ESTIMATION AND HYPOTHESIS TESTING FOR GENERALIZED MULTIVARIATE LINEAR MODELS

by

David G. Kleinbaum
Department of Statistics
Mimeo Series 669
University of North Carolina at Chapel Hill

FEBRUARY 1970

d

ESTIMATION AND HYPOTHESIS TESTING FOR GENERALIZED MULTIVARIATE LINEAR MODELS

by

David G. Kleinbaum

A thesis 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 Statistics

Chapel Hill

1970

Approved by:

____

Adviser

DAVID GEORGE KLEINBAUM. Estimation and Hypothesis Testing for Generalized Multivariate Linear Models. (Under the direction of NORMAN L. JOHNSON.)

This study investigates estimation and testing problems for multivariate linear models in which observations may be missing and/or in which different design matrices correspond to different response variates. Two new models called the More General Linear Multivariate (MGLM) model and the Generalized Growth Curve Multivariate (GGCM) model are defined to deal with these situations.

For the MGLM model, the problem of obtaining best linear unbiased estimates of estimable linear functions and sets of the unknown treatment parameters is discussed. Also, unbiased and consistent estimates of the variance-covariance matrix are obtained. Another method of estimation discussed is that of obtaining estimators which are best asymptotically normal (BAN). Several such estimators are proposed for the MGLM and GGCM models.

For the problem of testing linear hypotheses in the MGLM and GGCM models, a testing method for large samples is described which uses a test criterion given in general form by Wald. The asymptotic distribution of the test statistic is a central chi-square variable. A BAN estimator of a linear set of the treatment parameters and consistent estimators of the variance-covariance parameters are needed for its computation.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I wish to express particular gratitude to Dr. Norman L. Johnson and Dr. James E. Grizzle for the guidance that made this study possible. I also wish to thank the other members of my Doctoral Committee--Dr. P. K. Sen, Dr. I. M. Chakravarti, Dr. M. R. Leadbetter and Dr. A. R. Bolstein for their various contributions. In addition, I wish to thank Dr. S. John of the Australian National University for his advice during the early part of this study and Dr. A. W. Davis of CSIRO, Australia, for his help during the latter stages of this study. Also, I am grateful to Mr. James O. Kitchen who provided invaluable assistance for the computer work of Chapter Five.

Finally, a special tribute must be given my wife, Anna, who typed both the original and final drafts, provided encouragement and companionship, and generally was instrumental in helping me keep my sanity during my course of graduate study.

This research was supported in part by the Department of Biostatistics through the National Institutes of Health, Institute of General Medical Sciences Grant No. 12868-06, and in part by the Department of Statistics through Air Force Grant No. AFOSR-68-1415.

NOTATION AND ABBREVIATIONS

$\underline{a}(q \times 1)$	is a $(q \times 1)$ column vector, and $\underline{a}'$ is the corresponding $(1 \times q)$ row vector.
$A(p \times q) = ((a_{ij}))$	is a $(p \times q)$ matrix with a_{ij} as the element in the i-th row and j-th column.
$\langle A \rangle_{ij} = a_{ij}$	is the element in the i-th row and j-th column of the matrix A.
$A = ((A_{ij}))$	is a partitioned matrix in which A_{ij} is the sub-matrix in the i-th row and j-th column.
$R(A)$	is the rank of the matrix A.
$V(A)$	is the vector space spanned by the rows of A.
A'	is the transpose of A.
$\mathrm{tr}\ A$	is the trace of A.
A^{-1}	is the unique inverse of a square matrix A of full rank.
A^{-}	is any generalized inverse of the matrix A and is defined by $AA^{-}A = A$.
$ch_i(A)$	is the i-th largest characteristic root of A.
$A \otimes B$	is the Kronecker Product of the matrix A and B defined by $A \otimes B = ((a_{ij}B))$ where $A = ((a_{ij}))$.
I_q	is the identity matrix of order q.

$0_{p,q}$	is the $(p \times q)$ matrix of zeros.
$\underline{o}_p$	is the $(p \times 1)$ vector of zeros.
$\underline{1}_n$	is the $(n \times 1)$ column vector of all unities.
$\Omega^{\frac{1}{2}}$	is the square root of a symmetric matrix Ω defined by $\Omega^{\frac{1}{2}} = C'\Lambda C$ where C is an orthogonal matrix and Λ is a diagonal matrix such that $\Omega = C'\Lambda^2 C$.

For $\underline{x} = (x_1, \ldots, x_n)'$ and $\underline{y} = (y_1, \ldots, y_m)'$,

Cov $(\underline{x}, \underline{y})$	is the $(n \times m)$ matrix with $\text{Cov}(x_i, y_j)$ in the i-th row and j-th column;
$\text{Var}(\underline{x})$	is the $(n \times n)$ matrix $\text{Cov}(\underline{x}, \underline{x})$.

For $Y(n \times p) = ((y_{ij}))$, a matrix of random variables,

E(Y)	is the $(n \times p)$ matrix of expectations of the elements of Y, i.e., $E(Y) = ((Ey_{ij}))$;
Var(Y)	is the $(np \times np)$ variance-covariance matrix of the $(np \times 1)$ vector defined by putting the <u>rows</u> of Y underneath each other in a long column vector.
$\underline{x} \sim \mathcal{N}_p(\underline{\mu}, \Sigma)$	means that the random variable $\underline{x}$ has a p-variate multinormal distribution with mean vector $\underline{\mu}$ and variance covariance matrix Σ.
$x \sim \chi^2_c$	means that the scalar random variable x is distributed as a central chi-square variable with c degrees of freedom.

$\underline{x}_n \xrightarrow[L]{} \underline{x}$	means that the random variable $\underline{x}_n$ converges <u>in law</u> to the random variable $\underline{x}$.
$\mathcal{L}(\underline{x}_n) \to F$	means $\underline{x}_n \xrightarrow[L]{} \underline{x}$ where $\underline{x}$ has distribution function F.
$X_n \xrightarrow[p]{} A$	means that each element of the matrix X_n converges in probability to the corresponding element of A.
p.d.	means positive definite.
i.i.d.	means independently and identically distributed.

INTRODUCTION AND SUMMARY

The main purpose of this study is to investigate estimation and testing problems for multivariate linear models which describe experimental situations that cannot be analyzed under the standard parametric multivariate general linear model (hereafter referred to as the SM model).

Chapter One contains a discussion of the SM model and the growth curve multivariate (GCM) model, which are the two linear models that have been extensively dealt with in the literature and used in applications. Situations which cannot be described by these models are then discussed. These occur when there are missing data and when different response variates have different design matrices. A new model generalizing the SM model called the More General Linear Multivariate (MGLM) model is then proposed to incorporate the new situations. Special cases of the MGLM model which have been dealt with in the literature are described. In addition, the GCM model, which is not a special case of the MGLM model, is generalized to allow for gaps in the data.

Chapter Two considers the problem of BLUE estimation in the MGLM model of estimable linear functions and sets of the unknown treatment parameters $\underline{\xi} = (\underline{\xi}_1', \ldots, \underline{\xi}_p')'$ and unbiased and consistent estimation of the variance-covariance matrix Σ. In the discussion of BLUE estimation, necessary and sufficient conditions are given for the existence of a unique BLUE in terms of known linear functions (or sets) of the data for the entire experiment, that is, functions which do not depend on the unknown Σ. The method of proof simplifies and clarifies considerably the work of Srivastava [60][1] for some special cases of the MGLM model. An unbiased and consistent estimate of Σ is then obtained and used to determine an unbiased estimator of the variance of any linear function of the data.

Chapter Three deals with Best Asymptotically Normal (BAN) estimation of estimable linear sets $H'\underline{\xi}$. BAN estimators are usually non-linear functions of the data. They have variances which in large samples are the minimum that would be achieved by linear estimators if Σ were known. Also, they are needed for computing the Wald Statistic described in Chapter Four, which is used for testing linear hypotheses in the MGLM and Generalized Growth Curve (GGCM) models. Section 3.2 considers the calculation for the MGLM and GGCM models of the information matrix for $\underline{\xi}$ and the asymptotic variance matrix of any estimable set $H'\underline{\xi}$ where $H(M \times w)$ is of full rank w. Use of the rules of matrix differentiation (as described in the Appendix) simplifies considerably these derivations. In the next two sections the literature is reviewed for some important special cases of the MGLM model for which BAN estimators have been obtained. One of these is a general multivariate model for missing observations called the General Incomplete Multivariate (GIM) model. For a special case of the GIM model called the Hierarchal Incomplete (HIM) model, in which the missing data have a simple monotonic structure, maximum likelihood estimators are

[1] The numbers in square brackets refer to the bibliography.

obtained using a method of Anderson [7]. This result generalizes some work of Bhargava [13]. A third special case of the MGLM model allows different response variates to have different design matrices. This model is called the Multi-Design Multivariate (MDM) model. The MDM has been dealt with in the literature (e.g., [61], [72]) under the title "Seemingly Unrelated Regressions." In Section 3.5 a BAN estimator for the MGLM model is then proposed which generalizes an estimator obtained by Zellner [72] for the MDM model. Alternative estimators may be obtained by a method of iteration suggested by Zellner or by the method of scoring. Finally, a BAN estimator for the GGCM model is obtained.

In Chapter Four a method is proposed for obtaining test criteria for testing linear hypotheses for the MGLM model when the underlying data are normally distributed. The test statistics obtained are quadratic forms called Wald Statistics, and they are constructed from BAN estimators of estimable sets $H'\underline{\xi}$ and consistent estimators of the variance-covariance parameters. The asymptotic distribution of the Wald Statistic under the null hypothesis is a central chi-square variable. When applied to the special case of the SM model, the method yields Hotelling's trace criterion. In addition, the testing method is applied to growth curve models. For the GCM model, a competitor to Khatri's trace [31] which uses S, the sample covariance matrix, is obtained. For the GGCM model, two competitors are proposed, one being the Wald Statistic and the other, a generalization of Khatri's trace.

In Chapter Five some practical examples are analyzed to illustrate the proposed BAN estimator and the corresponding Wald Statistic.

Chapter Six contains recommendations for further research.

Some matrix results which are important for use in the main text are given in the Appendix. Although some of these results are new (in particular, those dealing with the differentiation of a matrix with respect to another matrix), all are easily established from the definitions of generalized inverses, Kronecker-delta products and matrix differentiation. Proofs are therefore not given.

CHAPTER I

MULTIVARIATE LINEAR MODELS

1.1 The Standard Multivariate General Linear Model (SM)

Consider an experiment in which there are n individuals or experimental units and on each individual p measurements are taken. Each measurement is a univariate random variable which represents the value of a particular characteristic (e.g., height, weight, blood pressure) of interest or the same characteristic measured under different conditions. The different characteristics or conditions, as the case may be, are frequently referred to as response variates. If we label the p response variates as $V_1, V_2, \ldots, V_p$ and denote the measurement of the j-th variate V_j on the i-th individual by y_{ij}, then the results of the experiment may be described as:

Individual \ Response Variate	V_1	V_2	. . .	V_p
1	y_{11}	y_{12}	. . .	y_{1p}
2	y_{21}	y_{22}	. . .	y_{2p}
⋮	⋮	⋮		⋮
n	y_{n1}	y_{n2}	. . .	y_{np}

TABLE 1.1
DATA FOR A MULTIVARIATE EXPERIMENT

The data obtained from the whole experiment may thus be compactly summarized by the $(n \times p)$ matrix $Y = ((y_{ij}))$. If we assume that the standard Multivariate General Linear Model (SM) describes the experimental situation, then we can write

$$(1.1.1) \qquad E(Y) = A\,\xi$$

$$\operatorname{Var} Y = I_n \otimes \Sigma$$

where A is a $(n \times m)$ known design matrix of rank $R(A) = r\ (\leq m \leq n)$

ξ is a $(m \times p)$ matrix of unknown parameters

and $\Sigma = ((\sigma_{rs}))$ is a $(p \times p)$ positive definite (p.d.) matrix of (usually) unknown parameters which represents the variance-covariance matrix of any row of Y.

It is clear from (1.1.1) that in the SM model, the measurements on different individuals are assumed to be uncorrelated whereas the measurements of the p response variates on the same individual may be correlated. The expectation of each measurement is assumed to be a linear function of the unknown parameters of ξ. For testing hypotheses and interval estimation in the SM model, it is usually assumed that the rows of Y are each distributed according to the multivariate normal distribution.

An alternative way to represent the SM model, which will provide some insight for analysis of the generalized models discussed later in the chapter, can be given if we look at the experiment variate-wise, and this will be referred to as the variate-wise representation of the model:

If we let $\underline{y}_s (n \times 1)$ denote the s-th column of Y and $\underline{\xi}_s (m \times 1)$

denote the s-th column of ξ, so that

$$Y = [\underline{y}_1\ \underline{y}_2\ \cdots\ \underline{y}_p] \quad \text{and } \xi = [\underline{\xi}_1\ \underline{\xi}_2\ \cdots\ \underline{\xi}_p] ,$$

then the SM model may be described as

$$(1.1.2) \qquad E(\underline{y}_s) = A\underline{\xi}_s , \ \mathrm{Var}(\underline{y}_s) = \sigma_{ss}I_n , \ s=1,\ldots,p ;$$

$$\mathrm{Cov}(\underline{y}_r, \underline{y}_s) = \sigma_{rs}I_n \quad \text{when } r \neq s .$$

Thus the variate-wise representation gives for each response variate, the univariate model that would describe the experiment if only this variate were observed. These p separate univariate models are related by the $\frac{p(p-1)}{2}$ covariances between the different variate pairs.

A third way to represent the SM model, which will be referred to as the <u>vector version</u>, results from rewriting Y and ξ as

$$\underline{y} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_p \end{bmatrix} \quad \text{and} \quad \underline{\xi} = \begin{bmatrix} \underline{\xi}_1 \\ \underline{\xi}_2 \\ \vdots \\ \underline{\xi}_p \end{bmatrix} , \text{ respectively.}$$

Then we have

$$(1.1.3) \qquad E(\underline{y}) = D\underline{\xi}$$

$$\mathrm{Var}\ \underline{y} = \Omega$$

where $\qquad D = I_p \otimes A \quad$ and $\quad \Omega = \Sigma \otimes I_n$.

If Ω is known, it is clear that the vector version is equivalent to the univariate weighted least squares model.

1.1.1 Estimation in the SM Model

In the SM model, it is usually of interest to estimate linear functions of the form $\underline{c}'\xi\underline{d}$ and linear sets of the form $C\,\xi D$ where $\underline{c}(m \times 1)$, $\underline{d}(p \times 1)$, $C(g \times m)$ and $D(p \times v)$ are known. The estimates are usually required to be linear functions or sets of the data and the usual estimation criteria are unbiasedness (estimability) and minimum variance. Under the assumption of normality on the rows of Y, a maximum likelihood approach may yield such estimates. It is also of interest to determine unbiased estimators of Σ and of the variances of the estimators of $\underline{c}'\xi\underline{d}$ and $C\xi D$.

If we write

$$(1.1.1.1)\qquad \underline{\xi}' = (\underline{\xi}_1' \ \underline{\xi}_2' \ \cdots \ \underline{\xi}_p')$$

where $\underline{\xi}_s$ is the s-th column of ξ, $s=1,\ldots,p$ and we let

$$\underline{d}' = (d_1, d_2, \ldots, d_p)$$

$$\underline{c}_s' = d_s\underline{c}', \ s=1,\ldots,p$$

it is easily seen that

$$\underline{c}'\xi\underline{d} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s = \underline{h}'\underline{\xi}$$

where

$$\underline{h}' = (\underline{c}_1' \ \underline{c}_2' \ \cdots \ \underline{c}_p') .$$

Thus we can consider the more general problem of estimating $\underline{h}'\underline{\xi}$ where $\underline{h}$ is arbitrary and apply the results obtained to the special case $\underline{c}'\xi\underline{d}$. Similarly, instead of $C\xi D$ we can consider the more general problem of estimating $H'\underline{\xi}$ where $H' = [C_1' \mid C_2' \mid \dots \mid C_p']$ is an arbitrary matrix of known constants, and we can apply the results obtained to the special case $C\xi D$ by letting

$$C_s'(gv \times m) = \underline{d}_s \otimes C \ , \ s=1,\dots,p$$

where $\underline{d}_s'(1 \times v)$ is the s-th row of D so that

$$H' = D' \otimes C \ .$$

Now it is well known (e.g., Roy [49], Srivastava [60]) that the conditions for estimability of $\underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$ are that $\underline{c}_s' \in V(A)$, $s=1,\dots,p$, where $V(A)$ is the vector space spanned by the rows of A. For the special case $\underline{c}'\xi\underline{d}$ the conditions simplify to $\underline{c}' \in V(A)$, without any requirement on $\underline{d}$. Similarly the conditions for estimability of $H'\underline{\xi} = \sum_{s=1}^{p} C_s'\underline{\xi}_s$ are that the rows of C_s' belong to $V(A)$ for each s. For the special case $C\xi D$ this simplifies to the condition that the rows of C belong to $V(A)$, without any requirement on D.

Rao [41], using generalized inverses, and Roy [49], using the reparameterization technique, have shown that the best linear unbiased estimate (BLUE) of $\underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$ when it is estimable, is given by the sum of the BLUE's for $\underline{c}_s'\underline{\xi}_s$ obtained separately from the univariate models $E(\underline{y}_s) = A\underline{\xi}_s$, Var $\underline{y}_s = \sigma_{ss}I_n$, $s=1,\ldots,p$. More specifically, the BLUE for $\underline{h}'\underline{\xi}$ is given by

$$(1.1.1.2) \qquad \sum_{s=1}^{p} \underline{c}_s'(A'A)^{-}A'\underline{y}_s$$

where $(A'A)^{-}$ denotes any generalized inverse of $A'A$.
Alternatively we can write this BLUE as

$$(1.1.1.3) \qquad \sum_{s=1}^{p} \underline{c}_{sI}'(A_I'A_I)^{-1} A_I'\underline{y}_s$$

where

$$A = (A_I \vdots A_D) \text{ and } A_I(n \times r) \text{ is a basis for } A$$

and

$$\underline{c}_s' = (\underline{c}_{sI}' \vdots \underline{c}_{sD}') \text{ and } \underline{c}_{sI}' \text{ is } (1 \times r).$$

The variance of this estimate is given by

$$\sum_{r=1}^{p} \sum_{s=1}^{p} \underline{c}_s'(A'A)^{-}\underline{c}_r\sigma_{rs} \quad \text{or} \quad \sum_{r=1}^{p} \sum_{s=1}^{p} \underline{c}_{sI}'(A_I'A_I)^{-1} \underline{c}_{rI}\sigma_{rs} .$$

For estimating an estimable linear set $H'\underline{\xi} = \sum_{s=1}^{p} C_s'\underline{\xi}_s$,

Roy [49] suggests using the sum of the BLUE's for the linear sets $C_s'\underline{\xi}_s$ obtained separately from the univariate models $E(\underline{y}_s) = A\underline{\xi}_s$, Var $\underline{y}_s = \sigma_{ss}I_n$, $s=1,\ldots,p$. The estimate is thus

$$(1.1.1.4) \qquad \sum_{s=1}^{p} C_s'(A'A)^{-}A'\underline{y}_s \text{ or } \sum_{s=1}^{p} C_{sI}'(A_I'A_I)^{-1} A_I'\underline{y}_s$$

where C'_{sI} is a matrix of r columns obtained from

$$C'_s = [C'_{sI} \mid C'_{sD}], \; s=1,\ldots,p.$$

The dispersion matrix of this estimate is given by

$$\sum_{r=1}^{p} \sum_{s=1}^{p} C'_s(A'A)^- C_r \sigma_{rs} \quad \text{or} \quad \sum_{r=1}^{p} \sum_{s=1}^{p} C'_{sI}(A'_I A_I)^{-1} C_{rI}\sigma_{rs} .$$

(Henceforth we will only use the generalized inverse description.)

Roy has shown that if $\sum_{s=1}^{p} B'_s \underset{\sim}{y}_s$ is any competitor to $\sum_{s=1}^{p} C'_s(A'A)^- A'\underset{\sim}{y}_s$,

then

(1.1.1.5) (i) $\mathrm{Var}(\sum_{s=1}^{p} B'_s \underset{\sim}{y}_s) - \mathrm{Var}(\sum_{s=1}^{p} C'_s(A'A)^- A'\underset{\sim}{y}_s)$ is positive semi-definite,

(ii) $ch_{max}[\mathrm{Var}(\sum_{s=1}^{p} B'_s \underset{\sim}{y}_s)] \geq ch_{max}[\mathrm{Var}(\sum_{s=1}^{p} C'_s(A'A)^- A'\underset{\sim}{y}_s)]$,

(iii) $tr[\mathrm{Var}(\sum_{s=1}^{p} B'_s \underset{\sim}{y}_s)] > tr[\mathrm{Var}(\sum_{s=1}^{p} C'_s(A'A)^- A'\underset{\sim}{y}_s)]$.

(iv) $\left| \mathrm{Var}(\sum_{s=1}^{p} B'_s \underset{\sim}{y}_s) \right| > \left| \mathrm{Var}(\sum_{s=1}^{p} C'_s(A'A)^- A'\underset{\sim}{y}_s) \right|$.

In fact, a strict inequality like (1.1.1.5) (iii) or (iv) would hold for all the symmetric functions of the characteristic roots such as the trace and determinant functions. Thus (1.1.1.4) is a unique solution in terms of the minimization of the trace and generalized variance criteria, but it is not a unique solution, though it still yields a minimum, in terms of the maximum root criterion.

If it is assumed for the SM model that the rows of Y are each

12

normally distributed, then maximum likelihood estimators of functions of ξ may be considered. In this case we can write the likelihood function for the SM model as

$$(1.1.1.6) \qquad \phi = (2\pi)^{-\frac{1}{2}np} \mid \Sigma \mid^{-\frac{n}{2}} \exp [-\frac{1}{2} \text{ tr } \Sigma^{-1} (Y - A\xi)'(Y - A\xi)].$$

A solution to the maximum likelihood equations obtained from ϕ is then given by

$$\hat{\xi} = (A'A)^{-}A'Y$$

and

$$\hat{\Sigma} = \frac{1}{n} Y'[I - A(A'A)^{-}A']Y.$$

It then follows that unique maximum likelihood estimators (MLE's) of estimable functions of ξ of the form $C\xi D$ and $H'\underline{\xi}$ are given by $C\hat{\xi}D$ and $H'\hat{\underline{\xi}}$, respectively, where $\hat{\underline{\xi}}$ is obtained by rolling out the columns of $\hat{\xi}$. Also $\hat{\Sigma}$ is the unique MLE of Σ and it is also a consistent estimator of Σ, i.e., $\hat{\Sigma} \underset{p}{\rightarrow} \Sigma$. $\hat{\Sigma}$ is biased for Σ however, although the bias can be corrected by replacing n by n - r. Note further that the MLE's $C\hat{\xi}D$ and $H'\hat{\underline{\xi}}$ are the same as the respective BLUE estimators given above.

1.1.2 Testing Linear Hypotheses in the SM Model

The general linear hypothesis for the SM model is usually given in the form

$$H_o: \; C\xi D = 0.$$

where $C(g \times m)$ is of full rank $g \leq m$, $D(p \times v)$ is of full rank v and $C\xi D$ is estimable. A number of different test criteria have been proposed to test H_o. Three of these--Wilks' Likelihood Ratio, Hotelling's Trace (T_o^2) and Roy's Largest Root--have been used in most applications. The test criteria are defined as follows:

Let

$$S_H = D'Y'A(A'A)^{-}C'[C(A'A)^{-}C']^{-1}C(A'A)^{-}A'YD,$$

$$S_E = D'Y'[I - A(A'A)^{-}A']YD,$$

and

$$\lambda_i = ch_i(S_H S_E^{-1}),\ i=1,\ldots,p.$$

Then Wilks' Likelihood Ratio Criterion is given by

$$(1.1.2.1) \qquad \Lambda = \frac{|S_E|}{|S_H + S_E|} = \prod_{p=1}^{p} (1 + \lambda_j)^{-1},$$

Hotelling's Trace (T_o^2) Criterion is given by

$$(1.1.2.2) \qquad T_o^2 = tr\ (S_H S_E^{-1}) = \sum_{j=1}^{p} \lambda_j \quad \text{and}$$

Roy's Largest Root Criterion is given by

$$(1.1.2.3) \qquad ch_1(S_H S_E^{-1}) = \lambda_1.$$

Some other proposed criteria are

$$(1.1.2.4) \qquad tr(S_H[S_E + S_H]^{-1}) = \sum_{j=1}^{p} \frac{\lambda_j}{1 + \lambda_j},$$

$$(1.1.2.5) \qquad \left[\sum_{i=1}^{p} \frac{1 + \lambda_j}{\lambda_j}\right]^{-1}$$

both due to Pillai [37], and

<u>Roy's Minimum Root Criterion</u>:

$$(1.1.2.6) \quad ch_p(S_H S_E^{-1}) = \lambda_p .$$

In most practical applications, the use of Wilks' Λ criterion has been restricted to large sample situations.[1] It is well known that $-[n - r - \frac{1}{2}(v - g + 1)]\log\Lambda$ is distributed as a chi-square variate with gv degrees of freedom as the sample size n tends to infinity. The above statistic can thus be used when the sample size is large by comparing it with the appropriate percentage point of χ^2_{gv}. Various researchers, in particular Box [14], have investigated closer approximations to the distribution of Λ. A comprehensive treatment of these results has been given by Anderson [8].

The distribution of T_o^2 has recently been obtained by Siotani [52] and Davis [18], although no tables have been published as yet. Use of T_o^2 has thus also been restricted to its large sample distribution. It can be easily seen that $n\,T_o^2$ is distributed as χ^2_{gv} as n tends to infinity.

To use Roy's Largest Root Criterion, one can use the charts of Heck [26]. Here however it is necessary to use

$\theta_1 = \dfrac{\lambda_1}{1 + \lambda_1}$ instead of λ_1 and to compute the parameters

$$s = \min\{g, v\}, \ \tilde{m} = \frac{|g - v| - 1}{2}, \ \tilde{n} = \frac{n - r - v - 1}{2}.$$

---

[1] Recently the exact distribution of Λ has been tabled by Schatzoff [51] and Pillai [38], but use of their results is not widespread at this time.

The acceptance region is

$$\theta_1 \le x_{\alpha,s,\tilde{m},\tilde{n}}$$ where $x_{\alpha,s,\tilde{m},\tilde{n}}$ is the upper 100α percentage point obtained from the appropriate Heck chart.

Although none of the other test criteria is usually used, Pillai's trace given in (1.1.2.4) has been tabled by Pillai [37].

At present, relatively little is known of the comparative power of the three test criteria. Nevertheless, in a recent study by Schatzoff [50], simulation techniques were used to compare the six test criteria given above with regard to power. It was concluded that Wilks' and Hotelling's criteria provide the best protection against a wide spectrum of alternatives, although Roy's criterion is best for certain restricted alternatives. Ito [29] has compared the power probabilities of the T_o^2 and Λ statistics for a simple class of alternatives and has shown that when $N - r = 100$ or 200, the powers differ at most in the second place and for $N - r = \infty$, the powers are identical for all alternatives.

1.2 <u>The Growth Curve Model (GCM)</u>

Experimental techniques which observe and record responses of an individual for a period of time, perhaps over a sequence of doses of some drug, are generally discussed under the title of <u>growth curve experiments</u>. Growth curve models have been studied extensively in the literature by Rao [40], [42], [43], [45], Khatri [31], Potthoff and Roy [39], Allen and Grizzle [2], and others (e.g., [15], [22], [23]). The model considered by all these authors may be given in general form as

$$(1.2.1) \qquad E(Y) = A\xi P$$

$$\mathrm{Var}\ (Y) = I_n \otimes \Sigma$$

where $Y(n \times q)$, $A(n \times m)$, $\xi(m \times p)$, $P(p \times q)$, $\Sigma(q \times q)$ is p.d., $R(A) = r(\leq m)$, $R(P) = p$, and the rows of Y are independently normally distributed. Implicit in the model is the assumption that each individual's response can be described by a linear model of the form

(1.2.2) $E(\underline{y}_i) = P'\underline{\beta}_i$ where $\underline{y}_i(q \times 1)$ is the observation vector on the i-th individual and $\underline{\beta}_i$ is a vector of unknown parameters. We call (1.2.2) the <u>within individual</u> model. It is also implicitly assumed that the expected value of each individual's response comes from the same family (e.g., it is a (p-1)-th degree polynomial).

1.2.1 Estimation in the GCM Model

As with the SM model, it is usually of interest to estimate for the GCM model linear functions and sets of ξ of the form

$\underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}'_s\underline{\xi}_s$ and $H'\underline{\xi} = \sum_{s=1}^{p} C'_s\underline{\xi}_s$ where $\underline{\xi}_s$ is the s-th column of ξ and $\underline{\xi}$ is the column vector defined in (1.1.1.1) by rolling out the columns of ξ into one long $(mp \times 1)$ vector. Of particular interest is the estimation of ξ itself (i.e., $H'\underline{\xi}$ when $H \equiv I_{mp}$).

If it is assumed that $R(A) = m$ (i.e., A is of full rank), then it is easy to see that ξ (i.e., $\underline{\xi}$) itself is estimable and an unbiased estimate of ξ can be given by

$$(1.2.1.1) \qquad \hat{\xi} = (A'A)^{-1}A'YG^{-1}P'(PG^{-1}P')^{-1}$$

where $G(q \times q)$ is any symmetric p.d. matrix either non-stochastic or independent of Y for which $PG^{-1}P'$ is of full rank (e.g., $G = I_q$ or $G = \Sigma$). The matrix G^{-1} is frequently referred to as the <u>weights</u> or the <u>weight matrix</u>. Use of G was first suggested by Potthoff and Roy [39]. Provided G is non-stochastic, it is easy to see that $\hat{\underline{\xi}}$ can be put in the form $F'\underline{y}$ where F is non-stochastic and $\underline{y}$ is defined by rolling out the columns of $Y = [\underline{y}_1 \ldots \underline{y}_q]$ into one long vector. Thus, $\hat{\underline{\xi}}$ is a linear set in $\underline{y}$. Since ξ is estimable when $R(A) = m$, it follows that $\underline{h}'\underline{\xi}$ and $H'\underline{\xi}$ are estimable for arbitrary $\underline{h}$ and H, and unbiased estimators are given by $\underline{h}'\hat{\underline{\xi}}$ and $H'\hat{\underline{\xi}}$ respectively, where $\hat{\underline{\xi}}$ is defined by (1.2.1.1).

If A is not of full rank then ξ is not estimable. In this case it is easy to see that if we want to estimate $\underline{c}'\xi\underline{d}$ and $C\,\xi D$ and if we require our estimators to be of the form $\underline{t}'Y\underline{w}$ and $T'YW$ where $\underline{t}$, $\underline{w}$, T and W are non-stochastic, then the estimability conditions are that $\underline{c}' \in V(A)$, $\underline{d} \in V(P')$ for $\underline{c}'\xi\underline{d}$ and that the rows of C belong to $V(A)$, the columns of D belong to $V(P')$ for $C\,\xi D$. More generally, if we consider arbitrary $\underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$ and $H'\underline{\xi} = \sum_{s=1}^{p} C_s'\underline{\xi}_s$ and we require our estimators to be of the form $\underline{f}'\underline{y} = \sum_{i=1}^{q} \underline{f}_i'\underline{y}_i$ and $F'\underline{y} = \sum_{i=1}^{q} F_i'\underline{y}_i$, respectively, where $Y = [\underline{y}_1 \ldots \underline{y}_q]$, then the estimability conditions become for $\underline{h}'\underline{\xi}$:

(1.2.1.2) $\quad \underline{c}_s' = \underline{p}_s'\, LA, \; s=1,\ldots,p,$

18

where L is arbitrary and $\underline{p}_s'$ is the s-th row of P, and for $H'\underline{\xi}$:

(1.2.1.3) $C_s' = L_s'A$, $s=1,\ldots,p$

where the i-th row of L_s is of the form $\underline{p}_s'\tilde{L}_i$ for arbitrary $\tilde{L}_i$. The conditions given in (1.2.1.2) are more restrictive than just $\underline{c}_s' \varepsilon V(A)$ alone, the estimability requirement for the SM model, since the matrix P is involved. Similarly, (1.2.1.3) is more restrictive than the requirement that the rows of C_s' belong to V(A). These conditions are obtained observing that if

$$F'\underline{y} = \sum_{i=1}^{p} F_i'\underline{y}_i \text{ is unbiased for } \sum_{s=1}^{p} C_s'\underline{\xi}_s$$

then

$$E\left[\sum_{i=1}^{q} F_i'\underline{y}_i\right] = \sum_{i=1}^{q} F_i'A\left(\sum_{s=1}^{p} p_{si}\underline{\xi}_s\right), \text{ where } P = ((p_{si}))$$

$$= \sum_{s=1}^{p}\left(\sum_{i=1}^{q} p_{si}F_i'A\right)\underline{\xi}_s$$

$$= \sum_{s=1}^{p} C_s'\underline{\xi}_s$$

so that

$$C_s' = \left(\sum_{i=1}^{q} p_{si}F_i'\right) A = L_s'A$$

as in (1.2.1.3).

In Roy's paper [49] it is pointed out that if $\underline{h}'\underline{\xi}$ is estimable with regard to functions of the form $\underline{f}'\underline{y}$, then

$$\sum_{s=1}^{p} \underline{c}_s'(A'A)^{-}A'\underline{x}_s ,$$

where $\underline{x}_s$ is the s-th column of $X = YP'(PP')^{-1}$, is the BLUE of $\underline{h}'\underline{\xi}$ in terms of linear functions of the elements of X. This can be extended to show that for any choice of G as in (1.2.1.1)

$$\sum_{s=1}^{p} \underline{c}_s'(A'A)^{-}A'\underline{x}_s^{*} ,$$

where $\underline{x}_s^{*}$ is the s-th column of $X^{*} = YG^{-1}P'(PG^{-1}P')^{-1}$, is the BLUE of $\underline{h}'\underline{\xi}$ in terms of linear functions of the elements of X^{*}. We would like to choose G to obtain a BLUE of $\underline{h}'\underline{\xi}$ in terms of linear functions of the elements of Y. We can do this by using Lagrange multipliers with respect to the estimability conditions (1.2.1.2); that is, we wish to find $\underline{f}$ to minimize $\mathrm{Var}(\underline{f}'\underline{y})$ subject to the estimability conditions that $\underline{c}_s' = \underline{p}_s'LA$, where $\underline{f}'\underline{y} = \sum_{i=1}^{p} \underline{f}_i'\underline{y}_i$ and $L' = [\underline{f}_1 \ldots \underline{f}_q]$.

Now

$$\mathrm{Var}(\underline{f}'\underline{y}) = \sum_{i=1}^{q} \sum_{j=1}^{q} \underline{f}_i'\underline{f}_j\sigma_{ij}$$

so that we must consider the Lagrange expression

$$\Psi = \sum_{i=1}^{q} \sum_{j=1}^{q} \underline{f}_i'\underline{f}_j\sigma_{ij} + \sum_{s=1}^{p} \underline{\lambda}_s'(\underline{c}_s - A'\sum_{i=1}^{q} p_{si}\underline{f}_i).$$

Taking derivatives with respect to $\underline{f}_j$ we get

$$\frac{\partial\Psi}{\partial\underline{f}_j} = 2\sum_{i=1}^{q} \underline{f}_i\sigma_{ij} - A\sum_{s=1}^{p} p_{sj}\underline{\lambda}_s, \; j=1,\ldots,q$$

or

$$\frac{\partial\Psi}{\partial L'} = 2L'\Sigma - A\Lambda P ,$$

where $\qquad \Lambda = [\underline{\lambda}_1 \ldots \underline{\lambda}_p]$.

20

Now $\frac{\partial\Psi}{\partial L'} = 0$ provided

$$2L'\Sigma = A\Lambda P$$

i.e., $$2A'L'\Sigma\Sigma^{-1}P' = A'A\Lambda P\Sigma^{-1}P'$$

i.e., $$2A'L'P'(P\Sigma^{-1}P')^{-1} = A'A\Lambda$$

i.e., $$2C'(P\Sigma^{-1}P')^{-1} = A'A\Lambda,$$

where $$C' = [\underline{c}_1\ \underline{c}_2 \cdots \underline{c}_p] .$$

Thus $$L'\Sigma = A(A'A)^{-}C'(P\Sigma^{-1}P')^{-1}P$$

i.e., $$L' = A(A'A)^{-}C'(P\Sigma^{-1}P')^{-1}P\Sigma^{-1}$$

and $$\underline{f}_i' = \underline{\rho}_i'C(A'A)^{-}A' ,$$

where $\underline{\rho}_i'$ is the i-th row of $\Sigma^{-1}P'(P\Sigma^{-1}P')^{-1}$.

Thus

$$\sum_{i=1}^{q} \underline{f}_i'\underline{y}_i = \sum_{i=1}^{q} \underline{\rho}_i'C(A'A)^{-}A'\underline{y}_i$$

$$= \sum_{i=1}^{q}\sum_{s=1}^{p} \rho_{is}\underline{c}_s'(A'A)^{-}A'\underline{y}_i,$$ where ρ_{is} is the s-th

element of $\underline{\rho}_i$

$$= \sum_{s=1}^{p} \underline{c}_s'(A'A)^{-}A'\left(\sum_{i=1}^{q} \rho_{is}\underline{y}_i\right)$$

$$= \sum_{s=1}^{p} \underline{c}_s'(A'A)^{-}A'\underline{x}_s^*$$

where

$$\underline{x}_s^* \text{ is the s-th column of } Y\Sigma^{-1}P'(P\Sigma^{-1}P')^{-1} .$$

Thus the optimum choice of G is given by $G = \Sigma$ and we have the following theorem:

<u>THEOREM 1.2.1.</u> For the GCM model, the BLUE of an estimable linear function of the form $\underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$ is given by

$$(1.2.1.4) \qquad \sum_{s=1}^{p} \underline{c}_s'(A'A)^- A'\underline{x}_s^*$$

where $\underline{x}_s^*$ is the s-th column of $Y\Sigma^{-1}P'(P\Sigma^{-1}P')^{-1}$.

Theorem 1.2.1 is unfortunate since Σ is usually unknown. This leads us to consider estimates of ξ which are not necessarily linear sets of Y alone. In particular, the theorem suggests that we use for G a good estimate of Σ based on the data or based on previous experiments. In fact, it has been shown (e.g., Khatri [31], Rao [43], Grizzle and Allen [2]) that if $C\,\xi D$ is estimable and normality is assumed on the rows of Y, then the maximum likelihood estimator of $C\,\xi D$ is given by $C\,\hat{\xi}D$ where

$$(1.2.1.5) \qquad \hat{\xi} = (A'A)^- A'YS^{-1}P'(PS^{-1}P')^{-1} ,$$

and S is the sample covariance matrix obtained from the data and defined by

$$(1.2.1.6) \qquad S = Y'[I - A(A'A)^- A']Y.$$

The maximum likelihood estimator of Σ is $\hat{\Sigma} = \frac{1}{n} S$ and use of $\hat{\Sigma}$ instead of S in (1.2.1.5) clearly does not change $\hat{\xi}$.

The dispersion matrix of $C\hat{\xi}D$ was derived by Rao [45] and Williams [70] and it is given by

$$(1.2.1.7) \qquad \operatorname{Var}(C\hat{\xi}D) = C\,(A'A)^{-}C' \otimes \frac{n-m-1}{n-m-(q-p)-1} D'(P\Sigma^{-1}P')^{-1}D \,.$$

1.2.2 Testing Hypotheses in the GCM Model

As for the SM model, hypotheses of interest for the GCM model are usually in the form

$$H_o: \; C\xi D = 0$$

where $C(g \times m)$ is of full rank g and $D(p \times v)$ is of full rank v and $C\xi D$ is estimable. In the analysis of Potthoff and Roy [39], H_o was tested by transforming the GCM model (1.2.1) into the SM model by use of a weights matrix G^{-1} where G is as described in the previous section, and then using either Wilks', Hotelling's or Roy's criterion as the test statistic. The transformed data matrix in this case is given by

$$(1.2.2.1) \qquad X = YG^{-1}P'(PG^{-1}P')^{-1}$$

and the hypothesis and error matrices are given by

$$(1.2.2.2) \qquad S_H = D'X'A(A'A)^{-}C'[C(A'A)^{-}C']^{-1}C(A'A)^{-}A'XD$$

and

$$(1.2.2.3) \qquad S_E = D'X'[I - A(A'A)^{-}A']XD \,.$$

Potthoff and Roy's approach did not allow G to be stochastic unless it was independent of Y. Thus G could be chosen as an estimate of $k\Sigma$ where k is any constant, provided this estimate was obtained independently of the experiment under consideration. Choosing $G^{-1} = S^{-1}$

where S is the sample covariance matrix based on the data in Y was therefore not allowed. Khatri [31] used the method of maximum likelihood to obtain test criteria in which S^{-1} is used as the weights. Khatri's hypothesis and error matrices simplify to

$$(1.2.2.4) \qquad S_H = D'X'A(A'A)^{-}C'[CRC']^{-1}C(A'A)^{-}A'XD$$

and

$$(1.2.2.5) \qquad S_E = D'[PS^{-1}P']^{-1}D$$

where X is defined by (1.2.2.1), S is defined by (1.2.1.6) and

$$(1.2.2.6) \qquad R = (A'A)^{-}+(A'A)^{-}A'Y[S^{-1}-S^{-1}P'(PS^{-1}P')^{-1}PS^{-1}]Y'A(A'A)^{-} .$$

Rao [43] criticized the use of G^{-1} by Potthoff and Roy by pointing out that not all of the information available in the sample Y would be utilized unless G^{-1} is chosen to be the unknown Σ^{-1}. Using the method of covariance adjustment Rao [42], [43] showed how the additional information could be utilized by incorporating into the model $(q - p)$ covariables which give a column basis of the matrix $[I - P'(PP')^{-1}P]$. (Note that $V([I - P'(PP')^{-1}P])$ is the within individual error space.) Rao also pointed out in [45] that weighting did not always produce shorter confidence intervals. Grizzle and Allen [2] unified the previous approaches and described how the results could be used on a computer in applications. Also they showed that Rao's results using covariance adjustment were identical to those of Khatri, and they described how to use stochastic weights other than S^{-1} if it was desired not to use all the vectors in the error space as covariables.

24

1.3 Experimental Situations to Which the SM Model Does Not Apply

The SM model as defined by (1.1.1), (1.1.2) or (1.1.3) contains two assumptions inherent to the experimental situation which are not always met in practice. These are

(i) each response variate is measured on each experimental unit (i.e., on each experimental unit is observed a p-variate vector)

(ii) the design matrix, A, is the same for each response variate (e.g., the same blocking system is applicable to each variate).

In general, condition (i) will not hold whenever it is physically impossible, uneconomical or inadvisable to observe each response variate on each experimental unit or, in other words, when there are missing observations, either by design or at random. Condition (ii) will not hold when different blocking systems are applicable to different response variates and when some of the response variates are insensitive to certain treatments. In addition, in a somewhat different context, a relaxation of (ii) is required for models which describe certain kinds of econometric problems which involve the estimation of a set of different regression equations which are not independent of each other (i.e., "seemingly unrelated regressions" as described by Zellner [72], Telser [61], and others).

Several examples of experiments in which (i) is not applicable are described by Trawinski [63], Trawinski and Bargmann [64], and Srivastava [57], [58]. In each of these, data is missing by design. For example, if the process of taking measurements is time consuming and the experimental material has a limited life or the experiment has a fixed time limit, then the experimenter may have to be content with measuring fewer than p response variates on each experimental unit.

As a second example, a psychologist may wish to administer 20 different tests to a group of subjects, but realizes that giving any subject more than, perhaps, four tests might result in a bias on later tests obtained from practice on earlier tests.

Thirdly, the experimenter may not be equally interested in the different response variates, and he might like to observe some variates on a larger number of experimental units than on others. This might be the case if there is a limited budget for the experiment and the measurement of responses is so costly that the experimenter cannot afford to measure each response variate on each experimental unit.

As a fourth example, suppose u experiments are conducted in different localities involving the same set of treatments. At the i-th locality, a subset of q_i out of p response variates is measured on each experimental unit.

Trawinski [63] describes an experiment in which there are 3600 individuals divided into two groups (boys and girls) of 1800 each, and there are four response variates of interest labeled V_1, V_2, V_3 and V_4. It is desired to administer three treatments (drugs, for example) on each group in such a way that each individual is observed on only two of the four variates, all variate pairs are observed, and an equal number of observations are made on each variate pair. Such an experiment requires choosing missing observations to provide a sort of <u>balance</u> among the variate pairs observed. In this example, the 1800 in each group are divided randomly into 3 subgroups of 600 individuals, each subgroup getting a different treatment. A given subgroup is then divided randomly into 6 sub-subgroups of 100 individuals each.

Individuals in a given sub-subgroup are observed for one of the 6 variate pairs (V_1, V_2), (V_1, V_3), (V_1, V_4), (V_2, V_3), (V_2, V_4) and (V_3, V_4). In this way, each variate pair is observed on the same number of individuals in the entire experiment, 600, the same number of individuals within each group, 300, and the same number of individuals for each treatment within each group, 100. Note that the variate design is a BIBD with $v = 4$, $b = 6$, $k = 2$, $r = 3$ and $\lambda = 1$. Other examples of variate designs are described by Trawinski.

Srivastava [57] gives a number of examples for which condition (ii) is inapplicable. For instance, consider an agricultural experiment in which seven varieties or treatments are applied to 28 experimental units according to the following plan:

Row Blocks \ Column Blocks	I	II	III	IV	V	VI	VII
I	3	1	5	2	7	4	6
II	6	3	7	5	1	2	4
III	7	4	1	6	2	3	5
IV	2	6	3	1	4	5	7

TABLE 1.2
SRIVASTAVA'S EXAMPLE

Suppose that two response variates V_1 and V_2 are measured on each experimental unit, but that the pattern of heterogeneity is such that the 7 column blocks are a good block system for V_1 and the 4 row blocks are a good block system for V_2. If we denote by $\underline{y}_1(28 \times 1)$

and $\underline{y}_2(28 \times 1)$ the vectors representing the data observed on V_1 and V_2, respectively, and let $Y = [\underline{y}_1\ \underline{y}_2]$, then the expectation model for the experiment is

$$E(Y) = [A_1\ \underline{\xi}_1 \mid A_2\underline{\xi}_2]$$

where $A_1(28 \times 14)$ is a design matrix for a BIB experiment with $v = b = 7$, $r = k = 4$, $\lambda = 1$,

$A_2(28 \times 11)$ is a design matrix for a Randomized Blocks experiment,

$$\underline{\xi}_1(14 \times 1) = \begin{bmatrix} \underline{\beta}_1 \\ \underline{\tau} \end{bmatrix},$$

$$\underline{\xi}_2(11 \times 1) = \begin{bmatrix} \underline{\beta}_2 \\ \underline{\tau} \end{bmatrix},$$

and

$\underline{\beta}_1(7 \times 1)$ denotes the column block effects,

$\underline{\beta}_2(4 \times 1)$ denotes the row block effects,

$\underline{\tau}(7 \times 1)$ denotes the treatment effects.

For another example, consider a $\frac{1}{4}$ fractional factorial experiment without blocks from a 2^5 factorial. Denote the factors by A, B, C, D, E and the eight treatment combinations used by (00000), (00101), (10110), (10011), (01010), (01111), (11100), (11001). Suppose two response variates V_1 and V_2 are observed on each experimental unit, but that V_1 is insensitive to levels of D and E, and V_2 is insensitive to levels of B and C. Suppose further that we are interested in

28

estimating effects up to and including two-factor interactions. Then the expectation model for the experiment is described by

$$E(Y) = [A_1 \underline{\xi}_1 \;\vdots\; A_2\underline{\xi}_2]$$

where $Y(8 \times 2)$ is the data matrix

$$\underline{\xi}_1(7 \times 1) = (\mu_1, A^1, B, C, A^1B, A^1C, BC)',$$

$$\underline{\xi}_2(7 \times 1) = (\mu_2, A^2, D, E, A^2D, A^2E, DE)',$$

$$A_1(8 \times 7) = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 \end{bmatrix}$$

and

$$A_2(8 \times 7) = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}.$$

Note that in this example $A_1 \neq A_2$.

Gourlay [24] and others have considered a model (See Table 1.4 below.) for an experimental design involving a complete set of 3×3 Latin Squares corresponding to the six possible orderings of three treatments A, B and C. The two squares are replicated n times so that each of the six groups consists of n individuals chosen at random from

Groups \ Ordinal Positions	1	2	3
1	A	B	C
2	B	C	A
3	C	A	B
4	A	C	B
5	B	A	C
6	C	B	A

TABLE 1.3
GOURLAY'S MODEL

a total of 6n individuals to whom the treatments are applied. Each individual goes through a sequence (row) of three treatments. It is assumed that the results for a given individual are correlated. Each of the n observations in any one of the eighteen cells is assumed to have the following model:

$$E(Y_{ij(k)q}) = Q_j + T_{(k)} + N_{ij}$$

$$Var(Y_{ij(k)q}) = \sigma^2$$

$$Cov(Y_{ij(k)q}, Y_{i'j'(k')q'}) = \begin{cases} \sigma_{jj'} \text{ if } j \neq j', i=i', q=q', \\ 0 \text{ if } i \neq i' \text{ and/or } q \neq q' \end{cases}$$

where $i = 1, \ldots, 6$ (groups)

$j = 1, 2, 3$ (ordinal positions)

$k = A, B, C$ (treatments)

$q = 1, \ldots, n$ (replications).

If the observations on each individual were not correlated, then the usual univariate analysis of variance model of the form

$$E(\underline{y}) = A\underline{\xi}$$

$$\mathrm{Var}(\underline{y}) = \sigma^2 I_{18n}$$

where A is an $(18n \times 24)$ design matrix of zeros and ones,

$$\underline{\xi} = (\underline{Q}', \underline{T}', \underline{N}'),$$

$$\underline{Q} = (Q_1, Q_2, Q_3)',$$

$$\underline{T} = (T_A, T_B, T_C)'$$

and

$$\underline{N} = (N_{11}, \ldots, N_{61}, N_{12}, \ldots, N_{62}, N_{13}, \ldots, N_{63})'$$

would be appropriate. However, because the rows are correlated we actually have a model of the form

$$E(\underline{y}) = A\underline{\xi}$$

$$\mathrm{Var}(\underline{y}) = \Sigma \otimes I_{6n}$$

where $\Sigma = ((\sigma_{jj'}))$ and $\sigma_{jj} = \sigma^2$ for all j.

By considering the ordinal positions as response variates V_1, V_2, and V_3, respectively, and by letting $\underline{y}_s(6n \times 1)$, s=1, 2, 3 denote the observations on V_s, the model can also be expressed as

$$E[\underline{y}_1 \vdots \underline{y}_2 \vdots \underline{y}_3] = [A_1 \underline{\xi}_1 \vdots A_2\underline{\xi}_2 \vdots A_3\underline{\xi}_3],$$

$$\mathrm{Var}(\underline{y}_1 \vdots \underline{y}_2 \vdots \underline{y}_3) = I_{6n} \otimes \Sigma$$

where $\underline{\xi}_s = (Q_s, \underline{T}', N_{1s}, N_{2s}, N_{3s}, N_{4s}, N_{5s}, N_{6s})'$

and $A_s(6n \times 10)$ is the design matrix corresponding to V_s, s=1, 2, 3.

The three design matrices A_1, A_2, and A_3 will all be different, although any one of these matrices can be obtained from any other by column permutations.

Zellner [72] has pointed out that there are many econometric situations which call for linear models in which different design matrices correspond to different response variates. The usual problem can be described as the estimation of a set of correlated regression equations. The different response variates would correspond to the different regression equations. Typical examples of such situations include analysis of temporal cross-section data, time series regression analysis of demands for a variety of goods, or a cross-section budget study involving regressions for several commodities[2]. Another application occurs when each regression equation refers to different points in space, as in Barten and Koerts' analysis of voters' transitions from party to party within various voting districts [10].

1.4 The More General Linear Multivariate Model (MGLM)

To deal with situations such as described in the previous section and also with situations which involve both missing observations and different design matrices corresponding to different response variates, we now define a model which we call the More General Linear Multivariate Model (MGLM). "More" is used instead of "Most" in the

---

[2]With regard to the latter, a set of data may have been obtained from a survey of household expenditures during some time period. The regression equation used to explain outlays on some commodity may generate residuals that are correlated with the residuals of another regression equation that was used to explain expenditures on a different commodity.

model name because there are meaningful models such as growth curve models which are not special cases of the MGLM model.

Let us assume that there are n experimental units and p response variates $V_1, \ldots, V_p$ in total. The n experimental units are divided into u disjoint sets of experimental units $S_1, S_2, \ldots, S_u$ with n_j units in S_j. On each unit in the set S_j, we measure $q_j (\le p)$ responses $V_{\ell_{j1}}, V_{\ell_{j2}}, \ldots, V_{\ell_{jq_j}}$. (The remaining $p - q_j$ response variates are not measured in S_j.)

Then the MGLM model is given by

$$(1.4.1)\quad E(Y_j) = [A_{j\ell_{j1}} \underline{\xi}_{\ell_{j1}} \mathrel{\vdots} A_{j\ell_{j2}} \underline{\xi}_{\ell_{j2}} \mathrel{\vdots} \cdots \mathrel{\vdots} A_{j\ell_{jq_j}} \underline{\xi}_{\ell_{jq_j}}] ,$$

$$\mathrm{Var}(Y_j) = I_{n_j} \otimes B_j' \Sigma B_j ,$$

$$j=1,\ldots,u;\ 1 \le q_j \le p, 1 \le \ell_{j1} < \ell_{j2} < \ldots < \ell_{jq_j} \le p$$

and for each s, $s=1,\ldots,p$ there exists at least one pair (j, r) such that $\underline{\xi}_{\ell_{jr}} = \underline{\xi}_s$, i.e., $\ell_{jr} = s$ (so that the total set of unknown parameters is $\{\underline{\xi}_1, \ldots, \underline{\xi}_p, \Sigma\}$)

Here $Y_j(n_j \times q_j)$ is the matrix of observations for the j-th set S_j,

$A_{j\ell_{jr}}(n_j \times m_{\ell_{jr}})$ is the design matrix for variate $V_{\ell_{jr}}$ in the set S_j, and $R(A_{j\ell_{jr}}) = a_{jr}$,

$\underline{\xi}_{\ell_{jr}}$ $(m_{\ell_{jr}} \times 1)$ is the parameter vector for variate $V_{\ell_{jr}}$

(and $m_{\ell_{jr}} = m_{\ell_{j'r'}}$ whenever $\ell_{jr} = \ell_{j'r'}$),

$B_j(p \times q_j)$ is the incidence matrix of rank $R(B_j) = q_j$ for the j-th set of experimental units. It consists of 0's and 1's and is defined by $B_j = ((b_{(j)_{sr}}))$ where

(1.4.2) $$b_{(j)_{sr}} = \begin{cases} 1 \text{ if } \ell_{jr} = s, \text{ i.e., variate } V_s \text{ is the r-th ordered variate measured in the j-th experimental set } S_j \\ 0 \text{ otherwise.} \end{cases}$$

And let us further assume that

(1.4.3) Y_j and Y_j' are independent if $j \neq j'$

and

(1.4.4) the rows of Y_j are also independent and distributed as a q_j-variate multinormal vector with variance-covariance matrix $B_j'\Sigma B_j$.

The MGLM model can alternatively be described by its <u>variate-wise representation</u> (as was done for the SM model in (1.1.2)), provided the following definitions are made:

Let $\underline{x}_s$, s=1,...,p, be the vector of length N_s, say, corresponding to all observations on V_s in the entire experiment.

Let $D_s(N_s \times m_s)$, s=1,...,p, be the design matrix corresponding to $\underline{x}_s$, i.e., D_s is determined by $E(\underline{x}_s) = D_s\underline{\xi}_s$.

Let $Q_{rs}(N_r \times N_s)$, $r < s$ denote the incidence matrix of 0's and 1's defined by $Q_{rs} = ((q_{ij_{(rs)}}))$ where

$$q_{ij_{(rs)}} = \begin{cases} 1 \text{ if the i-th component of } \underline{x}_r \text{ and the j-th} \\ \quad \text{component of } \underline{x}_s \text{ are observed on the same} \\ \quad \text{experimental unit} \\ 0 \text{ otherwise.} \end{cases} \tag{1.4.5}$$

Then the variate-wise representation of the MGLM model is given by

$$\begin{aligned} &E(\underline{x}_s) = D_s\underline{\xi}_s \,, \\ &\mathrm{Var}(\underline{x}_s) = \sigma_{ss}I_{N_s} \\ &\mathrm{Cov}(\underline{x}_r, \underline{x}_s) = \sigma_{rs}Q_{rs}, \; r < s \\ &\mathrm{Cov}(\underline{x}_r, \underline{x}_s) = \sigma_{rs}Q'_{sr}, \; r > s, \quad r,s=1,\ldots,p. \end{aligned} \tag{1.4.6}$$

Note that if we let

$$Y_j = [\underline{y}_{j\ell_{j1}}, \underline{y}_{j\ell_{j2}}, \ldots, \underline{y}_{j\ell_{jq_j}}] \,,$$

u_s = the number of sets S_i in which V_s is measured,

$S_{t_{si}}$, $i=1,\ldots,u_s$ denote the different sets in which V_s is measured where

$$1 \le t_{s1} < t_{s2} < \ldots < t_{su_s} \le u \,,$$

then $\underline{x}_s$, D_s, and N_s are given by

$$(1.4.7) \qquad \underline{x}_s(N_s \times 1) = \begin{bmatrix} \underline{y}_{t_{s1}s} \\ \underline{y}_{t_{s2}s} \\ \vdots \\ \underline{y}_{t_{su_s}s} \end{bmatrix}, \quad D_s(N_s \times m_s) = \begin{bmatrix} A_{t_{s1}s} \\ A_{t_{s2}s} \\ \vdots \\ A_{t_{su_s}s} \end{bmatrix},$$

and $\qquad N_s = \sum_{i=1}^{u_s} n_{t_{si}}$.

With the above definitions, we can also present the vector version of the MGLM model (as was similarly obtained for the SM model in (1.1.3)) as follows:

$$(1.4.8) \qquad E(\underline{x}) = D\underline{\xi}$$

$$\mathrm{Var}(\underline{x}) = \Omega$$

where

$$(1.4.9) \qquad \underline{x}(N \times 1) = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_p \end{bmatrix}, \quad D(N \times M) = \begin{bmatrix} D_1 & & 0 \\ & D_2 & \\ 0 & & \ddots \\ & & & D_p \end{bmatrix}, \quad \underline{\xi}(M \times 1) = \begin{bmatrix} \underline{\xi}_1 \\ \underline{\xi}_2 \\ \vdots \\ \underline{\xi}_p \end{bmatrix},$$

$$\Omega(N \times N) = \begin{bmatrix} \sigma_{11} I_{N_1} & \sigma_{12} Q_{12} & \cdots & \sigma_{1p} Q_{1p} \\ \sigma_{12} Q_{12}' & \sigma_{22} I_{N_2} & \cdots & \sigma_{2p} Q_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{1p} Q_{1p}' & \sigma_{2p} Q_{2p}' & \cdots & \sigma_{pp} I_{N_p} \end{bmatrix},$$

$$N = \sum_{s=1}^{p} N_s \text{ , and } M = \sum_{s=1}^{p} m_s \text{ .}$$

1.4.1 Special Cases of the MDM, GIM, HM and Other Models

The special cases of the MGLM model to be described in this section have all been dealt with to some degree in the literature. We will defer a thorough discussion of this literature until Chapters Two, Three and Four, which consider estimation and testing hypotheses.

Srivastava [55], [57], [60], Zellner [72], Zellner and Huang [74], Telser [61], and Kmenta and Gilbert [33] have considered the special case of the MGLM model where $u = 1$, $n_1 = n$, $q_1 = p$, $Y_1 = Y$ and $B_1 = I_p$. This reduces to a model given by

$$(1.4.1.1) \quad E(Y) = [A_1\underline{\xi}_1 \mid A_2\underline{\xi}_2 \mid \cdots \mid A_p\underline{\xi}_p]$$

$$\mathrm{Var}(Y) = I_n \otimes \Sigma.$$

Srivastava has called (1.4.1.1) the Multiple Design Multivariate Model (MDM). It is clear from (1.4.1.1) that for the MDM model different design matrices correspond to different response variates. The other authors have referred to (1.4.1.1) as seemingly unrelated regressions. In the latter context the MDM model may be put into the following variate-wise representation:

$$(1.4.1.2) \quad E(\underline{y}_s) = A_s\underline{\xi}_s,\ \mathrm{Cov}(\underline{y}_r, \underline{y}_s) = \sigma_{rs}I_n \ , \quad r,s=1,\ldots,p,$$

where $\quad Y = [\underline{y}_1\ \underline{y}_2\ \cdots\ \underline{y}_p]$.

Thus from (1.4.1.2) it can be seen that the MDM model describes a situation where there are p regression equations involving different regression coefficients, so that they are seemingly unrelated, and yet any two regressions are correlated according to the covariance structure defined in Σ.

The vector version of the MDM model is given by

$$(1.4.1.3) \qquad E(\underline{y}) = D\underline{\xi}, \quad Var(\underline{y}) = \Sigma \otimes I_n .$$

where

$$\underline{y}(np \times 1) = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_p \end{bmatrix}, \quad D(np \times \sum_{s=1}^{p} m_s) = \begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_p \end{bmatrix}$$

and

$$\underline{\xi}(\sum_{s=1}^{p} m_s \times 1) = \begin{bmatrix} \underline{\xi}_1 \\ \underline{\xi}_2 \\ \vdots \\ \underline{\xi}_p \end{bmatrix} .$$

Clearly D is of full rank $R(D) = \sum_{s=1}^{p} m_s$ if and only if each of the A_s matrices is of full rank.

A second special case of the MGLM model is a general multivariate linear model for missing observations which is obtained when $A_{j\ell_{jr}} \equiv A_j$, $r=1,\ldots,q_j$ and when $\underline{\xi}_s$ is $(m \times 1)$, where m is independent of s. This model is then given by

$$(1.4.1.4) \qquad E(Y_j) = A_j \xi B_j$$

$$\mathrm{Var}(Y_j) = I_{n_j} \otimes B_j'\Sigma B_j \ , \ j=1,\ldots,u.$$

Srivastava [57], [58], [60] has called (1.4.1.4) the General Incomplete Multivariate Model (GIM), and we will henceforth refer to it by this name.

Trawinski [63] and Trawinski and Bargmann [64] have considered the special case of the GIM model where $q_j \equiv q$, $n_j \equiv \frac{n}{u}$, $A_j \equiv A$, $j=1,\ldots,u$. Another special case of the GIM model which has been considered separately by Srivastava [57], [60] occurs when the missing observations are structured so that $q_j = j$, $u = p$ and $B_j(p \times j) = [I_j \mid 0_{j,p-j}]'$. We will refer to this latter model as the Hierarchal Incomplete Multivariate Model (HIM).

It should be noted that if there are no missing observations (i.e., the model is complete), then (1.4.1.4) reduces to the SM model where the design matrix A of (1.1.1) is given by

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_u \end{bmatrix} .$$

Thus for the GIM model it is clear that each of the u mutually exclusive sets of experimental units (e.g., S_j) corresponds to a unique subset of response variates (e.g., $V_{\ell_{j1}}, \ldots, V_{\ell_{jq_j}}$) which are observed together only on the given set of experimental units (S_j). The largest value that u can take is $2^p - 1$, which will be obtained when the sets of missing observations correspond to all possible combinations of the p variates taken 1, 2, ... and p at a time.

The HIM model described above has been dealt with in a more general context by Srivastava and Roy in [55]. By letting $u = p$, $q_j = j$ for $j=1,\dots,p$, $\ell_{jr} = r$ for $r=1,\dots,j$ and $B_j = [I_j \mid 0_{j,p-j}]'$ in the MGLM model, we obtain:

$$(1.4.1.5) \qquad E(Y_j) = [A_{j1}\underline{\xi}_1 \mid A_{j2}\underline{\xi}_2 \mid \cdots \mid A_{jj}\underline{\xi}_j]$$

$$Var(Y_j) = I_{n_j} \otimes B_j'\Sigma B_j \ , \quad j=1,\dots,p.$$

It is clear that (1.4.1.5) incorporates both missing observations and different design matrices for different response variates. Srivastava has called this model simply the <u>Hierarchal Multivariate</u> (HM) model. Obviously, when $A_{jr} = A_j$ for $r=1,\dots,j$ the HM becomes the HIM model. Note that if we let

U_s = the set of all experimental units on which response variate V_s is measured

then for the HM model we have

$$U_1 \supset U_2 \supset \cdots \supset U_p \ .$$

Thus the HM model deals with a situation in which response variates can be graded monotonically in descending order of importance. In addition, if V_r is more important than V_s, i.e., $r < s$, then V_r is observed on each experimental unit on which V_s is observed. The variate-wise representation in (1.4.6) and the vector versions in

(1.4.8) simplify considerably for the HM model since we have

$$(1.4.1.6) \qquad \underline{x}_s = \begin{bmatrix} \underline{y}_{ss} \\ \underline{y}_{s+1,s} \\ \vdots \\ \underline{y}_{ps} \end{bmatrix}, \quad A_s = \begin{bmatrix} A_{ss} \\ A_{s+1,s} \\ \vdots \\ A_{ps} \end{bmatrix},$$

$$Q_{rs} = [0_{N_s, N_r - N_s} \mid I_{N_s}]' \quad \text{for } r < s,$$

and

$$N_s = \sum_{r=s}^{p} n_r .$$

1.5 A Generalization of the Growth Curve Multivariate Model

Assume that there are n experimental units and q time points at which measurements are taken. The n experimental units are divided into u disjoint sets of experimental units $S_1, S_2, \ldots, S_u$ where S_j has n_j units. For $j=1,\ldots,u$ let $t_{j1}, \ldots, t_{jq_j}$ be the q_j ($\le q$) time points at which observations are taken in S_j. Let B_j be the $(q \times q_j)$ incidence matrix specifying which of the q time points is observed in S_j. Assume that each individual's response can be described by a linear model of the form

$$(1.5.1) \qquad E(\underline{y}_{jk}) = P_j' \underline{\beta}_{jk}$$

where $\underline{y}_{jk}(q_j \times 1)$ is the observation vector on the k-th individual from S_j, $P_j(p \times q_j)$ has rank p for all j, where $p \le q_j$, and $\underline{\beta}_{jk}(p \times 1)$ is a vector of unknown parameters. We call (1.5.1) the <u>within</u>

individual model. We assume that the expected value of each individual in the total experiment is from the same family so that $P_j = PB_j$ where $P(p \times q)$ is of rank p. The complete model expressing both the within individual and across individual design is given by

$$E(Y_j) = A_j \xi P B_j \tag{1.5.2}$$

$$\mathrm{Var}(Y_j) = I_{n_j} \otimes B_j' \Sigma B_j \ , \quad j=1,\ldots,u.$$

where $Y_j(n_j \times q_j)$ is the observation matrix for the j-th set S_j,

$A_j(n_j \times m)$ is the design matrix for the j-th set S_j,

$\xi(m \times p)$ is the matrix of unknown parameters,

$P(p \times q)$ is a regression matrix of rank p,

$B_j(q \times q_j)$ is the incidence matrix of 0's and 1's for S_j.

We will henceforth refer to model (1.5.2) as the Generalized Growth Curve Multivariate Model (GGCM). We will return to the GGCM model in Chapter Three when we discuss BAN estimation of estimable linear sets of the treatment parameters.

CHAPTER II

BLUE ESTIMATION FOR THE MGLM MODEL

2.1 Introduction

This chapter considers BLUE estimation for the MGLM model defined in (1.4.1), (1.4.6) or (1.4.8) of linear functions and linear sets of the treatment parameters $\underline{\xi}_1, \ldots, \underline{\xi}_p$ of the form

(2.1.1) $$\theta = \underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$$

and

(2.1.2) $$\underline{\theta} = H'\underline{\xi} = \sum_{s=1}^{p} C_s'\underline{\xi}_s .$$

We seek linear estimators of the form

(2.1.3) $$\hat{\theta} = \underline{f}'\underline{x} = \sum_{s=1}^{p} \underline{f}_s'\underline{x}_s$$

and

(2.1.4) $$\hat{\underline{\theta}} = F'\underline{x} = \sum_{s=1}^{p} F_s'\underline{x}_s ,$$

respectively, where

$\underline{x}_s(N_s \times 1)$, $s=1,\ldots,p$ is the observation vector for V_s alone as defined by (1.4.7),

$\underline{x}(N \times 1) = (\underline{x}_1' \; \underline{x}_2' \; \ldots \; \underline{x}_p')'$,

$\underline{f}(N \times 1)$ is non-stochastic and independent of the elements of Σ

and

$F(N \times w)$ is non-stochastic and independent of the elements of Σ.

Srivastava [60] has considered BLUE estimation for the special cases of the HIM, GIM and MDM models. The approach described below generalizes Srivastava's results to the MGLM model. The method of proof simplifies Srivastava's work considerably. In addition, the results obtained for linear functions are extended to linear sets $\underline{\theta}$. Furthermore, a consistent and unbiased estimate of the covariance matrix is obtained and may be used to determine unbiased estimators of $\mathrm{var}(\hat{\theta})$.

2.2 BLUE Estimation of Linear Functions $\underline{h}'\underline{\xi}$

Definition 2.2.1: (Srivastava [60]) For the MGLM model, a linear function of the form $\underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$ is said to be piecewise estimable if each component sum $\underline{c}_s'\underline{\xi}_s$ has a BLUE $\underline{g}_s'\underline{x}_s$ under the univariate model for V_s alone as given in (1.4.6):

$$E(\underline{x}_s) = D_s\underline{\xi}_s \ , \quad \mathrm{Var}(\underline{x}_s) = \sigma_{ss} I_{N_s} \ .$$

(This is equivalent to requiring that $\underline{c}_s' \in V(D_s)$ for each s.)

Lemma 2.2.1. (Conditions for the estimability of $\underline{h}'\underline{\xi}$): For the MGLM model, $\theta = \underline{h}'\underline{\xi}$ is estimable if and only if it is piecewise estimable.

Proof: The method of proof used by Srivastava [60] for the special cases of the HIM, GIM and MDM models can be directly extended to the MGLM model.

Theorem 2.2.1: If, for the MGLM model, the variance-covariance matrix $\Sigma = ((\sigma_{rs}))$ is known, then every estimable linear function $\theta = \underline{h}'\underline{\xi}$ has a unique BLUE given by

$$(2.2.1) \qquad \hat{\theta} = \underline{h}'\hat{\underline{\xi}} = \underline{h}'(D'\Omega^{-1}D)^{-}D'\Omega^{-1}\underline{x}$$

where D and Ω are the design matrix and variance-covariance matrix from the vector version definition of the MGLM model as given in (1.4.8). The variance of this BLUE is given by

$$(2.2.2) \qquad \mathrm{Var}(\hat{\theta}) = \underline{h}'(D'\Omega^{-1}D)^{-}\underline{h} \ .$$

Proof: Clearly Ω is known, symmetric and positive definite if Σ is known, symmetric and positive definite. Thus (1.4.8) is equivalent to the univariate weighted least squares regression model which yields the BLUE of (2.2.1) and variance of (2.2.2).

q.e.d.

Because Σ is assumed unknown, Ω must be unknown, too. Thus if we restrict our estimators to be known linear functions of $\underline{x}$, then we cannot use the $\hat{\theta}$ in (2.2.1) unless it is independent of Ω. Fortunately, when considering the SM model, the structure of D and Ω allows Σ to drop out of (2.2.1) so that this estimate becomes the sum of the BLUE's for $\underline{c}_s'\underline{\xi}_s$ obtained separately for each s as given in (1.1.1.2). This is shown as follows using the notation of (1.1.1), (1.1.2) and (1.1.3):

$$\underline{h}'(D'\Omega^{-1}D)^{-}D'\Omega^{-1}\underline{y} = \underline{h}'[(I_p \otimes A')(\Sigma^{-1} \otimes I_n)(I_p \otimes A)]^{-}(I_p \otimes A')(\Sigma^{-1} \otimes I_n)\underline{y}$$

$$= \underline{h}'[\Sigma^{-1} \otimes A'A]^{-}[\Sigma^{-1} \otimes A']\underline{y}$$

$$= \underline{h}'[\Sigma \otimes (A'A)^{-}][\Sigma^{-1} \otimes A']\underline{y}$$

$$= \underline{h}'[I_p \otimes (A'A)^{-}A']\underline{y}$$

$$= \sum_{s=1}^{p} \underline{c}_s'(A'A)^{-}A'\underline{y}_s \; .$$

For the MGLM model in general, Σ does not drop out of (2.2.1) so that if we restrict estimators of θ to be linear, we might have to be satisfied with a $\hat{\theta}$ that does not achieve the minimum variance (2.2.2). It is, therefore, of interest to determine when the minimum variance is attained by a known linear function of $\underline{x}$, that is, a linear function which does not depend on the elements of Σ. Srivastava [60] has obtained for the special cases of the HIM, GIM and MDM models necessary and sufficient conditions for $\underline{h}'\underline{\xi}$ as defined in (2.1.1) to have a BLUE of the form $\underline{f}'\underline{x}$ where $\underline{f}$ is known. These are conditions on the vector spaces spanned by the columns of the design matrices D_s corresponding to the different response variates. Srivastava showed that if the BLUE is not a function of the elements of Σ, then it is given by the sum of the BLUE's for $\underline{c}_s'\underline{\xi}_s$ obtained separately for each s from the univariate model given in (1.4.6) corresponding to V_s alone. This result is thus the same as that obtained for the SM model except that in the latter the BLUE is always independent of Σ if $\underline{h}'\underline{\xi}$

is estimable whereas in the GLM and MDM models other conditions aside from estimability conditions must be satisfied. These results of Srivastava extend to the MGLM model. The proofs given by Srivastava have been simplified considerably by this author by use of the following lemma:

Lemma 2.2.2: $\hat{\theta} = \sum_{s=1}^{p} \underline{g}_s'\underline{x}_s$ is a BLUE for $\theta = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$

if and only if $\mathrm{Cov}(\hat{\theta}, z) = 0$ whenever $z = \sum_{s=1}^{p} \underline{d}_s'\underline{x}_s$ satisfies $E(z) = 0$.

Proof: The reader is referred to Rao [41], page 257, for a proof of this lemma.

It is important to note that if for the MGLM model $\hat{\theta}$ is restricted to be a linear function of $\underline{x}$ which does not involve Σ, then any $\hat{\theta}$ satisfying Lemma 2.2.2 will be the same estimator given in (2.2.1) and thus will have minimum variance for all functions of the data.

Theorem 2.2.2: Let $\theta = \underline{h}'\underline{\xi}$ be an estimable linear function of $\underline{\xi}$. If the BLUE for θ is not a function of the elements of Σ, then it is given by

(2.2.4) $$\hat{\theta} = \sum_{s=1}^{p} \underline{c}_s'(D_s'D_s)^{-}D_s'\underline{x}_s$$

i.e., the sum of the BLUE's of $\underline{c}_s'\underline{\xi}_s$ obtained separately for each s from the univariate model for V_s alone.

Proof: Since the BLUE is independent of Σ, it must be given by the BLUE obtained when $\Sigma = I$. This is clearly (2.2.4).

q.e.d.

Lemma 2.2.3: If $\underline{w}(t \times 1)$ and $\underline{z}(q \times 1)$ are arbitrary random variables and if $\underline{\alpha}(t \times 1)$ and $\underline{\beta}(q \times 1)$ are arbitrary non-stochastic vectors, then $\mathrm{Cov}(\underline{\alpha}'\underline{w}, \underline{\beta}'\underline{z}) = \underline{\alpha}'\mathrm{Cov}(\underline{w}, \underline{z})\underline{\beta}$.

Proof: This is a well known result which can be proved by direct computation.

Lemma 2.2.4: If for the MGLM model $\underline{f}'\underline{x} = \sum_{s=1}^{p} \underline{f}_s'\underline{x}_s$ and $\underline{d}'\underline{x} = \sum_{s=1}^{p} \underline{d}_s'\underline{x}_s$ are any two linear functions of $\underline{x}$, then

$$(2.2.5)\quad \mathrm{Cov}(\underline{f}'\underline{x}, \underline{d}'\underline{x}) = \sum_{s=1}^{p} \underline{f}_s'\underline{d}_s\sigma_{ss} + \sum_{r<s}^{p}\sum \underline{f}_r'Q_{rs}\underline{d}_s\sigma_{rs} + \sum_{r>s}^{p}\sum \underline{d}_s'Q_{sr}\underline{f}_r\sigma_{rs}$$

where $Q_{rs}(N_r \times N_s)$ is defined by (1.4.5).

Proof: From the variate-wise representation of the MGLM model as given in (1.4.6), we know that $\mathrm{Var}(\underline{x}_s) = \sigma_{ss}I_{N_s}$,

$\mathrm{Cov}(\underline{x}_r, \underline{x}_s) = \sigma_{rs}Q_{rs}$, $r < s$ and $\mathrm{Cov}(\underline{x}_r, \underline{x}_s) = \sigma_{rs}Q_{sr}'$, $r > s$.

Thus, using Lemma 2.2.3 where $\underline{\alpha} = \underline{f}_r$, $\underline{\beta} = \underline{d}_s$, $\underline{w} = \underline{x}_r$ and $\underline{z} = \underline{x}_s$ we get

$$\mathrm{Cov}(\underline{f}_r'\underline{x}_r, \underline{d}_s'\underline{x}_s) = \underline{f}_r' \mathrm{Cov}(\underline{x}_r, \underline{x}_s)\underline{d}_s$$

$$= \begin{cases} \underline{f}_s'\underline{d}_s\sigma_{ss}, & \text{if } r=s \\ \underline{f}_r'Q_{rs}\underline{d}_s\sigma_{rs}, & \text{if } r<s \\ \underline{f}_r'Q_{sr}'\underline{d}_s\sigma_{rs}, & \text{if } r>s. \end{cases}$$

Using Lemma 2.2.3 again where $\underline{\alpha} = \underline{1}_p = \underline{\beta}$, $\underline{w} = (\underline{f}_1'\underline{x}_1, \ldots, \underline{f}_p'\underline{x}_p)'$ and $\underline{z} = (\underline{d}_1'\underline{x}_1, \ldots, \underline{d}_p'\underline{x}_p)'$ we obtain

$$\text{Cov}(\underline{f}'\underline{x},\ \underline{d}'\underline{x}) = \text{Cov}\left(\sum_{s=1}^{p} \underline{f}_s'\underline{x}_s,\ \sum_{s=1}^{p} \underline{d}_s'\underline{x}_s\right)$$

$$= \sum_{r=1}^{p}\sum_{s=1}^{p} \text{Cov}(\underline{f}_r'\underline{x}_r,\ \underline{d}_s'\underline{x}_s)$$

$$= \sum_{s=1}^{p} \underline{f}_s'\underline{d}_s\sigma_{ss} + \sum\sum_{r<s}^{p} \underline{f}_r'Q_{rs}\underline{d}_s\sigma_{rs} + \sum\sum_{r>s}^{p} \underline{d}_s'Q_{sr}\underline{f}_r\sigma_{rs}\ .$$

q.e.d.

Corollary 2.2.3: If $\theta = \underline{h}'\underline{\xi}$ has a BLUE of the form $\hat{\theta} = \underline{g}'\underline{x}$ where $\underline{g}$ is known, then

$$(2.2.6)\quad \text{Var}(\hat{\theta}) = \sum_{s=1}^{p} \sigma_{ss}\underline{c}_s'(D_s'D_s)^{-}\underline{c}_s + 2\sum\sum_{r<s}^{p} \sigma_{rs}\underline{c}_s'(D_s'D_s)^{-}D_s'Q_{rs}'D_r(D_r'D_r)^{-}\underline{c}_r$$

where Q_{rs} is defined by (1.4.5).

Proof: From Theorem 2.2.2 we know that $\hat{\theta}$ is given by

$\hat{\theta} = \sum_{s=1}^{p} \underline{g}_s'\underline{x}_s$ where $\underline{g}_s' = \underline{c}_s'(D_s'D_s)^{-}D_s'$. Clearly since

$\text{Var}\ \hat{\theta} = \text{Cov}\left(\sum_{s=1}^{p} \underline{g}_s'\underline{x}_s,\ \sum_{s=1}^{p} \underline{g}_s'\underline{x}_s\right)$, we can use (2.2.5) of Lemma 2.2.4 with $\underline{g}_s = \underline{f}_s = \underline{d}_s$ to obtain the result (2.2.6).

q.e.d.

We now prove a lemma which will enable us to give necessary and sufficient conditions for $\theta = \underline{h}'\underline{\xi}$ to have a unique BLUE in terms of known linear functions of $\underline{x}$.

Lemma 2.2.5: Suppose $\theta = \underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$ is piecewise estimable and $\underline{g}'\underline{x} = \sum_{s=1}^{p} \underline{g}_s'\underline{x}_s$, where $\underline{g}_s' = \underline{c}_s'(D_s'D_s)^{-}D_s'$, is the unique BLUE of θ.

Then $\text{Cov}(\underline{g}'\underline{x},\ \underline{d}'\underline{x}) = 0$ whenever $\underline{d}'\underline{x} = \sum_{s=1}^{p} \underline{d}_s'\underline{x}_s$ satisfies

$E(\underline{d}'\underline{x}) = 0$ for all $\underline{\xi}$ if and only if

(2.2.7) $$\underline{g}_r' Q_{rs} \underline{d}_s = 0,\ r < s$$

and

(2.2.8) $$\underline{d}_s' Q_{sr} \underline{g}_r = 0,\ r > s$$

both hold whenever $E(\underline{d}'\underline{x}) = 0$ for all $\underline{\xi}$.

<u>Proof</u>: From Lemma 2.2.4 we know that

$$\mathrm{Cov}(\underline{g}'\underline{x},\ \underline{d}'\underline{x}) = \sum_{s=1}^{p} \underline{g}_s'\underline{d}_s \sigma_{ss} + \sum\sum_{r<s} \underline{g}_r' Q_{rs} \underline{d}_s \sigma_{rs} + \sum\sum_{r>s} \underline{d}_s' Q_{sr} \underline{g}_r \sigma_{rs}$$

for any $\underline{g}$ and $\underline{d}$. Furthermore, $E(\underline{d}'\underline{x}) = \sum_{s=1}^{p} \underline{d}_s' D_s \underline{\xi}_s = 0$ if and only if $\underline{d}_s' D_s = 0$ for each s, i.e., $E(\underline{d}'\underline{x}) = 0$ if and only if $\underline{d}_s'$ is orthogonal to $V(D_s')$ for each s. Now since $\underline{g}_s' = \underline{c}_s'(D_s' D_s)^{-} D_s'$ we must have $\underline{g}_s' \in V(D_s')$ for each s. Therefore, we must have $\underline{g}_s$ orthogonal to $\underline{d}_s$, i.e., $\underline{g}_s'\underline{d}_s = 0$ for each s whenever $E(\underline{d}'\underline{x}) = 0$ and this is so regardless of whether $\mathrm{Cov}(\underline{g}'\underline{x},\ \underline{d}'\underline{x}) = 0$. Thus we need only show

$\sum\sum_{r<s} \underline{g}_r' Q_{rs} \underline{d}_s \sigma_{rs} + \sum\sum_{r>s} \underline{d}_s' Q_{sr} \underline{g}_r \sigma_{rs} = 0$ whenever $\underline{d}_s' D_s = 0$ for each s <u>if and only if</u> each term in the summation is zero whenever $\underline{d}_s' D_s = 0$ for each s.

<u>Sufficiency</u> follows trivially. As for <u>necessity</u>, we note first that, by assumption, the result should be true regardless of the actual values of $\Sigma = ((\sigma_{rs}))$ provided Σ is positive definite. Now suppose we fix $s \in \{1, \ldots, p\}$ and choose $\underline{d}_s$ to satisfy $\underline{d}_s' D_s = 0$. Clearly $E(\underline{d}_s'\underline{x}_s) = 0$ so that by the assumption of necessity we must have

$$\mathrm{Cov}\left(\sum_{t=1}^{p} \underline{g}_t'\underline{x}_t,\ \underline{d}_s'\underline{x}_s\right) = 0$$

i.e.,
$$\sum_{t=1}^{s-1} \underline{g}_t'Q_{ts}\underline{d}_s\sigma_{ts} + \sum_{t=s+1}^{p} \underline{d}_s'Q_{st}\underline{g}_t\sigma_{ts} = 0 .$$

Let us now choose $r \in \{1, \ldots, p\}$ such that $r < s$ and then choose Σ so that its diagonal elements are all non-zero and its off diagonal elements are all zero except for $\sigma_{rs} = \sigma_{sr}$ which satisfies $\sigma_{ss}\sigma_{rr} - \sigma_{sr}^2 \geq 0$. This Σ is positive definite and the above equation reduces to $\mathrm{Cov}(\sum_{t=1}^{p} \underline{g}_t'\underline{x}_t, \underline{d}_s'\underline{x}_s) = \underline{g}_r'Q_{rs}\underline{d}_s\sigma_{rs} = 0$. Thus for any $r < s$ we can find a Σ so that $\underline{g}_r'Q_{rs}\underline{d}_s = 0$. Since the result must hold true regardless of Σ we therefore must always have $\underline{g}_r'Q_{rs}\underline{d}_s = 0$, $r < s$ whenever $\underline{d}_s'D_s = 0$. Similarly, for fixed s we can show that $\underline{d}_s'Q_{sr}\underline{g}_r = 0$, $r > s$ always, whenever $\underline{d}_s'D_s = 0$. By allowing s to range from 1 to p we thus get the desired result, i.e., (2.2.7) and (2.2.8) are satisfied.

q.e.d.

<u>Theorem 2.2.4. (Main Theorem of BLUE Estimation)</u>: For the MGLM model let $\theta = \underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$ be piecewise estimable and let $\underline{g}_s' = \underline{c}_s'(D_s'D_s)^- D_s'$, $s=1,\ldots,p$. Then necessary and sufficient conditions for $\underline{g}'\underline{x} = \sum_{s=1}^{p} \underline{g}_s'\underline{x}_s$ to be the unique BLUE for θ in terms of <u>known</u> linear functions of $\underline{x}$ are given by

(2.2.9) $\quad \underline{g}_r'Q_{rs} \in V(D_s')$, $r < s$

and

(2.2.10) $\quad \underline{g}_r'\ Q_{sr}' \in V(D_s')$, $r > s$

where $Q_{rs}(N_r \times N_s)$, $r < s$ is defined by (1.4.5).

Proof: From Lemma 2.2.2 and Theorem 2.2.2 we know that $\sum_{s=1}^{p} \underline{g}_s'\underline{x}_s$ is the unique BLUE for θ if and only if $\mathrm{Cov}(\sum_{s=1}^{p} \underline{g}_s'\underline{x}_s, \sum_{s=1}^{p} \underline{d}_s'\underline{x}_s) = 0$ whenever $\underline{d}_s'D_s = 0$ for each s. From Lemma 2.2.5 we know that $\mathrm{Cov}(\sum_{s=1}^{p} \underline{g}_s'\underline{x}_s, \sum_{s=1}^{p} \underline{d}_s'\underline{x}_s) = 0$ whenever $\underline{d}_s'D_s = 0$ for each s if and only if $\underline{g}_r'Q_{rs}\underline{d}_s = 0$, $r < s$ and $\underline{d}_s'Q_{sr}\underline{g}_r = 0$, $r > s$ whenever $\underline{d}_s'D_s = 0$ for each s. Clearly $\underline{g}_r'Q_{rs}\underline{d}_s = 0$, $r < s$ whenever $\underline{d}_s'D_s = 0$ is equivalent to (2.2.9) and $\underline{d}_s'Q_{sr}\underline{g}_r = \underline{g}_r'Q_{sr}'\underline{d}_s = 0$, $r > s$ whenever $\underline{d}_s'D_s = 0$ is equivalent to (2.2.10).

q.e.d.

An example of a MGLM model which is not one of the special cases considered by Srivastava and for which the conditions of Theorem 2.2.4 are satisfied is described as follows:

Let $p=3$, $u=3$, $q_1 = q_2 = q_3 = 2$

and consider the expectation model

$$E[Y_1(12 \times 2)] = [A_{11}\,\underline{\xi}_1 \mid A_{12}\,\underline{\xi}_2]$$

$$E[Y_2(8 \times 2)] = [A_{21}\,\underline{\xi}_1 \mid A_{22}\,\underline{\xi}_3]$$

$$E[Y_3(8 \times 2)] = [A_{31}\,\underline{\xi}_2 \mid A_{32}\,\underline{\xi}_3]$$

where

$$A_{11}(12 \times 6) = \begin{bmatrix} \underline{1}_4 & \underline{0}_4 & \underline{0}_4 & \underline{1}_4 & \underline{0}_4 & \underline{0}_4 \\ \underline{0}_4 & \underline{1}_4 & \underline{0}_4 & \underline{0}_4 & \underline{1}_4 & \underline{0}_4 \\ \underline{0}_4 & \underline{0}_4 & \underline{1}_4 & \underline{0}_4 & \underline{0}_4 & \underline{1}_4 \end{bmatrix}, \quad A_{12}(12 \times 6) = \begin{bmatrix} \underline{0}_4 & \underline{1}_4 & \underline{0}_4 & \underline{0}_4 & \underline{1}_4 & \underline{0}_4 \\ \underline{1}_4 & \underline{0}_4 & \underline{0}_4 & \underline{1}_4 & \underline{0}_4 & \underline{0}_4 \\ \underline{0}_4 & \underline{0}_4 & \underline{1}_4 & \underline{0}_4 & \underline{0}_4 & \underline{1}_4 \end{bmatrix}$$

$$A_{21}(8\times 6) = \begin{bmatrix} \underset{\sim}{0}_4 & \underset{\sim}{1}_4 & \underset{\sim}{0}_4 & \underset{\sim}{1}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 \\ \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{1}_4 & \underset{\sim}{0}_4 & \underset{\sim}{1}_4 & \underset{\sim}{0}_4 \end{bmatrix}, \quad A_{22}(8\times 6) = \begin{bmatrix} \underset{\sim}{1}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{1}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 \\ \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{1}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{1}_4 \end{bmatrix}$$

$$A_{31}(8\times 6) = \begin{bmatrix} \underset{\sim}{1}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{1}_4 & \underset{\sim}{0}_4 \\ \underset{\sim}{0}_4 & \underset{\sim}{1}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{1}_4 \end{bmatrix}, \quad A_{32}(8\times 6) = \begin{bmatrix} \underset{\sim}{1}_4 & \underset{\sim}{1}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 \\ \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{0}_4 & \underset{\sim}{1}_4 & \underset{\sim}{1}_4 \end{bmatrix}.$$

The matrices D_1, D_2 and D_3 are given by

$$D_1 = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}, \quad D_2 = \begin{bmatrix} A_{12} \\ A_{31} \end{bmatrix}, D_3 = \begin{bmatrix} A_{22} \\ A_{32} \end{bmatrix}.$$

A piecewise estimable linear function of the treatment parameters can be given by $\theta = \underset{\sim}{c}_1' \underset{\sim}{\xi}_1 + \underset{\sim}{c}_2' \underset{\sim}{\xi}_2 + \underset{\sim}{c}_3' \underset{\sim}{\xi}_3$ where

$$\underset{\sim}{c}_1' = (2,\ 0,\ -3,\ 2,\ -3,\ 0)$$

$$\underset{\sim}{c}_2' = (3,\ -2,\ 0,\ 3,\ 0,\ -2)$$

$$\underset{\sim}{c}_3' = (0,\ 0,\ 0,\ 0,\ 0,\ 0).$$

Note that we can write

$$\underset{\sim}{c}_1' = \underset{\sim}{a}_1'D_1 \quad \text{where } \underset{\sim}{a}_1' = (\underset{\sim}{1}_2',\ \underset{\sim}{0}_{14}',\ -1\ \underset{\sim}{1}_3',\ 0)$$

$$\underset{\sim}{c}_2' = \underset{\sim}{a}_2'D_2 \quad \text{where } \underset{\sim}{a}_2' = (\underset{\sim}{0}_4',\ \underset{\sim}{1}_3',\ \underset{\sim}{0}_9',\ -1\ \underset{\sim}{1}_2',\ \underset{\sim}{0}_2')$$

$$\underset{\sim}{c}_3' = \underset{\sim}{a}_3'D_3 \quad \text{where } \underset{\sim}{a}_3' = \underset{\sim}{0}_{16}' \quad .$$

After some calculations we will obtain

$$\underset{\sim}{g}_1' = \underset{\sim}{c}_1'(D_1'D_1)^- D_1' = \frac{1}{4}\ (2\ \underset{\sim}{1}_4',\ \underset{\sim}{0}_4',\ \underset{\sim}{0}_4'\ \underset{\sim}{0}_4',\ -3\ \underset{\sim}{1}_4')$$

$$\underset{\sim}{g}_2' = \underset{\sim}{c}_2'(D_2'D_2)^- D_2' = \frac{1}{4}\ (\underset{\sim}{0}_4',\ 3\ \underset{\sim}{1}_4',\ \underset{\sim}{0}_4',\ \underset{\sim}{0}_4',\ -2\ \underset{\sim}{1}_4')$$

$$\underset{\sim}{g}_3' = \underset{\sim}{c}_3'(D_3'D_3)^- D_3' = \underset{\sim}{0}_{16}' \ .$$

Now $\quad Q_{12}(20 \times 20) = \begin{bmatrix} I_{12} & 0_{12,8} \\ 0_{8,12} & 0_{8,8} \end{bmatrix}$, $\quad Q_{13}(20 \times 16) = \begin{bmatrix} 0_{12,8} & 0_{12,8} \\ I_8 & 0_{8,8} \end{bmatrix}$

and $\quad Q_{23}(16 \times 16) = \begin{bmatrix} 0_{8,8} & 0_{8,8} \\ 0_{8,8} & I_8 \end{bmatrix}$.

Thus we have

$$\underline{g}_1' Q_{12} = (\tfrac{1}{2}\underline{1}_4', \underline{0}_4', \underline{0}_4', \underline{0}_4', \underline{0}_4')$$

$$\underline{g}_2' Q_{12}' = (\underline{0}_4', \tfrac{3}{4}\underline{1}_4', \underline{0}_4', \underline{0}_4', \underline{0}_4')$$

$$\underline{g}_1' Q_{13} = (\underline{0}_4', -\tfrac{3}{4}\underline{1}_4', \underline{0}_4', \underline{0}_4')$$

$$\underline{g}_2' Q_{23} = (\underline{0}_4', \underline{0}_4', \underline{0}_4', -\tfrac{1}{2}\underline{1}_4')$$

$$\underline{g}_3' Q_{13}' = \underline{0}_{20}'$$

and $$\underline{g}_3' Q_{23}' = \underline{0}_{16}' \,.$$

It is easy to see that $\underline{g}_1'Q_{12} \in V(D_2')$, $\underline{g}_2'Q_{12}' \in V(D_1')$, $\underline{g}_1'Q_{13} \in V(D_3')$ and $\underline{g}_2'Q_{23} \in V(D_3')$ since we can write

$$\underline{g}_1' Q_{12} = \underline{f}'D_2' \quad \text{where } \underline{f}' = \tfrac{1}{2}(0, 1, 1, 0, 0, -1)$$

$$\underline{g}_2' Q_{12}' = \underline{h}'D_1' \quad \text{where } \underline{h}' = \tfrac{3}{4}(1, 1, 0, -1, 0, 0)$$

$$\underline{g}_1' Q_{13} = \underline{a}'D_3' \quad \text{where } \underline{a}' = -\tfrac{3}{4}(0, 0, 1, 0, 0, 0)$$

$$\underline{g}_2' Q_{23} = \underline{b}'D_3' \quad \text{where } \underline{b}' = -\tfrac{1}{2}(0, 0, 0, 0, 1, 0) \,.$$

Thus, for this example, the conditions of Theorem 2.2.4 are satisfied and we may conclude that the BLUE of the chosen θ is given by $\underline{g}_1'\underline{x}_1 + \underline{g}_2'\underline{x}_2 + \underline{g}_3'\underline{x}_3$ where $\underline{g}_1$, $\underline{g}_2$ and $\underline{g}_3$ are given above.[1]

---

[1]It can be established later from Theorem 2.2.10 that for this model <u>every</u> estimable function will have a BLUE which does not depend on the elements of Σ.

It should be noted here that if conditions (2.2.9) and (2.2.10) are not satisfied for some (r,s) pair, then Theorem 2.2.4 tells us the variance of $\underline{g}'\underline{x}$ is not minimum in terms of known functions of $\underline{x}$. Theorem 2.2.4 does not contradict the well known result that the sum of minimum variance unbiased estimators is the minimum variance unbiased estimator of the sum of the expected values of each of the estimators. Whereas $\underline{c}_s'(D_s'D_s)^- D_s'\underline{x}_s$ is the BLUE for $\underline{c}_s'\underline{\xi}_s$ based on only those observations on V_s, it is not necessarily the BLUE for $\underline{c}_s'\underline{\xi}_s$ based on all the observations in the entire experiment unless the appropriate conditions of Theorem 2.2.4 are satisfied. This might seem startling at first, but we can see that this result makes intuitive sense if, for example, in our experiment p=2, V_1 represents height, V_2 represents weight, and there are missing observations on height. Assuming height and weight are correlated, then those weight measurements corresponding to missing heights would provide some additional information for estimating any estimable function of heights alone. Thus we might not be surprised if some function of the height measurements alone is not as good an estimate as a function of both heights and weights. Of course, from Theorem 2.2.1 the actual BLUE for $\underline{c}_s'\underline{\xi}_s$ would be given by a linear function involving Σ. It is thus clear that Theorem 2.2.1 together with Theorem 2.2.4 suggests that linear estimation might not be as worthwhile as nonlinear estimation involving estimates of Σ. This is dealt with in the next chapter on BAN estimation.

Furthermore, note that it is incorrect to claim that the BLUE of $\underline{c}_s'\underline{\xi}_s$ for the entire experiment can always be obtained by taking the weighted average of the BLUE's for $\underline{c}_s'\underline{\xi}_s$ obtained from each independent set (e.g., S_j) of experimental units. Clearly, for example, if for $\underset{\sim}{y}_1(n_j \times 1)$ and $\underset{\sim}{y}_2(n_j \times 1)$, we have that $\underset{\sim}{y} = (\underset{\sim}{y}_1' \ \ \underset{\sim}{y}_2')'$ satisfies $E(\underset{\sim}{y}) = A\underset{\sim}{\xi}_s$, $Var(\underset{\sim}{y}) = \sigma_{ss} I_{n_1+n_2}$, $A = [A_1' \mid A_1']'$, then the BLUE of an estimable $\underset{\sim}{c}_s'\underset{\sim}{\xi}_s$ is given by $\underset{\sim}{c}_s'\hat{\underset{\sim}{\xi}}_s$ where

$$\hat{\underset{\sim}{\xi}}_s = (A_1'A_1 + A_2'A_2)^- A_1'\underset{\sim}{y}_1 + (A_1'A_1 + A_2'A_2)^- A_2'\underset{\sim}{y}_2$$

and not in general by $\underset{\sim}{c}_s'\tilde{\underset{\sim}{\xi}}_s$ where

$$\tilde{\underset{\sim}{\xi}}_s = \frac{n_1}{n_1 + n_2} (A_1'A_1)^- A_1'\underset{\sim}{y}_1 + \frac{n_2}{n_1 + n_2} (A_2'A_2)^- A_2'\underset{\sim}{y}_2 \ .$$

Thus one cannot claim that for the GIM model the BLUE for $\underset{\sim}{c}_s'\underset{\sim}{\xi}_s$ is

$$\underset{\sim}{c}_s'\tilde{\underset{\sim}{\xi}}_s = \underset{\sim}{c}_s' \sum_{j=1}^{u} \frac{n_j}{N_s} (A_j'A_j)^- A_j'Y_j\underset{\sim}{b}_{(j)s}$$

where we are using the notation given by (1.4.1.4) and $\underset{\sim}{b}_{(j)s}'$ is defined to be the s-th row of B_j. Similarly, it is incorrect to claim that $\sum_{s=1}^{p} \underset{\sim}{c}_s'\tilde{\underset{\sim}{\xi}}_s$ is the BLUE for $\sum_{s=1}^{p} \underset{\sim}{c}_s'\underset{\sim}{\xi}_s$, where $\underset{\sim}{c}_s'\tilde{\underset{\sim}{\xi}}_s$ is the weighted average estimate for each s.

Continuing with our theory, it is easy to see that Theorem 2.2.4 incorporates the results obtained by Srivastava [60] for the HIM, GIM and MDM models separately. For the HIM and MDM models the theorem simplifies considerably:

Corollary 2.2.5. (Srivastava [60]): For the HIM model as given in Section 1.4.1, let $\theta = \underset{\sim}{h}'\underset{\sim}{\xi} = \sum_{s=1}^{p} \underset{\sim}{c}_s'\underset{\sim}{\xi}_s$ be estimable and $\underset{\sim}{g}_s' = \underset{\sim}{c}_s'(D_s'D_s)^- D_s'$, $s=1,\dots,p$. Then necessary and sufficient conditions for $\underset{\sim}{g}'\underset{\sim}{x} = \sum_{s=1}^{p} \underset{\sim}{g}_s'\underset{\sim}{x}_s$

to be the unique BLUE for θ are given by

(2.2.11) $\quad \underline{g}'_{r(s)} \in V(D'_s)$, $r \neq s$

where

$$\underline{g}_{r(s)} = \begin{cases} \text{the } (N_s \times 1) \text{ vector given by the last } N_s \text{ elements} \\ \quad \text{of } \underline{g}_r \text{ , } r < s \\ \text{the } (N_s \times 1) \text{ vector given by } [\underline{0}'_{N_s - N_r}, \underline{g}'_r]' \text{ , } r > s \text{ .} \end{cases}$$

<u>Proof</u>: For the HIM model the matrix $Q_{rs}(N_r \times N_s)$, $r < s$, is given by $Q_{rs} = [0_{N_s, N_r - N_s} \vdots I_{N_s}]'$. Thus $\underline{g}'_r Q_{rs} = \underline{g}'_{r(s)}$, $r < s$ and $\underline{g}'_r Q'_{sr} = \underline{g}'_{r(s)}$, $r > s$ by direct computation and the result follows from Theorem 2.2.4.

q.e.d.

<u>Corollary 2.2.6</u>. (Srivastava [60]): For the MDM model as given in (1.4.1.1) let $\theta = \underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}'_s \underline{\xi}_s$ be estimable and let $\underline{g}'_s = \underline{c}'_s (A'_s A_s)^{-} A'_s$, $s=1,\ldots,p$. Then necessary and sufficient conditions for $\underline{g}'\underline{x} = \sum_{s=1}^{p} \underline{g}'_s \underline{x}_s$ to be the unique BLUE for θ are given by

(2.2.12) $\quad \underline{g}'_r \in \bigcap_{s=1}^{p} V(A'_s)$, $r=1,\ldots,p$.

<u>Proof</u>: For the MDM model the matrix $Q_{rs}(n \times n) = I_n$, $r < s$. Thus $\underline{g}'_r Q_{rs} = \underline{g}'_r$, $r < s$ and $\underline{g}'_r Q'_{sr} = \underline{g}'_r$, $r > s$ and the result follows from Theorem 2.2.4.

q.e.d.

It is important to note that whereas Theorem 2.2.4 provides necessary and sufficient conditions for a <u>given</u> estimable function $\underline{h}'\underline{\xi}$ to have a BLUE independent of the elements of Σ, we may want to find necessary and sufficient conditions for <u>every</u> estimable function $\underline{h}'\underline{\xi}$ to have a BLUE independent of the elements of Σ. The latter is a more restrictive requirement than the former. Srivastava [60] has considered this problem for the HIM, GIM and MDM models. His results extend directly to the MGLM model. Before presenting them, we make the following definitions:

<u>Definition 2.2.3</u>. (Srivastava [60]): An <u>orthogonal multi-response design</u> (OMD) is a design (i.e., model) in which <u>every</u> piecewise estimable function $\theta = \underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$ has a BLUE which does not depend on the elements of Σ.

<u>Definition 2.2.4</u>: For the MGLM model, let $D_{rs}(N_{rs} \times m_r)$ $r \neq s$ denote the matrix consisting of those rows of D_r which correspond to those sets of experimental units in which <u>both</u> variates V_r and V_s are measured and let $D^*_{rs}((N_r - N_{rs}) \times m_r)$ be the matrix consisting of the remaining rows.

The following theorems due to Srivastava [60], give necessary and sufficient conditions for the MDM, HIM and GIM models to be OMD:

<u>Theorem 2.2.7</u>. (Srivastava [60]): A necessary and sufficient condition that a MDM design is OMD is that $A_1, \ldots, A_p$ have identical column spaces; i.e., $V(A_1') = V(A_2') = \ldots = V(A_p')$.[2]

[2] We note that these conditions are satisfied for the fractional factorial design of Srivastava and the Latin Squares design of Gourlay as described in Section 1.3. A design for which $A_r'A_s = 0$, $r \neq s$, however, will not be OMD.

Theorem 2.2.8. (Srivastava [60]): A necessary and sufficient condition that a GIM design is OMD is $R(D_r) = R(D_{rs}) + R(D^*_{rs})$, $r \neq s$, $r,s=1,\ldots,p$.

Corollary 2.2.9. (Srivastava [60]): An HIM design is OMD if and only if $R(A'_1 \mid A'_2 \mid \ldots \mid A'_p) = \sum_{i=1}^{p} R(A_i)$.

Theorems 2.2.7 and 2.2.8 can be directly extended to the MGLM model using Srivastava's method of proof to obtain the following theorem:

Theorem 2.2.10: Necessary and sufficient conditions for a MGLM design to be OMD are

(2.2.13) $\quad R(D_r) = R(D_{rs}) + R(D^*_{rs})$

and

(2.2.14) $\quad V(D'_{rs}) = V(D'_{sr})$

$r \neq s; \quad r,s = 1,\ldots,p.$[3]

2.3 BLUE Estimation of Linear Sets $H'\underline{\xi}$

We now extend the results obtained for linear functions to linear sets of the form $\underline{\theta} = H'\underline{\xi} = \sum_{s=1}^{p} C'_s\underline{\xi}_s$ as given in (2.1.2). We seek estimators of the form $\underline{\theta} = F'\underline{x} = \sum_{s=1}^{p} F'_s\underline{x}_s$ as given in (2.1.4).

Definition 2.3.1: For the MGLM model a linear set of the form $H'\underline{\xi} = \sum_{s=1}^{p} C'_s\underline{\xi}_s$ is said to be a piecewise estimable set if for each s each component of $C'_s\underline{\xi}_s$ has a BLUE under the univariate model for V_s

---

[3] Note that for the GIM model $D_{rs} = D_{sr}$, $r \neq s$ whereas for the MGLM model, in which different response variates have different design matrices, we may have $D_{rs} \neq D_{sr}$, $r \neq s$.

alone.[4]

Lemma 2.3.1. (Conditions for the Estimability of $H'\underline{\xi}$): $H'\underline{\xi}$ is estimable if and only if it is a piecewise estimable set.

Proof: This is a direct extension of Lemma 2.3.1.

Definition 2.3.2: For the MGLM model, $\hat{\underline{\theta}} = F'\underline{x}$ will be called a BLUE set for $\underline{\theta} = H'\underline{\xi}$ if $F'\underline{x}$ is known and unbiased for $H'\underline{\xi}$ and if whenever $\tilde{\underline{\theta}} = G'\underline{x}$ is known and unbiased for $H'\underline{\xi}$ we must have $\Lambda = \text{Var}(\tilde{\underline{\theta}}) - \text{Var}(\hat{\underline{\theta}})$ is non-negative definite.

It is well known (e.g., Rao [41]) that for any least squares model of the form (1.4.8), the components of any BLUE set must be BLUE's of their respective expected values. Furthermore, if $\hat{\underline{\theta}}$ is a BLUE set for $\underline{\theta}$ and if $\tilde{\underline{\theta}}$ is any other known estimable set for $\underline{\theta}$ we must have

$$ch_{max}[\text{Var}(\hat{\underline{\theta}})] \leq ch_{max}[\text{Var}(\tilde{\underline{\theta}})] ,$$

$$tr[\text{Var}(\hat{\underline{\theta}})] < tr[\text{Var}(\tilde{\underline{\theta}})] ,$$

and

$$| \text{Var}(\hat{\underline{\theta}}) | < | \text{Var}(\tilde{\underline{\theta}}) | .$$

Thus a BLUE set is a "good" estimate in the sense of minimization of the maximum root, the trace and the generalized variance.

Theorem 2.3.1: For the MGLM model,

let $\underline{\theta} = H'\underline{\xi} = \sum_{s=1}^{p} C'_s\underline{\xi}_s$ be estimable. If Σ is known then $H'\underline{\xi}$ has a unique BLUE set given by

---

[4] This is equivalent to requiring that each of the rows of C'_s belong to $V(D_s)$ for each s.

(2.3.1) $\quad \hat{\underline{\theta}} = H'\hat{\underline{\xi}} = H'(D'\Omega^{-1}D)^{-}D'\Omega^{-1}\underline{x}$

whose variance-covariance matrix is given by

(2.3.2) $\quad \mathrm{Var}(\hat{\underline{\theta}}) = H'(D'\Omega^{-1}D)^{-}H$.

Proof: This follows directly from Theorem 2.2.1 since each component of $\hat{\underline{\theta}}$ is the BLUE of its corresponding component of $\underline{\theta}$.

Theorem 2.3.2: For the MGLM model, let $\underline{\theta} = H'\underline{\xi} = \sum_{s=1}^{p} C_s'\underline{\xi}_s$ be estimable and let $G_s' = C_s'(D_s'D_s)^{-}D_s'$, $s=1,\dots,p$. Then necessary and sufficient conditions for $G'\underline{x} = \sum_{s=1}^{p} G_s'\underline{x}_s$ to be the unique BLUE set for $\underline{\theta}$ in terms of known linear sets are given by:

(2.3.3) $\quad$ each row of $G_r'Q_{rs} \in V(D_s')$, $r < s$ and

(2.3.4) $\quad$ each row of $G_r'Q_{sr}' \in V(D_s')$, $r > s$ where

$Q_{rs}(N_r \times N_s)$, $r < s$ is defined by (1.4.5).

Proof: From Theorem 2.2.4 for linear functions it is clear that each component of $G'\underline{x}$ is a BLUE for its corresponding component of $H'\underline{\xi}$ if and only if (2.3.3) and (2.3.4) both hold. But we know that $G'\underline{x}$ is a BLUE set if and only if each component of $G'\underline{x}$ is a BLUE for its expected value.

q.e.d.

Corollary 2.3.3: If an estimable set $\underline{\theta} = H'\underline{\xi} = \sum_{s=1}^{p} C_s'\underline{\xi}_s$ has a unique BLUE set $\hat{\underline{\theta}} = G'\underline{x}$ in terms of known linear sets, then

(2.3.5) $\quad \mathrm{Var}(\hat{\underline{\theta}}) = \sum_{s=1}^{p} \sigma_{ss}C_s'(D_s'D_s)^{-}C_s + 2\sum\sum_{r<s}\sigma_{rs}C_r'(D_r'D_r)^{-}D_r'Q_{rs}D_s(D_s'D_s)^{-}C_s$

where Q_{rs} is given by (1.4.5).

Proof: This follows from direct calculation of $Var(\hat{\underline{\theta}})$ using $\hat{\underline{\theta}} = \sum_{s=1}^{p} G_s'\underline{x}_s$ as given in Theorem 2.3.2. q.e.d.

It is obvious that $\underline{\xi}$ (i.e., $H'\underline{\xi}$ when $H = I_M$) is estimable if and only if each D_s is of full rank, and in this case an unbiased estimator is given by $\hat{\underline{\xi}} = (\hat{\underline{\xi}}_1' \ldots \hat{\underline{\xi}}_p')'$ where

$$(2.3.6) \qquad \hat{\underline{\xi}}_s = (D_s'D_s)^{-1} D_s'\underline{x}_s \text{ , } s=1,\ldots,p \text{ .}$$

Each element of $\hat{\underline{\xi}}_s$ is the BLUE of the corresponding element of $\underline{\xi}_s$, $s=1,\ldots,p$, based on the observations on V_s alone.

We then have the following corollary:

Corollary 2.3.4: Suppose D_s is of full rank for each s. Then $\hat{\underline{\xi}} = (\hat{\underline{\xi}}_1' \ldots \hat{\underline{\xi}}_p')'$ is a BLUE set for $\underline{\xi}$ where $\hat{\underline{\xi}}_s$, $s=1,\ldots,p$ is given by (2.3.6) if and only if (2.3.3) and (2.3.4) hold where

$$G' = [G_1' \mid G_2' \mid \ldots \mid G_p'] = \begin{bmatrix} (D_1'D_1)^{-1}D_1' & & & 0 \\ & (D_2'D_2)^{-1}D_2' & & \\ & & \ddots & \\ 0 & & & (D_p'D_p)^{-1}D_p' \end{bmatrix} .$$

2.4 Unbiased and Consistent Estimation of Σ

Lemma 2.4.1: Let $\underline{y}_1$ and $\underline{y}_2$ be $(n \times 1)$ random vectors and let $Z(n \times 2) = (\underline{y}_1, \underline{y}_2)$. Suppose that $Var(Z) = I_n \otimes \Gamma$ where $\Gamma(2 \times 2) = ((\gamma_{ij}))$ is positive definite. Then for any non-stochastic matrix $G(n \times n)$ we have

$$E(\underline{y}_1' G \underline{y}_2) = \gamma_{12} \operatorname{tr} G + E(\underline{y}_1')\, G\, E(\underline{y}_2) .$$

<u>Proof</u>: Let $\underline{y}_1 = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n} \end{bmatrix}$, $\underline{y}_2 = \begin{bmatrix} y_{21} \\ \vdots \\ y_{2n} \end{bmatrix}$, $G = ((g_{ij}))$.

Then

$$\begin{aligned} E(\underline{y}_1' G \underline{y}_2) &= \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} E(y_{1i} y_{2j}) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} [\operatorname{Cov}(y_{1i}, y_{2j}) + E(y_{1i}) E(y_{2j})] \\ &= \sum_{i=1}^{n} g_{ii} \gamma_{12} + \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} E(y_{1i}) E(y_{2j}) \end{aligned}$$

since $\operatorname{Var}(Z) = I_n \otimes \Gamma$ by assumption

$$= \gamma_{12} \operatorname{tr} G + E(\underline{y}_1')\, G\, E(\underline{y}_2) .$$

q.e.d.

<u>Theorem 2.4.1</u>: An unbiased estimator $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$ of Σ for the MGLM model is given by[5]

$$(2.4.1) \quad \hat{\sigma}_{rr} = \frac{1}{N_r - R(D_r)} \underline{x}_r' [I_{N_r} - D_r (D_r' D_r)^- D_r'] \underline{x}_r , \quad r=1,\ldots,p$$

and

$$(2.4.2) \quad \hat{\sigma}_{rs} = \frac{1}{\mu_{rs}} \underline{x}_{rs}' [I_{N_{rs}} - D_{rs} (D_{rs}' D_{rs})^- D_{rs}'] [I_{N_{rs}} - D_{sr} (D_{sr}' D_{sr})^- D_{sr}'] \underline{x}_{sr} ,$$

$r \neq s$

where

---

[5]It is shown in Chapter Five that the $\hat{\Sigma}$ of Theorem 2.4.1 is not necessarily positive definite. The implications of this on estimation of $\underline{\xi}$ using $\hat{\Sigma}$ and on testing hypotheses about linear sets $H'\underline{\xi}$ are discussed in Section 5.3 of Chapter Five.

(2.4.3) $\mu_{rs} = \mathrm{tr}[I_{N_{rs}} - D_{rs}(D'_{rs}D_{rs})^{-}D'_{rs}][I_{N_{rs}} - D_{sr}(D'_{sr}D_{sr})^{-}D'_{sr}]$,

$N_r (\geq 2)$ is the number of experimental units on which V_r is observed,

$N_{rs} (\geq 2)$ is the number of experimental units on which both V_r and V_s are observed together,

$\underline{x}_r (N_r \times 1)$ is the vector of all observations on V_r ,

$\underline{x}_{rs} (N_{rs} \times 1)$, $r \neq s$ is the vector of observations on V_r which correspond to experimental units on which both V_r and V_s are observed together,

$D_r (N_r \times m_r)$ is the design matrix corresponding to $\underline{x}_r$,

and $D_{rs} (N_{rs} \times m_r)$ is the design matrix corresponding to $\underline{x}_{rs}$,[6]

Proof: For $r \neq s$, we have that if $Z_{rs} = (\underline{x}_{rs}, \underline{x}_{sr})$ then

$E(Z_{rs}) = (D_{rs}\underline{\xi}_r \mid D_{sr}\underline{\xi}_s)$ and

$\mathrm{Var}(Z_{rs}) = I_{N_{rs}} \otimes \Sigma_{rs}$, where

$$\Sigma_{rs} = \begin{bmatrix} \sigma_{rr} & \sigma_{rs} \\ \sigma_{rs} & \sigma_{ss} \end{bmatrix} .$$

Letting $G = [I_{N_{rs}} - D_{rs}(D'_{rs}D_{rs})^{-}D'_{rs}][I_{N_{rs}} - D_{sr}(D'_{sr}D_{sr})^{-}D'_{sr}]$,

we may use Lemma 2.3.1 to obtain

$$E(\underline{x}'_{rs}G\underline{x}_{sr}) = \sigma_{rs}\mu_{rs} + E(\underline{x}'_{rs})GE(\underline{x}_{sr}) .$$

But

$$E(\underline{x}'_{rs})GE(\underline{x}_{sr}) = (D_{rs}\underline{\xi}_r)'[I_{N_{rs}} - D_{rs}(D'_{rs}D_{rs})^{-}D'_{rs}][I_{N_{rs}} - D_{sr}(D'_{sr}D_{sr})^{-}D'_{sr}]D_{sr}\underline{\xi}_s$$

---

[6] Note that in general $\underline{x}_{rs} \neq \underline{x}_{sr}$, $D_{rs} \neq D_{sr}$ but $N_{rs} = N_{sr}$, $r \neq s$.

$$= (D_{rs}\underline{\xi}_r)'[I_{N_{rs}} - D_{rs}(D'_{rs}D_{rs})^- D'_{rs}][D_{sr}\underline{\xi}_s - D_{sr}(D'_{sr}D_{sr})^- D'_{sr}D_{sr}\underline{\xi}_s]$$

$$= (D_{rs}\underline{\xi}_r)'[I_{N_{rs}} - D_{rs}(D'_{rs}D_{rs})^- D'_{rs}][D_{sr}\underline{\xi}_s - D_{sr}\underline{\xi}_s]$$

by virtue of the properties of $(D'_{sr}D_{sr})^-$

$$= 0.$$

Thus we have that

$$\sigma_{rs} = E(\frac{1}{\mu_{rs}} \underline{x}'_{rs} G\, \underline{x}_{sr}),\ r \neq s$$

i.e., $E(\hat{\sigma}_{rs}) = \sigma_{rs}$, $r \neq s$.

It is clear that $E(\hat{\sigma}_{rr}) = \sigma_{rr}$, since $\hat{\sigma}_{rr}$ is the usual estimate of error obtained for the univariate model:

$$E(\underline{x}_r) = D_r\underline{\xi}_r\ ,\ \mathrm{Var}(\underline{x}_r) = \sigma_{rr} I_{N_r}\ .$$

q.e.d.

It is important to note here that by requiring $N_{rs} \geq 2$, $r \neq s$, Theorem 2.4.1 restricts the use of the MGLM model to situations for which every pair of response variates is observed together for at least one set (S_i) of experimental units and for at least two units in total. This is clearly not a severe restriction, however.

<u>Corollary 2.4.2</u>: An alternative unbiased estimator of σ_{rs}, $r \neq s$ for the MGLM model is given by

$$(2.4.4) \qquad \tilde{\sigma}_{rs} = \hat{\sigma}_{rs} + k_1\{\hat{\sigma}_{ss} - \tilde{\sigma}_{ss}\} + k_2\{\hat{\sigma}_{rr} - \tilde{\sigma}_{rr}\}$$

where $\hat{\sigma}_{rr}$ and $\hat{\sigma}_{ss}$ are given by (2.4.1), $\hat{\sigma}_{rs}$ is given by (2.4.2),

$$\tilde{\sigma}_{rr} = \frac{1}{N_{rs} - R(D_{rs})} \underline{x}'_{rs}[I_{N_{rs}} - D_{rs}(D'_{rs}D_{rs})^- D'_{rs}]\underline{x}_{rs}\ ,$$

$$\tilde{\sigma}_{ss} = \frac{1}{N_{rs} - R(D_{sr})} \underline{x}'_{sr}[I_{N_{rs}} - D_{sr}(D'_{sr}D_{sr})^- D'_{sr}]\underline{x}_{sr}\ ,$$

and k_1, k_2 are arbitrary constants.

Proof: It is obvious that $E(\tilde{\sigma}_{tt}) = E(\hat{\sigma}_{tt}) = \sigma_{tt}$ for t=r,s.

Thus $$E(\tilde{\sigma}_{rs}) = E(\hat{\sigma}_{rs}) + 0 + 0$$
$$= \sigma_{rs} .$$ q.e.d.

Corollary 2.4.3: For the MGLM model, the $\hat{\Sigma}$ defined by (2.4.1) and (2.4.2) is a consistent estimator of Σ provided the following conditions both hold:

(2.4.5) $\lim_{n\to\infty} \frac{N_s}{n}$ exists and is non-zero, $s=1,\ldots,p$.

and

(2.4.6) $\lim_{n\to\infty} \frac{N_{rs}}{n}$ exists and is non-zero, $r\neq s$.

Proof: Using the assumption of normality on the data for the MGLM model, it is clear that for each r, $\hat{\sigma}_{rr}$ as given by (2.4.1) is the usual pooled estimate of error based on the N_s observations on V_s alone which is known (e.g., Rao [41], Anderson [8]) to be consistent for σ_{ss} as $N_s \to \infty$. Clearly if $N_s \to \infty$ as $n \to \infty$ according to condition (2.4.5), then $\hat{\sigma}_{ss}$ is consistent for σ_{ss} as $n \to \infty$. Similarly for $r\neq s$, $\hat{\sigma}_{rs}$ as given by (2.4.2) is the usual pooled estimate of covariance based on the N_{rs} observations on V_r and V_s together which is known (e.g., Rao [41], Anderson [8]) to be consistent for σ_{rs} as $N_{rs}\to \infty$. Clearly if $N_{rs}\to \infty$ as $n \to \infty$ according to condition (2.4.6), then $\hat{\sigma}_{rs}$ is consistent for σ_{rs} as $n \to \infty$.

q.e.d.

We now use the $\hat{\Sigma}$ given in Theorem 2.4.1 to obtain an unbiased estimator of the variance of the BLUE of an estimable linear function of the treatment parameters.

Corollary 2.4.4: For the MGLM model let $\theta = \underline{h}'\underline{\xi} = \sum_{s=1}^{p} \underline{c}_s'\underline{\xi}_s$ be estimable and let $\underline{g}_s' = \underline{c}_s'(D_s'D_s)^- D_s'$. Then an unbiased estimator of the variance of $\hat{\theta} = \sum_{s=1}^{p} \underline{g}_s'\underline{x}_s$ is given by

$$\text{(2.4.7)} \qquad \hat{\text{Var}}(\hat{\theta}) = \sum_{s=1}^{p} \hat{\sigma}_{ss}\underline{g}_s'\underline{g}_s + 2 \sum\sum_{r<s} \hat{\sigma}_{rs}\underline{g}_s'Q_{rs}\underline{g}_r$$

where

$\hat{\sigma}_{ss}$ and $\hat{\sigma}_{rs}$ are obtained from (2.4.1) and (2.4.2), respectively, of Theorem 2.4.1 and Q_{rs} , $r < s$ is given by (1.4.5).

Proof: This is obvious from Corollary 2.2.3 and Theorem 2.4.1.

We now apply the results of Theorem 2.4.1 and its corollaries to the special cases of the MDM, GIM and HIM models.

Corollary 2.4.5: For the MDM model, a consistent and unbiased estimate of Σ is given by $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$ where

$$\text{(2.4.8)} \qquad \hat{\sigma}_{ss} = \frac{1}{n - R(A_s)} \underline{y}_s'[I_n - A_s(A_s'A_s)^- A_s']\underline{y}_s \ , \ s=1,\dots,p$$

$$\text{(2.4.9)} \qquad \hat{\sigma}_{rs} = \frac{1}{\mu_{rs}} \underline{y}_s'[I_n - A_s(A_s'A_s)^- A_s'][I_n - A_r(A_r'A_r)^- A_r']\underline{y}_r \ , \ r \neq s$$

and

$$\mu_{rs} = \text{tr}[I_n - A_s(A_s'A_s)^- A_s'][I_n - A_r(A_r'A_r)^- A_r'] \ .$$

Corollary 2.4.6: For the GIM model, a consistent and unbiased estimate of Σ is given by $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$ where

$$(2.4.10) \qquad \hat{\sigma}_{ss} = \frac{1}{N_s - R(D_s)} \underline{x}_s'[I_{N_s} - D_s(D_s'D_s)^- D_s']\underline{x}_s \;, \; s=1,\ldots,p$$

and

$$(2.4.11) \qquad \hat{\sigma}_{rs} = \frac{1}{N_{rs} - R(D_{rs})} \underline{x}_{rs}'[I_{N_{rs}} - D_{rs}(D_{rs}'D_{rs})^- D_{rs}']\underline{x}_{sr} \;, \; r\neq s.$$ [7]

Corollary 2.4.7: For the HIM model a consistent and unbiased estimate of Σ is given by $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$ where $\hat{\sigma}_{ss}$, $s=1,\ldots,p$ is given by (2.4.10) and

$$(2.4.12) \qquad \hat{\sigma}_{rs} = \frac{1}{N_s - R(D_s)} \underline{x}_s'[I_{N_s} - D_s(D_s'D_s)^- D_s']Q_{rs}'\underline{x}_r \;, \; r < s$$

$$(2.4.13) \qquad \hat{\sigma}_{rs} = \frac{1}{N_r - R(D_r)} \underline{x}_r'[I_{N_r} - D_r(D_r'D_r)^- D_r']Q_{sr}'\underline{x}_s \;, \; r > s \;.$$ [8]

---

[7] Note that $D_{rs} = D_{sr}$, $r\neq s$ for the GIM model.

[8] Note that $\hat{\sigma}_{rs} = \hat{\sigma}_{sr}$ for $r\neq s$ and also $D_{rs} = D_s$, $\underline{x}_{rs} = Q_{rs}'\underline{x}_r$ $\underline{x}_{sr} = \underline{x}_s$ for $r < s$.

CHAPTER III

BAN ESTIMATION FOR THE MGLM AND GGCM MODELS

3.1 Introduction

We have seen in the previous chapter that for the MGLM model the unique BLUE of any estimable linear function or linear set of the treatment parameters is given by a linear function (2.2.1) or linear set (2.3.1), respectively, which involves the unknown parameters of the variance matrix Σ. By restricting linear estimates to be known functions which do not involve Σ we will obtain estimates whose variances are not minimum unless certain additional conditions on the model (given in Theorem 2.2.4 and Theorem 2.3.2) are satisfied. We have also seen in the first chapter that for the GCM model the unique BLUE of any estimable linear function of the treatment parameters is given by a linear function (1.2.1.4) of the data which involves the unknown variance matrix. These results naturally lead us to consider nonlinear methods of estimation for the MGLM and GGCM models which use estimates of Σ and which give variances that are, in large samples, the minimum that could be achieved by linear estimators if Σ were known.

In this chapter we consider for these models a method of estimation which generally yields nonlinear estimators whose variances are minimum in large samples. Estimators obtained by this method are called Best Asymptotically Normal (BAN) estimators. BAN estimators and their limiting variances may sometimes, as with many linear models, be determined without knowing the likelihood function for the data of the experiment. In general, however, the latter is needed. Specifically, we define a BAN estimator as follows:

<u>Definition 3.1.1</u>: An estimator $\tilde{\underset{\sim}{\theta}}_n(u \times 1)$ based on a sample of n observations is said to be a BAN estimator for the parameter $\underset{\sim}{\theta}(u \times 1) = (\theta_1, \ldots, \theta_u)'$ provided

$$(3.1.1) \qquad \mathcal{L}(\sqrt{n}\, B_n^{\frac{1}{2}} (\tilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o)) \to \mathcal{N}_u(\underset{\sim}{0}_u, I_u)$$

where

$$(3.1.2) \qquad B_n(u \times u) = \text{Fisher's Information Matrix}$$

$$= E_{\underset{\sim}{\theta}_o}\left[- \frac{1}{n} \frac{\partial^2 \log \phi_n}{\partial\, \underset{\sim}{\theta}^2} \Bigg|_{\underset{\sim}{\theta} = \underset{\sim}{\theta}_o} \right] ,$$

$\underset{\sim}{\theta}_o$ is the true value of $\underset{\sim}{\theta}$,

ϕ_n is the likelihood function for the sample,

and matrix differentiation is defined in the Appendix.

It is well known (e.g., Rao [41]) that in general, if B_n is positive definite, then $\frac{1}{n} B_n^{-1}$ is the multiparameter Cramer-Rao lower bound for the variance of an unbiased estimate of $\underset{\sim}{\theta}$, i.e., $\text{Var}(\underset{\sim}{\theta}^*_n) - \frac{1}{n} B_n^{-1}$ is non-negative definite whenever $\underset{\sim}{\theta}^*_n$ is any unbiased estimate of $\underset{\sim}{\theta}$. Furthermore, if $B_n^{-1} \to B^{-1}$ as $n \to \infty$, then $\text{Var}(\underset{\sim}{\theta}^*_n) - \frac{1}{n} B^{-1}$ is non-negative definite. It is in this sense that we consider a BAN estimator to have a minimum limiting variance

$\frac{1}{n} B_n^{-1}$ or $\frac{1}{n} B^{-1}$.

It is also well known (e.g., Cramer [17], Rao [41]) that when the data for an experiment consists of independent and identically distributed (i.i.d.) random variables, maximum likelihood estimators (MLE's) are BAN provided the underlying density function satisfies certain regularity conditions. In this case, if we let $\underline{x}_1, \ldots, \underline{x}_n$ denote the data and $p(\underline{x}, \underline{\theta})$ denote the density function where $\underline{\theta}(u \times 1)$ is an unknown parameter, then

$$\phi_n = \prod_{i=1}^{n} p(\underline{x}_i, \underline{\theta}) \text{ and } B_n(u \times u) = E_{\underline{\theta}_o}\left[- \frac{\partial^2 p(\underline{x}, \underline{\theta})}{\partial \underline{\theta}^2} \,\middle|\, \underline{\theta} = \underline{\theta}_o \right].$$

Now in many experiments (e.g., the SM and the MGLM models) the original data are independent but not identically distributed. Nevertheless, it may often be shown for such experiments that the solution to the MLE equations or any other statistic based on the data is BAN merely be calculating its limiting distribution. This can be done for many simply structured statistics and for iterative estimators based initially on BAN estimators. For the SM model, for example, it is well known that by transforming the original data of (1.1.1) to $X = Y - A\xi$, we obtain i.i.d. data. Thus the MLE of Σ is BAN. Furthermore, irrespective of this, the solution of the original untransformed maximum likelihood equations for ξ when estimable is given by $(A'A)^{-1}A'Y$, which is clearly normally distributed (if the rows of Y are normal), BLUE and therefore BAN. For the MGLM model it is clear from Theorem 2.3.1 that if Ω is known, then the MLE of an estimable set $\underline{\theta} = H'\underline{\xi}$ is given by the weighted least squares estimate (using the vector version notation)

(3.1.3) $\quad \hat{\underline{\theta}} = H'\hat{\underline{\xi}} = H'(D'\Omega^{-1}D)^{-}D'\Omega^{-1}\underline{x}$.

This is clearly normally distributed, since $\underline{x}$ is assumed normal, and efficient (i.e., minimum variance). If Ω is unknown but can be estimated consistently by $\hat{\Omega}$, it is also clear that the restricted MLE of $\underline{\theta}$ obtained by replacing Ω by $\hat{\Omega}$ in (3.1.3) is asymptotically efficient and normal, since $\hat{\Omega} \xrightarrow[p]{} \Omega$. If, furthermore, Ω is unknown and is estimated by $\Omega(\hat{\Sigma})$, where $\hat{\Sigma}$ is the solution to the MLE equations for Σ, then the restricted MLE of $\underline{\theta}$ using $\Omega(\hat{\Sigma})$ is asymptotically normal and efficient since $\Omega(\hat{\Sigma})$ is consistent from general MLE theory.

In this chapter, we describe for the MGLM and GGCM models a number of different methods for obtaining BAN estimators which use different consistent estimators of Ω in (3.1.3). Whereas it is clear from above that all these estimators are asymptotically equivalent, their small sample properties are different and ought to be considered whenever we are not completely sure that the sample size is large. We will later describe a computer study which has dealt with such small sample comparisons. We begin this chapter by deriving the limiting variance matrix for BAN estimators of estimable sets $\underline{\xi}$ and $H'\underline{\xi}$. This matrix must be obtained for the calculation of the Wald Statistic described in Chapter Four and used for testing hypotheses. The limiting variance matrix solution is fairly obvious when using the vector version notation. Nevertheless, it is of considerable practical utility because of the unwieldy size of Ω to express this result in terms of the original notation (involving much smaller matrices) given in (1.4.1). This can be done with some effort using a number of valuable matrix differentiation rules. We then describe several BAN estimators for $\underline{\xi}$

(when estimable) which have been proposed in the literature for the GIM and MDM models. In addition, we give an extension of some results of Bhargava [13] using a method of Anderson [8] to obtain solutions to the MLE equations for $\underline{\xi}$ in the HIM model. A BAN estimator for $H'\underline{\xi}$ for the MGLM model is then proposed and two methods of iteration which use this estimator initially are described. Finally, we consider BAN estimation of $H'\underline{\xi}$ for the GGCM model.

3.2 Derivation of the Asymptotic Variance Matrix of a BAN Estimator for the MGLM Model

Recalling the notation of (1.4.1) to (1.4.8) defining the MGLM model, we can write its likelihood function ϕ_n in two different ways:

$$(3.2.1) \qquad \log \phi_n = -\frac{1}{2} N \log 2\pi - \frac{1}{2} \log|\Omega| - \frac{1}{2} (\underline{x} - D\underline{\xi})'\Omega^{-1}(\underline{x} - D\underline{\xi})$$

where N, D, Ω and $\underline{\xi}$ are defined from the vector-version model (1.4.8), or

$$(3.2.2) \qquad \log \phi_n = -\sum_{j=1}^{u} \frac{n_j q_j}{2} \log 2\pi - \sum_{j=1}^{u} \frac{n_j}{2} \log |B_j'\Sigma B_j|$$

$$- \frac{1}{2} \sum_{j=1}^{u} \operatorname{tr}[B_j'\Sigma B_j]^{-1} [Y_j - E(Y_j)]'[Y_j - E(Y_j)]$$

using the notation of (1.4.1), where

$$E(Y_j) = [A_{j\ell_{j1}} \underline{\xi}_{\ell_{j1}} \mid A_{j\ell_{j2}} \underline{\xi}_{\ell_{j2}} \mid \cdots \mid A_{j\ell_{jq_j}} \underline{\xi}_{\ell_{jq_j}}] .$$

To obtain the information matrix for $\underline{\xi}$ as given in (3.1.4) we will use four of the matrix differentiation rules described in the Appendix. These are:

$$(3.2.3) \qquad \frac{\partial}{\partial \underline{\mu}} (\underline{y} - A\underline{\mu})'\Gamma(\underline{y} - A\underline{\mu}) = -2A'\Gamma(\underline{y} - A\underline{\mu}),$$

(3.2.4) $$\frac{\partial}{\partial \underline{\mu}} B\underline{\mu} = B' ,$$

(3.2.5) $$\frac{\partial}{\partial \Lambda} \operatorname{tr} U^{-1}P'P = -2B'PU^{-1}M' ,$$

where $U = M'\Gamma M$ and $P = X - B\Lambda M$,

and

(3.2.6) $$\frac{\partial}{\partial \Lambda} A\Lambda B = B' \otimes A' .$$

Now using (3.2.3) on (3.2.1) we obtain

(3.2.7) $$\frac{\partial \log \phi_n}{\partial \underline{\xi}} = D'\Omega^{-1}(\underline{x} - D\underline{\xi}).$$

We then use (3.2.4) on (3.2.7) to obtain

(3.2.8) $$\frac{\partial^2 \log \phi_n}{\partial \underline{\xi}^2} = \frac{\partial}{\partial \underline{\xi}} \left[\frac{\partial \log \phi_n}{\partial \underline{\xi}}\right] = -D'\Omega^{-1}D$$

Thus the <u>information matrix for $\underline{\xi}$</u> in the MGLM model is given by

(3.2.9) $$B_{n,\underline{\xi}} = \frac{1}{n} D'\Omega^{-1}D$$

where Ω is the true variance-covariance matrix. The <u>asymptotic variance matrix for $\underline{\xi}$</u> in the MGLM model is then given by

(3.2.10) $$\frac{1}{n} B^{-1}_{n,\underline{\xi}} = (D'\Omega^{-1}D)^{-1}$$

provided D is of full rank and Ω is positive definite. In terms of the notation used in the original definition of the MGLM model as given in (1.4.1), the information matrix $B_{n,\underline{\xi}}$ is not easy to present. We can, nevertheless, obtain $B_{n,\underline{\xi}}$ in the original notation using the likelihood function in (3.2.2) for the nicer cases of the GIM and MDM models and then inferring from the information matrices so obtained the information matrix for $\underline{\xi}$ in the MGLM model.

The <u>likelihood function for the GIM model</u> is given by:

$$(3.2.11)\qquad \log \phi_n = -\sum_{j=1}^{u} \frac{n_j q_j}{2} \log 2\pi - \sum_{j=1}^{u} \frac{n_j}{2} \log|B_j'\Sigma B_j|$$

$$- \frac{1}{2} \sum_{j=1}^{u} \operatorname{tr}[B_j'\Sigma B_j]^{-1}[Y_j - A_j\xi B_j]'[Y_j - A_j\xi B_j] ,$$

where $\xi = (\underline{\xi}_1\ \underline{\xi}_2\ \cdots\ \underline{\xi}_p)$.

Using (3.2.5) on (3.2.11) we obtain

$$\frac{\partial \log \phi_n}{\partial \xi} = \sum_{j=1}^{u} A_j'(Y - A_j\xi B_j)(B_j'\Sigma B_j)^{-1} B_j'$$

$$(3.2.12)\qquad = \sum_{j=1}^{u} A_j'Y(B_j'\Sigma B_j)^{-1} B_j' - \sum_{j=1}^{u} A_j'A_j\xi B_j(B_j'\Sigma B_j)^{-1} B_j'$$

We then use (3.2.6) on (3.2.12) to obtain

$$(3.2.13)\qquad \frac{\partial^2 \log \phi_n}{\partial \xi^2} = -\sum_{j=1}^{u} [B_j(B_j'\Sigma B_j)^{-1} B_j' \otimes A_j'A_j]$$

Thus the information matrix for $\underline{\xi}$ for the GIM model is given by

$$(3.2.14)\qquad B_{n,\underline{\xi}}(\text{GIM}) = \frac{1}{n} \sum_{j=1}^{u} [B_j(B_j'\Sigma B_j)^{-1} B_j' \otimes A_j'A_j]$$

The likelihood function for the MDM model can be given by

$$(3.2.15)\qquad \log \phi_n = -\frac{np}{2} \log 2\pi - \frac{n}{2} \log|\Sigma| - \frac{1}{2} \operatorname{tr}\Sigma^{-1}[Y-E(Y)]'[Y-E(Y)]$$

where $E(Y) = [A_1\underline{\xi}_1 \,\vdots\, A_2\underline{\xi}_2 \,\vdots\, \cdots \,\vdots\, A_p\underline{\xi}_p]$.

Letting $Y = [\underline{y}_1 \cdots \underline{y}_p]$ it is easy to see that the (r,s)-th element of $[Y - E(Y)]'[Y - E(Y)]$ is given by $(\underline{y}_r - A_r\underline{\xi}_r)'(\underline{y}_s - A_s\underline{\xi}_s)$ so that

$$\operatorname{tr}\Sigma^{-1}[Y - E(Y)]'[Y - E(Y)] = \sum_{r=1}^{p} \sum_{s=1}^{p} \sigma^{rs}(\underline{y}_r - A_r\underline{\xi}_r)'(\underline{y}_s - A_s\underline{\xi}_s) ,$$

where $\Sigma^{-1} = ((\sigma^{rs}))$. Now using (3.2.3) on (3.2.15) we obtain

$$\frac{\partial \log \phi_n}{\partial \underline{\xi}_t} = -\frac{1}{2}\frac{\partial}{\partial \underline{\xi}_t}\left[\sum_{r=1}^{p}\sum_{s=1}^{p} \sigma^{rs}(\underline{y}_r - A_r\underline{\xi}_r)'(\underline{y}_s - A_s\underline{\xi}_s)\right]$$

$$= -\frac{1}{2}[-2\sigma^{tt}A_t'(\underline{y}_t - A_t\underline{\xi}_t) -2\sum_{\substack{s=1\\ s\neq t}}^{p} \sigma^{ts}A_t'(\underline{y}_s - A_s\underline{\xi}_s)]$$

(3.2.16) $$= \sum_{s=1}^{p} \sigma^{ts}A_t'(\underline{y}_s - A_s\underline{\xi}_s) \ .$$

Using (3.2.4) on (3.2.16) we obtain

(3.2.17) $$-\frac{\partial^2 \log \phi}{\partial \underline{\xi}_t \partial \underline{\xi}_w} = \sigma^{tw}A_w'A_t \ ,$$

so that

(3.2.18) $$-\frac{\partial^2 \log \phi}{\partial \underline{\xi}^2} = ((\sigma^{rs}A_r'A_s)) \ .$$

Thus the <u>information matrix for $\underline{\xi}$ in the MDM model</u> is given by

(3.2.19) $$B_{n,\underline{\xi}}(\text{MDM}) = \frac{1}{n}((\sigma^{rs}A_r'A_s)) \ .$$

From (3.2.14) and (3.2.19) the information matrix for the MGLM model can be easily inferred. The result is given in the following theorem:

<u>Theorem 3.2.1</u>: For the MGLM model the information matrix for $\underline{\xi}$ as defined by (3.1.4) is given by the

$(M \times M)$ matrix:

(3.2.20) $$B_{n,\underline{\xi}} = \frac{1}{n} D'\Omega^{-1}D = \frac{1}{n}((\sum_{i=1}^{u_{rs}} \sigma_{t_{rsi}}^{rs} A'_{t_{rsi}r}A_{t_{rsi}s}))$$

where

D and Ω are given by the vector-version of the MGLM model,

u_{rs} is the number of sets S_j in which V_r and V_s are observed together,

t_{rsi} is the index of the i-th set $S_{t_{rsi}}$ in which both V_r and

V_s are observed together where

$1 \le t_{rs1} < t_{rs2} < \ldots < t_{rsu_{rs}} \le u$,

$A_{t_{rsi}r}$ is the design matrix for V_r in set $S_{t_{rsi}}$,

and $\sigma^{rs}_{t_{rsi}}$ is the element of $[B'_{t_{rsi}} \Sigma B_{t_{rsi}}]^{-1}$ which corresponds to

the variate pair (V_r, V_s) .

It is important to note that $\sigma^{rs}_{t_{rsi}} = \sigma^{sr}_{t_{rsi}}$ and that if variates V_r and V_s, $r \neq s$ are not measured together in the entire experiment, then we must replace

$$\sum_{i=1}^{u_{rs}} \sigma^{rs}_{t_{rsi}} A'_{t_{rsi}r} A_{t_{rsi}s}$$

by $O(m_r \times m_s)$ in (3.2.20).

Corollary 3.2.2: For the MGLM model, the limiting variance matrix of any BAN estimator of $\underline{\xi}$ is given by

$$(3.2.21) \qquad \frac{1}{n} B^{-1}_{n,\underline{\xi}} = (D'\Omega^{-1}D)^{-1} = \Big(\Big(\sum_{i=1}^{u_{rs}} \sigma^{rs}_{t_{rsi}} A'_{t_{rsi}r} A_{t_{rsi}s} \Big)\Big)^{-1}$$

provided D is of full rank.

If we wish to consider instead of $\underline{\xi}$ any estimable set $H'\underline{\xi}$, it is not necessary to assume that D is of full rank to obtain a unique limiting variance matrix. We first need to extend the definition of a BAN estimator to functions of unknown parameters:

<u>Definition 3.2.1</u>: Suppose $\underline{g}(\underline{\theta}) = (g_1(\underline{\theta}), \ldots, g_w(\underline{\theta}))'$ is a $(w \times 1)$ vector function of an unknown parameter $\underline{\theta}(u \times 1) = (\theta_1, \ldots, \theta_u)'$. Let $B_{n,\underline{\theta}_o}$ be the information matrix for $\underline{\theta}$ given by (3.1.2) and define Δ $(u \times w)$ by

$$\Delta = \frac{\partial\, \underline{g}\, (\underline{\theta})}{\partial \underline{\theta}} = ((\frac{\partial\, g_j\, (\underline{\theta})}{\partial \theta_i})) .$$

Then an estimator $\underline{g}(\underline{\tilde{\theta}}_n)$ based on a sample of n observations is said to be a BAN estimator for $\underline{g}(\underline{\theta})$ provided

$$(3.2.22) \qquad \mathcal{L}(\sqrt{n}\, C_n^{-\frac{1}{2}}[\underline{g}(\underline{\tilde{\theta}}_n) - \underline{g}(\underline{\theta}_o)]) \to \mathcal{N}_w(\underline{0}_w, I_w)$$

where $\quad \underline{\theta}_o$ is the true value of $\underline{\theta}$,

$$\Delta_o = \Delta \Big|_{\underline{\theta} = \underline{\theta}_o} ,$$

and $\quad C_n = \Delta_o' B^-_{n,\underline{\theta}_o} \Delta_o$ is symmetric, positive definite and of full rank w.

Rao [41], page 265, has shown that if $\underline{f}(\underline{\theta}_n^*)$ is any unbiased estimator of $\underline{g}(\underline{\theta})$, then $\mathrm{Var}(\underline{f}(\underline{\theta}_n^*)) - \frac{1}{n} C_n$ is non-negative definite. It is in this sense, then, that we consider a BAN estimator $\underline{g}(\underline{\tilde{\theta}}_n)$ to have minimum variance.

Now consider for the MGLM model an estimable linear set $H'\underline{\xi}$ where $H'(w \times M)$ is of full rank w. Clearly, if we let $\underline{g}(\underline{\xi}) = H'\underline{\xi}$, then $\Delta = \frac{\partial \underline{g}}{\partial \underline{\xi}} = H$. If we let $B^-_{n,\underline{\xi}}$ denote any generalized inverse of $B_{n,\underline{\xi}}$ (where we assume $\underline{\xi}$ to be the true parameter), then using

Theorem A.1.1. (iii) of the Appendix we have

$$C_n = \Delta' B^-_{n,\underline{\xi}} \Delta = H' B^-_{n,\underline{\xi}} H \text{ is unique, symmetric, positive}$$

definite and of full rank. Thus we have the following theorem:

Theorem 3.2.3: For the MGLM model, the asymptotic variance matrix of any BAN estimator of an estimable linear set $H'\underline{\xi}$,

where H $(M \times w)$ is of full rank w, is given by

$$(3.2.23) \qquad \frac{1}{n} C_n = H'(D'\Omega^{-1}D)^- H.$$

Note that $\frac{1}{n} C_n$ is the same as the variance matrix of the unique BLUE set (2.2.1) for $H'\underline{\xi}$ when Ω is known.

We turn now to the GGCM model defined by (1.5.2), and we derive the information matrix for $\underline{\xi}$ and the asymptotic variance matrix for $H'\underline{\xi}$ where $H(mp \times w)$. The likelihood function for the GGCM model is given by

$$(3.2.24) \qquad \log \phi_n = -\sum_{j=1}^{u} \frac{n_j q_j}{2} \log 2\pi - \sum_{j=1}^{u} \frac{n_j}{2} \log |B_j' \Sigma B_j|$$

$$- \frac{1}{2} \sum_{j=1}^{u} \operatorname{tr}[B_j' \Sigma B_j]^{-1} [Y_j - A\xi P B_j]'[Y_j - A\xi P B_j] .$$

Using (3.2.5) on (3.2.24) we obtain

$$(3.2.25) \qquad \frac{\partial \log \phi_n}{\partial \xi} = \sum_{j=1}^{u} A_j'(Y - A\xi P B_j)(B_j' \Sigma B_j)^{-1} B_j' P' .$$

Then using (3.2.6) on (3.2.25) we obtain

$$(3.2.26) \qquad \frac{\partial^2 \log \phi_n}{\partial \xi^2} = -\sum_{j=1}^{u} P B_j (B_j' \Sigma B_j)^{-1} B_j' P' \otimes A_j' A_j$$

and since $\frac{\partial^2 \log \phi_n}{\partial \xi^2} = \frac{\partial^2 \log \phi_n}{\partial \underline{\xi}^2}$ from Theorem A.2.1. (x) of the Appendix we thus obtain the following theorem:

Theorem 3.2.4: For the GGCM model, the information matrix for $\underline{\xi}$ is given by the (mp × mp) matrix:

$$(3.2.27) \qquad B_{n,\xi} = \frac{1}{n} \sum_{j=1}^{u} [PB_j(B_j'\Sigma B_j)^{-1} B_j'P' \otimes A_j'A_j] \,.$$

Using Definition 3.2.1 for the BAN estimator of any transformation of $\underline{\xi}$ we then obtain the following corollary:

Corollary 3.2.5: For the GGCM model, the limiting variance matrix of any BAN estimator of an estimable linear set $H'\underline{\xi}$, where H(mp × w) is of full rank w, is given by

$$(3.2.28) \qquad \frac{1}{n} C_n = H'[\sum_{j=1}^{u} PB_j(B_j'\Sigma B_j)^{-1} B_j'P' \otimes A_j'A_j]^{-}H \,.$$

3.3 Estimation for the GIM Model--A Review of the Literature with Some Extensions

In this section we review some methods proposed in the literature for estimating the parameters ξ and Σ of the GIM model defined in (1.4.1.4) of Chapter One. For convenience we repeat the definition of the model:

$$(3.3.1) \qquad E(Y_j) = A_j\xi B_j$$

$$Var(Y_j) = I_{n_j} \otimes B_j'\Sigma B_j \ , \ j=1,\ldots,u \ ,$$

where the rows of $Y_j(n_j \times q_j)$ are normally distributed,

$\xi(m \times p) = [\underline{\xi}_1 \ \underline{\xi}_2 \ \ldots \ \underline{\xi}_p]$ is unknown,

$\Sigma(p \times p) = ((\sigma_{rs}))$ is unknown and positive definite,

$A_j(n_j \times m)$ is known,

$B_j(p \times q_j)$ is the incidence matrix of 0's and 1's for the j-th set (S_j) of experimental units,

$$q_j \le p, \; j=1,\ldots,u \, ,$$

and
$$n = \sum_{j=1}^{u} n_j \, .$$

It is assumed throughout the discussion that $R(A_j) = m$, i.e., A_j is of full rank for each j, so that ξ is estimable.

As described in Chapter One, the GIM model is the general multivariate linear model for missing observations. The problem of missing observations in multivariate experiments has been discussed often in the literature. In a recent paper by Afifi and Elashoff [3], the literature is quite thoroughly reviewed and a complete bibliography through 1966 is given. Subsequent papers have been written by Afifi and Elashoff [4], [5], [6] and Hocking and Smith [28]. Most of the earlier papers (e.g., Wilks [69], Yates [71], Bartlett [11], Tocher [62], Buck [16] and Dear [19]) and the recent papers of Afifi and Elashoff consider randomly occurring missing observations. Except for Wilks, all these authors considered the regression of a preferred response variate on the rest of the variates and did not assume normality on the data. In some early papers by Lord [34], Edgett [21], Nicholson [36], Anderson [7] and more recently by Bhargava [13], Trawinski and Bargmann [63], Hocking and Smith [28], and Srivastava [58], [60] attention has focused on the problem of missing observations which follow some pattern. In all of these papers except those of Srivastava, who considers BLUE estimation only, and Trawinski and Bargmann, whose paper assumes $A_i = A$ for all i, the very special case of $A_i = \underline{1}_{n_i}$ and $\xi = \underline{\xi}'(1 \times p)$ is considered. We will discuss below the papers by Anderson [7] and Hocking and Smith [28], whose approaches were general enough to include many models discussed in other papers (e.g., Lord, Edgett, Nicholson). Anderson's method will be extended to the HIM model to obtain maximum likelihood estimators of ξ and Σ. Hocking and Smith's method can be used for the general missing observations model of (3.3.1), although it seems to be computationally unfeasible except for simply structured situations.

3.3.1 Anderson's Method

Anderson [7] has described a method for obtaining the MLE's of $\xi = \underline{\xi}'(p \times 1)$ and Σ when $A_j = \underline{1}_{n_j}$. The usual method for obtaining MLE's of given parameters involves solving the likelihood equations for the model. Using the likelihood function given in (3.2.11) and the rules of matrix differentiation as given in the Appendix, we obtain the following likelihood equations from

$\frac{\partial \log \phi}{\partial \xi} = 0$ and $\frac{\partial \log \phi}{\partial \Sigma} = 0$, respectively, for the GIM model:

$$\sum_{i=1}^{u} A_i'(Y_i - A_i\xi B_i)(B_i'\Sigma B_i)^{-1} B_i' = 0$$

$$\sum_{i=1}^{u} [n_i B_i(B_i'\Sigma B_i)^{-1}B_i' - B_i(B_i'\Sigma B_i)^{-1} (Y_i-A_i\xi B_i)'(Y_i-A_i\xi B_i) \cdot (B_i'\Sigma B_i)^{-1} B_i'] = 0 \,.$$

It is clear that these equations cannot be solved directly for the MLE's $\hat{\xi}$ and $\hat{\Sigma}$ and that a solution by iteration techniques using a computer, similar to that used for estimating factor loadings in the factor analysis model, would require a costly and laborious procedure. To avoid this difficulty, Anderson has given an alternative method for obtaining solutions to the maximum likelihood equations.

To use Anderson's method it is first necessary to be able to partition the p response variates $V_1, \ldots, V_p$ into, say, $k(\le p)$ mutually exclusive groups $W_1, W_2, \ldots, W_k$ so that the likelihood function ϕ for the entire experiment can be written as

$$\phi = \prod_{i=1}^{k} \phi_i(\Theta_i) ,$$

where $\phi_i(\Theta_i)$ is the likelihood function for the data observed on the variates in W_i conditioned on the data observed on the variates in $W_1, W_2, \ldots, W_{i-1}$ together, and Θ_i denotes the unknown (conditional) parameters of the i-th conditional likelihood. If the parameters $\Theta_1, \ldots, \Theta_k$ form a mutually independent set,[1] then ϕ can be maximized by finding $\hat{\Theta}_i$ which maximize $\phi_i(\Theta_i)$ separately for each i. The maximum likelihood solutions $\hat{\xi} = [\hat{\underline{\xi}}_1, \ldots, \hat{\underline{\xi}}_p]$ and $\hat{\Sigma}$ are then found, usually recursively, using the identities which relate $\Theta_1, \ldots, \Theta_k$ to ξ and Σ.

Anderson's method, unfortunately, does not apply to the GIM model in general. It works only for special cases in which the B_i are fairly simply structured. Such models are characterized by not having any pair of variates for which the experimental units common to both is a non-empty proper subset of each of the sets of experimental units of

---

[1]This will be the case provided none of the original parameters $\underline{\xi}_1, \ldots, \underline{\xi}_p$ appear in $\Theta_1, \ldots, \Theta_k$ both conditionally and unconditionally, and provided the total number of parameters in $\Theta_1, \ldots, \Theta_k$ equals the total number in ξ and Σ.

the variates separately. For example, we would have an unacceptable model if V_2 was the only variate observed in the j-th set and V_3 was the only variate observed in the k-th set, and yet both V_2 and V_3 were observed together in the ℓ-th set. If however, V_2 and V_3 were observed separately in different sets but were not observed together in any set, then the model would still be acceptable. Some examples of acceptable structures include:

a) The <u>HIM model</u>, where

$u=p,\ B_1 = I_p,\ B_i' = [I_{p-i+1} \mid 0_{p-i+1,i-1}]\ ,\ i=2,\ldots,p,$

$k=p,\ W_j = \{V_j\},\ j=1,\ldots,p.$

b) <u>Edgett's model</u>, where

$p=3,\ u=2,\ q_1 = 3,\ q_2 = 2,\ B_1 = I_3,\ B_2' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$

$k=2,\ W_1 = \{V_1, V_2\},\ W_2 = \{V_3\}.$

c) <u>Lord's model</u>, where

$p=3,\ u=2,\ q_1 = 2,\ q_2 = 2,\ B_1' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},\ B_2' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$

$k=3,\ W_1 = \{V_1\},\ W_2 = \{V_2\},\ W_3 = \{V_3\}.$

d) $p=5,\ u=3,\ q_1 = 3,\ q_2 = 3,\ q_3 = 2,$

$B_1' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},\ B_2' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},\ B_3' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix},$

$k=5,\ W_1 = \{V_1\},\ W_2 = \{V_2\},\ W_3 = \{V_4\},\ W_4 = \{V_5\},\ W_5 = \{V_3\}.$

Some unacceptable structures include:

e) $p=3,\ u=3,\ q_1 = 3,\ q_2 = 2,\ q_3 = 2,$

$$B_1 = I_3,\ B_2' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},\ B_3' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

f) $p=2,\ u=3,\ q_1 = 2,\ q_2 = 1,\ q_3 = 1,$

$$B_1 = I_2,\ B_2' = \begin{bmatrix} 1 & 0 \end{bmatrix},\ B_3' = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$

We can present these examples diagramatically as follows:[2]

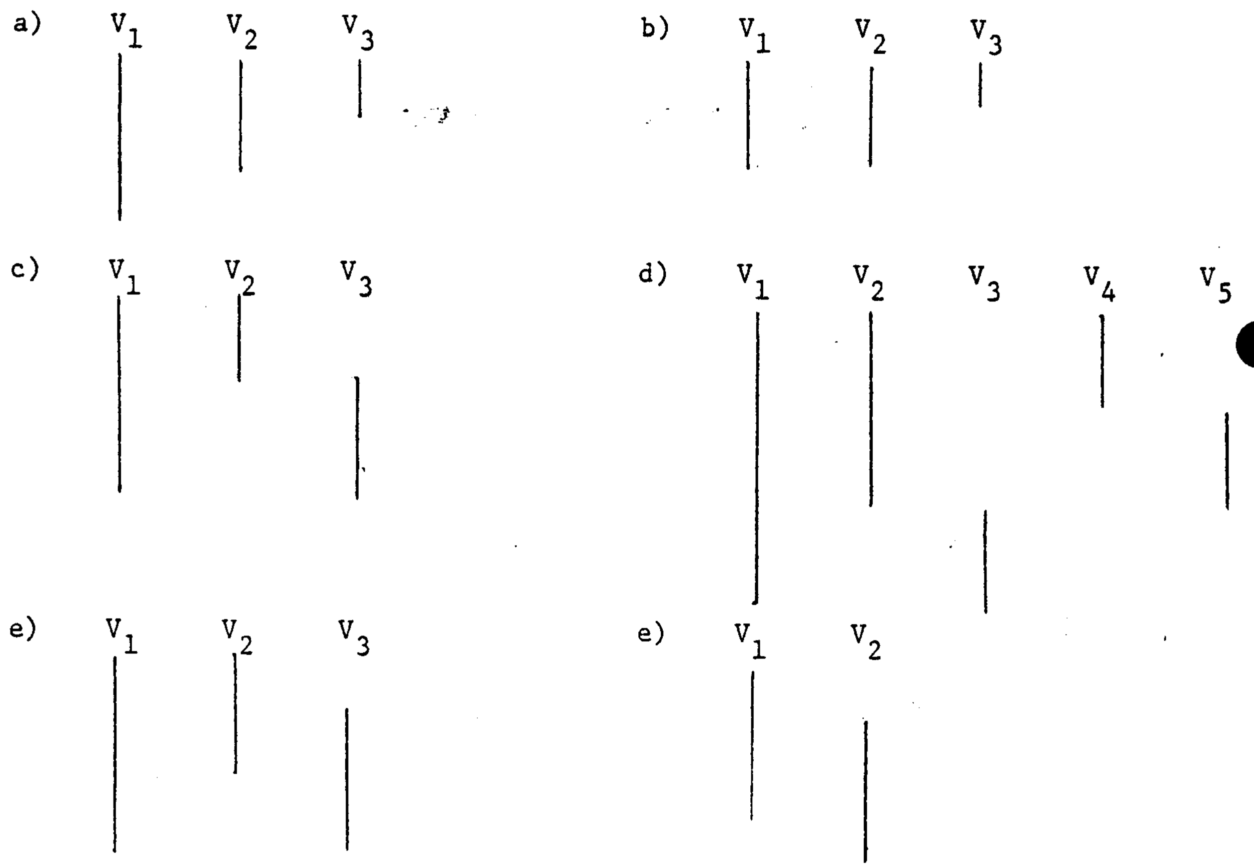

TABLE 3.1
EXAMPLES OF GIM SITUATIONS

---

[2] A vertical line in the diagrams represents the experimental units observed on the response variate above the line. Any group of response variates is observed together on the <u>same</u> experimental unit if a horizontal line can be drawn which intersects each of the vertical lines corresponding to each response variate in the group.

In his paper, Anderson did not give explicit results except for the HIM model when p=2 and for Edgett's and Lord's model . Bhargava [13], however, did obtain explicit, though recursive, solutions for the HIM model with general p and $A_i = \underline{1}_{n_i}$, $\xi = \underline{\xi}'(1 \times p)$. We now extend Bhargava's results to the HIM model:

3.3.2 Derivation of MLE's for the HIM Model

Noting that $B_j(p \times j) = [I_j \mid 0_{j,p-j}]'$, j=1,...,p in the HIM model, it is easy to see that $N_r > N_s$ if r < s, where

$N_s = \sum_{j=s}^{p} n_j$. Letting $\underline{x}_s(N_s \times 1)$ and $D_s(N_s \times m)$ denote the observation vector and design matrix, respectively, for response V_s alone, s=1,...,p, we have

$$\underline{x}_s(N_s \times 1) = \begin{bmatrix} \underline{y}_{ss} \\ \underline{y}_{s+1,s} \\ \vdots \\ \underline{y}_{ps} \end{bmatrix} \quad \text{and} \quad D_s(N_s \times m) = \begin{bmatrix} A_s \\ A_{s+1} \\ \vdots \\ A_p \end{bmatrix} ,$$

where

$$Y_j(n_j \times j) = [\underline{y}_{j1}\ \underline{y}_{j2} \cdots \underline{y}_{jj}] \ , \ j=1,\dots,u$$

and

$A_j(n_j \times m)$, j=1,...,u is the design matrix for the j-th set of experimental units.

We also define the following additional notation:

For s=1,...,p, let

$$\Sigma_s(s \times s) = B_s'\Sigma B_s \ ,$$

$$\xi_{(s)}(m \times s) = \xi B_s \ ,$$

and for $s=2,\ldots,p$, let

$$\underline{\sigma}_s((s-1)\times 1) = (\sigma_{s1}, \sigma_{s2}, \ldots, \sigma_{s,s-1})' ,$$

$$\delta_s((s-1)\times 1) = \Sigma_{s-1}^{-1}\underline{\sigma}_s = (\delta_{s\cdot 1}, \ldots, \delta_{s\cdot s-1})' ,$$

$$\sigma_{ss\cdot 12\ldots(s-1)} = \sigma_{ss} - \underline{\sigma}_s'\Sigma_{s-1}^{-1}\underline{\sigma}_s ,$$

$$\underline{\alpha}_s(m\times 1) = \underline{\xi}_s - \xi_{(s-1)}\underline{\delta}_s ,$$

and

$$X_s(N_s\times(s-1)) = \begin{bmatrix} \underline{y}_{s1} & \underline{y}_{s2} & \cdots & \underline{y}_{s,s-1} \\ \vdots & \vdots & & \vdots \\ \underline{y}_{p1} & \underline{y}_{p2} & \cdots & \underline{y}_{p,s-1} \end{bmatrix} .$$

Lemma 3.3.2.1: For $s=2,\ldots,p$ we have

$$\underline{x}_s \mid \underline{x}_1, \ldots, \underline{x}_{s-1} \sim \mathcal{N}_{N_s}(D_s\underline{\alpha}_s + X_s\underline{\delta}_s, \sigma_{ss\cdot 12\ldots(s-1)}I_{N_s}) .$$

Proof: Clearly, the distribution of $\underline{x}_s \mid \underline{x}_1, \ldots, \underline{x}_{s-1}$ is the same as that of $\underline{x}_s \mid X_s$ since the first $N_r - N_s$ rows of $\underline{x}_r$, $r < s$, are independent of $\underline{x}_s$ by assumption. Now

$Z_s = [X_s \mid \underline{x}_s]$ is a $(N_s \times s)$ random matrix whose rows are each s-variate normal and

$$E[Z_s] = [EX_s \mid E\underline{x}_s] = [D_s\xi_{(s-1)} \mid D_s\underline{\xi}_s] = D_s\xi_{(s)}$$

$$\mathrm{Var}[Z_s] = I_{N_s} \otimes \Sigma_s .$$

Since

$$\Sigma_s = \begin{bmatrix} \Sigma_{s-1} & \underline{\sigma}_s \\ \underline{\sigma}_s' & \sigma_{ss} \end{bmatrix},$$

we can use for each row of Z_s a well known theorem (e.g., Rao [41], page 441) which gives the conditional distribution of any subset of components of a multinormal vector. Noting that the rows of Z_s are independent, we then obtain

$$\underline{x}_s \mid X_s \sim \mathcal{N}_{N_s}\left(D_s\underline{\xi}_s + (X_s - D_s\xi_{(s-1)})\Sigma_{s-1}^{-1}\underline{\sigma}_s,\ (\sigma_{ss} - \underline{\sigma}_s'\Sigma_{s-1}^{-1}\underline{\sigma}_s)I_{N_s}\right).$$

Since $D_s\underline{\xi}_s + (X_s - D_s\xi_{(s-1)})\Sigma_{s-1}^{-1}\underline{\sigma}_s = D_s\underline{\alpha}_s + X_s\underline{\delta}_s$ and

$\sigma_{ss} - \underline{\sigma}_s'\Sigma_{s-1}^{-1}\underline{\sigma}_s = \sigma_{ss\cdot 12\ldots(s-1)}$ by definition, we have the desired result. q.e.d.

We can now write the likelihood function ϕ for our model as

$$\phi = \phi_1(\underline{\xi}_1, \sigma_{11})\phi_2(\underline{\alpha}_2, \underline{\delta}_2, \sigma_{22\cdot 1})\cdots\phi_p(\underline{\alpha}_p, \underline{\delta}_p, \sigma_{pp\cdot 12\ldots(p-1)})$$

where $\phi_1(\underline{\xi}_1, \sigma_{11})$ is the density function for $\mathcal{N}_{N_1}(D_1\underline{\xi}_1, \sigma_{11}I_{N_1})$ and $\phi_s(\underline{\alpha}_s, \underline{\delta}_s, \sigma_{ss\cdot 12\ldots(s-1)}I_{N_s})$ is the density function for $\mathcal{N}_{N_s}(D_s\underline{\alpha}_s + X_s\underline{\delta}_s, \sigma_{ss\cdot 12\ldots(s-1)}I_{N_s})$, $s=2,\ldots,p$. Using Anderson's approach, ϕ will be maximized by maximizing ϕ_s separately for each s. Clearly, the estimators of $\underline{\xi}_1$ and σ_{11} are $\hat{\underline{\xi}}_1 = (D_1'D_1)^{-1}D_1\underline{x}_1$ and $\sigma_{11} = \frac{1}{N_1}\underline{x}_1'(I - D_1(D_1'D_1)^{-1}D_1)\underline{x}_1$, since $\phi_1(\xi_1, \sigma_{11})$ is the likelihood function for univariate ANOVA.[3] Also $\phi_s(\underline{\alpha}_s, \underline{\delta}_s, \sigma_{ss\cdot 12\ldots(s-1)})$ is the likelihood function for univariate covariate ANOVA and thus we can obtain estimators of $\underline{\alpha}_s$, $\underline{\delta}_s$ and $\sigma_{ss\cdot 12\ldots(s-1)}$ as follows:

$$\begin{bmatrix} \hat{\underline{\alpha}}_s \\ \hat{\underline{\delta}}_s \end{bmatrix} = ([D_s \mid X_s]'[D_s \mid X_s])^{-1}[D_s \mid X_s]'\underline{x}_s$$

and $$\sigma_{ss\cdot 12\ldots(s-1)} = \frac{1}{N_s}\underline{x}_s'(I_{N_s} - [D_s|X_s]\{[D_s|X_s]'[D_s|X_s]\}^{-1}[D_s|X_s]')\underline{x}_s .$$

Now

---

[3]Since $\underline{\xi}$ is assumed to be estimable, the design matrices D_s must be of full rank for each s.

$$[D_s \mid X_s]'[D_s \mid X_s] = \begin{bmatrix} D_s'D_s & D_s'X_s \\ X_s'D_s & X_s'X_s \end{bmatrix} .$$

Assuming $[D_s \mid X_s]$ is of full rank, which will be the case with probability 1 when D_s is of full rank, we can write (Rao [41]):

$$(3.3.2.1)\quad \begin{bmatrix} D_s'D_s & D_s'X_s \\ X_s'D_s & X_s'X_s \end{bmatrix}^{-1} = \left[\begin{array}{c|c} (D_s'D_s)^{-1} + FE^{-1}F' & -FE^{-1} \\ \hline -E^{-1}F' & E^{-1} \end{array}\right]$$

where

$$(3.3.2.2)\quad F = (D_s'D_s)^{-1}D_s'X_s$$

$$(3.3.23)\quad E = X_s'X_s - X_s'D_s(D_s'D_s)^{-1}D_s'X_s = X_s'[I - D_s(D_s'D_s)^{-1}D_s']X_s$$

$$\stackrel{\text{def}}{=} N_s\tilde{\Sigma}_{s-1} .$$

Now

$$(D_s'D_s)^{-1} + FE^{-1}F' = (D_s'D_s)^{-1} + \frac{1}{N_s}(D_s'D_s)^{-1}D_s'X_s\tilde{\Sigma}_{s-1}^{-1}X_s'D_s(D_s'D_s)^{-1}$$

$$-E^{-1}F' = -\frac{1}{N_s}\tilde{\Sigma}_{s-1}^{-1}X_s'D_s(D_s'D_s)^{-1}$$

and

$$E^{-1} = \frac{1}{N_s}\tilde{\Sigma}_{s-1}^{-1} .$$

Thus

$$\underline{\hat{\alpha}}_s = (D_s'D_s)^{-1}D_s'\underline{x}_s + \frac{1}{N_s}(D_s'D_s)^{-1}D_s'X_s\tilde{\Sigma}_{s-1}^{-1}X_s'D_s(D_s'D_s)^{-1}D_s'\underline{x}_s$$

$$-\frac{1}{N_s}(D_s'D_s)^{-1}D_s'X_s\tilde{\Sigma}_{s-1}^{-1}X_s'\underline{x}_s$$

$$= (D_s'D_s)^{-1}D_s'\underline{x}_s - \frac{1}{N_s}(D_s'D_s)^{-1}D_s'X_s\tilde{\Sigma}_{s-1}^{-1}X_s'[I - D_s(D_s'D_s)^{-1}D_s']\underline{x}_s$$

and letting

$$(3.3.2.4) \qquad \underset{\sim}{\tilde{\sigma}}_s \overset{\text{def}}{=} \frac{1}{N_s} X_s'[I - D_s(D_s'D_s)^{-1} D_s']\underset{\sim}{x}_s \ ,$$

we can write $\hat{\underset{\sim}{\alpha}}_s = (D_s'D_s)^{-1} D_s'(\underset{\sim}{x}_s - X_s\tilde{\Sigma}_{s-1}^{-1}\underset{\sim}{\tilde{\sigma}}_s)$.

Also,

$$\hat{\underset{\sim}{\delta}}_s = -\frac{1}{N_s}\tilde{\Sigma}_{s-1}^{-1}X_s'D_s(D_s'D_s)^{-1}D_s'\underset{\sim}{x}_s + \frac{1}{N_s}\tilde{\Sigma}_{s-1}^{-1}X_s'\underset{\sim}{x}_s$$

$$= \frac{1}{N_s}\tilde{\Sigma}_{s-1}^{-1}X_s'[I - D_s(D_s'D_s)^{-1} D_s']\underset{\sim}{x}_s$$

$$= \tilde{\Sigma}_{s-1}^{-1}\underset{\sim}{\tilde{\sigma}}_s \ .$$

Now using the identity relationship $\underset{\sim}{\alpha}_s = \underset{\sim}{\xi}_s - \xi_{(s-1)}\underset{\sim}{\delta}_s$ we obtain a recursive formula for $\hat{\underset{\sim}{\xi}}_s$, the MLE of $\underset{\sim}{\xi}_s$, given by

$$\hat{\underset{\sim}{\xi}}_s = \hat{\underset{\sim}{\alpha}}_s + \hat{\xi}_{(s-1)}\hat{\underset{\sim}{\delta}}_s \quad \text{where } \hat{\xi}_{(s-1)} = [\hat{\underset{\sim}{\xi}}_1 \ \cdots \ \hat{\underset{\sim}{\xi}}_{s-1}] \ .$$

Thus to obtain $\hat{\underset{\sim}{\xi}}_s$, $s=2,\ldots,p$ we calculate $\hat{\underset{\sim}{\alpha}}_s$ and $\hat{\underset{\sim}{\delta}}_s$ and then use the previous values obtained for $\hat{\underset{\sim}{\xi}}_1, \ldots, \hat{\underset{\sim}{\xi}}_{s-1}$ in the recursive formula.

Using (3.3.2.1) through (3.3.2.4) we also have that

$$\hat{\sigma}_{ss\cdot 12\ldots(s-1)} = \frac{1}{N_s}\underset{\sim}{x}_s'[I-D_s(D_s'D_s)^{-1}D_s' + \frac{1}{N_s}D_s(D_s'D_s)^{-1}D_s'X_s\tilde{\Sigma}_{s-1}^{-1}X_s'\{I-D_s(D_s'D_s)^{-1}D_s'\}$$

$$- \frac{1}{N_s}X_s\tilde{\Sigma}_{s-1}^{-1}X_s'\{I-D_s(D_s'D_s)^{-1}D_s'\}]\underset{\sim}{x}_s$$

$$= \frac{1}{N_s}\underset{\sim}{x}_s'[I - D_s(D_s'D_s)^{-1}D_s']\underset{\sim}{x}_s$$

$$- \left(\frac{1}{N_s}\right)^2 \underset{\sim}{x}_s'[I-D_s(D_s'D_s)^{-1}D_s']X_s\tilde{\Sigma}_{s-1}^{-1}X_s'[I-D_s(D_s'D_s)^{-1}D_s']\underset{\sim}{x}_s$$

$$= \frac{1}{N_s}\underset{\sim}{x}_s'[I-D_s(D_s'D_s)^{-1}D_s']\underset{\sim}{x}_s - \underset{\sim}{\tilde{\sigma}}_s'\tilde{\Sigma}_{s-1}^{-1}\underset{\sim}{\tilde{\sigma}}_s$$

$$= \tilde{\sigma}_{ss} - \tilde{\underline{\sigma}}_s'\tilde{\Sigma}_{s-1}^{-1}\tilde{\underline{\sigma}}_s, \text{ where } \tilde{\sigma}_{ss} = \frac{1}{N_s}\underline{x}_s'[I - D_s(D_s'D_s)^{-1}D_s']\underline{x}_s .$$

We then obtain MLE's $\hat{\sigma}_{rs}$, $r < s$ and $\hat{\sigma}_{ss}$, $s=2,\dots,p$ recursively,

where $\hat{\sigma}_{rs}$, $r < s$ is the r-th element of $\hat{\underline{\sigma}}_s = \hat{\Sigma}_{s-1}\hat{\underline{\delta}}_s$

and $\hat{\sigma}_{ss} = \hat{\sigma}_{ss\cdot 12\dots(s-1)} - \hat{\underline{\sigma}}_s'\hat{\Sigma}_{s-1}^{-1}\hat{\underline{\sigma}}_s$.

and where $\hat{\Sigma}_{s-1}$ is the previously calculated maximum likelihood estimate of Σ_{s-1} . Here it is necessary to compute $\hat{\underline{\sigma}}_s = (\hat{\sigma}_{s1}, \dots, \hat{\sigma}_{s,s-1})'$ using $\hat{\Sigma}_{s-1}$ and $\hat{\underline{\delta}}_s$ before computing $\hat{\sigma}_{ss}$. Thus we have the following theorem:

Theorem 3.3.2.1: The maximum likelihood estimators of $\xi = [\underline{\xi}_1 \dots \underline{\xi}_p]$ and $\Sigma = ((\sigma_{rs}))$ for the HIM model with full rank design matrices D_s are given by

$$\hat{\underline{\xi}}_1 = (D_1'D_1)^{-1}D_1'\underline{x}_1, \quad \hat{\sigma}_{11} = \frac{1}{N_1}\underline{x}_1'[I - D_1(D_1'D_1)^{-1}D_1']\underline{x}_1 ,$$

$$\hat{\underline{\xi}}_s = \hat{\underline{\alpha}}_s + \hat{\xi}_{(s-1)}\hat{\underline{\delta}}_s \quad \text{for } s=2,\dots,p,$$

$$\hat{\sigma}_{ss} = \hat{\sigma}_{ss\cdot 12\dots(s-1)} + \hat{\underline{\sigma}}_s'\hat{\Sigma}_{s-1}^{-1}\hat{\underline{\sigma}}_s , \quad \text{for } s=2,\dots,p,$$

and $\hat{\sigma}_{rs}$, $r < s$ = the r-th element of $\hat{\underline{\sigma}}_s = \hat{\Sigma}_{s-1}\hat{\underline{\delta}}_s$

where $\hat{\xi}_{(s-1)} = [\hat{\underline{\xi}}_1 \dots \hat{\underline{\xi}}_{s-1}]$,

$\hat{\underline{\sigma}}_s = \hat{\Sigma}_{s-1}\hat{\underline{\delta}}_s$ and $\hat{\Sigma}_{s-1}$ are recursively calculated

maximum likelihood estimators of $\xi_{(s-1)}$, $\Sigma_{s-1}\underline{\delta}_s$ and Σ_{s-1}, respectively,

$$\hat{\underline{\alpha}}_s = (D_s'D_s)^{-1}D_s'(\underline{x}_s - X_s\tilde{\Sigma}_{s-1}^{-1}\tilde{\underline{\sigma}}_s) ,$$

$$\hat{\underline{\delta}}_s = \tilde{\Sigma}_{s-1}^{-1}\tilde{\underline{\sigma}}_s ,$$

$$X_s = \begin{bmatrix} \underset{\sim}{y}_{s1} & \underset{\sim}{y}_{s2} & \cdots & \underset{\sim}{y}_{s,s-1} \\ \vdots & \vdots & & \vdots \\ \underset{\sim}{y}_{p1} & \underset{\sim}{y}_{p2} & \cdots & \underset{\sim}{y}_{p,s-1} \end{bmatrix},$$

$$\tilde{\Sigma}_{s-1} = \frac{1}{N_s} X_s'[I - D_s(D_s'D_s)^{-1} D_s']X_s ,$$

$$\underset{\sim}{\tilde{\sigma}}_s = \frac{1}{N_s} X_s'[I - D_s(D_s'D_s)^{-1} D_s']\underset{\sim}{x}_s ,$$

$$\hat{\sigma}_{ss\cdot 12\ldots(s-1)} = \tilde{\sigma}_{ss} - \underset{\sim}{\tilde{\sigma}}_s'\tilde{\Sigma}_{s-1}^{-1}\underset{\sim}{\tilde{\sigma}}_s ,$$

$$\tilde{\sigma}_{ss} = \frac{1}{N_s} \underset{\sim}{x}_s'[I - D_s(D_s'D_s)^{-1} D_s]\underset{\sim}{x}_s , \text{ for } s=2,\ldots,p$$

and

$$N_s = \sum_{i=s}^{p} n_i ,$$

$$\underset{\sim}{x}_s = \begin{bmatrix} \underset{\sim}{y}_{ss} \\ \vdots \\ \underset{\sim}{y}_{ps} \end{bmatrix}, \quad D_s = \begin{bmatrix} A_s \\ A_{s+1} \\ \vdots \\ A_p \end{bmatrix}, \text{ for } s=1,\ldots,p.$$

3.3.3 Hocking and Smith's Approach

A second approach for obtaining BAN estimators of ξ and Σ for the GIM model was developed by Hocking and Smith [28]. Their general methodology applies to any GIM situation for which $B_j = I_p$ for some j, although it may be modified in case no set of experimental units is complete (i.e., without missing observations). For the special case of the HIM model Hocking and Smith have shown that their approach yields the maximum likelihood solutions obtained by Anderson's method. They have not established this for the general GIM model, however, but they have shown that their estimators of ξ and Σ are consistent and asymptotically efficient;i.e., they are BAN if we assume normality.

The estimation technique can be briefly summarized as follows:

(i) Initial estimates of the parameters are obtained from that group of observations with no missing observations (e.g. group j=1).

(ii) These initial estimators are modified by adjoining optimally the information in all the remaining groups in a sequential manner by the addition of linear combinations of zero expectations. By "optimally" it is meant that at each stage in the estimation procedure, the newly defined estimators will be consistent and asymptotically efficient when only the partial data used in making the new estimate is considered.

A severe drawback to Hocking and Smith's method,at least from a practical point of view, is the extremely cumbersome notation which would be required to write out for the general model the formulae for calculating the estimators at each stage. Thus a general computer program for obtaining the estimators will not only be costly to write, but might also be difficult for the applied statistician to use. In any case, such a computer program has not been developed as yet, and use of Hocking and Smith's results at the present must be restricted to simply structured models and to small u (e.g., u=2 or 3).

Because of the notational difficulties for the general model, Hocking and Smith were obliged in presenting their method to consider the Nested Multivariate (NM) model defined by

(3.3.3.1) $E(Y_j) = 0,\ \mathrm{Var}\,(Y_j) = I_{n_j} \otimes B_j'\Sigma B_j\ ,\ j=1,\ldots,u$

where $B_1 = I_p,\ \ B_j = [I_{q_j} \ \vdots\ 0_{q_j,p-q_j}]'\ ,\ j=2,\ldots,u$

and $\quad q_j > q_k$ if $j < k$.

(Note that when $q_j = p-j+1$ we have the HIM model variance structure.) We will not describe the estimator $\hat{\Sigma}$ obtained for Σ here, since the reader can refer to Hocking and Smith's paper. We will, however, describe how their method can be extended for the nested situation when the expectation structure is given by

(3.3.3.2) $\quad E(Y_j) = A_j \xi B_j$, $j=1,\ldots,u$

where A_j is of full rank for each j and B_j is defined by (3.3.1). We first write $\xi = [\underline{\xi}_1 \cdots \underline{\xi}_p]$ as $\underline{\xi} = (\underline{\xi}_1' \; \underline{\xi}_2' \; \ldots \; \underline{\xi}_p')'$. The <u>initial estimator</u> $\underline{\tilde{\xi}}\,(1)$ is then obtained by rolling out the columns of $\xi^* = (A_1'A_1)^{-1}A_1'Y_1$ into one long $(mp \times 1)$ vector. Additional data are then adjoined in stages so that the <u>j-stage estimator</u> $\underline{\tilde{\xi}}\,(j)$ is given by:

(3.3.3.3) $\quad \underline{\tilde{\xi}}\,(j) = \underline{\tilde{\xi}}\,(j-1) + \tilde{\Psi}_j' \, \underline{\tilde{\mu}}_j \;$, $j=2,\ldots,u$

where
$$\underline{\tilde{\mu}}_j(mq_j \times 1) = \begin{bmatrix} \underline{\tilde{\xi}}_1\,(j-1) - {}_j\underline{\xi}_1^* \\ \underline{\tilde{\xi}}_2\,(j-1) - {}_j\underline{\xi}_2^* \\ \vdots \\ \underline{\tilde{\xi}}_{q_j}\,(j-1) - {}_j\underline{\xi}_{q_j}^* \end{bmatrix},$$

$$[{}_j\underline{\xi}_1^* \;\; {}_j\underline{\xi}_2^* \;\; \cdots \;\; {}_j\underline{\xi}_{q_j}^*] = (A_j'A_j)^{-1}A_j'Y_j \; ,$$

$$\underline{\tilde{\xi}}\,(j-1) = \begin{bmatrix} \underline{\tilde{\xi}}_1\,(j-1) \\ \vdots \\ \underline{\tilde{\xi}}_p\,(j-1) \end{bmatrix}$$ is the previously obtained $(j-1)$-stage estimator of $\underline{\xi}$,

$$\tilde{\Psi}_j(mq_j \times mp) = \Psi_j \Big|_{\underline{\xi} = \tilde{\underline{\xi}}\,(j-1),\quad \Sigma = \tilde{\Sigma}\,(j-1)}$$

where $\tilde{\Sigma}$ (j-1) is the (j-1)-th stage estimator of Σ, and

$\Psi_j(mq_j \times mp)$ is a matrix function of the elements of Σ and ξ which is determined so that each of the mp diagonal elements of the (mp × mp) variance matrix

$$V_{(j)} = \mathrm{Var}(\tilde{\underline{\xi}}\,(j-1) + \Psi_j'\tilde{\underline{\mu}}_j)$$

are minimum variances.

Clearly we can write $V_{(j)}$ as

$$V_{(j)} = \mathrm{Var}(\tilde{\underline{\xi}}\,(j-1)) + \Psi_j'[E_j\Psi_j + 2F_j]$$

where $E_j(mq_j \times mq_j) = \mathrm{Var}(\tilde{\underline{\mu}}_j)$ and

$$F_j(mq_j \times mp) = \mathrm{Cov}(\tilde{\underline{\mu}}_j, \tilde{\underline{\xi}}\,(j-1)).$$

Thus the optimum value of Ψ_j is given by the solution to $E_j\Psi_j = -F_j$, i.e.,

$$\Psi_j = -E_j^{-1}F_j \quad (3.3.3.4)$$

Therefore, we can directly calculate Ψ_j and then substitute $\tilde{\underline{\xi}}$ (j-1) and $\tilde{\Sigma}$ (j-1) to obtain $\tilde{\Psi}_j$, provided we can calculate E_j and F_j. It is clear that as the number of sets S_j of experimental units increases, E_j and F_j will be increasingly difficult to compute. Thus even with the nice structure of the nested model, explicit formulae for the estimators of ξ and Σ would be very messy indeed. It is understandable that Hocking and Smith did not present such formulae even for Σ in their paper. So, despite the theoretical soundness of their approach, their method is not presently suitable for practical use in general GIM situations.

A simple example when p=2, $\xi = \underline{\xi}'(1 \times 2)$ and $A_j = \underline{1}_{n_j}$ which illustrates their method is presented in [53].

3.4 BAN Estimation for the MDM Model--A Literature Review

In this section, we discuss several BAN estimators for $\underline{\xi} = (\underline{\xi}_1' \ldots \underline{\xi}_p')'$ which have been proposed for the MDM model. For convenience we repeat definitions (1.4.1.1), (1.4.1.2) and (1.4.1.3) for the model:

(3.4.1) $$E(Y) = [A_1\underline{\xi}_1 \mid A_2\underline{\xi}_2 \quad \ldots \mid A_p\underline{\xi}_p]$$

$$\mathrm{Var}(Y) = I_n \otimes \Sigma$$

(3.4.2) $$E(\underline{y}_s) = A_s\underline{\xi}_s \; , \quad \mathrm{Cov}(\underline{y}_r, \underline{y}_s) = \sigma_{rs}I_n$$

(3.4.3) $$E(\underline{y}) = D\underline{\xi} \; , \; \mathrm{Var}(\underline{y}) = \Sigma \otimes I_n \; ,$$

where $\underline{\xi}_s(m_s \times 1)$ is unknown, s=1,...,p

$\Sigma(p \times p)$ is unknown

$A_s(n \times m_s)$ is known, s=1,...,p

$\underline{y}_s(n \times 1)$ is the s-th column of Y, s=1,...,p

$\underline{y}(np \times 1) = (\underline{y}_1' \; \underline{y}_2' \cdots \underline{y}_p')'$

and $$D(np \times \sum_{s=1}^{p} m_s) = \begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_p \end{bmatrix} .$$

It is assumed throughout this section that A_s is of full rank m_s for each s, so that $\underline{\xi}$ is estimable.

Four BAN estimators for $\underline{\xi}$ have been proposed and discussed in the literature and recently [33] have been compared by Monte Carlo methods. Two of these estimators are due to Zellner [72], [73], who was the first to deal with this subject. They are called <u>Zellner's-two stage Aitken estimator</u> (ZEF) and <u>Zellner's iterative Aitken estimator</u> (IZEF). A third estimator is given by Telser [61] and called <u>Telser's iterative estimator</u> (TIE). The last estimator is the <u>maximum likelihood estimator</u> (MLE) which has been obtained by Kmenta and Gilbert [33] using the assumption that the rows of Y come from a multivariate normal population.

(i) <u>ZEF</u>. To derive the ZEF estimator, one needs to use the vector version of the model given in (3.4.3). We have already established in Theorem 2.3.1 that if Σ were known, the BLUE of $\underline{\xi}$ would be given by

$$\underline{\xi}^* = [D'(\Sigma \otimes I_n)^{-1} D]^{-1} D'(\Sigma \otimes I_n)^{-1} \underline{y} .$$

Since Σ is assumed unknown, Zellner proposed replacing Σ by its consistent estimator $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$ defined by

$$\hat{\sigma}_{rs} = \frac{1}{n-\ell} \underline{y}_r'[I_n - A_r(A_r'A_r)^{-1} A_r'][I_n - A_s(A_s'A_s)^{-1} A_s']\underline{y}_s \tag{3.4.7}$$

where ℓ is the number of independent variables in each regression, which is thus assumed to be the same for each regression. (If this assumption is relaxed, it would be possible to use instead of ℓ either 0 or

$$\ell_{rs} = tr[I_n - A_r(A_r'A_r)^{-} A_r'][I_n - A_s(A_s'A_s)^{-1} A_s'].)$$

The ZEF can therefore be written

$$\hat{\underline{\xi}} = [D'(\hat{\Sigma}^{-1} \otimes I_n)D]^{-1} D'(\hat{\Sigma}^{-1} \otimes I_n)\underline{y}$$

or

$$(3.4.8) \quad \begin{bmatrix} \hat{\underline{\xi}}_1 \\ \hat{\underline{\xi}}_2 \\ \vdots \\ \hat{\underline{\xi}}_p \end{bmatrix} = \begin{bmatrix} \hat{\sigma}^{11}A_1'A_1 & \hat{\sigma}^{12}A_1'A_2 & \cdots & \hat{\sigma}^{1p}A_1'A_p \\ \hat{\sigma}^{12}A_2'A_1 & \hat{\sigma}^{22}A_2'A_2 & \cdots & \hat{\sigma}^{2p}A_2'A_p \\ \vdots & \vdots & & \vdots \\ \hat{\sigma}^{1p}A_p'A_1 & \hat{\sigma}^{2p}A_p'A_2 & \cdots & \hat{\sigma}^{pp}A_p'A_p \end{bmatrix}^{-1} \begin{bmatrix} A_1' \sum_{s=1}^{p} \hat{\sigma}^{1s}\underline{y}_s \\ A_2' \sum_{s=1}^{p} \hat{\sigma}^{2s}\underline{y}_s \\ \vdots \\ A_p' \sum_{s=1}^{p} \hat{\sigma}^{ps}\underline{y}_s \end{bmatrix}$$

where $\hat{\Sigma}^{-1} = ((\hat{\sigma}^{rs}))$.

Kakwani [30] has shown that the ZEF is unbiased even in small samples. Also, the asymptotic variance-covariance matrix of $\hat{\underline{\xi}}$ is given by

$$(D'[\Sigma^{-1} \otimes I_n]D)^{-1} = \begin{bmatrix} \sigma^{11}A_1'A_1 & \cdots & \sigma^{1p}A_1'A_p \\ \vdots & & \vdots \\ \sigma^{1p}A_p'A_1 & \cdots & \sigma^{pp}A_p'A_p \end{bmatrix}^{-1}$$

which is the limiting variance matrix already obtained for the MDM model. Thus $\hat{\underline{\xi}}$ is asymptotically efficient and since $\hat{\Sigma}$ is consistent for Σ, the ZEF is BAN.

(ii) <u>IZEF</u>. Zellner's ZEF can be used for calculating a new set of residuals leading to a new estimate of $\Omega = \Sigma \otimes I_n$, which in turn can be used for obtaining new estimates of $\underline{\xi}$ and so on. Thus if $\hat{\underline{\xi}}$ is the ZEF as given in (3.4.8), we would define $\hat{\Omega}_{(2)}$ as

$$\hat{\Omega}_{(2)} = (\underline{y} - D\hat{\underline{\xi}})(\underline{y} - D\hat{\underline{\xi}})'$$

and use as our first iterative estimate of $\underline{\xi}$,

$$\hat{\underline{\xi}}^{(2)} = [D'\hat{\Omega}_{(2)}^{-1}D]^{-1} D'\hat{\Omega}_{(2)}^{-1}\underline{y} .$$

Continuing in this way we would have for the (k-1)-th iteration:

$$\hat{\Omega}_{(k)} = (\underline{y} - D\hat{\underline{\xi}}^{(k-1)})(\underline{y} - D\hat{\underline{\xi}}^{(k-1)})'$$

and

$$\hat{\underline{\xi}}^{(k)} = [D'\hat{\Omega}^{-1}_{(k)}D]^{-1} D'\hat{\Omega}^{-1}_{(k)}\underline{y} .$$

The process would continue until the estimator of $\underline{\xi}$ converges and the final estimator would be the IZEF.

(iii) TIE. Letting $\underline{u}_s = \underline{y}_s - A_s\underline{\xi}_s$, $s=1,\ldots,p$ Telser showed that

$$(3.4.9) \qquad \underline{u}_s = \sum_{\substack{r=1 \\ r\neq s}}^{p} a_{sr}\underline{u}_r + \underline{v}_s , \quad s=1,\ldots,p$$

where $\quad E\underline{u}_r\underline{v}_s = 0$ for $r\neq s$ and $a_{sr} = -\dfrac{\sigma^{sr}}{\sigma^{ss}}$.

The main idea of the TIE is to substitute for $\underline{u}_s$ in each equation $\underline{y}_s = A_s\underline{\xi}_s + \underline{u}_s$ the right hand side of (3.4.9). Then estimates of $\underline{\xi}_s$ as well as of a_{sr} are obtained beginning with s=1 and proceeding to s=p and then repeating the cycle until the estimates of $\underline{\xi}_s$ converge.

(iv) MLE. Under normality assumptions on the rows of Y, the likelihood function for the MDM is given by

$$\phi = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\{-\tfrac{1}{2} \operatorname{tr}(\Sigma^{-1} \otimes I)(\underline{y} - D\underline{\xi})(\underline{y} - D\underline{\xi})'\} .$$

The maximum likelihood estimators $\hat{\underline{\xi}}$ and $\hat{\Sigma}$ were obtained by Kmenta an and Gilbert [33] using computer iteration techniques to solve the following equations:

$$\hat{\underline{\xi}} = [D'(\hat{\Sigma}^{-1} \otimes I)D]^{-1} D'(\hat{\Sigma}^{-1} \otimes I)\underline{y}$$

$$\hat{\Sigma} = [\underline{u}_1\ \underline{u}_2\ \cdots\ \underline{u}_p]'[\underline{u}_1\ \underline{u}_2\ \cdots\ \underline{u}_p]$$

where $\underset{\sim}{u}_s = \underset{\sim}{y}_s - A_s\hat{\underset{\sim}{\xi}}_s$, s=1,...,p .

3.4.1 Comparison of the Estimators by Monte Carlo Methods

Kmenta and Gilbert's study was organized around four experiments, each containing a number of different models. The experiments were distinguished from one another by the values of the explanatory variables, that is, the independent variables corresponding to the columns of the A_s matrices. The models within each experiment were characterized by having the same explanatory variables and were distinguished by the structures of the variance matrix Σ. Only design matrices of full rank (i.e., regression matrices) were used in the study. For every experiment and every model considered, the computer runs consisted of 100 samples of sizes 10, 20 and 100. The iterative process used in calculating the TIE, IZEF and MLE estimators was stopped when there was no change in the ninth decimal place. All calculations were carried out in double precision.

The most striking result obtained from the study was that the TIE, IZEF and MLE estimates were identical in every sample. Thus these three methods appear to be different ways of obtaining the same solution. In comparing the ZEF with the MLE, the IZEF and the TIE, the results were somewhat mixed. When the correlation between explanatory variables was low and the correlation between disturbances (i.e., $\rho_{rs} = \sigma_{rs}/\sigma_{rr}\sigma_{ss}$, $r \neq s$) was high, the MLE had smaller variances than the ZEF. Also, when the disturbances $\underset{\sim}{u}_s = \underset{\sim}{y}_s - A_s\underset{\sim}{\xi}_s$ were autoregressive and correlated across equations, the MLE had smaller variances regardless of the correlation between explanatory variables. In all other models, regardless of sample size, the ZEF estimator had variances smaller than or equal to the variances of the MLE estimator. Thus, overall, the results favor the ZEF over the other estimators with regard to efficiency in small samples. In addition, since the MLE, TIE and IZEF methods are considerably more time consuming than the procedure of calculating the ZEF, Kmenta and Gilbert recommended the latter. Furthermore, the ZEF is always unbiased, whereas the MLE, TIE and IZEF estimators may not be unbiased.

3.5 A BAN Estimator for the MGLM Model

The method used to obtain the ZEF estimator for the MDM model can be easily extended to the MGLM model. This is described in the following theorem:

Theorem 3.5.1: For the MGLM model, a BAN estimator which is unbiased for any estimable set $\underline{\theta} = H'\underline{\xi}$ is given by

$$\hat{\underline{\theta}}_n = H'\hat{\underline{\xi}} = H'(D'\hat{\Omega}^{-1}D)^- D'\hat{\Omega}^{-1}\underline{x} \tag{3.5.1}$$

where $\underline{x}$, D and Ω are given by (1.4.8) and $\hat{\Omega}$ is obtained from Ω by substituting the elements of $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$ given in Theorem 2.4.1 for the corresponding elements of $\Sigma = ((\sigma_{rs}))$.

Proof: We have already seen that the estimator $\underline{\theta}_n^*$ defined by

$$\underline{\theta}_n^* = H'\underline{\xi}^* = H'(D'\Omega^{-1}D)^- D'\Omega^{-1}\underline{x}$$

is BAN for $\underline{\theta}$. Of course, we cannot use $\underline{\theta}_n^*$ in practice, since Ω is unknown. But since $\hat{\Sigma} \xrightarrow[p]{} \Sigma$ from Corollary 2.4.3, we must have $\underline{\theta}_n^* - \hat{\underline{\theta}}_n \xrightarrow[p]{} 0$. Thus, using a well known result by Cramer [17], page 254, we can conclude that $\hat{\underline{\theta}}_n$ and $\underline{\theta}_n^*$ have the same asymptotic distribution; that is, $\hat{\underline{\theta}}_n$ is BAN for $\underline{\theta}$.

Unbiasedness follows from the argument of Kakwani [30], who showed for the MDM model that $\hat{\underline{\xi}}$ is symmetrically distributed about $\underline{\xi}$ whenever $\underline{y} - D\underline{\xi}$ follows a continuous symmetric probability law.

q.e.d.

It is important to note here that although (3.5.1) gives a BAN estimator, we are not claiming that it is the "best" BAN estimator in the sense of having a variance matrix which approaches the limiting variance matrix (3.2.23) at the fastest rate as the sample size increases. We shall describe below a method of iteration using (3.5.1) which will give another BAN estimator for the MGLM model. A computer study similar to the one made by Kmenta and Gilbert [33] for the MDM model would be needed to compare these different BAN estimators in small samples, although it can be said that the results for the MDM model indicate that the estimator (3.5.1) would compare favorably with iterative-type estimators.[4] In any case, it is clear that for large samples, (3.5.1) has minimum variance. We also show in the next chapter that (3.5.1) may be used in large sample tests of hypotheses for the MGLM model.

3.5.1 Special Cases

It is easily seen that if $\underline{\xi}$ itself is estimable and we consider the special case of the MDM model, then for $H = I_M$, (3.5.1) gives us Zellner's ZEF estimator. If we consider the SM model and we recall in Section 2.2 of Chapter Two that Σ, and therefore $\hat{\Sigma}$, drops out of

---

[4] It should also be pointed out that although Kmenta and Gilbert's study compared the small sample variances of different estimators, it did not directly compare the different estimators as to the rate at which they approached their limiting variances. A recent paper by Bahadur [9] discusses the concept of the rate at which an estimator approaches its limiting variance.

$(D'\Omega^{-1}D)^{-}D'\Omega^{-1}$, we see that (3.5.1) gives us the usual MLE and BLUE for the SM model. As for the GIM model, we will now show that we can write (3.5.1) in terms of the original data matrices Y_j by considering the original definition of the model:

$$E(Y_j) = A_j\xi B_j$$

$$\mathrm{Var}(Y_j) = I_n \otimes B_j'\Sigma B_j \ , \quad j=1,\ldots,u.$$

Letting $Y_j(n_j \times q_j) = [\underline{y}_{j1}, \ldots, \underline{y}_{jq_j}]$, we can define $\underline{y}_j(n_jq_j \times 1) = (\underline{y}_{j1}' \ \underline{y}_{j2}' \cdots \underline{y}_{jq_j}')'$. Clearly

$$\begin{aligned} E(\underline{y}_{jr}) &= A_j\xi\underline{b}_{jr} \text{ , where } \underline{b}_{jr} \text{ is the r-th column of } B_j \text{ ,} \\ &= [A_j\underline{\xi}_1 \mid A_j\underline{\xi}_2 \mid \cdots \mid A_j\underline{\xi}_p]\underline{b}_{jr} \\ &= [\underline{b}_{jr}' \otimes A_j]\underline{\xi} \ . \end{aligned}$$

Thus $$E(\underline{y}_j) = \begin{bmatrix} \underline{b}_{j1}' \otimes A_j \\ \vdots \\ \underline{b}_{jq_j}' \otimes A_j \end{bmatrix} \underline{\xi} = (B_j' \otimes A_j)\underline{\xi}.$$

Also $$\mathrm{Var}(\underline{y}_j) = B_j'\Sigma B_j \otimes I_{n_j} \ .$$

If we now define

$$\underline{y}(N \times 1) = (\underline{y}_1' \ \underline{y}_2' \cdots \underline{y}_u')'$$

we have

(3.5.1.1) $$E(\underline{y}) = \tilde{D}\underline{\xi} \ , \quad \mathrm{Var}(\underline{y}) = \tilde{\Omega}$$

where

(3.5.1.2) $$\tilde{D}(N \times mp) = \begin{bmatrix} B_1' \otimes A_1 \\ \vdots \\ B_u' \otimes A_u \end{bmatrix}$$

and

(3.5.1.3) $$\tilde{\Omega}(N \times N) = \begin{bmatrix} (B_1'\Sigma B_1) \otimes I_{n_1} & & 0 \\ & \ddots & \\ 0 & & (B_u'\Sigma B_u) \otimes I_{n_u} \end{bmatrix} .$$

Thus if $\underline{\theta} = H'\underline{\xi}$ is estimable, we can write the BAN estimator (3.5.1) as

$$\begin{aligned} \hat{\underline{\theta}}_n &= H'\hat{\underline{\xi}} \\ &= H'(\tilde{D}'\tilde{\Omega}^{-1}\tilde{D})^{-}\tilde{D}'\tilde{\Omega}^{-1}\underline{y} \end{aligned}$$

$$= H'[\sum_{j=1}^{u} (B_j \otimes A_j')\{(B_j'\hat{\Sigma}B_j)^{-1} \otimes I_{n_j}\}(B_j' \otimes A_j)]^{-} \sum_{j=1}^{u} (B_j \otimes A_j')\{(B_j'\hat{\Sigma}B_j)^{-1} \otimes I_{n_j}\}\underline{y}_j$$

$$= H'[\sum_{j=1}^{u} B_j(B_j'\hat{\Sigma}B_j)^{-1} B_j' \otimes A_j'A_j]^{-} \sum_{j=1}^{u} [B_j(B_j'\hat{\Sigma}B_j)^{-1} \otimes A_j']\underline{y}_j .$$

Hence we have the following corollary to Theorem 3.5.1:

<u>Corollary 3.5.1.1:</u> For the GIM model a BAN estimator for an estimable linear set $\underline{\theta} = H'\underline{\xi}$ is given by

(3.5.1.4) $$\hat{\underline{\theta}}_n = H'\hat{\underline{\xi}} = H'[\sum_{j=1}^{u} B_j(B_j'\hat{\Sigma}B_j)^{-1} B_j' \otimes A_j'A_j]^{-} \sum_{j=1}^{u} [B_j(B_j'\hat{\Sigma}B_j)^{-1} \otimes A_j']\underline{y}_j$$

where $\hat{\Sigma}$ is defined in Theorem 2.4.1, and $\underline{y}_j$ is formed by putting the columns of Y_j underneath each other.

3.5.2 Iterative Techniques

We now describe two iterative methods which use the estimator (3.5.1) to obtain alternative BAN estimators. The first is analogous to that used for the IZEF estimator described in Section 3.4. It uses (3.5.1) to calculate a new set of residuals leading to a new estimate of $\hat{\Omega}$. This estimate in turn is used for obtaining new estimates of $\hat{\underline{\theta}}$ and so on. Thus for $\hat{\underline{\theta}} = H'\hat{\underline{\xi}}$ as given in (3.5.1) we would define $\hat{\Omega}_{(2)}$ as

$$\hat{\Omega}_{(2)} = [\underline{x} - D\hat{\underline{\xi}}][\underline{x} - D\hat{\underline{\xi}}]'$$

and use as our first iterative estimate of $\underline{\theta}$:

$$\hat{\underline{\theta}}_{(2)} = H'(D'\hat{\Omega}^{-1}_{(2)}D)^{-}D'\hat{\Omega}^{-1}_{(2)}\underline{x} .$$

Continuing in this manner we would have for the (k-1)-th iteration

$$\text{(3.5.2.1)} \qquad \hat{\underline{\theta}}_{(k)} = H'\hat{\underline{\xi}}^{(k)} = H'(D'\hat{\Omega}^{-1}_{(k)}D)^{-}D'\hat{\Omega}^{-1}_{(k)}\underline{x}$$

where $$\hat{\Omega}_{(k)} = [\underline{x} - D\hat{\underline{\xi}}^{(k-1)}][\underline{x} - D\hat{\underline{\xi}}^{(k-1)}]' .$$

This process would continue until the estimator of $\underline{\theta}$ converges.

An alternative procedure, which should yield the same result as the above, uses the well-known (e.g., Rao [41]) method of <u>scoring</u>. In order to use this method, it is necessary to calculate $B_{n,\underline{\sigma}}$, the information matrix defined in (3.1.2) for the p(p-1)/2 independent parameters in $\Sigma = ((\sigma_{rs}))$, where

$$\underline{\sigma} = (\sigma_{11}, \sigma_{12}, \sigma_{22}, \sigma_{13}, \sigma_{23}, \sigma_{33}, \ldots, \sigma_{1p}, \sigma_{2p}, \ldots, \sigma_{pp})' .$$

We thus need the following result:

<u>Theorem 3.5.2.1</u>: For the MGLM model, the (r,s)-th element of $\frac{\partial \log \phi}{\partial \underline{\sigma}}$ is given by

$$\text{(3.5.2.2)} \qquad \frac{\partial \log \phi}{\partial \sigma_{rs}} = -\frac{1}{2}\delta_{rs}\sum_{j=1}^{u}\Big[n_j \left\langle B_j(B_j'\Sigma B_j)^{-1}B_j'\right\rangle_{rs} - \left\langle B_j(B_j'\Sigma B_j)^{-1}P_j'P_j(B_j'\Sigma B_j)^{-1}B_j'\right\rangle_{rs}\Big]$$

where $\delta_{rs} = \begin{cases} 2, & r \neq s \\ 1, & r = s \end{cases}$

and $P_j = Y_j - E(Y_j)$.

Also, the element of $B_{n,\underline{\sigma}}$ corresponding to σ_{rs} and σ_{wz} is given by the formula

$$(3.5.2.3) \qquad i_{rs,wz}(\underline{\sigma}) = \frac{1}{2n}\,\delta_{rs} \sum_{j=1}^{u} n_j \left\langle B_j (B_j' \Sigma B_j)^{-1} B_j' \frac{\partial \Sigma}{\partial \sigma_{wz}} B_j (B_j' \Sigma B_j)^{-1} B_j' \right\rangle_{rs}$$

where $\left\langle \dfrac{\partial \Sigma}{\partial \sigma_{wz}} \right\rangle_{mn} = \begin{cases} 1 \text{ if } (m,n) = (w,z) \text{ or } (z,w) \\ 0 \text{ otherwise.} \end{cases}$

<u>Proof</u>: (3.5.2.2) is obtained using (viii) and (xii) of Theorem A.2.1 in the Appendix.

Letting $\underline{b}'_{(j)r}(1 \times q_j)$ denote the r-th row of B_j we can write (3.5.2.2) as

$$\frac{\partial \log \phi}{\partial \sigma_{rs}} = -\frac{1}{2}\,\delta_{rs} \sum_{j=1}^{u} [n_j \underline{b}'_{(j)r} (B_j' \Sigma B_j)^{-1} \underline{b}_{(j)s} - \underline{b}'_{(j)r} (B_j' \Sigma B_j)^{-1} P_j' P_j (B_j' \Sigma B_j)^{-1} \underline{b}_{(j)s}] .$$

Differentiating again with respect to σ_{wz} using (i) and (ii) of Theorem A.2.1 gives

$$\begin{aligned}\frac{\partial^2 \log \phi}{\partial \sigma_{rs} \partial \sigma_{wz}} = \frac{1}{2} \sum_{j=1}^{u} \delta_{rs} [& n_j \underline{b}'_{(j)r} (B_j' \Sigma B_j)^{-1} B_j' \frac{\partial \Sigma}{\partial \sigma_{wz}} B_j (B_j' \Sigma B_j)^{-1} \underline{b}_{(j)s} \\ & - \underline{b}'_{(j)r} (B_j' \Sigma B_j)^{-1} B_j' \frac{\partial \Sigma}{\partial \sigma_{wz}} B_j (B_j' \Sigma B_j)^{-1} P_j' P_j (B_j' \Sigma B_j)^{-1} \underline{b}_{(j)s} \\ & - \underline{b}'_{(j)r} (B_j' \Sigma B_j)^{-1} P_j' P_j (B_j' \Sigma B_j)^{-1} B_j' \frac{\partial \Sigma}{\partial \sigma_{wz}} B_j (B_j' \Sigma B_j)^{-1} \underline{b}_{(j)s}] .\end{aligned}$$

Now $\quad E[P_j' P_j] = E[(Y_j - E(Y_j))'(Y_j - E(Y_j))]$

$$= n_j B_j' \Sigma B_j \; .$$

Thus

$$E\left[-\frac{1}{n}\frac{\partial^2 \log \phi_n}{\partial\sigma_{rs}\partial\sigma_{wz}}\right] = \frac{1}{2n}\,\delta_{rs}\sum_{j=1}^{u} n_j \underline{b}'_{(j)r}(B_j'\Sigma B_j)^{-1} B_j' \frac{\partial\Sigma}{\partial\sigma_{wz}} B_j(B_j'\Sigma B_j)^{-1}\underline{b}_{(j)s}$$

$$= \frac{1}{2n}\,\delta_{rs}\sum_{j=1}^{u} n_j \left\langle B_j(B_j'\Sigma B_j)^{-1} B_j' \frac{\partial\Sigma}{\partial\sigma_{wz}} B_j(B_j'\Sigma B_j)^{-1} B_j'\right\rangle_{rs} .$$

q.e.d.[5]

Using the results of Theorem 3.5.2.1, we can proceed with the method of scoring for the MGLM model as follows:

Let $\hat{\underline{\sigma}}$ be the estimate of $\underline{\sigma}$ which uses the $\hat{\Sigma}$ of Theorem 2.4.1. The first iterative estimator of $\underline{\sigma}$ is then given by

$$\hat{\underline{\sigma}}_{(2)} = \hat{\underline{\sigma}} + B^{-1}_{n,\hat{\underline{\sigma}}} \left.\frac{\partial \log \phi_n}{\partial \underline{\sigma}}\right|_{\underline{\sigma} = \hat{\underline{\sigma}},\quad \underline{\xi} = \hat{\underline{\xi}}} .$$

From $\hat{\underline{\sigma}}_{(2)}$ we define

$$\hat{\Omega}_{(2)} = \Omega\Big|_{\hat{\underline{\sigma}} = \hat{\underline{\sigma}}_{(2)}}$$

which we use to obtain the first iterative estimator of $\underline{\theta} = H'\underline{\xi}$. At the (k-1)-th iteration we thus have

$$\hat{\underline{\sigma}}_{(k)} = \hat{\underline{\sigma}}_{(k-1)} + B^{-1}_{n,\hat{\underline{\sigma}}_{(k-1)}} \left.\frac{\partial \log \phi_n}{\partial \underline{\sigma}}\right|_{\underline{\sigma} = \hat{\underline{\sigma}}_{(k-1)},\quad \underline{\xi} = \hat{\underline{\xi}}^{(k-1)}}$$

$$\hat{\Omega}_{(k)} = \Omega\Big|_{\underline{\sigma} = \hat{\underline{\sigma}}_{(k)}}$$

---

[5] Note also that

$$\left\langle B_j' \frac{\partial\Sigma}{\partial\sigma_{wz}} B_j\right\rangle_{tv} = \frac{1}{2}\,\delta_{wz}(b_{(j)wt}\, b_{(j)zv} + b_{(j)wv}\, b_{(j)zt})$$ where $b_{(j)cd}$ is the (c,d)-th element of B_j .

and $\hat{\underline{\xi}}_{(k)} = H'\hat{\underline{\xi}}^{(k)}$ given by (3.5.2.1).

Since Kmenta and Gilbert's study yields identical IZEF, TIE and MLE estimators, it would be expected that generalized versions of their estimators for the MGLM model would yield identical results. We thus do not suggest the use of a method analagous to the TIE method nor one which solves the likelihood equations directly, since both methods are difficult to use on the computer.

3.6 BAN Estimation for the GGCM Model

Recalling Section 1.5 of Chapter One, the GGCM model is given by

$$E(Y_j) = A_j \xi P B_j$$

$$\mathrm{Var}(Y_j) = I_{n_j} \otimes B_j' \Sigma B_j \ , \quad j=1,\ldots,u$$

where

Y_j is $(n_j \times q_j)$, A_j is $(n_j \times m)$, ξ is $(m \times p)$,

Σ is $(q \times q)$, $B_j(q \times q_j)$ is the incidence matrix for the j-th set S_j of experimental units

and $P(p \times q)$ is a regression matrix of full rank p .

We have obtained for the GGCM model in Corollary 3.2.4 of this chapter the limiting variance matrix of any BAN estimator of an estimable linear set $H'\underline{\xi}$, where $\underline{\xi} = (\underline{\xi}_1', \ldots, \underline{\xi}_p')'$ and $H(mp \times w)$ is of full rank w. The limiting variance matrix is given in (3.2.28) as

$$C_n = nH'[\sum_{j=1}^{u} PB_j (B_j'\Sigma B_j)^{-1} B_j'P' \otimes A_j'A_j]^{-}H \ .$$

To obtain an estimator which is both unbiased and BAN for $H'\underline{\xi}$ we can use a method similar to that used for the GIM model in Section 3.5.1.

We first give an unbiased and consistent estimator of Σ:[6]

Lemma 3.6.1: For the GGCM model, an unbiased and consistent estimator of $\Sigma(q \times q)$ is given by $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$, where

$$\hat{\sigma}_{rs} = \frac{1}{N_{rs} - R(D_{rs})} \underline{x}'_{rs}[I - D_{rs}(D'_{rs}D_{rs})^{-}D'_{rs}]\underline{x}_{sr} \tag{3.6.1}$$

$r,s=1,\ldots,q$

where $N_{rs}(\geq 2)$ is the number of experimental units in which both V_r and V_s are observed,

$\underline{x}_{rs}(N_{rs} \times 1)$ is the observation vector on V_r corresponding to those experimental units on which both V_r and V_s are observed,

and $D_{rs}(N_{rs} \times m)$ is the design matrix, consisting of a row of A_j matrices, corresponding to $\underline{x}_{rs}$.[7]

Proof: Using Lemma 2.4.1 of Chapter Two with $G= I - D_{rs}(D'_{rs}D_{rs})^{-}D'_{rs}$ and noting that $E(\underline{x}'_{rs})GE(\underline{x}_{sr}) = 0$, it is easily seen that

$$E(\hat{\sigma}_{rs}) = \frac{\text{tr } G}{N_{rs} - R(D_{rs})}\sigma_{rs} = \sigma_{rs} .$$

Consistency follows from the fact that $\hat{\sigma}_{rs}$ is the usual pooled estimate of covariance which is consistent for σ_{rs} provided $\lim\limits_{n\to\infty} \frac{N_{rs}}{n}$ exists and is non-zero for all r,s. q.e.d.

[6]As for the MGLM model, it is shown by example in Chapter Five that $\hat{\Sigma}$ is not necessarily positive definite, although it does have a positive definite limit.

[7]Note that when r=s, $N_{ss} = N_s$, $\underline{x}_{ss} = \underline{x}_s$, and $D_{ss} = D_s$ correspond to all those experimental units on which V_s is observed.

Theorem 3.6.1: For the GGCM model an unbiased BAN estimator of an estimable linear set $\underline{\theta} = H'\underline{\xi}$ is given by

$$(3.6.2)\quad \hat{\underline{\theta}}_n = H'\Big[\sum_{j=1}^{u} PB_j(B_j'\hat{\Sigma}B_j)^{-1}B_j'P' \otimes A_j'A_j\Big]^{-}\sum_{j=1}^{u}\big[PB_j(B_j'\hat{\Sigma}B_j)^{-1} \otimes A_j'\big]\underline{y}_j$$

where $\underline{y}_j(n_jq_j \times 1)$ is formed by putting the columns of Y_j underneath each other and $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$ is defined by (3.6.1).

Proof: Letting $Y_j = [\underline{y}_{j1}\ \underline{y}_{j2} \cdots \underline{y}_{jq_j}]$, we can write $\underline{y}_j = (\underline{y}_{j1}\ \underline{y}_{j2} \cdots \underline{y}_{jq_j})'$. Now if we let $PB_j = [\underline{p}_{j1}\ \underline{p}_{j2} \cdots \underline{p}_{jq_j}]$, then

$$\begin{aligned} E(\underline{y}_{jr}) &= A_j\xi\underline{p}_{jr} \\ &= [A_j\underline{\xi}_1 \mid A_j\underline{\xi}_2 \mid \cdots \mid A_j\underline{\xi}_p]\underline{p}_{jr} \\ &= [\underline{p}_{jr}' \otimes A_j]\underline{\xi}\ , \quad r=1,\ldots,q_j\ . \end{aligned}$$

Thus
$$E(\underline{y}_j) = \begin{bmatrix} \underline{p}_{j1}' \otimes A_j \\ \vdots \\ \underline{p}_{jq_j}' \otimes A_j \end{bmatrix}\underline{\xi} = (B_j'P' \otimes A_j)\underline{\xi}$$

Also
$$\mathrm{Var}(\underline{y}_j) = B_j'\Sigma B_j \otimes I_{n_j}\ .$$

If we now define $\underline{y}(N \times 1) = (\underline{y}_1'\ \underline{y}_2' \cdots \underline{y}_u')'$, then clearly

$$E(\underline{y}) = D\underline{\xi}\ ,\quad \mathrm{Var}(\underline{y}) = \Omega$$

where

$$D(N \times mp) = \begin{bmatrix} B_1'P' \otimes A_1 \\ \vdots \\ B_u'P' \otimes A_u \end{bmatrix}$$

and

$$\Omega(N \times N) = \begin{bmatrix} (B_1'\Sigma B_1) \otimes I_{n_1} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & (B_u'\Sigma B_u) \otimes I_{n_u} \end{bmatrix}.$$

Recalling the proof of Theorem 3.5.1 for the MGLM model it is obvious that

$$(3.6.3) \qquad \hat{\underline{\theta}}_n = H'\hat{\underline{\xi}} = H'(D'\hat{\Omega}^{-1}D)^{-}D'\hat{\Omega}^{-1}\underline{y}$$

is an unbiased BAN estimator of $\underline{\theta}$, where $\hat{\Omega}$ is obtained from Ω by replacing Σ by $\hat{\Sigma}$. Now by direct computation we obtain

$$(D'\hat{\Omega}^{-1}D)^{-}D'\hat{\Omega}^{-1}\underline{y}$$

$$= [\sum_{j=1}^{u} (PB_j \otimes A_j')\{(B_j'\hat{\Sigma}B_j)^{-1} \otimes I_{n_j}\} (B_jP' \otimes A_j)]^{-} \sum_{j=1}^{u} (PB_j \otimes A_j')[(B_j\hat{\Sigma}B_j)^{-1} \otimes I_{n_j}]\underline{y}_j$$

$$=[\sum_{j=1}^{u} PB_j(B_j'\hat{\Sigma}B_j)^{-1} B_j'P' \otimes A_j'A_j]^{-} \sum_{j=1}^{u} [PB_j(B_j'\hat{\Sigma}B_j)^{-1} \otimes A_j']\underline{y}_j$$

so that $\hat{\underline{\theta}}_n$ is given by (3.6.2).

q.e.d.

Alternative estimators to (3.6.2) can be obtained by iterative techniques which are directly analogous to those given in Section 3.5.2 for the MGLM model.

CHAPTER IV

TESTING LINEAR HYPOTHESES FOR THE MGLM AND GGCM MODELS

4.1 Introduction

In this chapter, a method is proposed for testing linear hypotheses in the MGLM and GGCM models when the underlying data is normally distributed. The test statistics are quadratic forms called Wald Statistics, and they are constructed from BAN estimators of linear functions of the treatment parameters and consistent estimators of the variance parameters. The asymptotic distribution of a Wald Statistic is a central chi-square variable. Thus when the sample size is large, the test criteria yield chi-square tests. Even for small sample sizes the test criterion may have desirable properties, since it can be regarded as a generalization of Hotelling's trace for the SM model. Nevertheless, to use the Wald Statistic in small samples, it is necessary to obtain its exact distribution. This has not yet been done and appears difficult.

4.2 Review of the Literature

A number of different approaches have been used in the literature to obtain tests of hypotheses for the MDM, HM and GIM models. The tests obtained are generally impractical with regard to computation and require more assumptions on the hypothesis and design matrices than the usual estimability requirement. For example, the necessary and sufficient conditions given by Srivastava in [57] for reparameterizing the MDM model to the SM model are that the vector spaces spanned by the columns of the different design matrices are the same. In [58], Srivastava restricts the GIM model[1] to be what he calls a "strongly regular" design. In [57], he proves to be not strongly regular a particular example of the kind of design likely to be considered in practice. In [58], however, he works through a GIM model which is strongly regular and specifies the general computational method for reparameterization. Nevertheless, the class of strongly regular designs needs to be better described in terms of GIM models likely to be used in practice. This problem seems formidable since the computations needed for validating the strongly regular property are quite complex. Furthermore, the reparameterized form of the model has a Σ matrix which is restricted by $|\rho_{ij}| \leq \lambda_{ij} \leq 1$ where λ_{ij} may be less than 1 for some $i \neq j$ and ρ_{ij} is the correlation between variates V_i and V_j. The problem of testing a hypothesis in a SM model with the above type of restriction on Σ remains to be solved and is likely to require a laborious computer procedure.

In [64], Trawinski and Bargmann summarize the main results of Trawinski's thesis [63] in which is obtained a likelihood ratio test for the special case of the GIM model in which $u=p$, $n_j \equiv n$, $q_j \equiv q$ and $A_j \equiv A$, $j=1,\ldots,p$. The maximum likelihood equations for their model require for solution an iterative technique based on the Newton-Raphson

---

[1] Srivastava's model in [58] allows $\underset{\sim}{\xi}_j$ to be $(m_j \times 1)$ and A_j to be $(n_j \times m_j)$, whereas we have assumed that $m_j \equiv m$ for all j.

method. However, their method cannot be directly generalized when not all the A_j matrices are the same. In a demonstration of their method for a particular set of data, the initial estimate Σ_o of Σ used in the iteration yielded an estimate of ξ which was comparable to the final estimate obtained after several iterations. Σ_o is a pooled estimate, in which the (r,s)-th element σ^o_{rs} is obtained by averaging estimates of σ_{rs} from all experimental units in which both V_r and V_s are measured together.[2]

In [59], Srivastava has suggested several different union-intersection type tests for the HM model, one of which was obtained in the earlier paper [55] and which involves a generalization of Roy's step-down procedure [47]. The testing procedure involves making a sequence of independent F tests and rejecting the hypothesis if any of the F tests are rejected. Srivastava points out, however, that such a procedure may only be used when the response variates can be ordered as in the case of the HM model.

The Wald Statistic proposed below was first suggested by Wald [66] and has been used by Allen [1] in connection with nonlinear multivariate models, and by Bhapkar [12] and others [25] for problems with

---

[2] Σ_o is the same as the $\hat{\Sigma}$ of Corollary 2.4.6. Thus σ^o_{rs} is given by

$$\sigma^o_{rs} = \frac{1}{N_{rs} - R(A)} \underline{x}'_{rs}[I_{N_{rs}} - D_{rs}(D'_{rs}D_{rs})^{-}D'_{rs}]\underline{x}_{sr}$$

where $D_{rs}(N_{rs} \times m) = [A' \ldots A']'$ and $\underline{x}_{rs}(N_{rs} \times 1)$ is the data vector for V_r corresponding to all experimental units on which both V_r and V_s are observed together.

categorical data. Allen showed that the likelihood ratio criterion and the Wald criterion for his nonlinear model are asymptotically equivalent. For small samples, however, the Wald Statistic was not a good approximation to a chi-square variable. Nevertheless, this may have been due to the use of a linear approximation of the test criterion, which is not necessary for the linear MGLM model.

4.3 The Wald Statistic

In Section 3.2 of the previous chapter we have defined a BAN estimator $\underline{g}(\hat{\underline{\theta}}_n)$ for $\underline{g}(\underline{\theta})$ by requiring

$$\mathcal{L}(\sqrt{n}\; C_n^{-\frac{1}{2}}\; [\underline{g}(\hat{\underline{\theta}}_n) - \underline{g}(\underline{\theta}_o)]) \to \mathcal{N}_w(\underline{0}_w,\; I_w)$$

where $\underline{\theta}_o$ is the true value of $\underline{\theta}$ $(u \times 1)$,

$\underline{g}(\underline{\theta})$ is a continuous $(w \times 1)$ vector function of $\underline{\theta}$ with continuous partial derivatives,

$C_n = \Delta_o' B^-_{n,\underline{\theta}_o} \Delta_o$ is symmetric, positive definite,

$$\Delta_o = \left.\frac{\partial\, \underline{g}(\underline{\theta})}{\partial\, \underline{\theta}}\right|_{\underline{\theta} = \underline{\theta}_o} ,$$

$$B_{n,\underline{\theta}_o} = E_{\underline{\theta}_o}\left[- \frac{1}{n} \left.\frac{\partial^2 \log \phi_n}{\partial\, \underline{\theta}^2}\right|_{\underline{\theta} = \underline{\theta}_o} \right]$$

and ϕ_n is the likelihood function for a sample of size n .

Now if $\underline{g}(\hat{\underline{\theta}}_n)$ is BAN for $\underline{g}(\underline{\theta})$, then it is well known, e.g., Cramer [17], that

$$(4.3.1) \qquad n[\underline{g}(\hat{\underline{\theta}}_n) - \underline{g}(\underline{\theta}_o)]'\; C_n^{-1}\; [\underline{g}(\hat{\underline{\theta}}_n) - \underline{g}(\underline{\theta}_o)] \xrightarrow[L]{} \chi^2_w .$$

Furthermore, if $\hat{\underline{\theta}}_n \xrightarrow[p]{} \underline{\theta}_o$, we can claim, e.g., using [17], that

$$(4.3.2) \qquad n[\underline{g}(\hat{\underline{\theta}}_n) - \underline{g}(\underline{\theta}_o)]' \, \hat{C}_n^{-1} \, [\underline{g}(\hat{\underline{\theta}}_n) - \underline{g}(\underline{\theta}_o)] \underset{L}{\to} \chi_w^2$$

where $\hat{C}_n$ is obtained from C_n by substituting $\hat{\underline{\theta}}_n$ for $\underline{\theta}_o$.

Now suppose we want to test the hypothesis

$$H_o : \underline{g}(\underline{\theta}_o) = 0 .$$

Wald [66] suggested using the statistic

$$(4.3.3) \qquad W_n = n[\underline{g}(\hat{\underline{\theta}}_n)]' \, \hat{C}_n^{-1} \, [\underline{g}(\hat{\underline{\theta}}_n)]$$

which is asymptotically χ_w^2 when H_o is true, i.e., we would reject H_o whenever W_n exceeded the $(1-\alpha)$-th percentage point of χ_w^2 , where α is a chosen level of significance. We call the statistic W_n the Wald Statistic, and we will use it to obtain test criteria for the MGLM and GGCM models. We note here that Wald suggested using W_n when the data in the sample were i.i.d. and when $\hat{\underline{\theta}}_n$ was the MLE of $\underline{\theta}$. It is clear from the above, however, that the asymptotic distribution of W_n will remain chi-square when $\underline{g}(\hat{\underline{\theta}}_n)$ is <u>any</u> BAN estimator for $\underline{g}(\underline{\theta})$, although the closeness of the approximation of W_n to χ_w^2 in small samples may vary with different BAN estimators. For the i.i.d. situation using MLE's, Wald showed that a test using W_n had a number of desirable asymptotic properties; namely, it had <u>asymptotically best average power</u>, <u>asymptotically best constant power</u>, and was <u>asymptotically most stringent</u>. It still remains to be shown that these properties can be carried over to W_n formed from independent but not i.i.d. data using arbitrary BAN estimators.

4.4 <u>Testing Linear Hypotheses in the MGLM Model</u>

We wish to test the hypothesis

$$H_o : H'\underline{\xi} = 0$$

where $H'\underline{\xi}$ is estimable, $H(M \times w)$ is of full rank $w(\leq M)$ and $\underline{\xi} = (\underline{\xi}_1' \ldots \underline{\xi}_p')'$ is the vector of unknown treatment parameters in the MGLM model. Since we have already obtained the asymptotic variance matrix of any BAN estimator for $H'\underline{\xi}$ in (3.2.23) of Chapter Three, we can form the corresponding Wald Statistic. Also, it is easily seen that if $\underline{g}(\underline{\xi}) = H'\underline{\xi}$ then $\Delta = \dfrac{\partial\, \underline{g}(\underline{\xi})}{\partial\, \underline{\xi}} = H$. Thus we have the following theorem:

Theorem 4.4.1: For the MGLM model, let $H'\underline{\xi}$ be estimable where $H(M \times w)$ is of full rank w. Then under

$$H_o : H'\underline{\xi} = 0$$

we have that

$$W_n = (H'\hat{\underline{\xi}})'[H'(D'\hat{\Omega}^{-1}D)^{-}H]^{-1}(H'\hat{\underline{\xi}}) \tag{4.4.1}$$

is asymptotically distributed as a central chi-square variable with w degrees of freedom, where

$$\hat{\Omega} = \Omega \Big|_{\Sigma = \hat{\Sigma}}$$

$\hat{\Sigma}$ is any positive definite consistent estimator of Σ,

Ω and D are defined by the vector version (1.4.8),

and $H'\hat{\underline{\xi}}$ is any BAN estimator of $H'\underline{\xi}$.

In particular we may use $H'\hat{\underline{\xi}}$ as given in Theorem 3.5.1 of Chapter Three and $\hat{\Sigma}$ as given in Theorem 2.4.1 of Chapter Two.[3,4]

[3] If $\hat{\Sigma}$ is not positive definite, however, then its use will lead to negative values of W_n , thus giving spurious results. In Chapter Five it is shown that the $\hat{\Sigma}$ of Theorem 2.4.1 is not always positive definite.

[4] $D'\Omega^{-1}D$ may be replaced by its equivalent form in terms of the original design matrices and $\hat{\Sigma}$ as given in Theorem 3.2.1 of Chapter Three.

To test the hypothesis $H_o : H'\underline{\xi} = 0$ we may thus reject H_o if $W_n \geq \chi^2_{w,1-\alpha}$ and accept otherwise. When the sample size is small W_n can be used as a test statistic <u>provided</u> its exact distribution can be obtained. The following corollary indicates that W_n might have desirable test properties even when n is small.

<u>Corollary 4.4.2</u>: The Wald Statistic obtained for the SM model is equivalent to Hotelling's Trace Criterion (T_o^2) when testing $H_o : C\xi D = 0$ where $C(g \times m)$ and $D(p \times v)$ are of full rank g and v, respectively.

<u>Proof</u>: Using the notation of Section 1.1 we have that

$$T_o^2 = \text{tr } S_H S_E^{-1} \text{ , where}$$

$$S_H = (C\hat{\xi}D)'[C(A'A)^- C']^{-1}(C\hat{\xi}D) \text{ ,}$$

$$S_E = D'Y'[I - A(A'A)^- A']YD$$

and $$\hat{\xi} = [\hat{\underline{\xi}}_1 \ldots \hat{\underline{\xi}}_p] = (A'A)^- A'Y \text{ .}$$

The Wald Statistic for testing $C\xi D = 0$ (i.e., $[D' \otimes C]\underline{\xi} = \underline{0}$) is clearly given by

$$W_n = [(D' \otimes C)\hat{\underline{\xi}}]'[(D' \otimes C)(\hat{\Sigma}^{-1} \otimes A'A)^-(D' \otimes C)']^{-1}[(D' \otimes C)\hat{\underline{\xi}}]$$

where $$\hat{\underline{\xi}} = (\hat{\underline{\xi}}_1' \ldots \hat{\underline{\xi}}_p')'$$

$$\hat{\Sigma} = \frac{1}{n} Y'[I - A(A'A)^- A']Y$$

and $$\Sigma^{-1} \otimes \frac{1}{n} A'A \text{ is the information matrix for } \underline{\xi} \text{ .}$$

Now using the properties of Kronecker Product multiplication and generalized inverses, we have that

$$W_n = [(D' \otimes C)\hat{\underline{\xi}}]'[(D' \otimes C)(\hat{\Sigma} \otimes (A'A)^-)(D \otimes C')]^{-1}[(D' \otimes C)\hat{\underline{\xi}}]$$

$$= [(D' \otimes C)\hat{\underline{\xi}}]'[D'\hat{\Sigma}D \otimes C(A'A)^{-}C']^{-1}[(D' \otimes C)\hat{\underline{\xi}}]$$

$$= n[(D' \otimes C)\hat{\underline{\xi}}]'[S_E^{-1} \otimes (C(A'A)^{-}C')^{-1}][(D' \otimes C)\hat{\underline{\xi}}]$$

$$= n \begin{bmatrix} C\hat{\xi}\underline{d}_1 \\ \vdots \\ C\hat{\xi}\underline{d}_v \end{bmatrix}' [S_E^{-1} \otimes (C(A'A)^{-}C')^{-1}] \begin{bmatrix} C\hat{\xi}\underline{d}_1 \\ \vdots \\ C\hat{\xi}\underline{d}_v \end{bmatrix}$$

where $\underline{d}_j$ is the j-th column of D

$$= n \sum_{i=1}^{v} \sum_{j=1}^{v} s_E^{ij} (C\hat{\xi}\underline{d}_i)'[C(A'A)^{-}C']^{-1}(C\hat{\xi}\underline{d}_j)$$

where s_E^{ij} is the (i,j)-th element of S_E^{-1}

$$= n\, tr(C\hat{\xi}D)'[C(A'A)^{-}C']^{-1}(C\hat{\xi}D)S_E^{-1}$$

$$= n\, T_o^2 .$$

q.e.d.

It is easily seen that even for the MGLM model, W_n can be written as a trace statistic:

$$W_n = tr[H'(D'\hat{\Omega}^{-1}D)^{-}H]^{-1}(H'\hat{\underline{\xi}})(H'\hat{\underline{\xi}})'$$

so that we may consider W_n to be a generalization of Hotelling's T_o^2. Since T_o^2 has been shown in the computer studies of Schatzoff [50] to provide good protection for a wide spectrum of alternative hypotheses, we might suspect that W_n will have this property, too. Nevertheless, work still needs to be done here and a study similar to that of Schatzoff might be used to compare different tests for the MGLM model.

We note in addition that if we consider the hypothesis $H_o : C\,\xi D = 0$ for the GLM model, then we obtain the following corollary:

Corollary 4.4.3: For the GIM model, the Wald Statistic for testing the hypothesis $H_o : C\xi D = 0$, where $C\xi D$ is estimable and $C(g \times m)$, $D(p \times v)$ are of full rank g and v, respectively, is given by

(4.4.2)

$$W_n = [(D' \otimes C)\hat{\underline{\xi}}]'[(D' \otimes C)\{\sum_{j=1}^{u} B_j(B_j'\hat{\Sigma}B_j)^{-1}B_j' \otimes A_j'A_j\}^{-}(D \otimes C')]^{-1}[(D' \otimes C)\hat{\underline{\xi}}]$$

where $(D' \otimes C)\hat{\underline{\xi}} = \begin{bmatrix} C\hat{\xi}\underline{d}_1 \\ \vdots \\ C\hat{\xi}\underline{d}_v \end{bmatrix}$,

$\underline{d}_j(p \times 1)$ is the j-th column of D ,

$(D' \otimes C)\hat{\underline{\xi}}$ is any BAN estimator of $(D' \otimes C)\underline{\xi}$

and $\hat{\Sigma}$ is any consistent positive definite estimator of Σ.

4.5 Testing Linear Hypotheses in the GGCM Model

For the GGCM model, we wish to test the hypothesis $H_o : H'\underline{\xi} = 0$ where $H'\underline{\xi}$ is estimable, $H(mp \times w)$ is of full rank $w(\le mp)$ and $\underline{\xi} = (\underline{\xi}_1', \ldots, \underline{\xi}_p')'$ is the vector of unknown treatment parameters. Since we have already obtained the asymptotic variance matrix of any BAN estimator of $H'\underline{\xi}$ in Theorem 3.2.4 of Chapter Three, we can form the corresponding Wald Statistic. We thus have the following theorem:

Theorem 4.5.1: For the GGCM model, let $H'\underline{\xi}$ be estimable where $H(mp \times w)$ is of full rank w. Then under $H_o : H'\underline{\xi} = 0$, we have that

$$(4.5.1)\quad W_n = (H'\hat{\underline{\xi}})'[H'\{\sum_{j=1}^{u} PB_j(B_j'\hat{\Sigma}B_j)^{-1}B_j'P' \otimes A_j'A_j\}^{-}H]^{-1}(H'\hat{\underline{\xi}})$$

is asymptotically distributed as χ^2_w, where

$\hat{\Sigma}$ is any consistent positive definite estimator of Σ

and $H'\hat{\underline{\xi}}$ is any BAN estimator of $H'\underline{\xi}$.

In particular, we may use $\hat{\Sigma}$ as given in Lemma 3.6.1[5] and $H'\hat{\underline{\xi}}$ as given in Theorem 3.6.1 of Chapter Three.

When restricted to the GCM model and the hypothesis $C\xi D = 0$, we obtain the following corollary:

<u>Corollary 4.5.2</u>: For the GCM model, let $C\xi D$ be estimable where $C(g \times m)$ and $D(p \times v)$ are of full rank g and v, respectively. Then the Wald Statistic is given by

$$W_n = n \operatorname{tr} S_H S_E^{-1} \tag{4.5.2}$$

where

$$S_H = (C\hat{\xi}D)'[C(A'A)^- C']^{-1}(C\hat{\xi}D) ,$$

$$S_E = D'(PS^{-1}P')^{-1} D ,$$

$$\hat{\xi} = (A'A)^- A'YS^{-1}P'(PS^{-1}P')^{-1} ,$$

and

$$S = Y'[I - A(A'A)^- A']Y .$$

<u>Proof:</u> We have already seen in Chapter One that $C\hat{\xi}D$ is BAN for $C\xi D$ and $\frac{1}{n} S_E = \hat{\Sigma}$ is consistent for Σ. Thus the Wald Statistic (4.5.1) becomes

$$W_n = [(D' \otimes C)\underline{\xi}]'[(D' \otimes C)(P\hat{\Sigma}^{-1}P' \otimes A'A)^-(D \otimes C')]^{-1}[(D' \otimes C)\hat{\underline{\xi}}]$$

$$= [(D' \otimes C)\hat{\underline{\xi}}]'[(D' \otimes C)\{(P\hat{\Sigma}^{-1}P')^{-1} \otimes (A'A)^-\}(D \otimes C')]^{-1}[(D' \otimes C)\hat{\underline{\xi}}]$$

$$= [(D' \otimes C)\hat{\underline{\xi}}]'[D'(P\hat{\Sigma}^{-1}P')^{-1} D \otimes C(A'A)^- C']^{-1}[(D' \otimes C)\hat{\underline{\xi}}]$$

$$= n[(D' \otimes C)\hat{\underline{\xi}}]'[S_E^{-1} \otimes (C(A'A)^- C')^{-1}][(D' \otimes C)\hat{\underline{\xi}}]$$

---

[5] $\hat{\Sigma}$ must be positive definite, however.

$$= n \begin{bmatrix} C\hat{\xi}\underline{d}_1 \\ \vdots \\ C\hat{\xi}\underline{d}_v \end{bmatrix}' [S_E^{-1} \otimes (C(A'A)^- C')^{-1}] \begin{bmatrix} C\hat{\xi}\underline{d}_1 \\ \vdots \\ C\hat{\xi}\underline{d}_v \end{bmatrix}$$

where $\underline{d}_j$ is the j-th column of D

$$= n \sum_{i=1}^{v} \sum_{j=1}^{v} s_E^{ij} (C\hat{\xi}\underline{d}_i)[C(A'A)^- C']^{-1} (C\hat{\xi}\underline{d}_i)$$

where s_E^{ij} is the (i,j)-th element of S_E^{-1}

$$= n \text{ tr } (C\hat{\xi}D)'[C(A'A)C']^{-1} (C\hat{\xi}D) S_E^{-1}$$

$$= n \text{ tr } S_H S_E^{-1} .$$

q.e.d.

Note that S_E is the same as that obtained by Khatri in (1.2.2.5) but that S_H differs from Khatri's (1.2.2.4) in that

$$R = (A'A)^- + (A'A)^- A'Y[S^{-1} - S^{-1}P'(PS^{-1}P')^{-1} PS^{-1}]Y'A(A'A)^-$$

is now replaced by $(A'A)^-$. Clearly Khatri's trace and $\frac{1}{n} W_n$ are asymptotically equivalent since they have the same asymptotic distribution. Also both of these statistics use the information in S, the sample covariance matrix. It is not obvious, however, which of these statistics is better in small samples, either as an approximation of a chi-square variable or with regard to protection against alternative hypotheses. A study similar to that of Schatzoff [50] would perhaps be needed to compare the two with regard to the latter.

Use of Khatri's trace for the GCM model clearly suggests an analogous competitor to (4.5.1) for the GGCM model. We can arrive at such a competitor by using the method of covariance adjustment. Since we have assumed that $P_j = PB_j$ is of rank p for each j=1,...,u , we can thus find matrices $F_j((q_j - p) \times q_j)$ of full rank $q_j - p$ for each j such that

$$(4.5.3) \qquad P_j F_j' = 0(p \times (q_j - p)),\ j=1,\ldots,u\ .$$

Note that we can choose F_j to be a column basis for $I_{q_j} - P_j'(P_jP_j')^{-1}P_j$.

We can now consider the following competitor to (4.5.1):

$$(4.5.4) \qquad W_n^* = (H'\hat{\underline{\xi}})'[H'(\sum_{j=1}^{u} PB_j(B_j'\hat{\Sigma}B_j)^{-1}B_j'P' \otimes Q_j)^{-}H]^{-1}(H'\hat{\underline{\xi}})$$

where

$$(4.5.5) \qquad Q_j = A_j'A_j - A_j'Y_jF_j'(F_jY_j'Y_jF_j')^{-1}F_jY_j'A_j\ ,\ j=1,\ldots,u\ .$$

To complete the analogy, we would like to show that Q_j is independent of the choice of F_j in the sense that Q_j has a generalized inverse $R_j = Q_j^-$ which is analogous to R. We first state without proof an important result due to Khatri:

<u>Lemma 4.5.1</u>. (Khatri [31]): Suppose $P(p \times q)$ has full rank $p(\le q)$ and $PF' = 0$ where $F((q - p) \times q)$ is of full rank. Then if $S(q \times q)$ is any positive definite matrix, we must have

$$(4.5.6) \qquad S^{-1} - S^{-1}P'(PS^{-1}P')^{-1}PS^{-1} = F'(FSF')^{-1}F\ .$$

We can now use Khatri's lemma to prove the following result:

<u>Lemma 4.5.2</u>: Let P and F be defined as in Lemma 4.5.1. Let $Y(n \times q)$ be of full rank q and let $S(q \times q) = Y'[I_n - A(A'A)^{-}A']Y$. Then a generalized inverse of

$$Q = A'A - A'YF'(FY'YF')^{-1}FY'A$$

is given by

$$(4.5.7) \qquad R = (A'A)^- + (A'A)^- A'Y[S^{-1} - S^{-1}P'(PS^{-1}P')^{-1} PS^{-1}]Y'A(A'A)^-$$

where $(A'A)^-$ is any generalized inverse of $A'A$. Furthermore, if A is of full rank m, then (4.5.7) gives a unique inverse of Q .

Proof: Let $U = A'A$, $V = A'YF'$, $W = FY'YF'$. Then, using Lemma 4.5.1, we have $R = U^- + U^-V(FSF')^{-1}V'U^-$. Clearly using Theorem A.1.1 (ii) of the Appendix

$$(4.5.8) \qquad UU^-V = (A'A)(A'A)^-A'YF' = A'YF' = V$$

always. Also

$$(4.5.9) \qquad VW^{-1}FSF' = V - VW^{-1}V'U^-V$$

since $\qquad FSF' = FY'[I - A(A'A)^-A']YF' = W - V'U^-V$.

$$\begin{aligned}
\text{Furthermore } QR &= [U - VW^{-1}V'][U^- + U^-V(FSF')^{-1}V'U^-] \\
&= UU^- - VW^{-1}V'U^- + [UU^- - VW^{-1}V'U^-]V(FSF')^{-1}V'U^- \\
&= UU^- - VW^{-1}V'U^- + [V - VW^{-1}V'U^-V](FSF')^{-1}V'U^- \\
&\qquad \text{using (4.5.8)} \\
&= UU^- - VW^{-1}V'U^- + VW^{-1}V'U^- \quad \text{using (4.5.9)} \\
&= UU^- .
\end{aligned}$$

$$\begin{aligned}
\text{Thus} \qquad QRQ &= UU^-[U - VW^{-1}V'] \\
&= UU^-U - (UU^-V)W^{-1}V' \\
&= U - VW^{-1}V' \\
&= Q
\end{aligned}$$

so that R is a generalized inverse of Q.

Now suppose A is of full rank, so that $A'A$ is of full rank and $(A'A)^- = (A'A)^{-1}$, the unique inverse of $A'A$. Then following the above proof it is clear that $QR = UU^{-} = I = UU^{-} = RQ$, so that (4.5.7) will be the unique inverse of Q .

q.e.d.

Using Lemma 4.5.2 we obtain the following theorem:

Theorem 4.5.3: For the GGCM model let $P_j = PB_j$,

$S_j = Y_j'[I_n - A_j(A_j'A_j)^{-}A_j']Y_j$, and define F_j to satisfy (4.5.3). Then a generalized inverse of Q_j as defined in (4.5.5) is given by

$$R_j = (A_j'A_j)^{-} + (A_j'A_j)^{-}A_j'Y_j[S_j^{-1} - S_j^{-1}P_j'(P_jS_j^{-1}P_j')^{-1}P_jS_j^{-1}]Y_j'A_j(A_j'A_j)^{-} .$$

Thus a competitor to (4.5.1) for testing $H'\underline{\xi} = 0$ is given by (4.5.4).

CHAPTER V

PRACTICAL EXAMPLES

5.1 Introduction

In this chapter, four examples of MGLM and GGCM models are presented to illustrate the kinds of results obtained when using the BAN estimators proposed in Theorems 3.5.1 and 3.6.1 of Chapter Three and the Wald tests of linear hypotheses described in Theorems 4.4.1 and 4.5.1 of Chapter Four. The four models used are (1) HIM (p=3), (2) MDM (p=2), (3) MDM (p=4) and (4) GGCM (p=7). The data for the first three of these models were generated using a normal random number generator. Thus the BAN estimators obtained for these models could be compared with the true values of the parameters since the latter (usually unknown) were known. Also, the results of tests of hypotheses could be compared with the actual truth or falsity of the hypotheses in question. Furthermore, the procedure used to generate the data enabled an experiment with a fixed set of underlying parameters $\{\underline{\xi}, \Sigma\}$ to be repeated to give additional data sets and thus provide some indication of the way in which the estimates and test statistics vary from sample to sample. It was also possible to vary Σ while keeping the model and the true $\underline{\xi}$ fixed to see how different covariance structures affected the estimates and test statistics.

The data for the fourth model were obtained by deleting observations from growth curve (GCM) data given in Allen and Grizzle's paper [2]. Thus although the unknown parameters in this case remain unknown, the BAN estimators and test results obtained could be compared with the results of Allen and Grizzle for the complete (i.e., without missing observations) data.

The numerical calculations for all the examples were made with the assistance of a GE-265 Time-Sharing Call-a-Computer (CAC). It is important to note that the CAC system does not allow calculations to be done in double precision. Thus it is possible that some of the results obtained were not as accurate as could be achieved from a computer such as the 360 which does allow double precision.

5.2 Example 1: An HIM Model[1]

The model used for Example 1 is given by

$$E(Y_j) = A\xi B_j$$

$$\mathrm{Var}(Y_j) = I_{15} \otimes B_j'\Sigma B_j \ , \quad j=1,2,3$$

where $A(15 \times 2) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -7 & -6 & -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{bmatrix}'$

and $B_1(3 \times 3) = I_3$, $B_2(3 \times 2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$, $B_3(3 \times 1) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

The parameters ξ and Σ used in the simulation are

---

[1] This model, except for different B_j's, was used by Trawinski in her thesis [63].

$$\xi(2 \times 3) = \begin{bmatrix} 10 & 15 & 20 \\ 3 & 2 & 1 \end{bmatrix} \quad \text{and} \quad \Sigma(3 \times 3) = \begin{bmatrix} 14 & 6 & -4 \\ 6 & 3 & 0 \\ -4 & 0 & 32 \end{bmatrix} .$$

The (15×3) data matrices Y_1, Y_2 and Y_3 were obtained by deleting the appropriate observations from the complete data matrix

$$Y^*(45 \times 3) = \begin{bmatrix} Y_1^* \\ Y_2^* \\ Y_3^* \end{bmatrix}$$

where $Y_j^* = A\xi + E_j$, $j=1,2,3$ and the rows of E_j are distributed as $N_3(\underline{0}, \Sigma)$. Four sample data sets were generated.

The estimators of Σ as given in Corollary 2.4.7 of Chapter Two are given in Table 5.1. Also given in Table 5.1 are the estimators of Σ obtained from the complete data.

The BAN estimators $\hat{\underline{\xi}}$ as given in Corollary 3.5.1.1 of Chapter Three are given in Table 5.2 together with the corresponding estimators for the complete data (which are BLUE as well as BAN). Also provided is $(\hat{\underline{\xi}} - \underline{\xi})'(\hat{\underline{\xi}} - \underline{\xi})$, the squared distance of $\hat{\underline{\xi}}$ from $\underline{\xi}$.

Four tests of hypotheses of the form $H'\underline{\xi} = 0$ were performed for each of the four samples using the Wald Statistic described in Chapter Four. The hypotheses tested are as follows:

H 1-1) $\quad H'(2 \times 6) = \begin{bmatrix} 3 & 0 & -2 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & -3 & 0 \end{bmatrix}$

or $\quad H' = D' \otimes C$

where $\quad C\ (1 \times 2) = (1, 0)$ and $D(3 \times 2) = \begin{bmatrix} 3 & -2 & 0 \\ 4 & 0 & -3 \end{bmatrix}'$.

HIM Data

$$\hat{\Sigma}(I) = \begin{bmatrix} 10.868 & 3.917 & -6.738 \\ 3.917 & 1.784 & -2.348 \\ -6.738 & -2.348 & 30.120 \end{bmatrix}$$

$$\hat{\Sigma}(II) = \begin{bmatrix} 15.035 & 7.481 & -11.319 \\ 7.481 & 3.360 & -3.729 \\ -11.309 & -3.729 & 34.213 \end{bmatrix}$$

$$\hat{\Sigma}(III) = \begin{bmatrix} 10.516 & 3.813 & 0.588 \\ 3.813 & 1.955 & 1.483 \\ 0.588 & 1.483 & 21.484 \end{bmatrix}$$

$$\hat{\Sigma}(IV) = \begin{bmatrix} 7.391 & 3.696 & 2.041 \\ 3.696 & 2.281 & 3.043 \\ 2.041 & 3.043 & 22.481 \end{bmatrix}$$

Complete Data

$$\tilde{\Sigma}(I) = \begin{bmatrix} 10.915 & 4.176 & -7.798 \\ 4.176 & 2.112 & -1.099 \\ -7.798 & -1.099 & 39.336 \end{bmatrix}$$

$$\tilde{\Sigma}(II) = \begin{bmatrix} 14.962 & 6.002 & -2.873 \\ 6.002 & 2.711 & 0.261 \\ -2.873 & 0.261 & 29.790 \end{bmatrix}$$

$$\tilde{\Sigma}(III) = \begin{bmatrix} 10.454 & 4.801 & -0.889 \\ 4.801 & 2.655 & 1.398 \\ -0.889 & 1.398 & 26.398 \end{bmatrix}$$

$$\tilde{\Sigma}(IV) = \begin{bmatrix} 7.518 & 3.308 & -0.326 \\ 3.308 & 1.930 & 1.760 \\ -0.326 & 1.760 & 24.665 \end{bmatrix}$$

TABLE 5.1

ESTIMATES OF $\Sigma = \begin{bmatrix} 14 & 6 & -4 \\ 6 & 3 & 0 \\ -4 & 0 & 32 \end{bmatrix}$

FOR SAMPLES I-IV OF EXAMPLE 1

Since $H'\underline{\xi} = (0, -20)' \neq (0, 0)'$, the test statistic should <u>reject</u> this hypothesis.

H 1-2) $$H'(2 \times 6) = \begin{bmatrix} 3 & 0 & -2 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & -2 & 0 \end{bmatrix}$$

or $$H' = D' \otimes C$$

$\underline{\xi}$	$\hat{\underline{\xi}}$(I)	$\tilde{\underline{\xi}}$(I)	$\hat{\underline{\xi}}$(II)	$\tilde{\underline{\xi}}$(II)	$\hat{\underline{\xi}}$(III)	$\tilde{\underline{\xi}}$(III)	$\hat{\underline{\xi}}$(IV)	$\tilde{\underline{\xi}}$(IV)
10	10.122	10.12	10.080	10.08	9.250	9.25	9.885	9.885
3	3.000	2.999	3.086	3.086	2.807	2.807	2.871	2.870
15	14.928	14.99	14.915	15.09	14.607	14.78	14.741	14.85
2	1.997	1.986	2.048	2.036	1.945	1.934	1.912	1.909
20	19.404	18.84	20.447	19.13	18.512	19.38	16.688	18.87
1	0.736	0.977	1.290	1.105	1.195	1.110	0.899	0.780
D^2	0.47	1.45	0.26	0.02	2.59	0.55	11.02	1.29

TABLE 5.2[2]
ESTIMATES OF $\underline{\xi}$ FOR SAMPLES I-IV OF EXAMPLE 1

where $\quad C\,(1 \times 2) = (1,\ 0)$ and $D(3 \times 2) = \begin{bmatrix} 3 & -2 & 0 \\ 4 & 0 & -2 \end{bmatrix}'$.

Since $H'\underline{\xi} = (0,\ 0)'$, the test statistic should <u>not reject</u> the hypothesis.

H 1-3) $\quad H'(4 \times 6) = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}$.

or $\quad H' = D' \otimes C$

where $\quad C\,(2 \times 2) = I_2$ and $D(3 \times 2) = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}'$.

---

[2] $\hat{\underline{\xi}}$ denotes the BAN estimate for the HIM data and $\tilde{\underline{\xi}}$ denotes the corresponding estimate for the complete data. Also, $D^2 = (\hat{\underline{\xi}}-\underline{\xi})'(\hat{\underline{\xi}}-\underline{\xi})$, where $\hat{\underline{\xi}}$ is replaced by $\tilde{\underline{\xi}}$ for the complete data.

This is equivalent to testing $\underline{\xi}_1 = \underline{\xi}_2 = \underline{\xi}_3$ where $\underline{\xi}_s$ is the s-th column of ξ. The test statistic should <u>reject</u> the hypothesis.

H 1-4) $$H'(4 \times 6) = \begin{bmatrix} 3 & 0 & -2 & 0 & 0 & 0 \\ 0 & 2 & 0 & -3 & 0 & 0 \\ 4 & 0 & 0 & 0 & -2 & 0 \\ 0 & 2 & 0 & 0 & 0 & -6 \end{bmatrix}.$$

This hypothesis matrix cannot be expressed as $H' = D' \otimes C$. Since $H'\underline{\xi} = (0, 0, 0, 0)'$, the test statistic should <u>not reject</u> the hypothesis.

The Wald Statistics obtained for each sample and each hypothesis are presented in Table 5.3. The values obtained were compared with percentage points of a chi-square random variable with the appropriate degrees of freedom and were deemed significant if they exceeded the 95th percentage point. Incorrect test results are circled. For the complete data sets, test statistics were also calculated.[3,4]

Tables 5.2 and 5.3 show that the estimators $\hat{\underline{\xi}}$ and test statistics compare well with the true $\underline{\xi}$ and the truth or falsity of the hypotheses, respectively. Furthermore, these results compare favorably with those for the complete data. Those cases in which the HIM data estimate is better than the complete data estimate may be attributed to sampling variability. The estimators of Σ in Table 5.1 for

[3]This was done using program MVGLM*, a library CAC program which could not handle hypothesis H 1-4 since it was not of the form $C\,\xi D=0$. Furthermore, only likelihood ratio statistics were provided for H 1-1 and H 1-2 . For H 1-3 both likelihood ratio and T_0^2 statistics were given. Thus the test results for the complete model could not be compared numerically with those of the HIM model except for H 1-3 , and they are therefore described only with regard to whether the hypothesis was rejected.

[4]For each of the complete data samples, hypotheses H 1-1 and H 1-3 were rejected with .01 significance and hypothesis H 1-2 was not rejected.

Sample / Hypothesis	I	II	III	IV	D.F.
H 1-1	21.59**	33.02**	23.97**	13.11**	2
H 1-2	0.26	0.66	2.94	9.00*	2
H 1-3	359.62**	902.71**	374.57**	754.41**	4
H 1-4	0.93	0.10	5.72	9.06	4

TABLE 5.3[5]
WALD STATISTICS FOR EXAMPLE 1

the HIM data are not very good, but since the estimators for the complete data are not accurate either, it is possible that these results can be attributed to sampling variability and too small a sample size. The results also indicate that the estimators $\hat{\underline{\xi}}$ are fairly accurate despite small differences in $\hat{\Sigma}$, although it should be remembered that only four simulations were made. Clearly $\hat{\Sigma}(IV)$ was the worst of the estimators of Σ and was considerably poorer than $\hat{\Sigma}(I)$ and $\hat{\Sigma}(II)$. Yet $\hat{\underline{\xi}}(IV)$ was inaccurate only in its estimate of 16.688 for 20.

5.3 Example 2: An MDM Model[6] with p=2

The model used for Example 2 is given by

$$E(Y) = [A_1\underline{\xi}_1 \;\vdots\; A_2\underline{\xi}_2]$$

$$Var(Y) = I_{30} \otimes \Sigma$$

---

[5] * denotes .05 significance but not .01 significance. ** denotes .01 significance. A circle denotes a test result that leads to an incorrect conclusion about the given hypothesis. D.F. denotes degrees of freedom.

[6] This model was used by Kmenta and Gilbert in their study [33].

where

$$A_1(30 \times 3) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 3 & 1 \\ 1 & 3 & 1 \\ 1 & 3 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 9 & 1 \\ 1 & 9 & 1 \\ 1 & 9 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 6 & 1 \\ 1 & 6 & 1 \\ 1 & 6 & 1 \\ 1 & 3 & 0 \\ 1 & 3 & 0 \\ 1 & 3 & 0 \\ 1 & 8 & 1 \\ 1 & 8 & 1 \\ 1 & 8 & 1 \\ 1 & 6 & 0 \\ 1 & 6 & 0 \\ 1 & 6 & 0 \\ 1 & 7 & 0 \\ 1 & 7 & 0 \\ 1 & 7 & 0 \end{bmatrix} \quad A_2(30 \times 3) = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 10 & 0 \\ 1 & 10 & 0 \\ 1 & 10 & 0 \\ 1 & 8 & 0 \\ 1 & 8 & 0 \\ 1 & 8 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 16 & 0 \\ 1 & 16 & 0 \\ 1 & 16 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 4 & 1 \\ 1 & 4 & 1 \\ 1 & 4 & 1 \\ 1 & 10 & 1 \\ 1 & 10 & 1 \\ 1 & 10 & 1 \\ 1 & 14 & 1 \\ 1 & 14 & 1 \\ 1 & 14 & 1 \end{bmatrix}$$

and the rows of $Y(30 \times 2)$ are each normally distributed.

The parameter $\underset{\sim}{\xi} = (\underset{\sim}{\xi}_1' \ \underset{\sim}{\xi}_2')'$ is given by

$$\underset{\sim}{\xi}_1(3 \times 1) = \begin{bmatrix} 10 \\ 2 \\ -5 \end{bmatrix} \text{ and } \underset{\sim}{\xi}_2(3 \times 1) = \begin{bmatrix} -10 \\ 6 \\ 3 \end{bmatrix} .$$

Five sample data sets were simulated. Samples I-III used

$$\Sigma = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} ,$$

Sample IV used

$$\Sigma = \begin{bmatrix} 5 & 5 \\ 5 & 10 \end{bmatrix}$$

and Sample V used

$$\Sigma = \begin{bmatrix} 20 & -18 \\ -18 & 29 \end{bmatrix}.$$

The estimators $\hat{\Sigma}$ of Σ as given in Corollary 2.4.5 of Chapter Two are presented in Table 5.4 and the BAN estimators $\hat{\underline{\xi}}$ of $\underline{\xi}$ as given by (3.4.8) of Chapter Three are presented in Table 5.5.

Four tests of hypotheses of the form $H'\underline{\xi} = 0$ were performed for each of the five samples using the Wald Statistic described in Chapter Four. The hypotheses tested are as follows:

H 1-1) $$H'(3 \times 6) = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}.$$

This hypothesis is equivalent to testing $\underline{\xi}_1 = \underline{\xi}_2$. The test statistic should reject the hypothesis.

Sample	Σ	$\hat{\Sigma}$
I	$\begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.708 & 0.474 \\ 0.474 & 0.851 \end{bmatrix}$
II	$\begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.853 & 0.368 \\ 0.368 & 0.599 \end{bmatrix}$
III	$\begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.105 & 0.370 \\ 0.370 & 0.873 \end{bmatrix}$
IV	$\begin{bmatrix} 5 & 5 \\ 5 & 10 \end{bmatrix}$	$\begin{bmatrix} 3.434 & 4.982 \\ 4.982 & 10.827 \end{bmatrix}$
V	$\begin{bmatrix} 20 & -18 \\ -18 & 29 \end{bmatrix}$	$\begin{bmatrix} 21.834 & -21.799 \\ -21.799 & 29.107 \end{bmatrix}$

TABLE 5.4
ESTIMATES OF Σ FOR EXAMPLE 2

$\underline{\xi}$	$\hat{\underline{\xi}}$(I)	$\hat{\underline{\xi}}$(II)	$\hat{\underline{\xi}}$(III)	$\hat{\underline{\xi}}$(IV)	$\hat{\underline{\xi}}$(V)
10	10.528	10.328	10.053	10.466	8.371
2	1.956	1.944	1.930	1.885	1.909
-5	-5.276	-5.164	-4.794	-4.637	-3.847
-10	-10.298	-9.701	-10.393	-9.430	-10.133
6	6.036	5.996	5.995	5.993	6.062
3	3.360	2.721	3.444	2.399	5.108
D^2	0.59	0.32	0.37	0.94	4.03

TABLE 5.5
BAN ESTIMATES OF $\underline{\xi}$ FOR EXAMPLE 2

H 2-2) $$H'(2 \times 6) = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 & -1 & 0 \end{bmatrix}.$$

Since $H'\underline{\xi} = (0, 0)'$ for this hypothesis, the test statistic should <u>not reject</u> the hypothesis.

H 2-3) $$H'(2 \times 6) = \begin{bmatrix} 0 & 5 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 \end{bmatrix}.$$

Here again $H'\underline{\xi} = (0, 0)'$ so the test statistic should <u>not reject</u> this hypothesis.

H 2-4) $$H'(2 \times 6) = \begin{bmatrix} 0 & 5 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.2 & -2 \end{bmatrix}.$$

Since for this hypothesis $H'\underline{\xi} = (0, 1.2)' \neq (0, 0)'$, the test statistic should <u>reject</u> the hypothesis.

The Wald Statistics obtained for all samples and for all hypotheses are presented in Table 5.6.

Examination of Table 5.5 shows that $\hat{\underline{\xi}}$ was very good for

Sample / Hypothesis	I	II	III	IV	V	D.F.
H 2-1	8433.98**	6672.53**	4429.45**	2114.31**	507.68**	3
H 2-2	1.63	2.35	2.77	2.63	1.64	2
H 2-3	2.63	2.64	1.75	0.48	2.62	2
H 2-4	3.92	13.35**	0.20	2.03	1.51	2

TABLE 5.6
WALD STATISTICS FOR EXAMPLE 2

Samples I-III, a little worse for Sample IV (where $|\Sigma| = 25$), and not very good for Sample V (where $|\Sigma| = 256$). These results may indicate that the estimates $\hat{\underline{\xi}}$ will decrease in accuracy as $|\Sigma|$ increases.

The results given in Table 5.6 provide good test performance except for the fourth hypothesis, for which four of the five samples did not give significance. It is not obvious why the results for the fourth hypothesis were poor since, when comparing $H'\hat{\underline{\xi}}$ for this hypothesis in all five samples, $H'\hat{\underline{\xi}}(II)$ does not have greatest distance from the origin. In fact $H'\hat{\underline{\xi}}\ (IV)$ and $H'\hat{\underline{\xi}}\ (V)$ have larger distances from the origin and yet the test statistic was not significant for these samples. The estimate of the information matrix clearly plays an important role here also.

5.3 Example 3: An MDM Model[7] with p=4

The model used for Example 3 is defined by

$$E(Y) = [A_1\underline{\xi}_1 \mid A_2\underline{\xi}_2 \mid A_3\underline{\xi}_3 \mid A_4\underline{\xi}_4]$$

$$Var(Y) = I_{20} \otimes \Sigma$$

[7] This model was also used by Kmenta and Gilbert [33].

where

$$A_1(20 \times 3) = \begin{bmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 1&1&3&3&0&0&9&9&1&1&6&6&3&3&8&8&6&6&7&7 \\ 0&0&1&1&1&1&1&1&0&0&1&1&0&0&1&1&0&0&0&0 \end{bmatrix}'$$

$$A_2(20 \times 3) = \begin{bmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 0&0&10&10&8&8&0&0&16&16&0&0&0&0&4&4&10&10&14&14 \\ 1&1&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1 \end{bmatrix}'$$

$$A_3(20 \times 3) = A_1$$

$$A_4(20 \times 3) = \begin{bmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 2&2&6&6&0&0&18&18&2&2&12&12&6&6&16&16&12&12&7&7 \\ 0&0&1&1&1&1&1&1&0&0&1&1&0&0&1&1&0&0&0&0 \end{bmatrix}'$$

and the rows of $\underset{\sim}{Y}(20 \times 4)$ are each normally distributed.

The parameter $\underline{\xi} = (\underline{\xi}_1'\underline{\xi}_2'\underline{\xi}_3'\underline{\xi}_4')'$ used in the simulation is given by

$$\underline{\xi}_1(3 \times 1) = \underline{\xi}_3(3 \times 1) = \begin{bmatrix} 10 \\ 2 \\ -5 \end{bmatrix} \text{ and } \underline{\xi}_2(3 \times 1) = \underline{\xi}_4(3 \times 1) = \begin{bmatrix} -10 \\ 6 \\ 3 \end{bmatrix}.$$

Five data sets were simulated. Samples I-III used

$$\Sigma = \begin{bmatrix} 1. & 0.64 & 0.64 & 0.64 \\ 0.64 & 1. & 0.64 & 0.64 \\ 0.64 & 0.64 & 1. & 0.64 \\ 0.64 & 0.64 & 0.64 & 1. \end{bmatrix}$$

and Samples IV and V used

$$\Sigma = \begin{bmatrix} 3 & 3 & -4 & 3 \\ 3 & 6 & -3 & 6 \\ -4 & -3 & 14 & -14 \\ 3 & 6 & -14 & 21 \end{bmatrix}.$$

The estimators $\hat{\Sigma}$ for Samples I-III are presented in Table 5.7

$$\hat{\Sigma}(I) = \begin{bmatrix} 0.534 & 0.416 & 0.381 & 0.236 \\ 0.416 & 1.300 & 0.449 & 0.361 \\ 0.381 & 0.449 & 0.810 & 0.120 \\ 0.236 & 0.361 & 0.120 & 0.610 \end{bmatrix}$$

$$\hat{\Sigma}(II) = \begin{bmatrix} 1.164 & 0.747 & 0.601 & 0.738 \\ 0.747 & 1.076 & 0.476 & 0.577 \\ 0.601 & 0.476 & 0.654 & 0.537 \\ 0.738 & 0.577 & 0.537 & 0.853 \end{bmatrix}$$

$$\hat{\Sigma}(III) = \begin{bmatrix} 0.570 & 0.414 & 0.490 & 0.366 \\ 0.414 & 0.713 & 0.406 & 0.386 \\ 0.490 & 0.406 & 0.908 & 0.416 \\ 0.366 & 0.386 & 0.416 & 0.610 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1. & 0.64 & 0.64 & 0.64 \\ 0.64 & 1. & 0.64 & 0.64 \\ 0.64 & 0.64 & 1. & 0.64 \\ 0.64 & 0.64 & 0.64 & 1. \end{bmatrix}$$

TABLE 5.7
ESTIMATES OF Σ FOR SAMPLES I-III OF EXAMPLE 3

and the estimators $\hat{\underline{\xi}}$, in Table 5.8.

Nine tests of hypotheses of the form $H'\underline{\xi} = 0$ were performed for each sample (including Samples IV and V, which will be dealt with later) using the Wald Statistic. The hypotheses tested are as follows:

H 3-1) $$H'(3 \times 12) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}.$$

$\underline{\xi}$	$\hat{\underline{\xi}}$(I)	$\hat{\underline{\xi}}$(II)	$\hat{\underline{\xi}}$(III)
10	9.692	9.823	9.615
2	2.160	2.000	1.967
-5	-5.723	-4.784	-4.346
-10	-10.700	-9.695	-9.907
6	6.028	5.975	6.016
3	3.367	2.509	2.669
10	10.325	9.951	10.152
2	1.989	1.948	1.944
-5	-5.863	-4.612	-4.563
-10	-9.800	-10.762	-9.912
6	6.040	5.992	5.963
3	1.865	4.253	3.479
D^2	3.02	2.92	1.22

TABLE 5.8
BAN ESTIMATES OF $\underline{\xi}$ FOR SAMPLES I-III OF EXAMPLE 3

This hypothesis is equivalent to testing $\underline{\xi}_1 = \underline{\xi}_3$. The test statistic should not reject the hypothesis.

H 3-2) $$H'(3 \times 12) = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

This is equivalent to testing $\underline{\xi}_2 = \underline{\xi}_4$. The test statistic should not reject the hypothesis.

H 3-3) $$H'(6 \times 12) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

This is equivalent to testing $\underline{\xi}_1 = \underline{\xi}_3$ and $\underline{\xi}_2 = \underline{\xi}_4$. The test statistic should not reject the hypothesis.

H 3-4) $$H'(9 \times 12) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This is equivalent to testing $\underline{\xi}_1 = \underline{\xi}_2 = \underline{\xi}_3 = \underline{\xi}_4$. The test statistic should reject the hypothesis.

H 3-5) $$H'(3 \times 12) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Here $H'\underline{\xi} = (0, 0, 0)'$ so that the test statistic should not reject the hypothesis.

H 3-6) $$H'(2 \times 12) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Here $H'\underline{\xi} = (-1,0)'$ so that the test statistic should reject the hypothesis.

H 3-7) $$H'(2 \times 12) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Here $H'\underline{\xi} = (0, -1)'$ so that the test statistic should __reject__ the hypothesis.

H 3-8) $$H'(2 \times 12) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Since $H'\underline{\xi} = (0, 0)'$, the test statistic should __not reject__ the hypothesis.

H 3-9) $$H'(4 \times 12) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Here $H'\underline{\xi} = (0, 0, 0, 2)'$ so that the test statistic should __reject__ the hypothesis.

The Wald Statistics obtained for Samples I-III are presented in Table 5.9.

The estimators of $\underline{\xi}$ given in Table 5.8 are fairly accurate with the noticable exception of the estimate of the last parameter for $\hat{\underline{\xi}}(I)$ and $\hat{\underline{\xi}}(II)$. $\hat{\underline{\xi}}(III)$ seems to be superior to the other two estimates. The accuracy of the test results of Table 5.9 appears to be correlated with the squared distance of $\hat{\underline{\xi}}$ from $\underline{\xi}$ given in Table 5.8 since the number of circles in the latter table decreases as the distance increases. The results for Sample I were particularly poor as four of the nine test statistics gave wrong answers.

Estimators of Σ for Samples IV and V are presented in Table 5.10 and estimators of $\underline{\xi}$ are presented in Table 5.11. Upon taking determinants it was shown that the estimators of Σ based on Corollary 2.4.5

Sample / Hypothesis	I	II	III	D.F.
H 3-1	11.23*	0.89	5.12	3
H 3-2	5.84	7.28	4.21	3
H 3-3	16.96**	9.32	9.38	6
H 3-4	9369.06**	17208.71**	13905.51**	9
H 3-5	9.45*	7.38	6.92	3
H 3-6	6.46*	41.84**	34.58**	2
H 3-7	10.63**	1.93	8.16*	2
H 3-8	9.31**	7.25*	5.65	2
H 3-9	2077.67**	1721.18**	19781.61**	4

TABLE 5.9
WALD STATISTICS FOR SAMPLES I-III OF EXAMPLE 3

of Chapter Two were not positive definite (since the determinants were negative). Although this does not seem to noticeably affect the accuracy of the BAN estimator $\hat{\underline{\xi}}$, it is clear that lack of positive definiteness could (and did) result in negative Wald Statistics. Thus for the purpose of testing hypotheses it is necessary to adjust the estimator $\hat{\Sigma}$ to be positive definite. Noting that the estimator is given by $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$ where

$$\hat{\sigma}_{rs} = \frac{1}{\mu_{rs}} \underline{y}_s'[I_n - A_s(A_s'A_s)^{-1} A_s'][I_n - A_r(A_r'A_r)^{-1} A_r']\underline{y}_r \ , \ r,s=1,\ldots,p$$

$$\mu_{rs} = tr[I_n - A_s(A_s'A_s)^{-1} A_s'][I_n - A_r(A_r'A_r)^{-1} A_r']$$

and $\underline{y}_s$ is the s-th column of the data matrix Y , it is clear that replacing μ_{rs} for each (r,s) pair by some constant would yield an

Sample IV | Sample V

Σ

$$\begin{bmatrix} 3 & 3 & -4 & 3 \\ 3 & 6 & -3 & 6 \\ -4 & -3 & 14 & -14 \\ 3 & 6 & -14 & 21 \end{bmatrix} \qquad \begin{bmatrix} 3 & 3 & -4 & 3 \\ 3 & 6 & -3 & 6 \\ -4 & -3 & 14 & -14 \\ 3 & 6 & -14 & 21 \end{bmatrix}$$

$$|\Sigma| = 36 \qquad |\Sigma| = 36$$

$\hat{\Sigma}$

$$\begin{bmatrix} 3.030 & 1.651 & -3.190 & 0.256 \\ 1.651 & 4.806 & 1.548 & 1.331 \\ -3.190 & 1.548 & 13.054 & -7.789 \\ 0.256 & 1.331 & -7.789 & 10.549 \end{bmatrix} \qquad \begin{bmatrix} 1.372 & 1.735 & -3.142 & 3.175 \\ 1.735 & 4.733 & -2.214 & 6.372 \\ -3.142 & -2.214 & 16.088 & -15.327 \\ 3.175 & 6.372 & -15.327 & 21.727 \end{bmatrix}$$

$$|\hat{\Sigma}| = -25.06 \qquad |\hat{\Sigma}| = -1.25$$

$\tilde{\Sigma}$

$$\begin{bmatrix} 2.576 & 1.285 & -2.712 & 0.216 \\ 1.285 & 4.085 & 1.204 & 1.036 \\ -2.712 & 1.204 & 11.096 & -6.576 \\ 0.216 & 1.036 & -6.576 & 8.967 \end{bmatrix} \qquad \begin{bmatrix} 1.166 & 1.350 & -2.671 & 2.677 \\ 1.350 & 4.023 & -1.723 & 4.962 \\ -2.671 & -1.723 & 13.675 & -12.941 \\ 2.677 & 4.962 & -12.941 & 18.468 \end{bmatrix}$$

$$|\tilde{\Sigma}| = 50.23 \qquad |\tilde{\Sigma}| = 0.54$$

Σ^*

$$\begin{bmatrix} 1.372 & 1.589 & -3.142 & 3.150 \\ 1.589 & 4.733 & -2.027 & 5.838 \\ -3.142 & -2.027 & 16.088 & -15.225 \\ 3.150 & 5.838 & -15.225 & 21.727 \end{bmatrix}$$

$$|\Sigma^*| = 0.46$$

TABLE 5.10[8]

ESTIMATES OF Σ FOR SAMPLES IV AND V OF EXAMPLE 3

---

[8] In this table, $\hat{\Sigma}$ is the estimator obtained from Corollary 2.4.5, $\tilde{\Sigma}$ is the adjusted estimator which uses n = 20 instead of μ_{rs}, and Σ^* is the adjusted estimator which uses n - 3 = 17 instead of μ_{rs}.

$\underset{\sim}{\xi}$	$\hat{\underset{\sim}{\xi}}$(IV)	$\tilde{\underset{\sim}{\xi}}$(IV)	$\hat{\underset{\sim}{\xi}}$(V)	$\tilde{\underset{\sim}{\xi}}$(V)	$\underset{\sim}{\xi}^*$(V)
10	10.849	10.876	10.191	10.179	10.179
2	1.954	1.943	2.083	2.082	2.082
-5	-5.092	-5.042	-5.342	-5.310	-5.310
-10	-9.872	-9.729	-9.641	-9.585	-9.585
6	6.040	6.039	6.026	6.030	6.030
3	3.296	3.060	2.634	2.505	2.505
10	10.048	10.030	8.748	8.978	8.978
2	2.091	2.092	2.010	1.952	1.952
-5	-6.032	-6.010	-3.787	-3.733	-3.733
-10	-10.661	-10.593	-8.314	-8.538	-8.538
6	5.938	5.924	5.905	5.934	5.934
3	4.549	4.638	1.784	1.762	1.762
D^2	4.51	4.86	7.93	6.93	6.93

TABLE 5.11[9]
BAN ESTIMATES OF $\underset{\sim}{\xi}$ FOR SAMPLES IV AND V OF EXAMPLE 3

estimate of Σ which was positive definite (since the estimate could then be written as XX' for some matrix X). To keep the adjusted estimate consistent (although it would clearly be biased) the obvious replacement for μ_{rs} would be n (in this case 20). Another possibility would be $n-\ell$ where ℓ was some positive constant between 0 and n . Table 5.10 contains for Samples IV and V adjusted estimators of Σ using n instead of μ_{rs} and for Sample V alone an adjusted estimator using

[9] $\hat{\underset{\sim}{\xi}}$(IV)and $\hat{\underset{\sim}{\xi}}$(V) are the BAN estimates using $\hat{\Sigma}$(IV) and $\hat{\Sigma}$(V), respectively. $\tilde{\underset{\sim}{\xi}}$(IV) and $\tilde{\underset{\sim}{\xi}}$(V) are the BAN estimates using $\tilde{\Sigma}$(IV) and $\tilde{\Sigma}$(V), respectively. $\underset{\sim}{\xi}^*$(V) is the BAN estimate using Σ^*(V).

n-3. (The value $\ell = 3$ was chosen because $\mu_{ss} = 17$ for each s.) The corresponding adjusted BAN estimators $\tilde{\underline{\xi}}$ and $\underline{\xi}^*$ were also calculated and presented in Table 5.11 with no noticeable change in accuracy from the original estimators $\hat{\underline{\xi}}$.

It should be noted here that neither Corollary 2.4.5 for the MDM model nor its parent Theorem 2.4.1 for the MGLM model claims that the estimator $\hat{\Sigma}$ will be positive definite. They do claim instead that this estimator will be positive definite in the limit (since Σ is assumed positive definite); that is, as the sample size n is increased, the estimator will eventually (in probability) be positive definite. Since the n chosen for Example 3 is 20, which is relatively small, it should not be surprising that positive definiteness was not obtained. Nevertheless, the problem remains in general to determine a method of adjustment of $\hat{\Sigma}$ whenever it does not turn out positive definite. Probably the most practical procedure is the use of scoring as described in Section 3.5.2 of Chapter Three to calculate succesive estimators of Σ until a positive definite $\hat{\Sigma}$ is obtained. Another possibility is the use of Hocking and Smith's procedure [28], although it remains to be shown that their estimator of Σ is positive definite.

Wald Statistics were calculated for Samples IV and V and hypotheses H 3-1 through H 3-9. These are presented in Table 5.12 for both the original and the first adjusted estimators.[10] Negative Wald Statistics were circled in addition to those leading to incorrect test conclusions.

[10] The second adjusted estimator was not used since $\tilde{\underline{\xi}}(V)$ and $\underline{\xi}^*(V)$ were equal.

Sample / Hypothesis	IV (original)	IV (adjusted)	V (original)	V (adjusted)	D.F.
H 3-1	0.77	0.85	1.47	1.72	3
H 3-2	1.28	3.63	-10.53	0.80	3
H 3-3	3.16	5.77	-6.22	5.54	6
H 3-4	-19090.99	7158.68**	-250830.9	7040.84**	9
H 3-5	1.59	1.93	1.61	0.58	3
H 3-6	-21.20	9.15*	-106.05	0.74	2
H 3-7	-22.56	5.40	-81.52	1.07	2
H 3-8	1.65	1.66	0.78	0.53	2
H 3-9	-882.75	1387.31**	-18130.69	2888.85**	4

TABLE 5.12
WALD STATISTICS FOR SAMPLES IV AND V OF EXAMPLE 3

5.5 Example 4: A GGCM Model

The model used in Example 4 was obtained by deleting some data from a growth curve (GCM) experiment described by Allen and Grizzle [2]. Their model is given by

$$E(Y) = A\xi P$$

$$\mathrm{Var}(Y) = I_{36} \otimes \Sigma$$

where

$$A(12 \times 4) = \begin{bmatrix} \underline{1}_9 & \underline{0}_9 & \underline{0}_9 & \underline{0}_9 \\ \underline{0}_{10} & \underline{1}_{10} & \underline{0}_{10} & \underline{0}_{10} \\ \underline{0}_8 & \underline{0}_8 & \underline{1}_8 & \underline{0}_8 \\ \underline{0}_9 & \underline{0}_9 & \underline{0}_9 & \underline{1}_9 \end{bmatrix}$$

$$P(4 \times 7) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ 5 & 0 & -3 & -4 & -3 & 0 & 5 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 \end{bmatrix}$$

$\xi(4 \times 4)$ is a matrix of unknown treatment parameters

and $\Sigma(7 \times 7)$ is the unknown covariance matrix.

The GGCM model formed from this data was obtained by deleting for a given individual the measurements on either the last three variates, the last variate or no variates. The model thus had a hierarchal pattern of missing observations and was given by

$$E(Y_j) = A_j \xi P B_j$$

$$\mathrm{Var}(Y_j) = I_{n_j} \otimes B_j' \Sigma B_j \ , \ j=1,2,3$$

where $n_1 = 14, \ n_2 = 12, \ n_3 = 10$

$$B_1(7 \times 7) = I_7 \ , \quad B_2(7 \times 6) = \begin{bmatrix} \underline{0}_6' \\ I_6 \end{bmatrix} \ , \quad B_3(7 \times 4) \quad \begin{bmatrix} 0_{3,4} \\ I_4 \end{bmatrix}$$

$$Y_1(14 \times 7) = \begin{bmatrix} 4.0 & 4.0 & 4.1 & 3.6 & 3.6 & 3.8 & 3.1 \\ 4.2 & 4.4 & 4.6 & 4.9 & 5.3 & 5.6 & 4.9 \\ 3.7 & 3.9 & 3.9 & 4.8 & 5.2 & 5.4 & 4.2 \\ 3.4 & 3.4 & 3.5 & 3.1 & 3.1 & 3.7 & 3.3 \\ 3.0 & 3.2 & 3.0 & 3.0 & 3.1 & 3.2 & 3.1 \\ 3.8 & 3.9 & 4.0 & 3.9 & 3.5 & 3.5 & 3.4 \\ 4.2 & 4.0 & 4.2 & 4.1 & 4.2 & 4.0 & 4.0 \\ 4.1 & 4.2 & 4.3 & 4.3 & 4.2 & 4.0 & 4.2 \\ 3.1 & 3.3 & 3.2 & 3.1 & 3.2 & 3.1 & 3.1 \\ 3.5 & 3.9 & 5.8 & 5.4 & 4.9 & 5.3 & 5.6 \\ 3.4 & 3.4 & 3.5 & 3.3 & 3.4 & 3.2 & 3.4 \\ 4.0 & 4.6 & 4.8 & 4.9 & 5.4 & 5.6 & 4.8 \\ 4.1 & 4.1 & 3.7 & 4.0 & 4.1 & 4.6 & 4.7 \\ 3.5 & 3.6 & 3.6 & 4.2 & 4.8 & 4.9 & 5.0 \end{bmatrix}$$

$$A_1(14 \times 4) = \begin{bmatrix} \underline{1}_3 & \underline{0}_3 & \underline{0}_3 & \underline{0}_3 \\ \underline{0}_5 & \underline{1}_5 & \underline{0}_5 & \underline{0}_5 \\ \underline{0}_2 & \underline{0}_2 & \underline{1}_2 & \underline{0}_2 \\ \underline{0}_2 & \underline{0}_2 & \underline{0}_2 & \underline{1}_2 \end{bmatrix}$$

$$Y_2(12 \times 6) = \begin{bmatrix} 4.3 & 4.7 & 4.7 & 4.8 & 5.0 & 5.2 \\ 4.4 & 5.3 & 5.6 & 5.9 & 5.9 & 5.3 \\ 4.6 & 4.4 & 4.6 & 5.4 & 5.9 & 5.6 \\ 3.1 & 3.2 & 3.0 & 3.3 & 3.0 & 3.0 \\ 3.3 & 3.3 & 3.4 & 3.6 & 3.1 & 3.1 \\ 4.4 & 4.3 & 4.5 & 4.3 & 4.4 & 4.4 \\ 3.3 & 3.8 & 3.8 & 4.4 & 4.2 & 4.0 \\ 3.4 & 3.5 & 4.6 & 4.9 & 5.2 & 4.4 \\ 4.0 & 4.4 & 4.2 & 4.6 & 4.8 & 5.4 \\ 3.5 & 3.5 & 3.2 & 3.0 & 3.0 & 3.2 \\ 3.2 & 3.6 & 3.7 & 3.7 & 4.2 & 4.4 \\ 3.9 & 4.5 & 4.7 & 3.9 & 3.8 & 3.7 \end{bmatrix}$$

$$A_2(12 \times 4) = \begin{bmatrix} \underline{1}_3 & \underline{0}_3 & \underline{0}_3 & \underline{0}_3 \\ \underline{0}_3 & \underline{1}_3 & \underline{0}_3 & \underline{0}_3 \\ \underline{0}_3 & \underline{0}_3 & \underline{1}_3 & \underline{0}_3 \\ \underline{0}_3 & \underline{0}_3 & \underline{0}_3 & \underline{1}_3 \end{bmatrix}$$

$$Y_3(10 \times 4) = \begin{bmatrix} 4.3 & 4.5 & 5.8 & 5.4 \\ 5.2 & 5.3 & 4.2 & 4.1 \\ 5.2 & 5.6 & 5.4 & 4.7 \\ 3.2 & 3.3 & 3.1 & 3.1 \\ 3.1 & 3.0 & 3.0 & 3.0 \\ 3.7 & 3.7 & 3.6 & 3.7 \\ 5.7 & 4.9 & 4.0 & 4.0 \\ 5.0 & 5.4 & 4.4 & 3.9 \\ 4.3 & 3.9 & 3.4 & 3.5 \\ 4.3 & 3.6 & 3.8 & 3.7 \end{bmatrix}$$

$$A_3(10 \times 4) = \begin{bmatrix} \underline{1}_3 & \underline{0}_3 & \underline{0}_3 & \underline{0}_3 \\ \underline{0}_2 & \underline{1}_2 & \underline{0}_2 & \underline{0}_2 \\ \underline{0}_3 & \underline{0}_3 & \underline{1}_3 & \underline{0}_3 \\ \underline{0}_2 & \underline{0}_2 & \underline{0}_2 & \underline{1}_2 \end{bmatrix} .$$

As defined in (3.6.1) of Chapter Three, a consistent unbiased estimator of Σ is given by $\hat{\Sigma} = ((\hat{\sigma}_{rs}))$ where

$$\hat{\sigma}_{rs} = \frac{1}{N_{rs} - R(D_{rs})} \underline{x}_{rs}[I - D_{rs}(D'_{rs}D_{rs})^{-1} D'_{rs}]\underline{x}_{sr} ,$$

$r,s=1,\ldots,7$.

In this case

$$((N_{rs})) = \begin{bmatrix} 14 & 14 & 14 & 14 & 14 & 14 & 14 \\ 14 & 26 & 26 & 26 & 26 & 26 & 26 \\ 14 & 26 & 26 & 26 & 26 & 26 & 26 \\ 14 & 26 & 26 & 36 & 36 & 36 & 36 \\ 14 & 26 & 26 & 36 & 36 & 36 & 36 \\ 14 & 26 & 26 & 36 & 36 & 36 & 36 \\ 14 & 26 & 26 & 36 & 36 & 36 & 36 \end{bmatrix}$$

$$D_{1s} = A_1 \quad \text{for } s=1,\ldots,7$$

$$D_{rs} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \quad \text{for } r=2,3 \text{ and } s=r,\ldots,7$$

and

$$D_{rs} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} \quad \text{for } r=4,5,6,7 \text{ and } s=r,\ldots,7.$$

Also, of course, $D_{rs} = D_{sr}$ for all r,s.

The estimator $\hat{\Sigma}$ calculated from the GGCM data is given by

$$\hat{\Sigma} = \begin{bmatrix} 0.1577 & 0.1545 & 0.2130 & 0.1993 & 0.1735 & 0.1698 & 0.1892 \\ 0.1545 & 0.1785 & 0.2078 & 0.2108 & 0.1966 & 0.2187 & 0.2300 \\ 0.2130 & 0.2078 & 0.4162 & 0.3674 & 0.2778 & 0.2983 & 0.3416 \\ 0.1993 & 0.2108 & 0.3674 & 0.4407 & 0.3690 & 0.2871 & 0.2583 \\ 0.1735 & 0.1966 & 0.2778 & 0.3690 & 0.4337 & 0.3734 & 0.3178 \\ 0.1698 & 0.2187 & 0.2983 & 0.2871 & 0.3734 & 0.5235 & 0.4607 \\ 0.1892 & 0.2300 & 0.3416 & 0.2583 & 0.3178 & 0.4607 & 0.5131 \end{bmatrix}$$

This compares fairly well with the estimator obtained by Allen and Grizzle [2] for the complete data, which is given by

$$\tilde{\Sigma} = \begin{bmatrix} 0.2261 & 0.1721 & 0.1724 & 0.2054 & 0.1705 & 0.1958 & 0.1817 \\ 0.1721 & 0.1696 & 0.1840 & 0.1919 & 0.1628 & 0.1700 & 0.1644 \\ 0.1724 & 0.1840 & 0.3917 & 0.3473 & 0.2370 & 0.1876 & 0.2194 \\ 0.2054 & 0.1919 & 0.3473 & 0.4407 & 0.3689 & 0.2870 & 0.2582 \\ 0.1705 & 0.1628 & 0.2370 & 0.3689 & 0.4337 & 0.3733 & 0.3178 \\ 0.1958 & 0.1700 & 0.1876 & 0.2870 & 0.3733 & 0.5235 & 0.4606 \\ 0.1817 & 0.1644 & 0.2194 & 0.2582 & 0.3178 & 0.4606 & 0.5131 \end{bmatrix}$$

Whereas it is clear that Σ must be positive definite by definition, it was found that $|\hat{\Sigma}| = -3.35 \times 10^{-9}$ which implies that $\hat{\Sigma}$ is not positive definite. Thus as in the previous example, use of $\hat{\Sigma}$ for calculating Wald Statistics might lead to negative values. For purposes of comparison, therefore, it was decided to use $\tilde{\Sigma}$ as an adjusted estimator.

BAN estimators of ξ were obtained using the formula given in (3.6.2) of Chapter Three. The estimators obtained using both $\hat{\Sigma}$ and $\tilde{\Sigma}$ are presented in Table 5.13 along with the two estimators obtained by Allen and Grizzle using the complete data. These estimators are given in the form $PP'\hat{\xi}$.

Six tests of hypotheses of the form $C\ \xi = 0$ (or $H'\underline{\xi} = 0$ where $H' = I_4 \otimes C$ and $\underline{\xi}$ is obtained from ξ in the usual way) were performed using the Wald test of hypotheses. The hypotheses used and the Wald Statistics obtained are described in Table 5.14. Negative values are circled. Since Allen and Grizzle did not present any test results in their paper, no comparisons were made. As described earlier, the negative values in the table are the result of $\hat{\Sigma}$'s not being positive definite.

5.6 Conclusions

It is not possible because of the limited scope of this computer study to make any definite statements with regard to the general

Source	$PP'\hat{\xi}$			
$\hat{\Sigma}$ (GGCM Data)	32.62	3.61	-6.40	-0.67
	24.90	-0.97	-1.01	0.09
	28.09	5.74	-5.97	0.01
	27.29	2.54	-0.70	-0.05
$\tilde{\Sigma}$ (GGCM Data)	33.00	3.69	-4.89	-0.98
	25.11	-0.83	-0.69	0.04
	28.36	5.00	-4.24	-0.31
	27.08	2.00	-1.69	0.07
Un-weighted (Complete Data)	32.57	4.48	-3.63	-0.98
	24.87	-0.56	-0.78	0.11
	29.05	3.49	-4.54	0.23
	27.58	1.72	-1.79	0.16
Weighted by $\tilde{\Sigma}$ (Complete Data)	32.70	4.14	-4.68	-1.14
	25.07	-0.69	-1.00	0.11
	28.59	4.10	-2.90	-0.48
	27.12	1.91	-1.55	0.04

TABLE 5.13
ESTIMATES OF $PP'\xi$ FOR EXAMPLE 4

reliability of the BAN estimators and the Wald test of linear hypotheses. Nevertheless, for all four examples the BAN estimators of $\underline{\xi}$ appeared to be fairly accurate and when the estimator of Σ was positive definite, the Wald test also seemed on the whole reliable. If the BAN estimator of $H'\underline{\xi}$ was poor, however, the test results were not likely

Hypothesis Number	Description[11]	C Matrix	Wald Statistic Using $\hat{\Sigma}$	Wald Statistic Using $\tilde{\Sigma}$	D.F.
H 4-1	$\underline{\xi}_1 = \underline{\xi}_4$	(1 0 0 -1)	-41.87	21.96**	4
H 4-2	$\underline{\xi}_1 = \underline{\xi}_3$	(1 0 -1 0)	36.94**	14.50**	4
H 4-3	$\underline{\xi}_1 = \underline{\xi}_2$	(1 -1 0 0)	23.19**	33.26**	4
H 4-4	$\underline{\xi}_2 = \underline{\xi}_3$	(0 1 -1 0)	42.62**	17.36**	4
H 4-5	$\underline{\xi}_2 = \underline{\xi}_4$	(0 1 0 -1)	-22.22	5.27	4
H 4-6	$\underline{\xi}_3 = \underline{\xi}_4$	(0 0 1 -1)	-52.12	5.65	4

TABLE 5.14
TEST OF HYPOTHESES USING WALD STATISTICS
FOR EXAMPLE 4

to be lead to correct conclusions, although this does not completely explain the reason for bad results (as was seen with hypothesis H 2-4 of Example 2). If the estimator of Σ was not positive definite, the BAN estimators of $\underline{\xi}$ still appeared accurate, but the reliability of tests using Wald Statistics was obviously spurious. It is clearly important therefore to plan one's experiment with a large enough n to insure that $\hat{\Sigma}$ is positive definite, but until further studies are made it is not clear what n to pick. Thus it is also important to determine a method for adjusting $\hat{\Sigma}$ to make it positive definite if in fact it is not. This can be easily done for the MDM model. In general, the most practical procedure would probably be the use of the method of <u>scoring</u> until a positive definite estimator is obtained.

---

[11] As usual, $\underline{\xi}_s$ is the s-th column of ξ.

CHAPTER VI

RECOMMENDATIONS FOR FURTHER RESEARCH

From the discussion in the preceding chapters, it is evident that a number of important problems requiring both theoretical and computational research still remain to be studied.

In Chapter Two we have obtained for the MGLM model necessary and sufficient conditions for the unknown variance matrix Σ to drop out of the BLUE estimate $H'(D\Omega^{-1}D)^{-}D'\Omega^{-1}\underline{x}$ for $H'\underline{\xi}$. Furthermore, it is easily seen when the conditions hold that $H'(D'D)^{-}D'\underline{x}$ is the MLE of $H'\underline{\xi}$ since the likelihood equations for $\underline{\xi}$ will then have a solution independent of the MLE for Σ. These conditions are restrictions on the design matrices that can be used in the experiment. A study which describes more precisely the type of designs that can be used will be extremely helpful to the experimenter in planning an experiment to which the MGLM model applies.

In Chapter Three we have obtained for the MGLM model a BAN estimator $H'\hat{\underline{\xi}}$ for $H'\underline{\xi}$ where $\hat{\underline{\xi}} = (D'\hat{\Omega}^{-1}D)^{-}D'\hat{\Omega}^{-1}\underline{x}$ and $\hat{\Omega}$ is an estimate of Ω using a consistent unbiased estimate $\hat{\Sigma}$ of Σ. We have also described an IZEF-type iterative procedure for obtaining additional BAN estimators. It is easily seen that at each iteration the likelihood equation for $\underline{\xi}$ is satisfied by $\hat{\underline{\xi}}$. A theoretical proof that the limiting estimator obtained by the iteration technique is indeed the MLE for $H'\underline{\xi}$ would corroborate the computational results of Kmenta and Gilbert [33] for the MDM model and would allow for direct computation of the MLE using the IZEF approach. Also, a study similar to that of Kmenta and Gilbert for the MGLM and GGCM models would be helpful in comparing different BAN estimators with regard to their small sample variances.

We have described in Chapter Three Hocking and Smith's approach [28] for obtaining asymptotically efficient estimators of ξ and Σ for the GIM model. At present because of notational and computational difficulties use of their results must be restricted to simply structured models and to small u (e.g., u=2 or 3). Although this author is somewhat skeptical as to the possibility of writing a general program for large u and arbitrary structure for the GIM model, it is neverthe- recommended here that an attempt be made to explore the possibility of such a program.

In Chapter Four the Wald Statistic has been proposed for testing linear hypotheses of the form $H'\underline{\xi} = 0$ for the MGLM and GGCM models. A computer study similar to that of Schatzoff [50] is recommended for comparing Wald Statistics using different BAN estimators with regard to small sample power considerations. In particular a comparison of the proposed estimators for the GCM and GGCM models would be of interest. In addition, a computer study similar to that of Kmenta and Gilbert [33] would be desirable for determining the rate of convergence to a chi-square variable of any Wald Statistic.

It has been shown in Chapter Four that the Wald Statistic for the MGLM model is a generalization of Hotelling's T_o^2. It would be desirable to obtain the exact distribution of the Wald Statistic which uses the BAN estimator or the proposed MLE given in Chapter Three, for we would then have an exact test for the MGLM model. The exact distribution appears to be quite complex, however.

It can be easily seen that when considering the GCM model, the Wald Statistic determined after covariance adjustment is the same as Khatri's trace described in Chapter One and at the end of Chapter Four. In particular, the BAN estimator of $C\xi D$ obtained from the conditional model is independent of the choice of error matrix F and consequently its conditional and unconditional distributions are the same. If it can be shown for the GGCM model that the conditional BAN estimator for $H'\underline{\xi}$ is independent of the choice of the F_j matrices defined in (4.5.3), then with some difficulty it can be established that the Wald Statistic determined after covariance adjustment is the same as the generalization of Khatri's trace given in (4.5.4).

Another approach to testing hypotheses for the MGLM model might possibly use the concept of sufficiency. It is therefore recommended that such an approach be explored and compared to the Wald Statistic method.

In Chapter Five it has been shown that the estimators $\hat{\Sigma}$ given in Theorem 2.4.1 and Lemma 3.6.1 may not be positive definite. It is therefore necessary to determine an alternative estimator to $\hat{\Sigma}$ which will always be positive definite and consistent for Σ. The most practical approach here appears to be the method of scoring as described in Section 3.5.2 of Chapter Three, but this needs to be demonstrated. Another possibility is the use of the method proposed by Hocking and Smith [28].

REFERENCES

[1] ALLEN, D. M. (1967). Multivariate analysis of nonlinear models. Institute of Statistics Mimeo Series No. 558. University of North Carolina. Chapel Hill, N. C.

[2] ALLEN, D. M. and GRIZZLE, J. E. (1968). Analysis of growth and dose response curves. Institute of Statistics Mimeo Series No. 576. University of North Carolina. Chapel Hill, N. C.

[3] AFIFI, A. and ELASHOFF, R. M. (1966). Missing observations in multivariate statistics I. Review of the literature. J. Amer. Statist. Assoc. 61 595-604.

[4] AFIFI, A. and ELASHOFF, R.M. (1967). Missing observations in multivariate statistics II. Point estimation in simple linear regression. J. Amer. Statist. Assoc. 62 10-29.

[5] AFIFI, A. and ELASHOFF, R. M. (1969). Missing observations in multivariate statistics III. Large sample analysis of a simple linear regression. J. Amer. Statist. Assoc. 64 337-358.

[6] AFIFI, A. and ELASHOFF, R. M. (1969). Missing observations in multivariate statistics IV. A note on a simple linear regression. J. Amer. Statist. Assoc. 64 359-365.

[7] ANDERSON, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. J. Amer. Statist. Assoc. 52 200-203.

[8] ANDERSON, T. W. (1958). An Introduction to Multivariate Statistical Analysis, Wiley and Sons, New York.

[9] BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. Ann. Math. Statist. 38 303-324.

[10] BARTEN, A. P. and KOERTS, J. (1962). Transition frequencies in voting behavior. Unpublished paper.

[11] BARTLETT, M. S. (1937). Some examples of statistical methods of research in agriculture. J. Roy Statist. Soc. Supplement 4 137-183.

[12] BHAPKAR, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. J. Amer. Statist. Assoc. 61 228-235.

[13] BHARGAVA, R. (1962). Multivariate tests of hypotheses with incomplete data. Applied Mathematics and Statistical Laboratories, Technical Report 3.

[14] BOX, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. Biometrika 36 317-346.

[15] BOX, G. E. P. (1950). Problems in the analysis of growth and wear curves. Biometrics 6 362-389.

[16] BUCK, S. F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. J. Roy Statist. Soc. Ser. B 22 302-307.

[17] CRAMER, H. (1946). Mathematical Methods of Statistics, Princeton University Press, Princeton, N. J.

[18] DAVIS, A. W. (1968). A system of linear differential equations for the distribution of Hotelling's generalized T_o^2. Ann. Math. Statist. 39 815-832.

[19] DEAR, R. E. (1959). A principal-component missing-data method for multiple regression models. System Development Corporation Report SP-86. Santa Monica, California.

[20] DWYER, P. S. (1967). Some applications of matrix derivatives in multivariate analysis. J. Amer. Statist. Assoc. 62 607-625.

[21] EDGETT, G. L. (1956). Multiple regression with missing observations among the independent variables. J. Amer. Statist. Assoc. 51 122-131.

[22] ELSTON, R. C. and GRIZZLE, J. E. (1962). Estimation of time response curves and their confidence bands. Biometrics 18 148-159.

[23] GLESER, L. and OLKIN, I. (1966). A K sample regression model with covariance. Multivariate Analysis (Edited by P. R. Krishnaiah.), Academic Press, New York.

[24] GOURLAY, N. (1955). F-test bias for experimental designs of the latin-square type. Psychometrika 20 273-287.

[25] GRIZZLE, J. E., STARMER, C. F. and KOCH, G. G. (1969). Analyses of categorical data by linear models. Biometrics 25 489-503.

[26] HECK, D. L. (1960). Charts of some upper percentage points of the distribution of the largest characteristic root. Ann. Math. Statist. 31 625-642.

[27] HOADLEY, A. B. (1968). Asymptotic properties of MLE's when the observations are independent but not iid with applications to estimation under variable censoring. Bell Telephone Laboratories Memorandum, Holmdel, N. J. Unpublished.

[28] HOCKING, R. R. and SMITH, W. B. (1968). Estimation of parameters in the multivariate normal distribution with missing observations. J. Amer. Statist. Assoc. 63 159-173.

[29] ITO, K. (1962). A comparison of the powers of two multivariate analysis of variance tests. Biometrika 49 455-462.

[30] KAKWANI, N. C. (1967). The unbiasedness of Zellner's seemingly unrelated regression equations estimators. J. Amer. Statist. Assoc. 62 141-142.

[31] KHATRI, C. G. (1966). A note on a MANOVA model applied to problems in growth curves. Ann. Inst. Statist. Math. 18 75-86.

[32] KLEINBAUM, D. G. (1969). A general method for obtaining test criteria for multivariate linear models with more than one design matrix and/or incomplete in response variates. Institute of Statistics Mimeo Series No. 614. University of North Carolina. Chapel Hill, N. C.

[33] KMENTA, J. and GILBERT, R. F. (1967). Small sample properties of alternative estimators of seemingly unrelated regressions. J. Amer. Statist. Assoc. 62 1180-1200.

[34] LORD, F. M. (1955). Estimation of parameters from incomplete data. J. Amer. Statist. Assoc. 50 870-876.

[35] MORRISON, D. F. (1967). Multivariate Statistical Methods, McGraw-Hill, Inc., New York.

[36] NICHOLSON, G. E. (1957). Estimation of parameters from incomplete multivariate samples. J. Amer. Statist. Assoc. 52 423-426.

[37] PILLAI, K. C. S. (1960). Statistical Tables for Tests of Multivariate Hypotheses. The Statistical Center, University of

the Philippines. Manila.

[38] PILLAI, K. C. S. and GUPTA, A. K. (1969). On the exact distribution of Wilks' criterion. Biometrika 56 109-118.

[39] POTTHOFF, R. F. and ROY, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika 51 122-127.

[40] RAO, C. R. (1959). Some problems involving linear hypotheses in multivariate analysis. Biometrika 46 49-58.

[41] RAO, C. R. (1965). Linear Statistical Inference and Its Applications, Wiley and Sons, New York.

[42] RAO, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to growth curves. Biometrika 52 447-458.

[43] RAO, C. R. (1966). Covariance adjustment and related problems in multivariate analysis. Multivariate Analysis (Edited by P. R. Krishnaiah). Academic Press, New York, 87-103.

[44] RAO, C. R. (1966). Generalized inverses for matrices and its applications in mathematical statistics. Research Papers in Statistics (Edited by F. N. David). Wiley, London, 263-269.

[45] RAO, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. Proceedings of the Fifth Berkeley Symposium 1 355-372.

[46] RAO, C. R. (1967). Calculus of generalized inverses of matrices. Part I: General theory. Sankhya Series A 29 317-342.

[47] ROY, J. (1958). Step-down procedure in multivariate analysis. Ann. Math. Statist. 29 1177-1187.

[48] ROY, S. N. (1957). Some Aspects of Multivariate Analysis. Wiley and Sons, New York.

[49] ROY, S. N. (1964). A report on some results in simultaneous or joint linear (point) estimation. Institute of Statistics, Mimeo Series No. 394. University of North Carolina. Chapel Hill, N. C.

[50] SCHATZOFF, M. (1966). Sensitivity comparisons among tests of the general linear hypothesis. J. Amer. Statist. Assoc. 61 415-435.

[51] SCHATZOFF, M. (1966). Exact distributions of Wilks' likelihood ratio criterion. Biometrika 53 347-351.

[52] SIOTANI, M. (1968). Some methods for asymptotic distributions in the multivariate analysis. Institute of Statistics Mimeo Series No. 595. University of North Carolina, Chapel Hill, N. C.

[53] SMITH, W. B. (1968). Bivariate samples with missing values, II. Answer to Query 29. Technometrics 10 867-868.

[54] SRIVASTAVA, J. N. (1965). A multivariate extension of the Gauss-Markov theorem. Ann. Inst. Statist. Math. 17 63-66.

[55] SRIVASTAVA, J. N. and ROY, S. N. (1965). Hierarchal and p-block multiresponse designs and their analysis. Mahalanobis Dedicatory Volume (edited by C. R. Rao). Pergamon Press, New York.

[56] SRIVASTAVA, J. N. (1966). Incomplete multiresponse designs. Sankhya Ser. A 28 377-388.

[57] SRIVASTAVA, J. N. (1966). Some generalizations of multivariate analysis of variance. Multivariate Analysis (Edited by

P. R. Krishnaiah). Academic Press, New York. 129-145.

[58] SRIVASTAVA, J. N. (1968). On a general class of designs for multiresponse experiments. Ann. Math. Statist. 39 1825-1843.

[59] SRIVASTAVA, J. N. (1968). Some studies of intersection tests in multivariate analysis of variance. Paper presented at the Symposium on Multivariate Analysis at Dayton, Ohio. Unpublished.

[60] SRIVASTAVA, J. N. (1967). On the extension of Gauss-Markov theorem to complex multivariate linear models. Ann. Inst. Statist. Math. 19 417-437.

[61] TELSER, L. G. (1964). Iterative estimation of a set of linear regression equations. J. Amer. Statist. Assoc. 59 842-862.

[62] TOCHER, K. D. (1952). The design and analysis of block experiments. J. Roy Statist. Soc., Ser. B 14 45-100.

[63] TRAWINSKI, I. M. (1961). Incomplete-variable designs. Unpublished thesis. Virginia Polytechnic Institute. Blacksburg, Virginia.

[64] TRAWINSKI, I. M. and BARGMANN, R. E. (1964). Maximum likelihood estimation with incomplete multivariate data. Ann. Math. Statist. 35 647-658.

[65] WALD, A. (1941). Asymptotically most powerful tests for statistical hypotheses. Ann. Math. Statist. 12 1-19.

[66] WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc.54 426-482.

[67] WALD, A. (1948). Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic

process. Ann. Math. Statist. 19 40-46.

[68] WALD, A. (1949). A note on the consistency of the maximum likelihood estimate. Ann. Math. Statist. 20 595-601.

[69] WILKS, S. S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples. Ann. Math. Statist. 3 163-195.

[70] WILLIAMS, J. S. (1967). The variance of weighted regression estimators. J. Amer. Statist. Assoc. 62 1290-1301.

[71] YATES, F. (1933). The analysis of replicated experiments when the field results are incomplete. Empire J. Exper. Agric. 1 129-142.

[72] ZELLNER, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. J. Amer. Statist. Assoc. 57 348-368.

[73] ZELLNER, A. (1963). Estimators for seemingly unrelated regression equations: some exact finite sample results. J. Amer. Statist. Assoc. 58 977-992.

[74] ZELLNER, A. and HUANG, D. S. (1962). Further properties of efficient estimators for seemingly unrelated regression equations. Internat. Econ. Rev. 3 300-313.

APPENDIX

We present below (without proof) a number of important matrix results used in the main text dealing with generalized inverses, matrix derivatives and Kronecker products. These topics have been treated by a number of authors, notably Rao [41], [44], [46], Dwyer [20] and Anderson [8].

A.1 Generalized Inverses

Definition A.1.1: A^- is a generalized inverse of A if and only if $AA^-A = A$.

Theorem A.1.1: Given $A(p \times q)$, $\Gamma(p \times p)$ and $C(v \times q)$ such that $R(C) = v$, $C = L'A$ for some $L(p \times v)$ and Γ is symmetric and positive definite, then

(i) $A(A'A)^-A'$ and $I_p - A(A'A)^-A'$ are uniquely determined, symmetric idempotent, $R(A[A'A]^-A') = R(A)$ and $R(I_p - A[A'A]^-A') = p - R(A)$.

(ii) $(A'\Gamma A)(A'\Gamma A)^-A' = A'$.

(iii) $C(A'\Gamma A)^-C'$ is unique and positive definite .

(iv) $C(A'\Gamma A)^-(A'\Gamma A)(A'\Gamma A)^-C' = C(A'\Gamma A)^-C'$.

(v) $S = Y'[I - A(A'A)^-A']Y$ is of full rank with probability one if Y is a matrix random variable whose rows are normally distributed and $E(Y) = A\mu$ for some matrix μ .

A.2 Matrix Differentiation

Definition A.2.1: Let ∂ denote partial differentiation. Let $A(p \times q) = ((a_{ij}))$, $B(r \times s) = ((b_{ij})) = [\underline{b}_1 \ldots \underline{b}_s]$ and α be any scalar function. Then

$$\frac{\partial \alpha}{\partial B}\ (r \times s) = ((\frac{\partial \alpha}{\partial b_{ij}}))\ , \quad \frac{\partial A}{\partial \alpha}\ (p \times q) = ((\frac{\partial a_{ij}}{\partial \alpha}))$$

$$\frac{\partial A}{\partial B}\ (rq \times ps) = [A_1' \mid A_2' \mid \dots \mid A_s'] \quad \text{when } r > 1 \text{ or } s > 1\ ,$$

where $A_\ell\ (p \times rq) = ((\frac{\partial a_{ij}}{\partial \underline{b}_\ell'}))\ , \ \ell = 1,\dots,s\ ,$

and $\dfrac{\partial^2 \alpha}{\partial A \partial B} = \dfrac{\partial}{\partial B}\ [\dfrac{\partial \alpha}{\partial A}]\ .$

<u>Theorem A.2.1</u>: The following is a list of matrix differentiation rules:

(i) $\dfrac{\partial}{\partial \alpha}\ (AB) = A\ \dfrac{\partial B}{\partial \alpha} + \dfrac{\partial A}{\partial \alpha}\ B.$

(ii) $\dfrac{\partial B^{-1}}{\partial \alpha} = -B^{-1}\ \dfrac{\partial B}{\partial \alpha}\ B^{-1}\ .$

(iii) $\dfrac{\partial B\underline{\mu}}{\partial \underline{\mu}} = B'\ .$

(iv) $\dfrac{\partial}{\partial \underline{\mu}}\ (\underline{y} - A\underline{\mu})'\Gamma(\underline{y} - A\underline{\mu}) = -\ 2A'\Gamma(\underline{y} - A\underline{\mu})$ whenever Γ is symmetric.

(v) The (i,j)-th element of $\dfrac{\partial}{\partial \Gamma}\ \underline{x}'\Gamma\underline{x}$ where Γ is symmetric is given by $\delta_{ij}\left\langle \underline{x}\underline{x}' \right\rangle_{ij}$ where $\delta_{ij} = \begin{cases} 2, & i \neq j \\ 1, & i = j \end{cases}$.

(vi) The (i,j)-th element of $\dfrac{\partial}{\partial \Gamma}\ \underline{x}'\Gamma^{-1}\underline{x}$ where Γ is symmetric is given by $\delta_{ij}\left\langle -\Gamma^{-1}\underline{x}\underline{x}'\Gamma^{-1} \right\rangle_{ij}$ where δ_{ij} is defined in (v).

(vii) $\dfrac{\partial}{\partial \mu}\ \mathrm{tr}\ U^{-1}P'P = -\ 2B'PU^{-1}M'$ where $U = M'\Gamma M$ is of full rank, $P = X - B\Lambda M$ and Γ is symmetric.

(viii) The (i,j)-th element of $\dfrac{\partial}{\partial \Gamma}\ \ln |U|$ where $U = M'\Gamma M$ is of full rank and Γ is symmetric is given by

$\delta_{ij}\left\langle MU^{-1}M'\right\rangle_{ij}$ where δ_{ij} is defined in (v).

(ix) $\frac{\partial}{\partial\mu}A\mu B = B' \otimes A'$ where the elements of μ are mutually independent (e.g., μ is not symmetric).

(x) $\frac{\partial^2\alpha}{\partial\mu^2} = \frac{\partial^2\alpha}{\partial\underline{\mu}^2}$ where $\mu = [\underline{\mu}_1, \ldots, \underline{\mu}_p]$ and $\underline{\mu} = (\underline{\mu}_1', \ldots, \underline{\mu}_p')'$ whenever α is of the form $\alpha = \text{tr}\ \Gamma^{-1}(B\mu M)'(B\mu M)$, where Γ is symmetric and B and M are arbitrary.

(xi) $\frac{\partial^2}{\partial\mu^2}\ \text{tr}\ U^{-1}P'P = 2\ MU^{-1}M' \otimes B'B$ where $U = M'\Gamma M$ is of full rank, $P = X - B\mu M$ and Γ is symmetric.

(xii) The (i,j)-th element of $\frac{\partial\ \text{tr}\ U^{-1}P'P}{\partial\Gamma}$ is given by $-\delta_{ij}\left\langle MU^{-1}P'PU^{-1}M'\right\rangle_{ij}$ where Γ, U, M and P are as in (xi) and δ_{ij} is defined in (v).

A.3 <u>Kronecker Products</u>

<u>Definition A.3.1:</u> $A \otimes B = ((a_{ij}))$ is the Kronecker product of the matrices $A = ((a_{ij}))$ and B.

<u>Theorem A.3.1</u>: The following are some results involving $\otimes$:

(i) $(A \otimes B) + (A \otimes C) = A \otimes (B + C)$.

(ii) $(A \otimes B) + (D \otimes B) = (A + D) \otimes B$.

(iii) $(A \otimes B)(C \otimes D) = (AC \otimes BD)$ if the products AC and BD are defined.

(iv) $(A \otimes B)^- = A^- \otimes B^-$.

(v) $A\xi B$ and $(B' \otimes A)\underline{\xi}$ where $\xi = [\underline{\xi}_1 \ldots \underline{\xi}_p]$ and $\underline{\xi} = (\underline{\xi}_1', \ldots, \underline{\xi}_p')'$ consist of the same set of elements.