$. Since $\hat{\gamma}_k$ is a simple average of consistent estimators of γ_k, $\hat{\gamma}_k$ is also consistent for γ_k. By Lemma 2.3.1, $\hat{\underset{\sim}{\Sigma}}_M$ is consistent for $\underset{\sim}{\Sigma}_M$.

q.e.d.

In the definition of $\hat{\underset{\sim}{B}}_M$, a simple average of the $\hat{\gamma}_{k_{(j)}}$ is used to estimate γ_k. If some of the data groups contain a small number of subjects, then it may be more reasonable to take a weighted average of the $\hat{\gamma}_{k_{(j)}}$ since the estimates based upon large n_j should be more reliable. Two possibilities would be to use $n_j / \sum_{j=1}^{u} n_j$ or $n_j^2 / \sum_{j=1}^{u} n_j^2$ as weights. It is obvious from the proof that $\hat{\gamma}_k$ will be consistent for γ_k if any positive weight is used.

As for the AR(1) model, $\hat{\underset{\sim}{B}}_M$ will be a BAN estimator if independent subsequences are used to estimate $\underset{\sim}{\Sigma}_M$ and $\underset{\sim}{B}_M$. For small sample sizes, iterative estimation is possible. The $h_{kk_{(j)}}$ are the "degrees of freedom" available for estimating $\hat{\gamma}_{k_{(j)}}$ in the j^{th} data group and will not vary from step to step, but $c_{k_{(j)}}$ is a function of the residuals. To see this note that

$$(5.2.20) \qquad \underset{\sim}{Q}_j \underset{\sim}{Y}_j = [\underset{\sim}{I}_{n_j} \otimes \underset{\sim}{D}_j (\underset{\sim}{D}_j' \underset{\sim}{D}_j)^{-1} \underset{\sim}{D}_j'] \underset{\sim}{Y}_j$$

which is just the vector of OLS residuals for $\underset{\sim}{Y}_j$. Given an estimate of $\underset{\sim}{B}_M$, WLS residuals can be computed as in (3.1.6) and used to produce refined estimates of $c_{k_{(j)}}$ and $\gamma_{k_{(j)}}$ by substituting the vector of residuals for $\underset{\sim}{Q}_j \underset{\sim}{Y}_j$ in the definition of $c_{k_{(j)}}$. This process can be repeated until the estimates converge.

The consistent estimate, $\hat{\underset{\sim}{\Sigma}}_M$, can be used to estimate the asymptotic variance matrix of $\hat{\underset{\sim}{B}}_M$ as discussed in Chapter 2. If $\hat{\underset{\sim}{b}}_M$ denotes the vector-version of $\hat{\underset{\sim}{B}}_M$ then

$$(5.2.21) \qquad \widehat{\mathrm{Var}}(\hat{\underset{\sim}{b}}_M) = [\sum_{j=1}^{u} \underset{\sim}{K}_j (\underset{\sim}{K}_j' \hat{\underset{\sim}{\Sigma}}_M \underset{\sim}{K}_j)^{-1} \underset{\sim}{K}_j' \otimes \underset{\sim}{D}_j' \underset{\sim}{D}_j]^{-1} .$$

Wald statistics can be computed from (5.2.21) to test hypotheses about the model parameters, $\underset{\sim}{B}_M$, in the GIM with a moving average error process provided that the conditions of Theorem 2.4.1 are met.

5.3 An Example

In this section estimation for the GIM assuming a moving average error structure is illustrated. For this purpose, observation vectors consisting of three responses were generated for 150 hypothetical subjects. A moving average model of order two was assumed as follows:

(5.3.1) $$u_{it} = e_{it} + 0.9\ e_{i,t-1} + 0.5\ e_{i,t-2}, \qquad i=1,\ldots,150,\ t=1,2,3,$$

where $e_{it} \sim N(0, .25)$

and $E(e_{it}\ e_{is}) = 0.$

The population variance components can be computed as follows:

(5.3.2) $$\gamma_0 = Var(u_{it}) = 0.25\ (1 + 0.9^2 + 0.5^2) = 0.515$$

$$\gamma_1 = Cov(u_{it}, u_{i,t-1}) = 0.25\ (0.9 + 0.9(0.5)) = 0.3375$$

$$\gamma_2 = Cov(u_{it}, u_{i,t-2}) = 0.25\ (0.5) = 0.125$$

Descriptive statistics for the generated errors are given in Table 5.3.1.

Responses at the three time points were generated according to the linear model

$$\underset{\sim}{Y} = [\underset{\sim}{y}_1\ \underset{\sim}{y}_2\ \underset{\sim}{y}_3] = \underset{\sim}{X}\ \underset{\sim}{B} + \underset{\sim}{u}$$

where $\underset{\sim}{X} = [\underset{\sim}{X}_1\ \underset{\sim}{X}_2]$ is the design matrix, $\underset{\sim}{X}_1 = \underset{\sim}{1}$, $\underset{\sim}{X}_2$ consists of elements ranging from 10 to 50, and

$$\underset{\sim}{B} = \begin{bmatrix} 1.0 & 2.0 & 3.0 \\ 0.2 & 0.8 & 1.0 \end{bmatrix}.$$

The model parameters can be interpreted as an intercept and a slope (with respect to X_2) at each time point.

Table 5.3.1 Descriptive Statistics for Errors at Three Time Points Generated from an MA(2) Process with $a_1 = 0.9$ and $a_2 = 0.5$

	N	Mean	Std. Deviation
u_1	150	0.021	0.726
u_2	150	0.001	0.693
u_3	150	0.029	0.743

Corr \ cov	u_1	u_2	u_3
u_1	0.5274	0.3232	0.0900
u_2	0.6419	0.4806	0.3309
u_3	0.1668	0.6422	0.5523

The variance of each row of $\underset{\sim}{Y}$ has the form specified in (5.2.6), i.e.,

$$\underset{\sim}{\Sigma} = \sum_{k=0}^{M} \gamma_k \underset{\sim}{V}_k$$

where

$$\underset{\sim}{V}_0 = \underset{\sim}{I}_3 ,$$

$$\underset{\sim}{V}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} ,$$

$$\underset{\sim}{V}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} .$$

For computational simplicity, only two patterns of missing data were introduced. The responses at time t = 2 were deleted in one-third of the observations, and the responses at time t = 3 were deleted in another third of the observations. The GIM parameters are given by

$$u = 3,$$

$$n_j = 50 \;,\; j=1,\ldots,3 \;,$$

$$\underset{\sim}{K}_1 = \underset{\sim}{I}_3 ,$$

$$\underset{\sim}{K}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} ,$$

$$\underset{\sim}{K}_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} .$$

It is clear from the definition of $\underset{\sim}{K}_j$, j=1,2,3, that γ_1 can be estimated from data groups 1 and 3 only, and that γ_2 can be estimated from groups 1 and 2 only. A software package written to compute Seely's estimates of the variance components for the vector-version of the GIM was used to produce estimates of $\underset{\sim}{\gamma}$, pooled across the data groups. We have

$$\hat{\gamma}_0 = 0.5497 ,$$
$$\hat{\gamma}_1 = 0.3442 ,$$
$$\hat{\gamma}_2 = 0.1720 .$$

The weighted least squares estimate of $\underset{\sim}{B}_M$ was computed with the GIM software substituting $\hat{\underset{\sim}{\Sigma}}_M = \underset{\sim}{\Sigma}_M(\hat{\underset{\sim}{\gamma}})$. These estimates and their estimated standard errors are given in Table 5.3.2.

For comparison purposes, a standard GLMM analysis was performed on the complete dataset prior to deleting responses. The results of this analysis are given in Table 5.3.3. Both the population parameters and the complete data estimates are within 95% confidence intervals of the two-stage WLS estimator proposed herein. As with the AR(1) model, use of the entire sample for estimation of $\underset{\sim}{\Sigma}_M$ and $\underset{\sim}{B}_M$ produced reasonable results.

5.4 An Alternate Estimator

In cases of particularly sparse data, computation of $\hat{\underset{\sim}{B}}_M$ as proposed in section 5.2 may not be possible. The estimation of $\hat{\gamma}_{k_{(j)}}$ requires the computation of an inverse for the matrix $\underset{\sim}{K}_j' \otimes \underset{\sim}{D}_j$. If n_j is small this matrix may be less than full rank. In the extreme case of unique times for each subject ($n_j = 1$ for $j = 1,\ldots,u$), this estimation process is not practical. In this section an estimation is introduced that involves pooling information from all subjects in the computation of h_{kk} and c_k before computing $\hat{\gamma}_k$, thereby eliminating the problem of small sample sizes among the missing data groups. Let

$$\underset{\sim}{Y} = \begin{bmatrix} \underset{\sim}{Y}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{Y}_u \end{bmatrix} , \tag{5.4.1}$$

Table 5.3.2 WLS Estimates for the GIM with Moving Average Errors

$\hat{\underset{\sim}{B}}_M$:	1.062	1.991	2.923
	0.199	0.800	1.004
$SE(\hat{\underset{\sim}{B}}_M)$:	0.169	0.186	0.197
	0.005	0.006	0.006

Table 5.3.3 GLMM Estimates for the Complete Dataset

$\hat{\underset{\sim}{B}}$:	1.062	2.030	3.006
$SE(\hat{\underset{\sim}{B}})$:	0.166	0.158	0.170
	0.005	0.005	0.005

$$\underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{K}_1' & x & \underset{\sim}{D}_1 \\ & \vdots & \\ \underset{\sim}{K}_u' & x & \underset{\sim}{D}_u \end{bmatrix} , \qquad \underset{\sim}{W}_k = \begin{bmatrix} \underset{\sim}{W}_{k_{(j)}} & & \underset{\sim}{0} \\ & \ddots & \\ \underset{\sim}{0} & & \underset{\sim}{W}_{k_{(u)}} \end{bmatrix} .$$

Then $\underset{\sim}{y}$ satisfies the following model:

(5.4.2) $$E(\underset{\sim}{y}) = \underset{\sim}{X}\,\underset{\sim}{b}$$

$$\mathrm{Var}(\underset{\sim}{y}) = \sum_{k=0}^{M} \gamma_k \underset{\sim}{W}_k$$

where $\underset{\sim}{b}$ is the vector-version of $\underset{\sim}{B}$ and $\underset{\sim}{K}_j$, $\underset{\sim}{D}_j$, $\underset{\sim}{y}_j$, and $\underset{\sim}{W}_{k_{(j)}}$ are as specified in (5.2.3) - (5.2.9). Seely's method of variance component estimation can be used to estimate the γ_k. For this model, the elements of $\underset{\sim}{c}$ in (5.2.7) are given by

(5.4.3) $$c_k = \underset{\sim}{y}' \,(\underset{\sim}{I}_N - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\,\underset{\sim}{X}')\,\underset{\sim}{W}_k(\underset{\sim}{I}_N - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\,\underset{\sim}{X}')\,\underset{\sim}{y}$$

where $N = \sum_{j=1}^{u} n_j m_j$. The matrices in this expression can become quite large. However, some simplification is possible. Let $\hat{\underset{\sim}{B}}_0$ denote the OLS estimate of $\underset{\sim}{B}$ and $\hat{\underset{\sim}{y}}$, the OLS predicted values, i.e.,

$$\hat{y} = \underset{\sim}{X}\,\hat{\underset{\sim}{B}}_0 = \begin{bmatrix} \hat{\underset{\sim}{y}}_1 \\ \vdots \\ \hat{\underset{\sim}{y}}_u \end{bmatrix} .$$

Then

(5.4.4) $$c_k = \underset{\sim}{y}'\,\underset{\sim}{W}_k\,\underset{\sim}{y} - \underset{\sim}{y}'\,\underset{\sim}{W}_k\,\hat{\underset{\sim}{y}} - \hat{\underset{\sim}{y}}'\,\underset{\sim}{W}_k\,\underset{\sim}{y} + \hat{\underset{\sim}{y}}'\,\underset{\sim}{W}_k\,\hat{\underset{\sim}{y}}$$

$$= \sum_{j=1}^{u} \left[\underset{\sim}{y}_j'\,\underset{\sim}{W}_{k_{(j)}}\,\underset{\sim}{y}_j - \underset{\sim}{y}_j'\,\underset{\sim}{W}_{k_{(j)}}\,\hat{\underset{\sim}{y}}_j - \hat{\underset{\sim}{y}}_j'\,\underset{\sim}{W}_{k_{(j)}}\,\underset{\sim}{y}_j' + \hat{\underset{\sim}{y}}_j'\,\underset{\sim}{W}_{k_{(j)}}\,\hat{\underset{\sim}{y}}_j\right] .$$

The elements of the matrix $\underset{\sim}{H}$ in (5.2.7) are given by

$$h_{k\ell} = \text{trace}\ (\underset{\sim}{Q}\ \underset{\sim}{W}_k\ \underset{\sim}{Q}\ \underset{\sim}{W}_\ell) \tag{5.4.5}$$

where $$\underset{\sim}{Q} = \underset{\sim}{I}_N - \underset{\sim}{X}\ (\underset{\sim}{X}'\ \underset{\sim}{X})^{-1}\ \underset{\sim}{X}'\ .$$

Due to the structure of $\underset{\sim}{W}_k$, the argument given in section 5.2 showing that $h_{k\ell_{(j)}} = 0$ for $k \neq \ell$ applies to $h_{k\ell}$ in (5.4.5). The diagonal elements h_{kk}, $k = 0,\ldots,M$, can be expressed in terms of the u data groups to lessen the computational burden.

Note that

$$\begin{aligned} \underset{\sim}{Q} &= \underset{\sim}{I}_N - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\ \underset{\sim}{X}' \\ &= \underset{\sim}{I}_N - \begin{bmatrix} X_1(X'X)^{-1}X_1' & X_1(X'X)^{-1}\ X_2' & \cdots & \underset{\sim}{X}_1(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}_u' \\ \underset{\sim}{X}_2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\ \underset{\sim}{X}_1' & & & \vdots \\ \vdots & & & \vdots \\ \underset{\sim}{X}_u(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}_1' & \cdot & \cdot\ \cdot & \underset{\sim}{X}_u(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}_u' \end{bmatrix} \end{aligned} \tag{5.4.6}$$

where $\underset{\sim}{X}_j = \underset{\sim}{K}_j' \times \underset{\sim}{D}_j$, $j=1,\ldots,u,$ and

$$\begin{aligned} (\underset{\sim}{X}'\underset{\sim}{X})^{-1} &= \left[\sum_{j=1}^{u} \underset{\sim}{X}_j'\ \underset{\sim}{X}_j\right]^{-1} \\ &= \left[\sum_{j=1}^{u} \underset{\sim}{K}_j\ \underset{\sim}{K}_j' \otimes \underset{\sim}{D}_j'\ \underset{\sim}{D}_j\right]^{-1} \\ &= \left[\sum_{j=1}^{u} \underset{\sim}{K}_j\ \underset{\sim}{K}_j'\right]^{-1} \otimes \left[\sum_{j=1}^{u} \underset{\sim}{D}_j'\ \underset{\sim}{D}_j\right]^{-1} \end{aligned} \tag{5.4.7}$$

Expanding (5.4.5) yields

$$\begin{aligned} h_{kk} = \text{trace}\ [&\underset{\sim}{W}_k\ \underset{\sim}{W}_k - \underset{\sim}{W}_k\ \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\ \underset{\sim}{X}'\underset{\sim}{W}_k - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\ \underset{\sim}{X}'\ \underset{\sim}{W}_k\ \underset{\sim}{W}_k \\ &+ \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\ \underset{\sim}{X}'\ \underset{\sim}{W}_k\ \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\ \underset{\sim}{X}'\ \underset{\sim}{W}_k]\ . \end{aligned} \tag{5.4.8}$$

Substituting the expression for $\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$ in (5.4.6) into (5.4.8), after much simplifying, we get

$$(5.4.9) \quad h_{kk} = \sum_{j=1}^{u} \text{trace}\, \underset{\sim}{W}_{k_{(j)}} \underset{\sim}{W}_{k_{(j)}}$$

$$- \sum_{j=1}^{u} \text{trace}(\underset{\sim}{X}_j(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}_j' \underset{\sim}{W}_{k_{(j)}} \underset{\sim}{W}_{k_{(j)}})$$

$$- \sum_{j=1}^{u} \text{trace}(\underset{\sim}{W}_{k_{(j)}} \underset{\sim}{X}_j(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}_j' \underset{\sim}{W}_{k_{(j)}} \underset{\sim}{W}_{k_{(j)}})$$

$$+ \sum_{i=1}^{u} \sum_{j=1}^{u} \text{trace}(\underset{\sim}{X}_i(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}_j' W_{k_{(j)}} \underset{\sim}{X}_j(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}_i' \underset{\sim}{W}_{k_{(i)}}) .$$

Estimates of γ_k are then given by

$$(5.4.10) \quad \hat{\gamma}_k = c_k / h_{kk} \quad , \quad k = 0,\ldots,M.$$

The computation of these estimates is independent of n_j, $j=1,\ldots,u$. Repeating the argument given in the proof of Theorem 5.2.1 it is easily seen from the above expression that h_{kk} and c_k are of the order $(N - q)$ and therefore the variance of $\hat{\gamma}_k$ is of the order $1/(N - q)$. As the number of subjects increases to infinity for a fixed number of time points, $\hat{\gamma}_k$ converges in probability to γ_k, $k = 0,\ldots,M$. Thus this estimation procedure provides an alternative to that presented in section 5.2. However, it should be noted that the burden of computing $\hat{\gamma}_k$ is much greater than that of computing $\hat{\gamma}_{k_{(j)}}$ for $j = 1,\ldots,u$. Due to the form of the h_{kk} it is impossible to process one data group at a time. Direct computation using the matrices defined in (5.4.3) and (5.4.5) will typically require a large amount of core for storage, while accumulation of the various components over the u data groups in (5.4.4) and (5.4.9) will require a large amount of computing time. Thus if sample sizes permit, the estimation method given in section 5.2 is preferred. The data analysis presented in Chapter 7 is an example of a study design for which the techniques presented in this section are required.

CHAPTER 6

MULTIPLE RESPONSE VARIABLES

6.1 Introduction and Model Formulation

It is frequently the case in longitudinal studies that more than one measurement is taken on a subject at a given test date. The measurements may all be related to some biological or psychological process and therefore highly correlated. In this chapter, linear model analysis of incomplete longitudinal data is considered in which multiple dependent variables are measured through time on each subject. We have seen in Chapters 4 and 5 that the assumption of a time series model for the error terms with respect to a single response variable is not only a reasonable assumption but also provides a method of analyzing particularly sparse data. The logical extension of those estimation methods is to combine several time series models on each subject in such a way that the estimation of linear model parameters takes into account both the correlations across time for a given response variable and the correlations between two response variables measured on a given subject. It will be seen that the assumption of the particular time series models selected for each response variable necessarily determines the between-response-variable correlation structure through time.

First the GIM notation must be extended. Assume that v dependent variables are measured at p time points on n subjects in the hypothetical complete data case. The model for the k^{th} response variable, $k = 1,\ldots,v$, is assumed to be the GLMM ($\underset{\sim}{Y}_k$; $\underset{\sim}{D}\ \underset{\sim}{B}_k$, $\underset{\sim}{\Sigma}_{kk}$). Note that the same design matrix, $\underset{\sim}{D}$, is assumed for each dependent variable for simplicity. Observations on the v dependent variables can be combined as follows:

$$(6.1.1)\qquad [\underset{\sim}{Y}_1\ \underset{\sim}{Y}_2 \cdots \underset{\sim}{Y}_v] = \underset{\sim}{D}\ [\underset{\sim}{B}_1\ \underset{\sim}{B}_2 \cdots \underset{\sim}{B}_v] + [\underset{\sim}{U}_1\ \underset{\sim}{U}_2 \cdots \underset{\sim}{U}_v]$$

where $\underset{\sim}{U}_k$ denotes the matrix of error terms for the k^{th} variable. The n rows of the matrix on the left are assumed to be independent with covariance structure given by

$$(6.1.2)\qquad \mathrm{Var}(i^{th} \text{ row of } [\underset{\sim}{Y}_1 \cdots \underset{\sim}{Y}_v] = \underset{\sim}{\Omega}$$

$$= \begin{bmatrix} \underset{\sim}{\Sigma}_{11} & \underset{\sim}{\Sigma}_{12} & \cdots & \underset{\sim}{\Sigma}_{1v} \\ \underset{\sim}{\Sigma}'_{12} & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \underset{\sim}{\Sigma}'_{1v} & & & \underset{\sim}{\Sigma}_{vv} \end{bmatrix}.$$

The (p x p) matrix $\underset{\sim}{\Sigma}_{k\ell}$ consists of the covariance parameters for the k^{th} and ℓ^{th} dependent variables across time. For example, if $\underset{\sim}{\Sigma}_{k\ell}$ = $((\sigma_{rs}))_{k\ell}$ and $\underset{\sim}{Y}_k = ((y_{it}))_k$ then

$$(6.1.3)\qquad \mathrm{cov}(y_{it_{(k)}}, y_{it'_{(\ell)}}) = \sigma_{tt'_{(k\ell)}} .$$

To incorporate missing data in this model, additional notation must be specified. As in the GIM, it is assumed that the subjects are grouped according to missing data patterns with respect to the v response variables. Let $\underset{\sim}{Y}_{jk}$ denote the (n_j x m_{jk}) matrix of observations on the k^{th} dependent variable for the j^{th} missing data group, k = 1,...,v, j = 1,...,u. Let $\underset{\sim}{y}_j$ denote the vector version (rolled out by columns) of $[\underset{\sim}{Y}_{j1}\ \underset{\sim}{Y}_{j2} \cdots \underset{\sim}{Y}_{jv}]$. Then the following model can be defined for $\underset{\sim}{y}_j$:

(6.1.4) $$E(\underset{\sim}{y}_j) \equiv \underset{\sim}{Z}_j\, \underset{\sim}{b} = \begin{bmatrix} \underset{\sim}{K}'_{j1} \otimes \underset{\sim}{D} \\ \vdots \\ \underset{\sim}{K}'_{jv} \otimes \underset{\sim}{D} \end{bmatrix} \begin{bmatrix} \underset{\sim}{b}_1 \\ \vdots \\ \underset{\sim}{b}_v \end{bmatrix}$$

and

(6.1.5) $$\mathrm{Var}(\underset{\sim}{y}_j) \equiv \underset{\sim}{\Omega}_j$$

$$= \begin{bmatrix} \underset{\sim}{K}'_{j1}\ \underset{\sim}{\Sigma}_{11}\ \underset{\sim}{K}_{j1} \otimes \underset{\sim}{I}_{n_j} & \cdots & \underset{\sim}{K}'_{j1}\ \underset{\sim}{\Sigma}_{1v}\ \underset{\sim}{K}_{jv} \otimes \underset{\sim}{I}_{n_j} \\ \vdots & & \\ \underset{\sim}{K}'_{jv}\ \underset{\sim}{\Sigma}'_{1v}\ \underset{\sim}{K}_{j1} \otimes \underset{\sim}{I}_{n_j} & \cdots & \underset{\sim}{K}'_{jv}\ \underset{\sim}{\Sigma}_{vv}\ K_{jv} \otimes \underset{\sim}{I}_{n_j} \end{bmatrix},$$

where

- $\underset{\sim}{b}_k$ denotes the (pq x 1) vector of columns of $\underset{\sim}{B}_k$,
- $\underset{\sim}{D}_j$ denotes the (nj x q) design matrix for the j^{th} group,
- $\underset{\sim}{K}_{jk}$ denotes the (p x m_{jk}) matrix indicating the missing data pattern for the j^{th} data group with respect to the k^{th} dependent variable, j = 1,...,u, k = 1,...,v.

This model notation simplifies somewhat if the only cause for missing data is a missed interview or examination in which case all response variables are missing for that time point. Then $\underset{\sim}{K}_{jk} = \underset{\sim}{K}_j$ for k = 1,...,v.

If an unconstrained formulation is assumed for $\underset{\sim}{\Sigma}_{k\ell}$, k,ℓ = 1,...,v, the estimation of $\underset{\sim}{b}$ requires the estimation of approximately $v^2p(p + 1)/2$ variance and covariance parameters, with the exact number depending on the missing data patterns. As in the GIM, pairwise deletion techniques could be employed based on the OLS residuals to consistently estimate these parameters. However, if v or p is large relative to n and there is missing data, it may be necessary to impose constraints on the structure of $\underset{\sim}{\Sigma}_{k\ell}$ in order to guarantee the estimability of all parameters involved. If, for k,ℓ = 1,...,v, $\underset{\sim}{\Sigma}_{k\ell} = \underset{\sim}{\Sigma}_{k\ell}\,(\underset{\sim}{\phi}_{k\ell})$ such that the elements of $\underset{\sim}{\Sigma}_{k\ell}$ are continuous functions of the elements of $\underset{\sim}{\phi}_{k\ell}$, then by Lemma

2.3.1, if consistent estimates of $\phi_{k\ell}$ exist, they can be substituted in (6.1.5) to produce a consistent estimate of Ω_j.

Clearly the model defined in (6.1.4) and (6.1.5) satisfies the assumptions of the LMI. If consistent estimates of $\phi_{k\ell}$ can be computed from one subsequence of the subjects, then estimates of b computed from an independent subsequence will be BAN estimates by Theorem 2.3.2. For small sample sizes it may be more desirable to compute WLS estimates of b from the entire sample. The estimator is defined as

$$(6.1.6) \qquad \hat{b} = \left[\sum_{j=1}^{u} Z_j' \hat{\Omega}_j^{-1} Z_j \right]^{-1} \sum_{j=1}^{u} Z_j' \hat{\Omega}_j^{-1} y_j$$

where $\hat{\Omega}_j$ is a consistent estimator of Ω_j.

An estimate of the variance-covariance matrix of $\hat{b}$ is given by

$$(6.1.7) \qquad \hat{\text{Var}}\,(\hat{b}) = \sum_{j=1}^{u} Z_j' \hat{\Omega}^{-1} Z_j \; .$$

As in the GIM, Wald statistics can be computed to test hypotheses about the model parameters b based on this estimate of the variance-covariance matrix for $\hat{b}$. If the conditions of Theorem 2.4.1 are met, the Wald statistics will follow an asymptotic chi-square distribution.

6.2 Multiple Autoregressive Models

We have seen in Chapter 4 that an autoregressive process of order one is often a reasonable model for a subject's series of residuals about the population regression line. Furthermore, consistent estimation of the process parameters is possible in the presence of missing data. In this section, a model is proposed that assumes an AR(1) process for the errors of each response variable observed on a subject and includes between response variable correlations across time.

Referring to the model specified in (6.1.1) for the complete data case, let $\underset{\sim}{u}_{i_{(k)}}$ denote the vector of errors for the i^{th} subject's measurements on the k^{th} response variable. If $\underset{\sim}{u}_{i_{(k)}} \sim AR(1)$ then for $k = 1,\ldots,v$,

(6.2.1) $$u_{it_{(k)}} = \rho_k \; u_{i,t-1_{(k)}} + e_{it_{(k)}} \quad , \qquad \begin{array}{l} i = 1,\ldots,n, \\ t = 2,\ldots,p, \end{array}$$

where

(6.2.2) $$e_{it_{(k)}} \sim N(0,\phi_{kk}),$$

$$\mathrm{cov}(e_{it_{(k)}}, e_{it'_{(k)}}) = 0 \quad , \qquad t \neq t'.$$

Assuming the following initial condition,

(6.2.3) $$\mathrm{var}(u_{i0_{(k)}}) = \phi_{kk} \,/\, (1 - \rho_k^2) \quad , \qquad i = 1,\ldots,n,$$

it was shown in section 4.1 that

(6.2.4) $$\mathrm{Var}(\underset{\sim}{u}_{i_{(k)}}) = \frac{\phi_{kk}}{1 - \rho_k^2} \begin{bmatrix} 1 & \rho_k & \cdots & \rho_k^{p-1} \\ \rho_k & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \rho_k^{p-1} & & \cdots & 1 \end{bmatrix} , \qquad i = 1,\ldots,n.$$

In order to incorporate between response variable correlations into the model and at the same time minimize the number of covariance parameters that must be estimated, the following additional assumptions are made:

(6.2.5) (i) $$\mathrm{cov}(e_{it_{(k)}}, e_{is_{(\ell)}}) = \begin{cases} \phi_{k\ell} & \text{if } t = s, \\ 0 & \text{if } t \neq s, \end{cases} \qquad \begin{array}{l} i = 1,\ldots,n, \\ k,\ell = 1,\ldots,v, \\ t,s = 1,\ldots,p. \end{array}$$

(ii) $$\mathrm{cov}(u_{i0_{(k)}}, u_{i0_{(\ell)}}) = \phi_{k\ell} \,/\, (1 - \rho_k \, \rho_\ell) \; , \qquad \begin{array}{l} i = 1,\ldots,n, \\ k,\ell = 1,\ldots,v. \end{array}$$

For a fixed time $t > 0$, the covariance between any two response variables is then given by

(6.2.6)

$$\operatorname{cov}(u_{it_{(k)}}, u_{it_{(\ell)}}) = E[(\rho_k u_{i,t-1_{(k)}} + e_{it_{(k)}})(\rho_\ell u_{i,t-1_{(\ell)}} + e_{it_{(\ell)}})], \quad i = 1,\ldots,n, \quad k,\ell = 1,\ldots,v.$$

By assumption (i) above, $u_{i,t-1_{(k)}}$ is independent of $e_{it_{(\ell)}}$, therefore

(6.2.7)

$$\operatorname{cov}(u_{it_{(k)}}, u_{it_{(\ell)}}) = \phi_{k\ell} (1 + \rho_k\rho_\ell + E(\rho_k^2 u_{i,t-2_{(k)}} \rho_\ell^2 u_{i,t-2_{(\ell)}})) .$$

By successively substituting in this manner for the $u_{i,t-2}$ terms, the right hand side becomes the expansion of the series for $\phi_{k\ell}/(1 - \rho_k \rho_\ell)$, thus giving the rationale for assumption (ii) in (6.2.5). Note that the covariances between response variables at the same time point are assumed to be constant over time.

The covariance between two response variables at different times can be derived in a similar manner. For $s > t$,

(6.2.8)
$$\operatorname{cov}(u_{it_{(k)}}, u_{is_{(\ell)}}) = E[u_{it_{(k)}} (\rho_\ell u_{i,s-1_{(\ell)}} + e_{is_{(\ell)}})]$$
$$= E[\rho_\ell u_{it_{(k)}} u_{i,s-1_{(\ell)}}] .$$

If $s - 1 > t$ another substitution can be made for $u_{i,s-1_{(\ell)}}$. Noting that $e_{i,s-1_{(\ell)}}$ is independent of $u_{it_{(k)}}$,

(6.2.9)
$$\operatorname{cov}(u_{it_{(k)}}, u_{is_{(\ell)}}) = E(\rho_\ell^2 u_{it_{(k)}} u_{i,s-2_{(\ell)}}).$$

If $s - m = t$ for some positive integer m, then exactly m substitutions can be made for $u_{is_{(\ell)}}$. The general form for the covariance is then given by

$$(6.2.10)\quad \operatorname{cov}(u_{it_{(k)}}, u_{is_{(\ell)}}) = \begin{cases} \rho_\ell^m \phi_{k\ell} / (1 - \rho_k \rho_\ell), & s = t + m, \\ \rho_k^m \phi_{k\ell} / (1 - \rho_k \rho_\ell), & t = s + m, \end{cases}$$

$$i = 1,\ldots,n, \quad k,\ell = 1,\ldots,v, \quad m > 0.$$

This implies that the covariance between response variables decreases with time at the same rate as the covariance for measurements on a single variable.

In terms of the matrix notation in (6.1.2),

$$(6.2.11)\quad \underset{\sim}{\Sigma}_{k\ell} = \underset{\sim}{\Sigma}_{k\ell}(\rho_k, \rho_\ell, \phi_{k\ell})$$

$$= \frac{\phi_{k\ell}}{1-\rho_k\rho_\ell} \cdot \begin{bmatrix} 1 & \rho^\ell & \rho_\ell^2 & \cdots & \rho_\ell^{p-1} \\ \rho_k & 1 & \rho_\ell & \cdots & \rho^{p-2} \\ \rho_k^2 & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \rho_k^{p-1} & \cdots & & & 1 \end{bmatrix}, \quad k,\ell = 1,\ldots,v.$$

A consistent estimate of $\underset{\sim}{\Sigma}_{k\ell}$ can be produced if consistent estimators of ρ_k, ρ_ℓ, and $\phi_{k\ell}/(1 - \rho_k\rho_\ell)$ are available. The k^{th} autocorrelation parameter can be estimated from the data on the k^{th} response variable by the formula given in step (v) of (4.2.8), since that data satisfies the assumptions of a GIM with an AR(1) error process. Let $\hat{\rho}_k$ denote this estimate. Pairwise deletion techniques can then be used to estimate

$$(6.2.12)\quad \sigma_{tt_{(k\ell)}} = \phi_{k\ell}/(1 - \rho_k\rho_\ell), \qquad t = 1,\ldots,p, \quad k,\ell = 1,\ldots,v.$$

For the j^{th} data group, let

$$(6.2.13)\quad \underset{\sim}{R}_{jk} = \underset{\sim}{Y}_{jk} - \underset{\sim}{D}_j \hat{\underset{\sim}{B}}_k^{(0)} \underset{\sim}{K}_j$$

where $\hat{\underset{\sim}{B}}_k^{(0)}$ denotes the OLS estimate of $\underset{\sim}{B}_k$. Let

$$(6.2.14) \qquad w_{k\ell} = \sum_{j=1}^{u} \sum_{i=1}^{n_j} \underset{\sim}{r}'_{i(jk)} \underset{\sim}{K}'_{j(k)} \underset{\sim}{K}_{j(\ell)} \underset{\sim}{r}_{i(j\ell)} / N_{k\ell}$$

where $\underset{\sim}{r}'_{i(jk)}$ denotes the i^{th} row of $\underset{\sim}{R}_{jk}$ and $\underset{\sim}{N}_{k\ell}$ denotes the number of pairs of observations on the k^{th} and ℓ^{th} response variable taken at the same time. Then $w_{k\ell}$ is the sum over time $(t = 1,\ldots,p)$ of pairwise deletion estimators of $\sigma_{tt(k\ell)}$. As the number of subjects on which the k^{th} and ℓ^{th} response variable were observed at the same time increase, $w_{k\ell}$ is consistent for $\phi_{k\ell} / (1 - \rho_k \rho_\ell)$. Substituting $\hat{\underset{\sim}{\Sigma}}_{k\ell} = \underset{\sim}{\Sigma}_{k\ell} (\hat{\rho}_k, \hat{\rho}_\ell, w_{k\ell})$ for $\underset{\sim}{\Sigma}_{k\ell}$ in (6.1.6) yields WLS estimates for $\underset{\sim}{b}$ when multiple AR(1) error models are assumed.

6.3 Multiple Moving Average Models

As discussed in Chapter 5, the assumption of a moving average process is often reasonable when modeling the errors on a subject's observations with respect to a single response variable. When multiple moving average error processes are considered jointly in a model with multiple response variables, the variance component estimation procedure presented in section 5.2 is not feasible due to the complex covariance structures involved. In this section a model is proposed that assumes an MA(1) process for the errors of each response variable observed on a subject and includes between response variable correlations. Consistent estimators for the covariance parameters are presented. Results for multiple higher order moving average processes have not been readily attainable.

Let $\underset{\sim}{u}_{i(k)}$ denote the vector of errors for the i^{th} subject's responses on the k^{th} dependent variable. If $\underset{\sim}{u}_{i(k)} \sim \text{MA}(1)$ then

(6.3.1) $$u_{it_{(k)}} = e_{it_{(k)}} + a_k\, e_{i,t-1_{(k)}} \quad , \quad i = 1,\ldots,n,\; k = 1,\ldots,v,\; t \in (-\infty, \infty).$$

It is usually assumed that $|a_k| < 1$. This guarantees that $u_{it_{(k)}}$ is expressible as a convergent infinite series in $u_{i,t-m_{(k)}}$, $m = 1,2,\ldots,\infty$ (e.g., Box and Jenkins, 1976). Assuming that $e_{it_{(k)}} \sim N(0, \phi_{kk})$ and the $e_{it_{(k)}}$ are independent, we have from (5.2.2) that

(6.3.2) $$\mathrm{var}(u_{it_{(k)}}) = \phi_{kk}\,(1 + a_k^2)$$

and

$$\mathrm{cov}(u_{it_{(k)}}, u_{is_{(k)}}) = \begin{cases} a_k\, \phi_{kk} & \text{if } |t - s| < 1, \\ 0 & \text{otherwise.} \end{cases}$$

In order to incorporate between response variable correlations the following additional assumption is made:

(6.3.3) $$\mathrm{cov}(e_{it_{(k)}}, e_{is_{(\ell)}}) = \begin{cases} \phi_{k\ell} & \text{if } t = s, \\ 0 & \text{if } t \neq s. \end{cases}$$

Then the following relationships can be derived:

(6.3.4) $$\mathrm{cov}\ (u_{it_{(k)}}, u_{it_{(\ell)}}) = \phi_{k\ell}\,(1 + a_k\, a_\ell),$$

$$\mathrm{cov}\ (u_{it_{(k)}}, u_{i,t-1_{(\ell)}}) = \phi_{k\ell}\, a_k \quad ,$$

$$\mathrm{cov}\ (u_{i,t-1_{(k)}}, u_{it_{(\ell)}}) = \phi_{k\ell}\, a_\ell,$$

$$\mathrm{cov}\ (u_{it_{(k)}}, u_{is_{(\ell)}}) = 0 \quad , \quad |t - s| > 1,$$

In terms of the parameter matrices defining $\underset{\sim}{\Omega}$ in (6.1.2), we have

(6.3.5) $$\underset{\sim}{\Sigma}_{k\ell} = \phi_{k\ell} \cdot \begin{bmatrix} 1 + a_k a_\ell & a_\ell & 0 & \ldots & 0 \\ a_k & 1 + a_k a_\ell & a_\ell & \ldots & 0 \\ 0 & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & & \ldots & & 1 + a_k a_\ell \end{bmatrix} .$$

Under this model, the first order autocorrelation is given by

(6.3.6) $$\rho_k = a_k/(1 + a_k^2), \quad k = 1,\ldots,v.$$

The restriction that $|a_k| < 1$ implies that $|\rho_k| < 0.5$. An estimate of ρ_k can be computed from data on the k^{th} response variable by the formula given in step (v) of (4.2.8). This estimate is consistent and, provided $|\hat{\rho}_k| < 0.5$, can be used to produce a consistent estimate of a_k as follows:

(6.3.7) $$\hat{a}_k = \frac{1 \pm \sqrt{1 - 4\hat{\rho}_k^2}}{2\rho_k}, \quad k = 1,\ldots,v.$$

The estimate w_k defined in (6.2.14) is a consistent estimate of $\phi_{k\ell}/(1 + a_k a_\ell)$, $k,\ell = 1,\ldots,v$, under this model as well as the AR(1) model. Thus $\hat{\underset{\sim}{\Sigma}}_{k\ell} = \underset{\sim}{\Sigma}_{k\ell}(\hat{a}_k, \hat{a}_\ell, w_{k\ell})$ is consistent for $\underset{\sim}{\Sigma}_{k\ell}$ defined in (6.3.5). Substituting $\hat{\underset{\sim}{\Sigma}}_{k\ell}$ in (6.1.6) yields a WLS estimate of $\underset{\sim}{b}$ for models with multiple MA(1) error processes. If $|\rho_k| > 0.5$, no real solution exists for a_k by this method.

CHAPTER 7

THE ANALYSIS OF A LONGITUDINAL STUDY OF SPIROMETRY IN CHILDREN

7.1 Description of the Data

Spirometric measurements were taken on young black and white children over a period of eight years as part of a longitudinal study conducted at the Frank Porter Graham Child Development Center, UNC, Chapel Hill, N.C. Subjects for the study were selected prior to birth from expectant parents living in the area. Six to 12 subjects were recruited per year. The children were trained to perform forced vital capacity (FVC) maneuvers as early as two and one-half years of age. At the satisfactory conclusion of the training period, spirometry was planned to be performed on each subject at set time intervals. Three of the measures obtained from spirometry are forced vital capacity (FVC), peak expiratory flow rate (PEF), and the maximum expiratory flow rate after 75% of the FVC has been exhaled ($\dot{V}max_{75\%}$). The FVC is measured in liters and the flows in liters per second.

The dataset available for analysis at the time of this research consisted of spirometric measurements on 72 children collected from Nov. 1, 1972 through Dec. 31, 1980. The subjects ranged in age from three to 12 years and consisted of 29 black females, 26 black males, nine white females, and eight white males. The data had been carefully examined by a pulmonary physiologist and only technically acceptable spirograms from non-obese, non-asthmatic children were included.

7.2 Analysis Strategies

The analysis of the spirometry data was performed using the method of estimation and hypothesis testing proposed in this research in order to illustrate the application of same to an incomplete longitudinal dataset. Of primary interest to the principal investigators was to predict population values of each spirometric measure for the four race and sex groups, to determine if significant growth occurred with respect to these measurements, and to determine if any significant differences existed between the groups.

The data collection was fairly haphazard with respect to calendar time. As a result, the data for each subject at each year of age from three to 12 years were used for this analysis. If several examinations were conducted on a subject in one year, only the examination occurring nearest to a birthday was included.

As described above, children were continually entering the study by design. In addition the study was subject to a fair amount of attrition due to families moving from the area. Once in the study, most children had examinations at least once a year until they left. However in several cases, entire years were missed due to illness or other causes of absence. For these reasons the analysis was complicated by a large amount of missing data. For one spirometric measure the hypothetical complete data matrix would consist of 10 yearly observations on 72 subjects. The actual dataset analyzed contained only 317 observations per spirometric measure.

Plots were produced of values of each spirometric measure versus age. Due to significant heteroscedasticity, a logarithmic transformation of the data was required. A listing of the logarithmic values of FVC at ages three to 12 years (LFVC3 - LFVC12), log values of PEF at ages three to 12 years (LPEF3 - LPEF12), and log values of $\dot{V}max_{75\%}$ at ages three to 12 years (LV75M3 - LV75M12) are given in Appendix A. The subjects have been grouped according to missing data patterns in these tables.

As a first step, the measurements on each spirometric parameter were analyzed separately. Due to the missing data, a GIM analysis was employed. There were 36 groups of like missing data patterns, many consisting of only one subject. For the k^{th} spirometric variable, the following model was assumed:

$$E(\underset{\sim}{Y}_{j(k)}) = \underset{\sim}{D}_j \underset{\sim}{B}_k \underset{\sim}{K}_j$$

$$\text{Var}(i^{th} \text{ row of } \underset{\sim}{Y}_{j(k)}) = \underset{\sim}{\Sigma}_{kk} \quad , \quad k = 1,2,3, \; j = 1,\ldots,36,$$

where

$\underset{\sim}{Y}_{j(k)}$ is the $(n_j \times m_j)$ matrix of responses on the k^{th} spirometric variable for the j^{th} data group,

$\underset{\sim}{D}_j$ is the $(n_j \times 4)$ design matrix for $\underset{\sim}{Y}_{j(k)}$

$\underset{\sim}{B}_k$ is the (4×10) matrix of model parameters,

$\underset{\sim}{K}_j$ is the $(10 \times m_j)$ missing data indicator matrix.

The model parameters correspond to four race by sex means at each year of age.

It was decided that sufficient data did not exist to produce reliable estimates of the 55 unknown variance and covariance parameters in $\underset{\sim}{\Sigma}_{kk}$ by pairwise deletion. Preliminary analyses of ordinary least squares residuals were conducted to determine reasonable covariance models for $\underset{\sim}{\Sigma}_{kk}$. Examination of the OLS residuals revealed that measurements on a subject were highly correlated over time, as would be expected. This was particularly true for FVC. In order to determine the adequacy of the AR(1) covariance model, plots of individual residuals were produced.

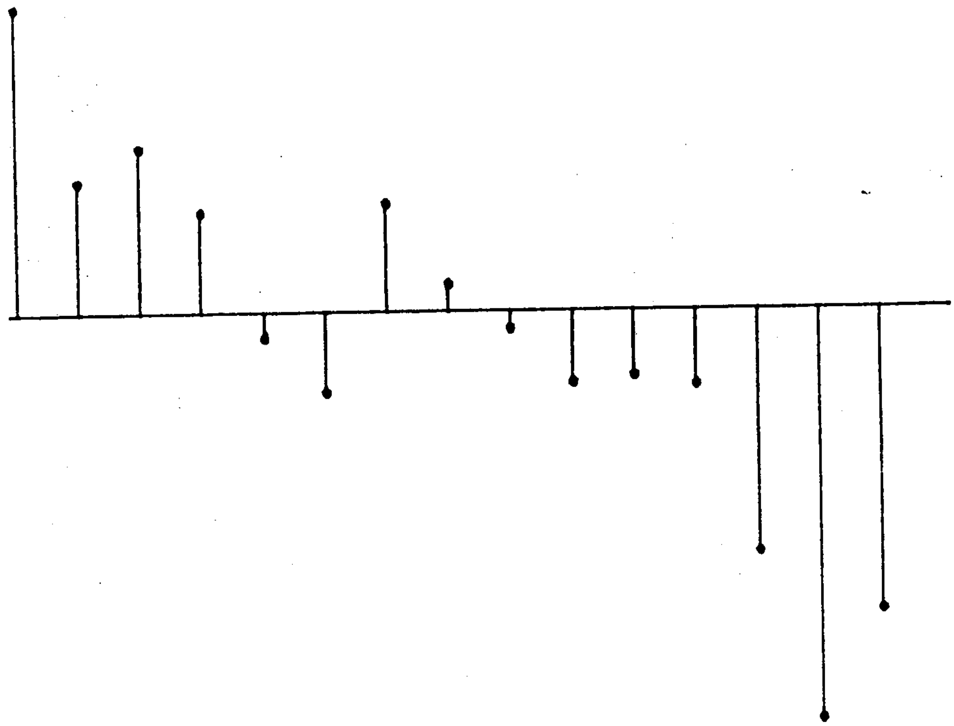

Figure 7.2.1 Plot of Hypothetical Residuals Illustrating a Typical AR(1) Process

Figure 7.2.1 illustates a typical hypothetical residual plot from an AR(1) process. Note that most of the information required to predict the residual at time t is contained in the residual at time t - 1. Several example plots of adjacent log FVC OLS residuals are given in Figure 7.2.2 for selected subjects. It can been seen from this figure as well as from the remainder of the 72 plots, that an autoregressive process of order higher than one would probably provide a better fit. Nevertheless, an AR(1) process was assumed for the first model fit to the data, since an AR(1) process is often a good approximation of an AR(p) process for p > 1 (e.g., Kmenta, 1971).

Finite moving average covariance models were also considered. The largest lag that could be estimated from the data was seven. No subjects remained in the study for longer than seven years. Furthermore, the number of subjects with data lags of five, six, and seven years were 16, nine, and eight respectively. Estimates of moving average parameters for lags larger than four or five years would probably be unreliable. Thus it was decided that a GIM would be fit to the data with an AR(1) process and then with an MA(4) process for the errors.

The second step of the analysis consisted of combining the three dependent variables into one model as described in Chapter 6. The MA(1) error model appeared totally unsatisfactory, so multiple AR(1) processes were assumed. The purpose of this analysis step was to examine the correlations among response variables.

7.3 Results

Estimates of the autocorrelation and variance parameters for an AR(1) process were produced according to (4.2.8) for each of the three spirometric variables. These estimates are given in Table 7.3.1. The

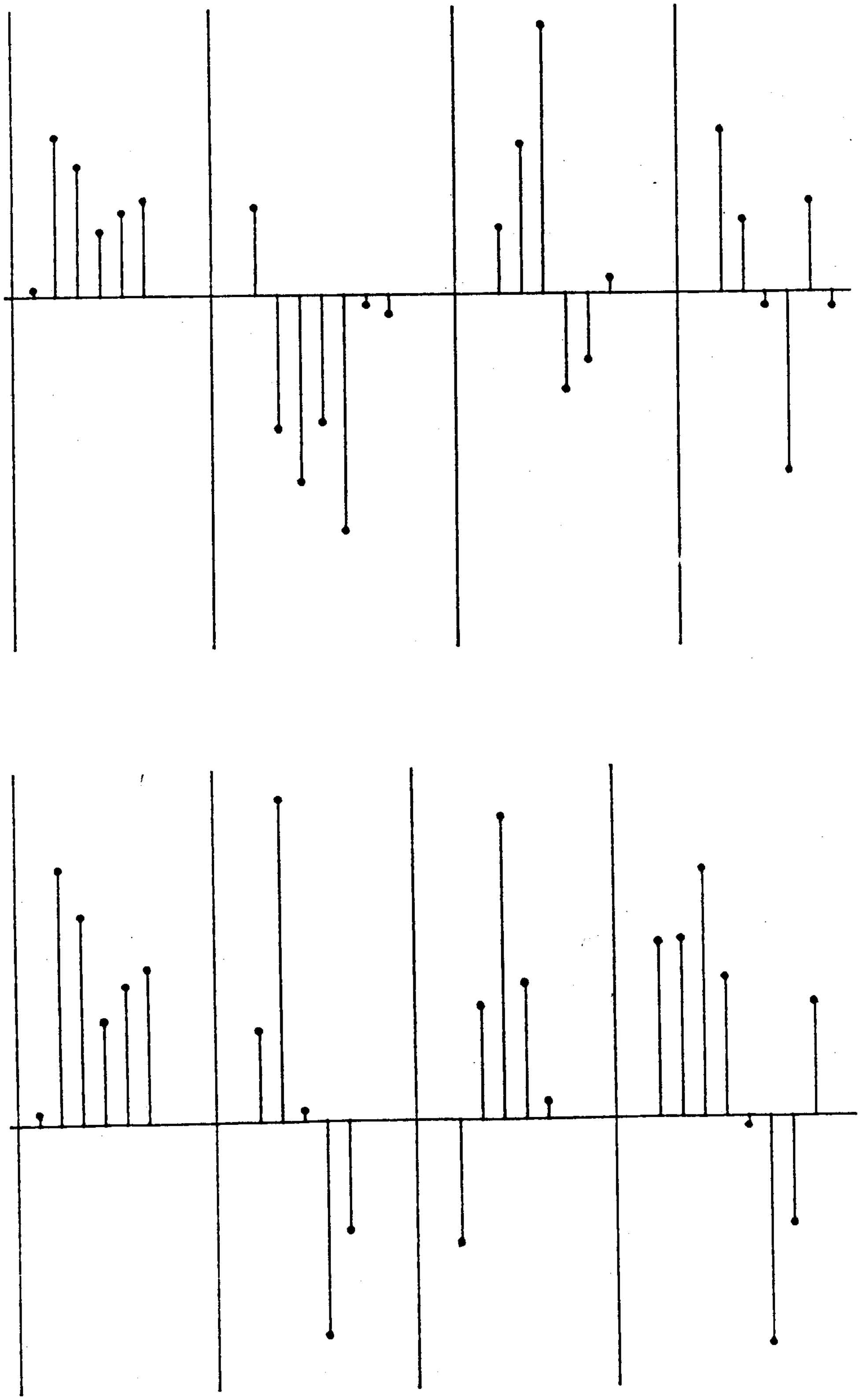

Figure 7.2.2 Plots of Residuals of Log FVC for Selected Subjects

Table 7.3.1 AR(1) Process Parameter Estimates

	Log FVC	Log PEF	Log $\dot{V}max_{75\%}$
$\hat{\rho}$	0.79	0.57	0.46
$\hat{\sigma}_u^2$	0.045	0.046	0.166

autocorrelation estimates were based on 167 pairs of consecutive residuals, and the variance estimates were based on all 317 observations, thus the asymptotic properties established for these estimators in Theorem 4.2.1 should be valid. A GIM was then fit to the data for each of the three dependent variables assuming an AR(1) error process. Estimates of the model parameters, $\hat{\Sigma}_s$ and $\hat{B}_s$, were computed according to (4.2.8) and are given in Tables 7.3.2 - 7.3.4.

Although there were observations on subjects at age three, only one of these was a white female and one a white male. Thus the estimates of these two cell means are unreliable. All hypotheses will be tested using only the parameter estimates for ages four through 12. For notational convenience, it will be assumed in defining the test statistics that elements of B corresponding to age three have been deleted.

The first hypothesis of interest is that of no differences among the four race by sex groups for any age, namely

$$H_1:\ C_1 B_s U_1 = 0 \tag{7.3.1}$$

where

$$C_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \quad \text{and } U_1 = I_9.$$

Table 7.3.2 Estimates of Σ_s, B_s, and S.E.($\hat{B}_s$) for Log FVC

$\hat{\Sigma}_s$	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12
LFVC3	0.0446	0.0353	0.0280	0.0222	0.0176	0.0139	0.0110	0.0087	0.0069	0.0055
LFVC4	0.0353	0.0446	0.0353	0.0280	0.0222	0.0176	0.0139	0.0110	0.0087	0.0069
LFVC5	0.0280	0.0353	0.0446	0.0353	0.0280	0.0222	0.0176	0.0139	0.0110	0.0087
LFVC6	0.0222	0.0280	0.0353	0.0446	0.0353	0.0280	0.0222	0.0176	0.0139	0.0110
LFVC7	0.0176	0.0222	0.0280	0.0353	0.0446	0.0353	0.0280	0.0222	0.0176	0.0139
LFVC8	0.0139	0.0176	0.0222	0.0280	0.0353	0.0446	0.0353	0.0280	0.0222	0.0176
LFVC9	0.0110	0.0139	0.0176	0.0222	0.0280	0.0353	0.0446	0.0353	0.0280	0.0222
LFVC10	0.0087	0.0110	0.0139	0.0176	0.0222	0.0280	0.0353	0.0446	0.0353	0.0280
LFVC11	0.0069	0.0087	0.0110	0.0139	0.0176	0.0222	0.0280	0.0353	0.0446	0.0353
LFVC12	0.0055	0.0069	0.0087	0.0110	0.0139	0.0176	0.0222	0.0280	0.0353	0.0446
$\hat{B}_s$	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12
WF	-.5553	-.1046	0.1103	0.1812	0.4180	0.5738	0.6108	0.7301	0.8033	0.8891
BF	-.4541	-.2114	0.0170	0.1833	0.3229	0.4239	0.6464	0.7318	0.7868	1.0214
WM	-.3294	0.0384	0.1566	0.3121	0.4050	0.6559	0.6938	0.8162	0.9322	0.8732
BM	-.5709	-.2964	-.0112	0.1433	0.2636	0.3479	0.5832	0.5925	0.7166	0.7682
SE($\hat{B}_s$)	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12
WF	0.1600	0.1196	0.0978	0.0802	0.0763	0.0742	0.0734	0.0750	0.0794	0.0853
BF	0.0489	0.0422	0.0409	0.0426	0.0453	0.0519	0.0641	0.0721	0.0803	0.1111
WM	0.1566	0.1123	0.0828	0.0831	0.0831	0.0828	0.0861	0.0881	0.0950	0.1182
BM	0.0528	0.0451	0.0440	0.0481	0.0552	0.0664	0.0767	0.0900	0.0868	0.0942

Table 7.3.3 Estimates of $\underset{\sim}{\Sigma}_s$, $\underset{\sim}{B}_s$, and S.E.($\hat{\underset{\sim}{B}}_s$) for Log PEF

$\hat{\underset{\sim}{\Sigma}}_s$	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
LPEF3	0.0464	0.0264	0.0151	0.0086	0.0049	0.0028	0.0016	0.0009	0.0005	0.0003
LPEF4	0.0264	0.0464	0.0264	0.0151	0.0086	0.0049	0.0028	0.0016	0.0009	0.0005
LPEF5	0.0151	0.0264	0.0464	0.0264	0.0151	0.0086	0.0049	0.0028	0.0016	0.0009
LPEF6	0.0086	0.0151	0.0264	0.0464	0.0264	0.0151	0.0086	0.0049	0.0028	0.0016
LPEF7	0.0049	0.0086	0.0151	0.0264	0.0464	0.0264	0.0151	0.0086	0.0049	0.0028
LPEF8	0.0028	0.0049	0.0086	0.0151	0.264	0.0464	0.0264	0.0151	0.0086	0.0049
LPEF9	0.0016	0.0028	0.0049	0.0086	0.0151	0.0264	0.0464	0.0264	0.0151	0.0086
LPEF10	0.0009	0.0016	0.0028	0.0049	0.0086	0.0151	0.0264	0.0464	0.0264	0.0151
LPEF11	0.0005	0.0009	0.0016	0.0028	0.0049	0.0086	0.0151	0.0264	0.0464	0.0264
LPEF12	0.0003	0.0005	0.0009	0.0016	0.0028	0.0049	0.0086	0.0151	0.0264	0.0464
$\hat{\underset{\sim}{B}}_s$	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
WF	0.5192	0.9255	1.1359	1.1395	1.2619	1.4380	1.3926	1.5297	1.5971	1.6190
BF	0.6824	0.8114	1.0248	1.1143	1.1943	1.2271	1.4605	1.5143	1.7784	1.8933
WM	0.8593	1.0478	1.1604	1.0923	1.3271	1.4185	1.4092	1.4923	1.5860	1.5885
BM	0.5774	0.7851	0.9464	1.0159	1.2037	1.2881	1.8085	1.4781	1.5858	1.5579
SE($\hat{\underset{\sim}{B}}_s$)	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
WF	0.1943	0.1408	0.1132	0.0856	0.0805	0.0786	0.0777	0.0793	0.0852	0.0929
BF	0.0553	0.0445	0.0424	0.0458	0.0502	0.0596	0.0750	0.0829	0.0913	0.1355
WM	0.1929	0.1347	0.0872	0.0908	0.0908	0.0872	0.0921	0.0937	0.1034	0.1383
BM	0.0599	0.0478	0.0461	0.0535	0.0633	0.0792	0.0899	0.1088	0.0951	0.1038

Table 7.3.4 Estimates of $\underset{\sim}{\Sigma}_s$, $\underset{\sim}{B}_s$, and S.E.$(\hat{\underset{\sim}{B}}_s)$ for Log $\dot{V}max_{75\%}$

$\hat{\underset{\sim}{\Sigma}}_s$	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7	LV75M8	LV75M9	LV75M10	LV75M11	LV75M12
LV75M3	0.1662	0.0767	0.0354	0.0164	0.0076	0.0035	0.0016	0.0007	0.0003	0.0002
LV75M4	0.0767	0.1662	0.0767	0.0.354	0.0164	0.0076	0.0035	0.0016	0.0007	0.0003
LV75M5	0.0354	0.0767	0.1662	0.0767	0.0364	0.0164	0.0076	0.0035	0.0016	0.0007
LV75M6	0.0164	0.0354	0.0767	0.1662	0.0767	0.0354	0.0164	0.0076	0.0035	0.0016
LV75M7	0.0076	0.0164	0.0354	0.0767	0.1662	0.0767	0.0354	0.0164	0.0076	0.0035
LV75M8	0.0035	0.0076	0.0164	0.0354	0.0767	0.1662	0.0767	0.0354	0.0164	0.0076
LV75M9	0.0016	0.0035	0.0076	0.0164	0.0354	0.0767	0.1662	0.0767	0.0354	0.0164
LV75M10	0.0007	0.0016	0.0035	0.0076	0.0164	0.0354	0.0767	0.1662	0.0767	0.0354
LV75M11	0.0003	0.0007	0.0016	0.0035	0.0076	0.0164	0.0354	0.0767	0.1662	0.0767
LV75M12	0.0002	0.0003	0.0007	0.0016	0.0035	0.0076	0.0164	0.0354	0.0767	0.01662
$\hat{\underset{\sim}{B}}_s$	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7	LV75M8	LV75M9	LV75M10	LV75M11	LV75M12
WF	0.0277	-.1453	-.1615	0.1851	0.2473	0.2350	0.3870	0.4287	0.4962	0.6049
BF	-.3760	-.0631	0.0145	0.0504	-.0197	0.0698	0.0536	0.2807	0.5355	0.6530
WM	0.1487	0.2321	0.1553	-.0508	0.3944	-.1880	0.1556	0.2525	0.3801	0.5926
WM	-.2682	-.1215	-.0510	-.0698	-.2264	-.0781	-.0166	0.0792	0.03352	0.3059
SE$(\hat{\underset{\sim}{B}}_s)$	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7	LV75M8	LV75M9	LV75M10	LV75M11	LV75M12
WF	0.3833	0.2755	0.2220	0.1637	0.1534	0.1506	0.1492	0.1517	0.1634	0.1785
BF	0.1077	0.0852	0.0807	0.0882	0.0975	0.1166	0.1467	0.1608	0.1766	0.2684
WM	0.3820	0.2669	0.1658	0.1756	0.1756	0.1658	0.1769	0.1791	0.1988	0.2717
BM	0.1169	0.091	0.0880	0.1040	0.1235	0.1562	0.1753	0.2161	0.1815	0.1993

In addition, hypotheses of specific two-way comparisons between groups were tested as follows:

(7.3.2) $\quad H_2$ (WF vs. BF) : $\underset{\sim}{C}_2 \underset{\sim}{B}_s \underset{\sim}{U}_1 = \underset{\sim}{0}$

where $\quad \underset{\sim}{C}_2 = [1 \quad -1 \quad 0 \quad 0]$,

(7.3.3) $\quad H_3$ (WF vs. WM) : $\underset{\sim}{C}_3 \underset{\sim}{B}_s \underset{\sim}{U}_1 = \underset{\sim}{0}$

where $\quad \underset{\sim}{C}_3 = [1 \quad 0 \quad -1 \quad 0]$,

(7.3.4) $\quad H_4$ = (BF vs. BM) : $\underset{\sim}{C}_4 \underset{\sim}{B}_s \underset{\sim}{U}_1 = \underset{\sim}{0}$

where $\quad \underset{\sim}{C}_4 = [0 \quad 1 \quad 0 \quad -1]$,

and

(7.3.5) $\quad H_5$ (BM vs. WM) : $\underset{\sim}{C}_5 \underset{\sim}{B}_s \underset{\sim}{U}_1 = \underset{\sim}{0}$

where $\quad C_5 = [0 \quad 0 \quad 1 \quad -1]$.

Each of these pairwise differences between groups were also tested at each year of age by setting $\underset{\sim}{U}$ equal to the t^{th} column of $\underset{\sim}{I}_9$ for the t^{th} test, $t = 1,\ldots,9$.

The hypotheses described above were tested by constructing a Wald statistic and comparing the result to a central chi-square distribution with w degrees of freedom, where w equals the number of rows in $\underset{\sim}{U}' \otimes \underset{\sim}{C}$. It should be noted that because the entire sample was used to estimate $\underset{\sim}{\Sigma}_s$ and $\underset{\sim}{B}_s$, the conditions of Theorem 2.4.1 are not met. Although the resulting test statistics are only approximately distributed as chi-square random variables, it was decided that this procedure was good practice.

For testing H: $\underset{\sim}{C}\,\underset{\sim}{B}\,\underset{\sim}{U} = \underset{\sim}{0}$, let

(7.3.6) $\quad \underset{\sim}{F} = \underset{\sim}{U}' \otimes \underset{\sim}{C}$.

Then

(7.3.7) $\quad \underset{\sim}{W} = (\underset{\sim}{F}\,\hat{\underset{\sim}{b}}_s)' (\underset{\sim}{F}\,\hat{\text{Var}}(\hat{\underset{\sim}{b}}_s)\,\underset{\sim}{F}')^{-1}\,\underset{\sim}{F}\,\hat{\underset{\sim}{b}}_s$

where $\hat{\text{Var}}(\hat{\underset{\sim}{b}}_s)$ is given in (4.2.16). Approximate probability levels for

Table 7.3.5 Approximate p-values of Tests for Age-Specific Group Differences in Mean Log FVC Assuming an AR(1) Error Model

	Any Age	4	5	6	7	8	9	10	11	12
Overall	0.13	0.08	0.26	0.38	0.31	0.01	0.79	0.36	0.41	0.39
WF vs. BF	0.24	0.40	0.38	0.98	0.28	0.10	0.71	0.99	0.88	0.35
WF vs. WM	0.58	0.38	0.72	0.26	0.91	0.46	0.46	0.46	0.30	0.91
BF vs. BM	0.78	0.47	0.64	0.53	0.41	0.37	0.53	0.23	0.55	0.08
WM vs. BM	0.03	0.01	0.07	0.08	0.16	0.00	0.34	0.08	0.09	0.49

Table 7.3.6 Approximate p-values for Tests of Age-Specific Group Differences in Mean Log PEF Assuming an AR(1) Error Model

	Any Age	4	5	6	7	8	9	10	11	12
Overall	0.42	0.26	0.10	0.48	0.58	0.11	0.64	0.98	0.37	0.23
WF vs. BF	0.15	0.44	0.36	0.08	0.48	0.03	0.53	0.89	0.15	0.10
WF vs. WM	0.99	0.53	0.86	0.71	0.59	0.87	0.89	0.76	0.93	0.85
BF vs. BM	0.28	0.69	0.21	0.16	0.91	0.54	0.19	0.79	0.14	0.05
WM vs. BM	0.58	0.07	0.03	0.47	0.26	0.27	0.43	0.92	0.99	0.86

Table 7.3.7 Approximate p-values for Tests of Age-Specific Group Differences in Mean Log $\dot{V}max_{75\%}$ Assuming an AR(1) Error Model

	Any Age	4	5	6	7	8	9	10	11	12
Overall	.37	.64	.62	.56	.01	.24	.28	.61	.84	.64
WF vs. BF	.79	.78	.46	.46	.14	.39	.11	.50	.87	.88
WF vs. WM	.26	.33	.25	.33	.53	.06	.32	.45	.65	.97
BF vs. BM	.94	.64	.58	.38	.19	.45	.76	.45	.43	.30
WM vs. BM	.06	.21	.27	.93	0.00	.63	.49	.54	.87	.39

the tests of the hypotheses concerning race and sex differences are presented in Tables 7.3.5 - 7.3.7 for each of the three spirometric variables.

Significant differences ($p > 0.05$) in mean log FVC between white and black males were detected for most age groups. With respect to log PEF, differences were found between young (four and five years) black and white males. Only seven-year-old black and white males differed with respect to log $\dot{V}max_{75\%}$. No other significant differences were found.

It was also of interest to determine if any growth occurred with respect to the three spirometric variables and to test for different profiles of growth among the four race and sex groups. To determine if any growth occurred, define the hypothesis

(7.3.) $$H_6:\ \underset{\sim}{C}_6\ \underset{\sim}{B}_s\ \underset{\sim}{U}_6 = 0$$

where

$$\underset{\sim}{C}_6 = \underset{\sim}{I}_4 \text{ and } \underset{\sim}{U}_6 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

The four group-specific hypotheses of zero growth are then defined by

$H_{6i}:\ \underset{\sim}{C}_{6i}\ \underset{\sim}{B}_s\ \underset{\sim}{U}_6$ where $\underset{\sim}{C}_{6i}$ is the i^{th} row of $\underset{\sim}{C}_6$, $i = 1,\ldots,4$.

The hypothesis of equal growth profiles among the four groups (i.e., no growth by group interaction) is given by

(7.3.9) $$H_7:\ \underset{\sim}{C}_1\ \underset{\sim}{B}_s\ \underset{\sim}{U}_6$$

where $\underset{\sim}{C}_1$ is defined in (7.3.1). Repeating (7.3.2) through (7.3.5), substituting $\underset{\sim}{U}_6$ for $\underset{\sim}{U}_1$ defines the hypotheses of equal growth profiles for the four pairwise group comparisons. The results of tests of hypotheses are given in Table 7.3.8.

Table 7.3.8 Approximate p-values for Tests of Growth (vs. (Age) for Each Race and Sex Group and Differences in Growth Profiles Between the Groups Assuming an AR(1) Error Model

	Log FVC	Log PEF	Log $\dot{V}max_{75\%}$
Growth:			
Overall	0.00	0.00	0.03
WF	0.00	0.00	0.35
BF	0.00	0.00	0.11
WM	0.00	0.01	0.02
BM	0.00	0.00	0.38
Differences in Growth:			
Overall	0.25	0.53	0.55
WF vs. BF	0.18	0.11	0.81
WF vs. BM	0.56	0.98	0.21
BF vs. BM	0.83	0.35	0.99
WM vs. BM	0.15	0.90	0.09

Significant growth ($p < 0.01$) as defined by (7.3.8) was detected for all race by sex groups with respect to log FVC and log PEF, and only for white males with respect to log $\dot{V}max_{75\%}$. The only significant difference ($p = 0.09$) between growth profiles occurred for white males versus black males with respect to log $\dot{V}max_{75\%}$.

Assuming a moving average process of order four for the errors in the GIM, estimates of variance components were produced using the estimation procedure described in section 5.4. It was noted earlier that many of the missing data groups consisted of only one subject, thus it was not possible to compute $\hat{y}_{k_{(j)}}$ as in (5.2.15) for all 36 groups. The moving average parameters were therefore computed from the formulae given in (5.4.4) and (5.4.8).

A GIM was fit to the data for each of the three spirometric variables assuming an MA(4) error model. Estimates of the model parameters, $\underset{\sim}{\Sigma}_M$ and $\underset{\sim}{B}_M$, and the standard errors of $\hat{\underset{\sim}{B}}_M$ are given in Tables 7.3.9 - 7.3.11 for these spirometric varibles. The estimate of $\underset{\sim}{\Sigma}_M$ for log FVC was nonpositive definite, thus producing unreliable standard error estimates for four of the race and sex cell means. These are indicated by dashed lines in Table 7.3.9. Typically the moving average process provides a good fit for data with autocorrelations less than 0.50 (see section 6.2). As a result, no hypotheses were tested for log FVC with this model.

The hypotheses specified above were each tested under the MA(4) model assumptions for log PEF and log $\dot{V}max_{75\%}$. Results for the tests concerning differences in age-specific means among the race by sex groups are given in Tables 7.3.12 and 7.3.13. Results for the tests concerning growth profiles are given in Table 7.3.14.

Table 7.3.9 Estimates of $\hat{\underset{\sim}{\Sigma}}_M$, $\underset{\sim}{B}_M$, and Standard Errors of $\hat{\underset{\sim}{B}}_M$ for Log FVC

$\hat{\underset{\sim}{\Sigma}}_M$	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12
LFVC3	0.0446	0.0342	0.0309	0.0377	0.442	0.0000	0.0000	0.0000	0.0000	0.0000
LFVC4	0.0342	0.0446	0.0342	0.0309	0.0377	0.0442	0.0000	0.0000	0.0000	0.0000
LFVC5	0.0309	0.0342	0.0446	0.0342	0.0309	0.0377	0.0442	0.0000	0.0000	0.0000
LFVC6	0.0377	0.0309	0.0342	0.0446	0.0342	0.0309	0.0377	0.0442	0.0000	0.0000
LFVC7	0.0442	0.0337	0.0309	0.0342	0.0446	0.0342	0.0309	0.0377	0.0442	0.0000
LFVC8	0.0000	0.0442	0.0377	0.0309	0.0342	0.0446	0.0342	0.0309	0.0377	0.0442
LFVC9	0.0000	0.0000	0.0442	0.0377	0.0309	0.0342	0.0446	0.0342	0.0309	0.0377
LFVC10	0.0000	0.0000	0.0000	0.0442	0.0377	0.0309	0.0342	0.0446	0.0342	0.0309
LFVC11	0.0000	0.0000	0.0000	0.0000	0.0442	0.0377	0.0309	0.0342	0.0446	0.0342
LFVC12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0442	0.0377	0.0309	0.0342	0.0446
$\hat{\underset{\sim}{B}}_M$	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12
WF	-.3188	0.5902	0.0369	-.1913	0.3646	0.6487	0.7646	0.8736	0.5996	0.6451
BF	-.3996	-.2665	0.0080	0.2621	0.3072	0.5950	0.6090	0.7353	0.8796	1.6469
WM	-.3115	0.0793	0.2683	0.2094	0.4125	0.6351	0.6178	0.9577	0.8904	0.8424
BM	-.6056	-.3065	-.0059	0.1505	0.2926	0.5279	0.6482	0.7187	0.8097	0.7373
SE($\hat{\underset{\sim}{B}}_M$)	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12
WF	(----)	0.1833	0.0890	0.1264	0.0761	0.0746	0.0824	0.0848	0.0775	0.1045
BF	0.0415	0.0400	0.0418	0.0451	0.0421	0.0299	0.0687	0.1120	0.1003	(----)
WM	(----)	0.1659	0.0770	0.1072	0.0807	0.0792	0.0774	0.1213	0.0913	(----)
BM	0.0480	0.0418	0.0434	0.0476	0.0468	0.0328	0.0745	0.1064	0.0832	0.1211

Table 7.3.10 Estimates of $\hat{\underset{\sim}{\Sigma}}_M$, $\underset{\sim}{B}_M$, and S.E.$(\hat{\underset{\sim}{B}}_M)$ for LOG PEF

$\hat{\underset{\sim}{\Sigma}}_M$	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
LPEF3	0.0464	0.0266	0.0213	0.0158	0.0207	0.0000	0.0000	0.0000	0.0000	0.0000
LPEF4	0.0266	0.0464	0.0266	0.0213	0.0158	0.0207	0.0000	0.0000	0.0000	0.0000
LPEF5	0.0213	0.0266	0.0464	0.0266	0.0213	0.0158	0.0207	0.0000	0.0000	0.0000
LPEF6	0.0158	0.0213	0.0266	0.0464	0.0266	0.0213	0.0158	0.0207	0.0000	0.0000
LPEF7	0.0207	0.0158	0.0213	0.0266	0.0464	0.0266	0.0213	0.0158	0.0207	0.0000
LPEF8	0.0000	0.0207	0.0158	0.0213	0.0266	0.0464	0.0266	0.0213	0.0158	0.0207
LPEF9	0.0000	0.0000	0.0207	0.0158	0.0213	0.0266	0.0464	0.0266	0.0213	0.0158
LPEF10	0.0000	0.0000	0.0000	0.0207	0.0158	0.0213	0.0266	0.0464	0.0266	0.0213
LPEF11	0.0000	0.0000	0.0000	0.0000	0.0207	0.0158	0.0213	0.0266	0.0464	0.0266
LPEF12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0207	0.0158	0.0213	0.0266	0.0464
$\hat{\underset{\sim}{B}}_M$	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
WF	0.5626	0.8923	1.1771	1.1519	1.2915	1.4295	1.3851	1.5234	1.6040	1.5649
BF	0.7056	0.8281	1.0091	1.1333	1.1757	1.2418	1.5129	1.4764	1.8024	1.9495
WM	0.9661	1.0662	1.1745	1.1008	1.3266	1.4138	1.3934	1.4901	1.5601	1.5892
BM	0.5819	0.7987	0.9601	1.0122	1.2175	1.2638	1.3048	1.4535	1.5152	1.5158
SE$(\hat{\underset{\sim}{B}}_M)$	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
WF	0.1587	0.1239	0.1066	0.0814	0.0780	0.0778	0.0765	0.0775	0.0807	0.0864
BF	0.0525	0.0437	0.0420	0.0451	0.0488	0.0551	0.0674	0.0731	0.0811	0.1113
WM	0.1597	0.1192	0.0845	0.0867	0.0894	0.0857	0.0893	0.0896	0.0977	0.1206
BM	0.0576	0.0471	0.0455	0.0523	0.0601	0.0728	0.0814	0.0973	0.0877	0.0963

Table 7.3.11 Estimates of $\hat{\Sigma}_M$, B_M, and S.E. $(\hat{B}_M)$ for Log $\dot{V}max_{75\%}$

$\hat{\Sigma}_M$	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7	LV75M8	LV75M9	LV75M10	LV75M11	LV75M12
LV75M3	0.1662	0.0663	0.0620	0.0441	0.0350	0.0000	0.0000	0.0000	0.0000	0.0000
LV75M4	0.0663	0.1662	0.0663	0.0620	0.0441	0.0350	0.0000	0.0000	0.0000	0.0000
LV75M5	0.0620	0.0663	0.1662	0.0663	0.0620	0.0441	0.0350	0.0000	0.0000	0.0000
LV75M6	0.0441	0.0620	0.0663	0.1662	0.0663	0.0620	0.0441	0.0350	0.0000	0.0000
LV75M7	0.0350	0.0441	0.0620	0.0663	0.1662	0.0663	0.0620	0.0441	0.0350	0.0000
LV75M8	0.0000	0.0350	0.0441	0.0620	0.0663	0.1662	0.0663	0.0620	0.0441	0.0350
LV75M9	0.0000	0.0000	0.0350	0.0441	0.0620	0.0663	0.1662	0.0663	0.0620	0.0441
LV75M10	0.0000	0.0000	0.0000	0.0350	0.0441	0.0620	0.0663	0.1662	0.0663	0.0620
LV75M11	0.0000	0.0000	0.0000	0.0000	0.0350	0.0441	0.0620	0.0663	0.1662	0.0663
LV75M12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0350	0.0441	0.0620	0.0663	0.1662
$\hat{B}_M$	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7	LV75M8	LV75M9	LV75M10	LV75M11	LV75M12
WF	0.0974	-.1804	-.1468	0.1730	0.2540	0.2251	0.4026	0.4429	0.4518	0.6053
BF	-.4184	-.0521	0.0290	0.0718	-.0288	0.0200	0.0410	0.2062	0.4294	0.5754
WM	0.1477	0.1649	0.1166	-.0987	0.3996	-.1825	0.1833	0.2471	0.3853	0.5770
BM	-.2419	-.1027	-.0488	-.0812	-.2216	-.0253	-.0146	0.0026	0.2515	0.2467
SE$(\hat{B}_M)$	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7	LV75M8	LV75M9	LV75M10	LV75M11	LV75M12
WF	0.3700	0.2666	0.2195	0.1619	0.1505	0.1498	0.1498	0.1507	0.1616	0.1755
BF	0.1067	0.0851	0.0806	0.0881	0.0969	0.1149	0.1444	0.1564	0.1727	0.2619
WM	0.3673	0.2662	0.1637	0.1746	0.1752	0.1640	0.1767	0.1781	0.1964	0.2660
BM	0.1162	0.0915	0.0876	0.1037	0.1210	0.1533	0.1703	0.2149	0.1773	0.1976

Table 7.3.12 Approximate p-values for Tests of Age-Specific Group Differences in Mean Log PEF Assuming an MA(4) Error Model

	AGE (Years)									
	Any Age	4	5	6	7	8	9	10	11	12
Overall	0.00	0.20	0.06	0.29	0.38	0.12	0.25	0.94	0.08	0.02
WF vs. BF	0.00	0.62	0.14	0.84	0.21	0.05	0.21	0.66	0.08	0.01
WF vs. WM	0.94	0.31	0.98	0.67	0.77	0.89	0.94	0.78	0.73	0.87
BF vs. BM	0.00	0.65	0.43	0.08	0.59	0.81	0.05	0.85	0.02	0.00
WM vs. BM	0.57	0.04	0.03	0.38	0.31	0.18	0.46	0.78	0.73	0.63

Table 7.3.13 Approximate p-values for Tests of Age-Specific Group Differences in Mean Log $\dot{V}max_{75\%}$ Assuming an MA(4) Error Model

	AGE (Years)									
	Any Age	4	5	6	7	8	9	10	11	12
Overall	0.49	0.77	0.71	0.45	0.01	0.33	0.22	0.39	0.79	0.54
WF vs. BF	0.67	0.65	0.45	0.58	0.16	0.28	0.08	0.28	0.79	0.92
WF vs. WM	0.42	0.36	0.34	0.25	0.53	0.07	0.34	0.40	0.68	0.93
BF vs. BM	0.93	0.69	0.51	0.26	0.21	0.81	0.80	0.44	0.47	0.32
WM vs. BM	0.10	0.34	0.37	0.93	0	0.49	0.42	0.39	0.61	0.32

Table 7.3.14 Approximate p-values for Tests of Growth (vs. Age) for Each Race and Sex Group and Differences in Growth Profiles Assuming an MA(4) Error Model

	Log PEF	Log $\dot{V}max_{75\%}$
Growth:		
Overall	0.00	0.08
WF	0.00	0.28
BF	0.00	0.21
WM	0.00	0.03
BM	0.00	0.55
Differences in Growth:	0.00	0.71
WF vs. BF	0.90	0.35
WF vs. WM	0.00	0.98
BF vs. BM	0.92	0.14
WM vs. BM	0.00	0.66

A comparison of the results from these two models indicates that while the estimation of $\underset{\sim}{B}$ is not very sensitive to the choice of covariance model, the hypothesis test results are. In particular, significant differences between white females and black females and between black females and black males were detected with respect to log PEF assuming an MA(4) model, while no significant group differences were detected with respect to log PEF assuming an AR(1) model. The tests with respect to $\dot{V}max_{75\%}$ yielded similar results under the two covariance models, however. The fact that the Wald test is very sensitive to the estimation of an unconstrained covariance matrix has been documented recently in the literature (e.g., Leeper, et. al., 1982). The results of this analysis indicate that the choice of the covariance model, in addition to the computational stability of the estimate, have a large impact on the test results produced.

As a second step, th data for the three spirometric variables were analyzed jointly by the methods described in Chapter 6. It was obvious from the previous analyses that an MA(1) model for the errors would not be adequate since positive correlations appear to exist between measurements separated by more than one year of age. Thus a model was fit to the data assuming multiple AR(1) error processes.

Estimates of autocorrelation parameters for each dependent variable were available from the previous analyses. The between measurement covariance parameters were estimated as described in (6.2.14). These estimates are given in Table 7.3.15. Referring to the notation in (6.1.2), $\hat{\underset{\sim}{\Sigma}}_{kk}$, k = 1,2,3, are given by the estimates of $\underset{\sim}{\Sigma}_s$ in Tables 7.3.2 - 7.3.4. Estimates of $\underset{\sim}{\Sigma}_{12}$, $\underset{\sim}{\Sigma}_{13}$, and $\underset{\sim}{\Sigma}_{23}$ are given in Table 7.3.16. Table 7.3.17 contains estimates of $\underset{\sim}{B}$.

Log FVC and log PEF appear to be highly correlated ($\hat{\rho}_{12} = 0.56$) as do log PEF and log $\dot{V}max_{75\%}$ ($\hat{\rho}_{23} = 0.35$). A slight positive correlation was found between log FVC and log $\dot{V}max_{75\%}$ ($\hat{\rho}_{13} = 0.13$). The estimates of $\underset{\sim}{B}$ for the joint model are similar to those computed for the three variables modeled separately, as are the standard errors of $\underset{\sim}{B}$.

For the most part, the results from the analysis are consistent with physiological explanations of pulmonary function in children. The analysis has provided an illustration of the application of modeled covariance matrices to incomplete longitudinal data. The major advantages of the methods employed are the ability to formulate the problem

Table 7.3.15 Estimates of Covariances and Correlations Between Spirometric Variables for Fixed Time

Cov. / Corr.	Log FVC	Log PEF	Log $\dot{V}max_{75\%}$
Log FVC	0.045	0.026	0.006
Log PEF	0.56	0.046	0.031
Log $\dot{V}max_{75\%}$	0.14	0.35	0.166

in terms of a reasonable number of estimable parameters and the use of all available information in the estimation process.

Several disadvantages are also apparent. The selection of a covariance model is subjective and often a result of trial and error. Examination of residual plots and preliminary pairwise deletion estimates can be useful in this choice. It would also be helpful to have some underlying biological or physiological model upon which to base the selection. The AR(1) model is often preferrable due to the small number of covariance parameters involved and the fact that correlations do not necessarily vanish with time under this model. Fewer assumptions are required for the MA(M) model, but often the covariance parameters corresponding to the different lags are estimated using sample sizes that differ dramatically. In this particular example, 68 subjects contributed to the estimation of γ_1, while only eight would have contributed to γ_7 if the MA(7) model had been assumed. It may be preferable, as in this example, to assume zero correlations for measurements separated by long lags than to assume that the consistency properties of the estimators

Table 7.3.16 Estimates of $\hat{\Sigma}_{12}$, $\hat{\Sigma}_{13}$, and $\hat{\Sigma}_{23}$ Assuming an AR(1) Error Model for Each Spirometric Variable

$\hat{\Sigma}_{12}$	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7	LV75M8	LV75M9	LV75M10	LV75M11	LV75M12
LFVC3	0.0062	0.0029	0.0013	0.0006	0.0003	0.0001	0.0001	0.0000	0.0000	0.0000
LFVC4	0.0049	0.0062	0.0029	0.0013	0.0006	0.0003	0.0001	0.0001	0.0000	0.0000
LFVC5	0.0030	0.0049	0.0062	0.0029	0.0013	0.0006	0.0003	0.0001	0.0001	0.0000
LFVC6	0.0031	0.0039	0.004	0.0062	0.0029	0.0013	0.006	0.0003	0.0001	0.0001
LFVC7	0.0024	0.0031	0.0039	0.0049	0.0062	0.002o	0.0013	0.0006	0.0003	0.0001
LFVC8	0.0019	0.0024	0.0031	0.0039	0.0049	0.0062	0.0029	0.0013	0.0006	0.0003
LFVC9	0.0015	0.0019	0.0024	0.0031	0.0038	0.0049	0.0062	0.0029	0.0013	0.0006
LFVC10	0.0012	0.0015	0.0019	0.0024	0.0031	0.0039	0.0049	0.0062	0.0029	0.0013
LFVC11	0.0010	0.0012	0.0015	0.0019	0.0024	0.0031	0.0039	0.0049	0.0062	0.0029
LFVC12	0.0008	0.0010	0.0012	0.0015	0.0019	0.0024	0.0031	0.0039	0.0049	0.0062
$\hat{\Sigma}_{13}$	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
LFVC3	0.0255	0.0145	0.0083	0.0047	0.0027	0.0015	0.0009	0.0005	0.0003	0.0002
LFVC4	0.0202	0.0255	0.0145	0.0083	0.0047	0.0027	0.0015	0.0009	0.0005	0.0003
LFVC5	0.0160	0.0202	0.0255	0.0145	0.0083	0.0047	0.0027	0.0015	0.0009	0.0005
LFVC6	0.0127	0.0160	0.0202	0.0255	0.0145	0.0083	0.0047	0.0027	0.0015	0.0009
LFVC7	0.0100	0.0127	0.0160	0.0202	0.0255	0.0145	0.0083	0.0047	0.0027	0.0015
LFVC8	0.0080	0.0100	0.0127	0.0160	0.0202	0.0255	0.0145	0.0083	0.0047	0.0027
LFVC9	0.0063	0.0080	0.0100	0.0127	0.0160	0.0202	0.0255	0.0145	0.0083	0.0047
LFVC10	0.0050	0.0063	0.0080	0.0100	0.0127	0.0160	0.0202	0.0255	0.0145	0.0083
LFVC11	0.0040	0.0050	0.0063	0.0080	0.0100	0.0127	0.0160	0.0202	0.0255	0.0145
LFVC12	0.0031	0.0040	0.0050	0.0063	0.0080	0.0100	0.0127	0.0160	0.0202	0.0255
$\hat{\Sigma}_{23}$	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
LV75M3	0.0305	0.0174	0.0099	0.0056	0.0032	0.0018	0.0010	0.0006	0.0003	0.0002
LV75M4	0.0141	0.0305	0.0174	0.0099	0.0056	0.0032	0.0018	0.0010	0.0006	0.0003
LV75M5	0.0065	0.0141	0.0305	0.0174	0.0099	0.0056	0.0032	0.0018	0.0010	0.0006
LV75M6	0.0030	0.0065	0.0141	0.0305	0.0174	0.0099	0.0056	0.0032	0.0018	0.0010
LV75M7	0.0014	0.0030	0.0065	0.0141	0.0305	0.0174	0.0099	0.0056	0.0032	0.0018
LV75M8	0.0006	0.0014	0.0030	0.0065	0.0141	0.0305	0.0174	0.0099	0.0056	0.0032
LV75M9	0.0003	0.0006	0.0014	0.0030	0.0065	0.0141	0.0305	0.0174	0.0099	0.0056
LF75M10	0.0001	0.0003	0.0006	0.0014	0.0030	0.0065	0.0141	0.0305	0.0174	0.0099
LF75M11	0.0001	0.0001	0.0003	0.0006	0.0014	0.0030	0.0065	0.0141	0.0305	0.0174
LF75M12	0.0000	0.0001	0.0001	0.0003	0.0006	0.0014	0.0030	0.0065	0.0141	0.0305

Table 7.3.17 Estimates of $\underset{\sim}{B}$ and S.E.$(\hat{\underset{\sim}{B}})$ Assuming An AR(1) Error Model for Each Spirometric Variable

$\hat{\underset{\sim}{B}}$	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7
	LV75M8	LV75M9	U75M10	V75M11	V75M12	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
WF	-.5451	-.0925	0.1097	0.1781	0.4096	0.5675	0.6118	0.7305	0.8036	0.8893	0.0430	-.1455	-.1609	0.1854	0.2493
	0.2372	0.3879	0.4286	0.4961	0.6049	0.5751	0.9474	1.1407	1.1404	1.2569	1.4341	1.3954	1.5497	1.5971	1.6190
BF	-.4504	-.2101	0.0198	0.1869	0.3274	0.4287	0.6387	0.7251	0.7821	1.0177	-.3761	-.0627	0.0139	0.0491	-.0212
	0.0665	0.0552	0.2831	0.5356	0.6530	0.6862	0.8119	1.0252	1.1146	1.1961	1.2288	1.4591	1.5158	1.7774	1.8927
WM	-.3483	0.0288	0.1436	0.2999	0.3983	0.6583	0.6938	0.8180	0.9336	0.8743	0.1602	0.2357	0.1586	-.0458	0.3986
	-.1882	0.1570	0.2523	0.5925	0.8709	1.0626	1.1567	1.0894	1.3252	1.4188	1.4083	1.4926	1.4926	1.5861	1.5885
BM	-.5692	-.2610	-.0153	0.1384	0.2633	0.3491	0.5916	0.5998	0.7184	0.7696	-.2700	-.1205	-.0501	-.0687	-.2249
	-.0751	-.0159	0.0745	0.3350	0.3058	0.5687	0.7836	0.9449	1.0114	1.2057	1.2934	1.3184	1.4770	1.5849	1.5574
SE$(\hat{\underset{\sim}{B}})$	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7
	LV75M8	LV75M9	U75M10	V75M11	V75M12	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
WF	0.1559	0.1171	0.0964	0.0796	0.0759	0.0740	0.0734	0.0750	0.0794	0.0853	0.3831	0.2753	0.2219	0.1637	0.1533
	0.1506	0.1492	0.1517	0.1634	0.1785	0.1910	0.1387	0.1120	0.0851	0.0801	0.0783	0.0776	0.0793	0.0852	0.0929
BF	0.0484	0.0420	0.0408	0.0425	0.0452	0.0510	0.0640	0.0720	0.0803	0.1111	0.1077	0.0852	0.0807	0.0882	0.0975
	0.1166	0.1467	0.1608	0.1766	0.2684	0.0548	0.0444	0.0423	0.0456	0.0501	0.0596	0.0750	0.0829	0.0913	0.1355
WM	0.1525	0.1103	0.0823	0.0828	0.0830	0.0827	0.0861	0.0881	0.0950	0.1182	0.3817	0.2667	0.1658	0.1756	0.1756
	0.1658	0.1768	0.1791	0.1989	0.2717	0.1897	0.1330	0.0867	0.0904	0.0906	0.0872	0.0921	0.0937	0.1034	0.1383
BM	0.0522	0.0449	0.0439	0.0480	0.0551	0.0663	0.0764	0.0898	0.0868	0.0942	0.1168	0.0916	0.0880	0.1040	0.1235
	0.1561	0.1753	0.2160	0.1815	0.1993	0.0594	0.0477	0.0460	0.0534	0.0632	0.0790	0.6896	0.1086	0.0951	0.1038

established in Theorem 5.2.1 are valid for small sample sizes. This problem is likely to arise in any longitudinal study due to attrition.

Another disadvantage of the MA(M) model assumption is the computational burden of the variance component estimation. The software developed as part of this research is executed under SAS PROC MATRIX. The computation of c_k and h_{kk} in (5.4.4) and (5.4.8) proved to be costly with respect to both storage and computing time. Gains in efficiency would probably be realized if these computations were programmed in PL/1 or Fortran and then linked to the SAS GIM software.

CHAPTER 8

SUGGESTIONS FOR FUTURE RESEARCH

Methods have been presented here for two classes of covariance models thought to be frequently occurring in longitudinal data. One obvious extension of this research is to consider other suitable models for the error structure. The model formulation and estimation theory established in Chapter 2 should be easily adapted to other time series models as well as to general patterns of covariance matrices appearing in the literature for which consistent estimation is possible. For example, small sample estimation of the parameters of an AR(2) process exists for complete data on a single time series (e.g., Box and Jenkins, 1976). Although the covariance structure associated with this model is more complicated than that of an AR(1) process, it may be possible to incorporate this model into the GIM.

Another extension of these methods is to determine maximum likelihood estimation of the covariance parameters for the models proposed here that are applicable to the missing data case. The EM algorithm may provide a suitable framework for such a procedure.

Further work is also needed in deriving suitable covariance constraints for the joint model described in Chapter 6. This problem involves minimizing the number of elements in $\underset{\sim}{\Omega}$ that must be estimated while keeping the underlying assumptions to a reasonable set.

Only large sample methods have been proposed here. A thorough study of the small sample properties of the estimators and test statistics is therefore suggested. Two simulation studies are particularly needed. The first would examine the reliability of the estimation and efficiency of the test statistics when one covariance model is assumed and another covariance model is actually present. A second study would determine the distribution (small sample) of the Wald statistic assuming that the covariance model is correct. It is hopefully the case that the imposition of covariance constraints will lessen the problems documented in the literature for the Wald tests in the unconstrained model (e.g., Leeper, et. al., 1982).

Andersen, A. H., Jensen, E. B. and Schou, G. Two-way analysis of variance with correlated errors. International Statistical Review, 1981, 49, 153-167.

Anderson, T. W. The Statistical Analysis of Time Series. New York: Wiley, 1971.

Anderson, T. W. Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. JASA, 1957, 52, 200-203.

Bhargava, R. Multivariate tests of hypotheses with incomplete data. Technical Report 3, Applied Mathematics and Statistics Laboratories, Stanford University, Stanford, CA, 1962.

Box, G. E. P. and Jenkins, G. M. Time Series Analysis Forecasting and Control. San Francisco, CA: Holden-Day, 1976.

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effectis of inequality of variance and of correlation between errors in the two-way classification. Annals of Mathematical Statistics, 1954, 25, 484-498.

Buck, S. F. A method of estimation of missing values in multivariate data suitable for use with an electronic computer. Journal of the Royal Statistical Society (B), 1960, 22, 302-307.

Cramer, H. Mathematical Methods of Statistics. Princeton, NJ: Princeton University Press, 1946.

Da Silva, J. G. C. The analysis of cross-sectional time series data. North Carolina State University Institute of Statistics Mimeo Series, No. 1011, 1975.

Danford, M. B., Hughes, H. M. and McNee, R. C. On the analysis of repeated-measurements experiments. Biometrics, 1960, 16, No. 4, 547-565.

Dear, R. E. A principal-component missing-data method for multiple regression models. System Development Corporation Report SP-86, 1959, Santa Monica, California.

Dempster, A. P., Rubin, D. B. and Tsutakawa, R. K. Estimates in covariance components models. Journal of the American Statistical Association, 1981, 26, 341-353.

Dempster, A. P., Laird, N. M. and Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society (B), 1977, 39, 1-22.

Dielman, T. E. Pooled cross-sectional and time series data: a survey of current statistical methodology. The American Statistician, 1983, 37, 111-122.

Drummond, D. J. and Gallant, A. R. TSCSREG: A SAS procedure for the analysis of time series cross-sectional data. North Carolina State University Institute of Statistics Mimeo Series, No. 1107, 1977.

Fuller, W. and Battese, G. E. Estimation of linear models with crossed-error structure. Journal of Econometrics, 1974, 2, 67-78.

Geisser, S. Multivariate analysis of variance for a special covariance case. JASA, 1963, 58, 660-669.

Glasser, M. Linear regression analysis with missing observations among the independent variables. Journal of the American Statistical Association, 1964, 59, 834-844.

Gleason, T. C. and Staelin, R. A proposal for handling missing data. Psychometrika, 1975, 40, 229-252.

Grizzle, J. E. and Allen, D. M. Analysis of growth and dose response curves, 1969, Biometrics, 25, 357-382.

Hannan, E. J. Regression for time series. Proceedings of the Symposium on Time Series Analysis. New York: Wiley, 1963.

Hartley, H. O. and Hocking, R. R. The analysis of incomplete data (with discussion). Biometrics, 1971, 27, 783-808.

Hocking, R. M. and Smith, W. B. Estimation of parameters in the multivariate normal distribution with missing observations. Journal of the American Statistical Association, 1968, 63, 159-173.

Hosking, J. D. Missing data in multivariate linear models: A comparison of several estimation techniques. Ph.D. Dissertation. University of North Carolina, Chapel Hill, NC, 1980.

Jöreskog, K. G. A general method for analysis of covariance structures. Biometrika, 1970, 57, 239-251.

Khatri, C. G. A note on MANOVA model applied to problems in growth curves. Amn. Inst. Stat. Math., 1966, 18, 75-86.

Kleinbaum, D. G. Testing linear hypotheses in generalized multivariate linear models. Communications in Statistics, 1973, 1, 433-457.

Kleinbaum, D. G. Estimation and hypothesis testing for generalized multivariate linear models. University of North Carolina Institute of Statistics Mimeo Series, No. 669, 1970.

Kmenta, J. Elements of Econometrics, New York: The Macmillan Co., 1971.

Laird, N. M. and Ware, J. H. Random effects models for longitudinal data. Biometrics, 1982, 38, 963-974.

Lee, Sik-Yum. Constrained estimation in covariance structure analysis. Biometrika, 1979, 66, 539-545.

Leeper, J. D. and Woolson, R. F. Testing hypotheses for the growth curve model when the data are incomplete. Journal of Statistical Computing and Simulation, 1982, 15, 97-107.

Machin, D. On design problems in growth studies. Biometrics, 1975, 31, 749-753.

Morrison, D. E. Multivariate Statistical Methods. New York: McGraw-Hill, 1976.

Morrison, D. E. The analysis of a single sample of repeated measurements. Biometrics, 1972, 28, 55-71.

Nicholson, G. E. Estimation of parameters from incomplete multivariate samples. JASA, 1957, 52, 523-526.

Orchard, T. and Woodbury, M. A. A missing information principal: Theory and applications. Sixth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1972, 697-715.

Parks, R. Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. Jounral of the American Statistical Association, 1967, 62, 500-509.

Potthoff, R. F. and Roy, S. N. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika, 1964, 51, 313-326.

Rao, C. R. Linear Statistical Inference and Its Applications. Second Edition. New York: Wiley, 1973.

Rao, C. R. Estimation of variance and covariance components in linear models. Journal of the American Statistical Association, 1972, 67, 112-115.

Rao, M. N. and Rao, C. R. Linked cross-sectional study for determining norms and growth rates - a pilot survey on Indian school-going boys. Sankhya, 1966, 28, 237-258.

Rubin, D. B. Inferences and missing data. Biometrika, 1976, 63, 581-590.

Rubin, D. B. Characterizing the estimation of parameters in incomplete data problems. JASA, 1974, 69, 467-474.

Schwertman, N. C. and Allen, D. M. The smoothing of an indefinite variance-covariance matrix. Journal of Statistical Computing and Simulation, 1979, 9, 183-194.

Searle, S. R. Topics in variance component estimation. Biometrics, 1971, 21, 1-76.

Seely, J. Estimation in finite-dimensional vector spaces with application to the mixed linear model. Ph.D. Dissertation, Department of Statistics, Iowa State University, Ames, IA, 1969.

Theil, H. Principals of Econometrics. New York: Wiley, 1971.

Timm, N. H. Estimation of the variance-covariance and correlation matrices from incomplete data. Psychometrika, 1970, 35, 415-437.

Trawinski, I. M. and Bargmann, R. E. Maximum likelihood estimation with incomplete multivariate data. Annals of Mathematical Statistics, 1964, 35, 647-657.

Wald, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Society, 1943, 54, 426-482.

Wilks, S. S. Moments and distributions of population parameters from fragmentary samples. Annals of Mathematical Statistics, 1932, 3, 163-193.

Wilks, S. S. Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. Amn. Math. Statist., 1946, 17, 257-281.

Woodbury, M. A. and Hasselblad, V. Maximum likelihood estimates of the variance-covariance matrix from the multivariate normal. Paper presented at the SHARE national meeting, Denver, March 1970.

Woolson, R. F. and Leeper, J. D. Growth curve analysis of complete and incomplete longitudinal data. Communications in Statistics, 1980, 14, 1491-1513.

Woolson, R. F., Leeper, J. D. and Clarke, W. R. Analysis of incomplete data from longitudinal and mixed longitudinal studies. Journal of the Royal Statistical Society (A), 1978, 141, 242-252.

APPENDIX A

A LISTING OF THE SCOR SPIROMETRY DATASET

OBS	RACE	SEX	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12
1	B	F	.	.	.	.	.	.	0.44	0.65	0.57	.
2	W	F	.	.	.	.	.	.	0.86	1.06	1.16	1.13
3	W	F	.	.	.	.	.	.	0.35	0.31	0.59	0.44
4	B	M	.	.	.	.	.	.	0.55	0.63	0.72	0.85
5	W	M	.	.	.	.	.	0.63	.	0.85	0.93	.
6	B	F	.	.	.	.	.	0.42	0.79	0.93	0.87	.
7	W	F	.	.	.	.	0.52	0.72	0.85	0.90	0.78	.
8	B	M	.	.	.	.	0.27	0.41	0.72	0.59	0.64	0.68
9	W	M	.	.	.	.	0.45	0.83	0.84	0.90	1.00	0.95
10	B	M	.	.	.	0.54	.	.	.	.	1.16	.
11	W	F	.	.	.	0.09	0.29	0.40	0.46	0.52	0.77	0.85
12	W	F	.	.	.	0.10	0.28	0.40	0.48	0.67	0.62	0.93
13	W	F	.	.	.	0.22	0.51	0.66	0.66	0.77	0.78	1.01
14	B	F	.	.	-0.01	.	0.25	.	.	.	.	.
15	W	M	.	.	0.00	0.08	.	.	.	.	.	.
16	B	F	.	.	0.00	0.22	.	.	.	.	.	.
17	B	M	.	.	0.03	0.03	.	.	.	.	.	.
18	W	F	.	.	0.28	0.28	0.49	0.78	.	.	.	.
19	W	M	.	.	0.21	0.49	0.45	0.55	0.65	.	.	.
20	B	F	.	.	0.12	0.22	0.64	0.67	0.82	.	1.09	.
21	B	M	.	.	-0.71	-0.56	-0.37	-0.54	-0.08	.	-0.14	-0.08
22	W	M	.	.	0.19	0.28	0.28	0.65	0.56	0.76	0.95	.
23	W	M	.	.	0.31	0.25	0.34	0.58	0.63	0.73	0.82	0.77
24	B	M	.	.	0.25	0.46	0.59	0.57	0.69	0.83	1.14	1.23
25	B	F	.	.	0.12	0.29	0.46	0.55	0.66	0.58	0.77	1.10
26	B	F	.	.	0.05	0.19	0.43	0.49	0.67	0.81	0.85	0.97
27	B	F	.	-0.46	.	.	.	.	.	.	.	.
28	W	M	.	0.05	0.17	.	.	.	.	.	.	.
29	B	F	.	.	.	.	.	.	.	.	.	.
30	B	M	.	-0.45	-0.02	.	.	.	.	.	.	.
31	B	F	.	-0.01	-0.14	.	.	.	.	.	.	.
32	B	M	.	-0.02	0.24	.	.	.	.	.	.	.
33	B	M	.	-0.04	0.22	.	.	.	.	.	.	.
34	B	F	.	0.17	0.36	.	.	.	.	.	.	.
35	B	M	.	-0.37	0.02	.	.	.	.	.	.	.

OBS	RACE	SEX	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12
36	B	M	.	-0.31	0.04	0.16	.	.	.	.	.	.
37	B	M	.	-0.48	-0.09	0.08	.	.	.	.	.	.
38	B	F	.	-0.29	0.01	0.14	.	.	.	.	.	.
39	B	F	.	-0.48	-0.12	-0.02	0.05	.	.	.	.	.
40	B	M	.	-0.21	0.10	0.25	0.41	.	.	.	.	.
41	B	F	.	-0.34	-0.08	0.25	0.28	.	.	.	.	.
42	B	F	.	-0.29	0.09	0.36	0.39	0.48	.	.	.	.
43	B	M	.	-0.22	0.02	0.17	0.31	0.58	.	.	.	.
44	W	F	.	0.05	0.09	0.37	0.47	0.57	.	0.95	.	.
45	B	M	.	-0.11	0.04	0.10	0.04	0.35	0.45	.	.	.
46	B	F	.	0.14	0.12	0.41	0.21	0.29	0.70	0.81	.	.
47	B	M	-0.45	.	.	.	.	.	.	.	.	.
48	B	M	-0.22	-0.27	.	.	.	.	.	.	.	.
49	B	F	-0.34	-0.04	.	0.43	0.41	.	.	.	.	.
50	B	M	-0.67	-0.43	-0.22	.	.	.	.	.	.	.
51	B	M	-0.51	0.12	0.16	.	.	.	.	.	.	.
52	B	F	-0.87	-0.42	-0.24	.	.	.	.	.	.	.
53	B	M	-0.69	-0.33	-0.04	.	.	.	.	.	.	.
54	B	M	-0.45	-0.25	0.12	.	.	.	.	.	.	.
55	B	F	-0.36	-0.22	0.03	.	0.44	.	.	.	.	.
56	B	F	-0.31	-0.13	0.22	0.34	.	.	.	.	.	.
57	B	M	-0.48	-0.07	0.04	0.09	.	.	.	.	.	.
58	B	F	-0.42	-0.30	0.12	0.27	.	.	.	.	.	.
59	B	F	-0.27	-0.12	0.01	0.07	.	0.55	.	.	.	.
60	B	F	-0.63	-0.45	-0.12	0.14	.	0.44	0.57	0.49	.	.
61	B	F	-0.65	-0.30	-0.31	-0.13	-0.15	.	.	.	.	.
62	B	M	-0.33	-0.02	0.04	0.27	0.48	.	.	.	.	.
63	B	F	-0.36	-0.22	-0.02	0.22	0.33	.	.	.	.	.
64	B	F	-0.54	-0.18	0.17	-0.01	0.21	.	.	.	.	.
65	B	M	-0.97	-0.62	-0.45	-0.09	-0.02	.	.	.	.	.
66	B	M	-0.73	-0.54	-0.15	-0.01	0.17	.	.	.	.	.
67	B	F	-0.51	-0.33	-0.07	0.10	0.26	0.32	.	.	.	.
68	B	F	-0.48	-0.29	-0.29	-0.02	0.30	0.45	.	.	.	.
69	B	M	-0.80	-0.27	-0.22	0.07	0.18	0.24	.	.	.	.

OBS	RACE	SEX	LFVC3	LFVC4	LFVC5	LFVC6	LFVC7	LFVC8	LFVC9	LFVC10	LFVC11	LFVC12
70	B	F	-0.36	-0.02	0.03	0.16	0.51	0.53	.	.	.	.
71	W	F	-0.76	-0.36	0.01	-0.05	0.36	0.54	0.39	.	.	.
72	W	M	-0.40	-0.05	0.04	0.43	0.71	0.84	0.90	0.97	.	.

OBS	RACE	SEX	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7	LV75M8	LV75M9	LV75M10	LV75M11	LV75M12
1	B	F	.	.	.	.	.	.	-0.37	0.91	0.39	.
2	W	F	.	.	.	.	.	.	0.40	0.56	0.74	0.65
3	W	F	.	.	.	.	.	.	0.49	0.13	0.42	0.25
4	B	M	.	.	.	.	.	.	-0.21	-0.25	0.12	0.40
5	W	M	.	.	.	.	.	-0.12	.	0.27	0.39	.
6	B	F	.	.	.	.	.	0.92	0.41	0.28	0.38	.
7	W	F	.	.	.	.	-0.14	-0.36	-0.08	-0.12	-0.12	.
8	B	M	.	.	.	.	0.76	0.51	0.42	0.28	0.77	0.51
9	W	M	.	.	.	.	0.25	-0.73	0.12	-0.09	0.26	0.58
10	B	M	.	.	.	0.77	.	.	.	.	.	.
11	W	F	.	.	.	-0.02	0.10	0.15	0.39	0.50	0.77	0.81
12	W	F	.	.	.	-0.24	0.12	0.32	0.32	0.85	0.61	1.08
13	W	F	.	.	.	0.00	0.31	0.34	0.33	0.18	0.34	0.42
14	B	F	.	.	-0.02	.	-0.69	.	.	.	.	.
15	W	M	.	.	.	0.32	.	.	.	.	.	.
16	B	F	.	.	-0.01	0.41	.	.	.	.	.	.
17	B	M	.	.	-0.58	-0.58	.	.	.	.	.	.
18	W	F	.	.	-0.25	0.34	0.33	0.28	.	.	.	.
19	W	M	.	.	0.17	-1.31	0.48	-0.63	-0.36	.	.	.
20	B	F	.	.	-0.33	0.00	0.18	0.37	0.05	.	0.69	.
21	B	M	.	.	-0.97	0.11	-0.39	0.26	0.35	.	-0.30	-0.20
22	W	M	.	.	0.20	0.31	0.75	0.26	0.60	0.43	0.48	.
23	W	M	.	.	0.16	0.17	0.26	0.49	0.24	0.41	0.39	0.56
24	B	M	.	.	-0.37	-0.29	-0.21	-0.06	0.01	0.34	0.18	0.25
25	B	F	.	.	0.16	0.71	0.44	0.47	0.39	0.11	0.73	0.70
26	B	F	.	.	0.17	-0.20	0.20	0.41	0.73	0.71	0.59	0.72
27	B	F	.	0.26	.	.	.	.	.	.	.	.
28	W	M	.	0.11	0.10	.	.	.	.	.	.	.
29	B	F	.	0.22	0.03	.	.	.	.	.	.	.
30	B	M	.	0.00	0.28	.	.	.	.	.	.	.
31	B	F	.	0.18	0.13	.	.	.	.	.	.	.
32	B	M	.	0.25	0.39	.	.	.	.	.	.	.
33	B	M	.	0.54	0.35	.	.	.	.	.	.	.
34	B	F	.	-0.36	-0.39	.	.	.	.	.	.	.
35	B	M	.	-0.46	-0.45	.	.	.	.	.	.	.

OBS	RACE	SEX	LV75M3	LV75M4	LV75M5	LV75M6	LV75M7	LV75M8	LV75M9	LV75M10	LV75M11	LV75M12
36	B	M	.	-0.01	0.25	0.22	.	.	.	.	.	.
37	B	M	.	-0.15	0.06	0.00	.	.	.	.	.	.
38	B	F	.	0.27	0.15	0.23	.	.	.	.	.	.
39	B	F	.	-0.24	0.06	-0.26	0.28	.	.	.	.	.
40	B	M	.	0.18	0.48	-0.49	-0.42	.	.	.	.	.
41	B	F	.	-0.42	-0.43	-0.80	-0.62	.	.	.	.	.
42	B	F	.	0.14	-0.11	-0.67	-0.45	0.04	.	.	.	.
43	B	M	.	-0.04	0.08	-0.04	-0.04	-0.12	.	.	.	.
44	W	F	.	-0.37	0.53	0.59	0.52	0.59	.	0.79	.	.
45	B	M	.	-0.82	0.14	-0.22	-0.43	-0.33	-0.49	.	.	.
46	B	F	.	-0.20	0.10	-0.07	0.02	-0.25	0.00	0.25	.	.
47	B	M	-0.33	.	.	.	.	.	.	.	.	.
48	B	M	-0.49	-0.36	.	.	.	.	.	.	.	.
49	B	F	-0.92	-1.90	.	-0.43	-0.37	.	.	.	.	.
50	B	M	-0.36	-0.54	0.06	.	.	.	.	.	.	.
51	B	M	-0.22	-0.03	0.04	.	.	.	.	.	.	.
52	B	F	0.05	0.23	-0.21	.	.	.	.	.	.	.
53	B	M	-0.84	-0.12	-0.76	.	.	.	.	.	.	.
54	B	M	-0.01	-0.04	-0.30	.	.	.	.	.	.	.
55	B	F	0.20	0.46	0.53	.	0.53	.	.	.	.	.
56	B	F	-0.09	-0.30	0.18	0.22	.	.	.	.	.	.
57	B	M	0.48	0.33	0.57	0.48	.	.	.	.	.	.
58	B	F	-0.36	-0.19	-0.05	-0.33	.	.	.	.	.	.
59	B	F	-0.84	0.10	0.01	0.44	.	-0.42	.	.	.	.
60	B	F	0.21	0.23	0.17	0.29	.	0.67	0.09	0.51	.	.
61	B	F	0.18	0.18	0.31	0.53	0.36	.	.	.	.	.
62	B	M	-0.06	-0.29	0.06	0.58	0.02	.	.	.	.	.
63	B	F	0.14	-0.15	-0.04	-0.34	-0.40	.	.	.	.	.
64	B	F	-0.33	0.20	0.29	0.33	0.29	.	.	.	.	.
65	B	M	-0.62	-0.49	-0.39	-0.87	-1.11	.	.	.	.	.
66	B	M	-0.54	0.10	-0.46	-0.84	-0.69	.	.	.	.	.
67	B	F	-2.21	-0.02	0.00	0.38	-0.27	-0.40	÷	÷	÷	÷
68	B	F	-0.99	0.30	0.29	0.23	-0.04	-0.27	.	.	.	.
69	B	M	0.03	0.22	-0.02	-0.17	-0.29	-0.37	.	.	.	.
70	B	F	-0.12	-0.39	0.03	0.23	-0.43	-0.40	.	.	.	.
71	W	F	0.25	0.33	-0.31	0.60	0.45	0.27	0.57	.	.	.
72	W	M	0.25	0.44	0.42	0.53	0.46	0.44	0.45	0.47	.	.

OBS	RACE	SEX	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
1	B	F	.	.	.	.	.	.	0.84	1.31	1.50	.
2	W	F	.	.	.	.	.	.	1.55	1.70	1.93	1.75
3	W	F	.	.	.	.	.	.	1.46	1.47	1.61	1.32
4	B	M	.	.	.	.	.	.	1.46	1.52	1.64	1.64
5	W	M	.	.	.	.	.	1.57	.	1.71	1.68	.
6	B	F	.	.	.	.	.	1.50	1.85	2.06	1.79	.
7	W	F	.	.	.	.	1.10	1.25	1.52	1.40	1.24	.
8	B	M	.	.	.	.	1.53	1.53	1.82	1.69	1.71	1.70
9	W	M	.	.	.	.	1.37	1.38	1.61	1.53	1.64	1.60
10	B	M	.	.	.	1.25	.	.	.	.	1.77	.
11	W	F	.	.	.	1.12	1.25	1.57	1.33	1.38	1.55	1.64
12	W	F	.	.	.	1.10	0.91	1.45	1.31	1.61	1.63	1.80
13	W	F	.	.	.	1.08	1.37	1.59	1.37	1.59	1.60	1.78
14	B	F	.	.	1.02	.	0.98	.	.	.	.	.
15	W	M	.	.	0.97	0.88	.	.	.	.	.	.
16	B	F	.	.	0.75	1.28	.	.	.	.	.	.
17	B	M	.	.	0.95	0.87	.	.	.	.	.	.
18	W	F	.	.	1.11	0.94	1.32	1.21	.	.	.	.
19	W	M	.	.	1.12	1.14	1.40	1.29	1.24	.	.	.
20	B	F	.	.	1.32	1.26	1.80	1.72	1.87	.	2.12	.
21	B	M	.	.	0.53	0.71	0.96	1.36	0.90	.	1.05	0.78
22	W	M	.	.	1.28	1.00	1.24	1.38	1.09	1.21	1.45	.
23	W	M	.	.	1.24	1.11	1.37	1.47	1.51	1.60	1.63	1.63
24	B	M	.	.	1.05	1.26	1.64	1.45	1.39	1.75	1.90	2.08
25	B	F	.	.	1.04	1.25	1.49	1.49	1.46	1.38	1.69	1.93
26	B	F	.	.	1.18	1.36	1.23	1.21	1.38	1.23	1.93	1.89
27	B	F	.	0.78	.	.	.	.	.	.	.	.
28	W	M	.	1.11	1.20	.	.	.	.	.	.	.
29	B	F	.	0.90	1.30	.	.	.	.	.	.	.
30	B	M	.	0.69	1.10	.	.	.	.	.	.	.
31	B	F	.	1.06	1.12	.	.	.	.	.	.	.
32	B	M	.	1.07	1.04	.	.	.	.	.	.	.
33	B	M	.	1.11	1.27	.	.	.	.	.	.	.
34	B	F	.	0.73	1.05	.	.	.	.	.	.	.
35	B	M	.	0.33	0.77	.	.	.	.	.	.	.

OBS	RACE	SEX	LPEF3	LPEF4	LPEF5	LPEF6	LPEF7	LPEF8	LPEF9	LPEF10	LPEF11	LPEF12
36	B	M	.	0.54	0.91	0.97	.	.	.	.	.	.
37	B	M	.	0.69	0.99	1.05	.	.	.	.	.	.
38	B	F	.	0.77	1.07	0.95	.	.	.	.	.	.
39	B	F	.	0.73	0.97	1.12	1.02	.	.	.	.	.
40	B	M	.	0.66	1.16	0.95	1.21	.	.	.	.	.
41	B	F	.	0.64	1.07	1.24	1.15	.	.	.	.	.
42	B	F	.	0.81	0.98	1.19	1.33	1.34	.	.	.	.
43	B	M	.	0.92	1.08	1.28	1.26	1.50	.	.	.	.
44	W	F	.	1.07	1.23	1.40	1.59	1.44	.	1.72	.	.
45	B	M	.	0.82	0.99	0.81	0.91	1.02	1.11	.	.	.
46	B	F	.	0.80	1.04	0.73	0.81	0.76	1.44	1.39	.	.
47	B	M	0.89	.	.	.	.	.	.	.	.	.
48	B	M	0.41	0.56	.	.	.	.	.	.	.	.
49	B	F	0.36	0.90	.	1.15	1.05	.	.	.	.	.
50	B	M	0.34	0.62	0.69	.	.	.	.	.	.	.
51	B	M	0.59	1.25	1.26	.	.	.	.	.	.	.
52	B	F	0.68	0.88	0.95	.	.	.	.	.	.	.
53	B	M	0.25	0.58	0.84	.	.	.	.	.	.	.
54	B	M	0.77	0.84	1.01	.	.	.	.	.	.	.
55	B	F	0.88	0.90	1.07	.	1.29	.	.	.	.	.
56	B	F	0.77	0.85	1.23	1.13	.	.	.	.	.	.
57	B	M	0.98	1.15	0.85	0.87	.	.	.	.	.	.
58	B	F	0.57	0.59	1.05	1.17	.	.	.	.	.	.
59	B	F	0.50	0.65	1.02	1.00	.	1.18	.	.	.	.
60	B	F	0.62	0.87	0.84	1.12	.	1.26	1.50	1.47	.	.
61	B	F	0.95	1.08	1.11	1.15	1.25	.	.	.	.	.
62	B	M	0.94	0.90	1.01	1.31	1.32	.	.	.	.	.
63	B	F	0.90	0.85	0.96	1.18	1.12	.	.	.	.	.
64	B	F	0.67	0.93	1.22	0.99	1.28	.	.	.	.	.
65	B	M	0.10	0.30	0.70	1.16	1.04	.	.	.	.	.
66	B	M	0.60	0.94	0.71	0.44	0.90	.	.	.	.	.
67	B	F	0.61	0.49	0.80	0.82	0.97	1.11	.	.	.	.
68	B	F	0.74	0.79	0.79	0.97	1.10	1.21	.	.	.	.
69	B	M	0.62	0.81	0.74	0.86	1.12	0.91	.	.	.	.
70	B	F	0.63	0.77	0.66	1.11	1.18	1.23	.	.	.	.
71	W	F	0.48	0.85	1.17	1.24	1.22	1.41	1.25	.	.	.
72	W	M	0.82	0.97	1.10	1.23	1.28	1.48	1.47	1.52	.	.