TESTING IN THE MULTIVARIATE GENERAL LINEAR MODEL

TESTING IN THE MULTIVARIATE GENERAL LINEAR MODEL

BY
TAKEAKI KARIYA

ECONOMIC RESEARCH SERIES
NO. 22
THE INSTITUTE OF ECONOMIC RESEARCH
HITOTSUBASHI UNIVERSITY
KUNITACH, TOKYO 186

KINOKUNIYA COMPANY LTD.
Tokyo, Japan

Printed in 1985

Published by Kinokuniya Company Ltd.
17-7 Shinjuku 3-chome, Shinjuku-ku,
Tokyo 160-91, Japan

(In U.S.A.)
Distributed by New York Kinokuniya Bookstores
10 West 49th Street,
New York, NY10020
U.S.A.

ISBN 4-314-00450-9

Printed by
Kenkyusha Printing Co., Ltd.
Tokyo, Japan

Dedicated to
SHIZUKO

cat. as
sep.

PREFACE

This book is a monograph on some hypothesis testing problems in the General MANOVA (Multivariate Analysis of Variance) model or general multivariate regression model and its extended models. The General MANOVA model was originally formulated as a general model for describing the growth curves of animals and it is often called a growth curve model. However, the model includes the classical MANOVA model and various multivariate regression models (see Introduction) and has a broader applicability in many fields. In fact, some applications to economic phenomena are also made in this book. In addition, the incompleteness of the model provides a theoretical interest or challenge for analysis.

We first treat the problem of testing the General MANOVA hypothesis which we simply call the GMANOVA problem. This problem contains as special cases many problems treated in the growth curves, the testing problem on means with covariates, etc. as well as the usual MANOVA problem. Then the problem is extended to cover some other interesting problems in the General MANOVA model such as the problem of tesing on means with missing data, the problem of testing on regression coefficients in an SUR (seemingly unrelated regression) model, etc. All the problems here deal with linear hypotheses on the regression coefficient matrices of the models.

On the other hand, the problems of testing some hypotheses on the covariance structure of the models are also treated. A hypothesis of our main concern is what is called Rao's covariance structure and the problem of testing the hypothesis in the GMANOVA model is regarded as an extension of the problem of testing independence in the MANOVA model or as the problem of choice between the ordinary least squares estimator and the generalized least squares estimator in the model. It is also closely related to the problem of testing independence with missing data and the problem of testing independence in an SUR model, which are to be treated in this book.

In the analysis of the problems, we attached a considerably great importance to application and tried to provide many practical examples though we analyze the problems in a systematic way based on the invariance principle. Consequently we had to save some topics and proofs of theoretical interest.

With deep gratitude I acknowledge the support of others in the past. First of all, I sincerely appreciate Professors Osamu Isono, Seiji Nabeya, Clifford Hildreth and Morris L. Eaton for their instructions and encouragement while I was a student. Secondly I wish to thank Professors P.R. Krishnaiah, R.A. Wijsman and Akio Kudo for supports and encouragement in the second stage of my academic life. My personal thanks go to my friends the Treichels, Hiroshi Ogura and Masayoshi Mizuno for their ceaseless sincere encouragement. I also thank Ms. Keiko Horii for her typing.

Finally I acknowledge the American Statistical Association, the Institute of Mathematical Statistics and Professor L. Banken for their permissions of citation.

Portions of the work have been supported by the Ministry of Education of Japan and some portions are supported by the Ministry of Education under General C 59540104 and by Nihon Keizai Kenkyu Shorei Zaidan (Japan Foudation for Economic Research).

March 1985

Takeaki Kariya

CONTENTS

INTRODUCTION

1. Brief Summary of the Problems in this Book

The purpose of the present book is to analyze the following problems by using the principle of invariance and make the results applicable to practical problems.

(1) *GMANOVA problem.* The general MANOVA (multivariate analysis of variance) problem, which we concern in Chapter 3, is the problem of testing a hypothesis of the form

$$(1.1) \qquad H: X_3BX_4=X_0 \quad \text{versus} \quad K: X_3BX_4\neq X_0$$

in the generalized multivariate regression model or growth curve model

$$(1.2) \qquad Y=X_1BX_2+E, \quad E\sim N(O, I_n\otimes\Omega)$$

where the notation here is made clear in Chapter 1. We call the model (1.2) the GMANOVA model and the problem (1.1) the GMANOVA problem. In this rather simple formulation, many hypothesis testing problems commonly treated in applications are included. For examples, when $X_2=I$ and $X_4=I$, the problem is well known as the MANOVA problem, which alone covers a lot of interesting problems in applications such as the problems of testing on means, regression coefficients, effects of treatments, etc. in univariate or multivariate models. On the other hand, some important problems which are not in the framework of the MANOVA problem are included in the above GMANOVA problem. Such examples are problems of testing on growth curves used in biometrics, econometrics etc., of testing on means with prior information, of testing the equality of means in a classification model.

The GMANOVA problem also entails some features of theoretical interest which the MANOVA problem does not provide. The features are in fact created by the presence of the matrices X_2 and X_4 in (1.2) and (1.1). One of the features is that the information provided by X_2 and X_4 produces an ancillary statistic—a statistic which is a part of a sufficient statistic but whose marginal distribution is parameter-free—and that the LRT (likelihood ratio test) ignores the ancillary statistic completely so that the LRT is inadmissible. This is rather surprising because no other such examples have ever been found in practical problems under normal models. Some other features are stated in Chapter 3.

(2) *Extended GMANOVA problem.* An extension of the GMANOVA problem is made by imposing additional restrictions on the coefficient matrix $\boldsymbol{B}$ as prior information;

$$P_iBQ_i = R_i \quad (i=1,\dots,k). \tag{1.3}$$

and then the problem is to test (1.1) under the model (1.2) with (1.3). This is called the extended GMANOVA problem in this book. By this extension, the formulation has a wider scope to cover such important problems in applications as problems of testing on means with incomplete data, of testing on regression coefficients in a SUR (seemingly unrelated regression) model, of testing the equality of means in a classification model with covariates etc. The analysis of the extended GMANOVA problem is concerned in Chapter 4.

(3) *Testing Independence in the GMANOVA model.* In the GMANOVA model (1.2), the problem of testing the hypothesis that the covariance matrix Ω is of the form

$$\Omega = X_2'\Gamma X_2 + Z_2'\Delta Z_2 \quad (X_2'Z_2 = 0), \tag{1.4}$$

which is known as Rao's covariance structure, corresponds to the problem of testing independence in the MANOVA problem with extra data. In fact, an invariance reduction of the problem of testing (1.4) yields the problem of testing independence between two sets of variates with extra data on the first set. Hence the analysis of the problem also solves the problem of testing independence with incomplete or missing data. In addition, the problem of testing independence in a classification model with covariates is a special case of the problem here. On the other hand, the problem of testing (1.4) in itself is interesting because it is considered the problem of a choice between the GLSE (generalized least squares estimator) and the OLSE (ordinary LSE) in the GMANOVA model. This is a theme of Chapter 5.

(4) *Testing independence in a SUR model.* Chapter 6 deals with the problem of testing independence in a two equations SUR model, which is regarded as a special case of testing independence in an extended GMANOVA model. The problem here also is regarded as the problem of a choice between the GLSE and the OLSE in the SUR model.

In Chapter 1, the explanations and implications of these problems are given in details with many practical examples and in Chapter 2, the Neyman-Pearson testing theory is reviewed with a special emphasis on invariance and optimalities of tests. In all the problems, LBI (locally best invariant) tests are derived and made applicable by providing approximate null distributions.

2. Notation

Some notation we use throughout the book is listed here. By R^n, we denote an n-dimmensional Eucledian space, and by $|A|$, tr$A$ and A' the determinant, trace and transpose of a matrix A respectively. For matrices and vectors, gothic letters are used and A: $n \times k$ means that A is an $n \times k$ matrix. In addition, we use the notation:

$\mathcal{GL}(n)=\{A : n\times n \mid |A| \neq 0\}$: the group of $n\times n$ nonsingular matrices,

$\mathcal{O}(n)=\{A\in\mathcal{GL}(n) \mid A'A=I_n\}$: the group of $n\times n$ orthogonal matrices,

$\mathcal{S}(n)=\{A : n\times n \mid A'=A\}$: the set of symmetric matrices,

$\mathcal{S}_+(n)=\{A\in\mathcal{S}(n) \mid A$ is positive definite$\}$, and

$\mathcal{GU}(n)=\{A\in\mathcal{GL}(n) \mid A=(a_{ij}),\ a_{ij}=0$ for $i>j\}$: the group of $n\times n$ nonsingular upper triangular matrices.

For $A\in\mathcal{S}_+(n)$, $A^{1/2}$ denotes the symmetric square root of A, i.e. $(A^{1/2})^2=A$, and diag $\{a_1,\dots,a_n\}$ denotes the diagonal matrix with diagonal elements a_i's in the given order, while DIAG$\{A_1,\dots,A_n\}$ the block diagonal matrix with A_i's as diagonal blocks. Further, for $A\in\mathcal{S}(n)$ $\mathrm{ch}_i(A)$ denotes the i-th largest characteristic root of A.

Chapter 1

TESTING PROBLEMS IN GMANOVA AND EXTENDED GMANOVA MODELS

1. GMANOVA and SUR Models.

1.1. *Classification of regression models.* Whichever univariate or multivariate, a linear regression model is formally expressed as

$$y = X\beta + \varepsilon \tag{1.1}$$

where X is an $N\times K$ matrix and ε is an $N\times 1$ error term vector with mean $E(\varepsilon)=0$ and covariance matrix $\mathrm{Cov}(\varepsilon)=E(\varepsilon\varepsilon')=\Omega$. As will be seen, what is called a seemingly unrelated regression model (SUR model), a multivariate regression model or a MANOVA (multivariate analysis of variance) model, a growth curve model etc, are special cases of the model (1.1). In fact, when the model (1.1) is written as a natural form according to the structure of X and Ω, it is called under different names and, needless to say, different approaches are taken to the problems of estimation and testing according to the structure of the model. On the other hand, a linear restriction on the coefficient vector β also differentiates the model. That is, when a general linear restriction of the form

$$(1.2) \qquad R\beta = r, \quad R : r\times k, \quad \mathrm{rank}(R)=r,$$

is imposed on β as prior information, the structure of the matrix R sometimes differentiates the model (1.1). Such a restriction is often encountered in applications as a result of a formal treatment of the model in question. Here we shall consider some representative models produced by the differences of the structure of matrices X or Ω or R. Incidentally if no information on the structure of Ω in (1.1) is available, Ω is not estimable so that no efficient methods for estimation or testing can be provided in general.

(A) *Heteroscedastic model.* Let

$$X := \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}, \quad X_i : n_i\times k, \quad \mathrm{rank}(X_i)=k \ (i=1,2)$$
$$N=n_1+n_2, \quad K=2k$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} : 2k\times 1, \quad \text{and} \quad \Omega = \begin{pmatrix} \omega_{11}I_{n_1} & 0 \\ 0 & \omega_{22}I_{n_2} \end{pmatrix}$$

In economics, this model is assumed when a structural change in an economy is suspected and it is often the case that the hypothesis

$$H : \beta_1=\beta_2 \quad \text{versus} \quad K : \beta_1 \neq \beta_2$$

is considered under the assumption $\omega_{11}\neq\omega_{22}$. Of course, it is also interesting to test $H : \omega_{11}=\omega_{22}$ in this model.

(B) *SUR model.* In the model (1.1), let

$$(1.3) \qquad X = \begin{pmatrix} X_1 & & & 0 \\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_p \end{pmatrix}, \quad X_i : n\times k_i, \quad \mathrm{rank}(X_i)=k_i$$
$$N=np, \quad K=\textstyle\sum_{i=1}^{p} k_i$$

and

$$\Omega = [\Sigma\otimes I_n], \quad \Sigma : p\times p$$

($i=1, \cdots, p$), where $A\otimes B$ denotes the Kronecker product of two matrices A and B. The model (1.1) with these X and Ω is called an SUR model or Zellner's model. This model is usually expressed as p different regression models with cross correlation,

$$(1.4) \qquad y_i = X_i\beta_{ii} + \varepsilon_i,$$
$$\mathrm{Cov}(y_i, y_j) = E(\varepsilon_i\varepsilon_j') = \sigma_{ij}I_n \quad (i,j=1,\cdots,p)$$

where y, β and ε in (1.1) are decomposed as

$y'=(y_1', \cdots, y_p')$ with $y_i : n\times 1$, $\varepsilon'=(\varepsilon_1', \cdots, \varepsilon_p')$ with $\varepsilon_i : n\times 1$,

$\beta'=(\beta_{11}', \cdots, \beta_{pp}')$ with $\beta_{ii} : k_i\times 1$ and $\Sigma=(\sigma_{ij}) : p\times p$.

The model is also rewritten as a multivariate regression model with prior information on the coefficient matrix:

$$(1.5) \qquad Y = X\tilde{B} + E, \qquad \mathrm{Cov}(E) = I_n\otimes\Sigma$$

where $Y=[y_1,\cdots,y_p] : n\times p$, $X=[X_1, \cdots, X_p] : n\times K$, which may not be of full rank, $E=[\varepsilon_1,\cdots,\varepsilon_p] : n\times p$ and

$$\tilde{B} = \begin{pmatrix} \beta_{11} & & 0 \\ & \ddots & \\ 0 & & \beta_{pp} \end{pmatrix} : K\times p.$$

It should be noted that the information $\beta_{ij}=0$ $(i\neq j)$ in the structure of this coefficient matrix $\tilde{B}$ cannot be expressed as a single general linear restriction

$$(1.6) \qquad R_1BR_2=R_0, \quad R_1 : r_1\times K, \quad R_2 : p\times r_2, \quad \mathrm{rank}(R_i)=r_i$$

where $B=(\beta_{ij})$ with $\beta_{ij} : k_i\times 1$, though the model (1.5) is a multivariate regression model. In this sense, an SUR model is considered different from a multivariate regression model or a general multivariate regression model introduced below. Some references on estimation or testing problems under this model are

Zellner (1962, 1963), Mehta and Swamy (1976), Revankar (1974, 1976), Srivastava (1970), Kariya (1981 a, b), Kariya and Maekawa (1981), Kariya, Fujikoshi and Krishnaiah (1984) etc. In the last paper, the model (1.4) is extended to a multivariate version : $Y_i = X_i B_{ii} + E_i (i=1,\cdots,p)$.

(C) *Multivariate regression model.* Let $X_1 = X_2 = \cdots = X_p \equiv X_o : n \times k$ in (1.4) and $B = [\beta_{11}, \cdots \beta_{pp}] : k \times p$. Then the model (1.1) with (1.4) becomes a well known multivariate regression model

$$(1.7) \qquad Y = X_0 B + E, \qquad \mathrm{Cov}(E) = I_n \otimes \Sigma.$$

In this sense, an SUR model may be regarded as an extension of the multivariate regression model (1.7), while (1.7) in turn includes (1.4) as a special case though X in (1.5) may not be of full rank.

In the model (1.7), most linear restrictions on the coefficient matrix B are expressed as a general linear restriction of the form

$$(1.8) \qquad R_1 B = R_0, \quad R_1 : r_1 \times k, \quad \mathrm{rank}(R_1) = r_1.$$

The problem of testing this linear restriction (1.8) under the model (1.7) is called the MANOVA (multivariate analysis of variance) problem and there (1.8) is called a general linear hypothesis. There are many specific problems put in this form and a lot of references on the MANOVA problem. Typical examples are the problem of testing on regression coefficients, the problem of testing the equality of mean vectors in two populations known as Hotelling's T^2-problem etc. Textbooks such as Anderson (1958), Giri (1977), Eaton (1983) etc. will be useful as references on the MANOVA problem and other testing problems.

The linear restriction (1.8) is further generalized to the form of (1.6). But the problem of testing the linear restriction (1.6) under the model (1.7) is more difficult to treat than that of testing

(1.8) under (1.7) because the structure of R_2 and the structure of Σ become associated. In fact, this problem is a special case of the GMANOVA (general MANOVA) problem, as will be seen next.

(D) *General multivariate regression model.* In (1.1), let

$$(1.9) \qquad \begin{cases} X = X_1 \otimes X_2', & X_1 : n \times k, \quad X_2 : q \times p, \\ \mathrm{rank}(X_1) = k, & \mathrm{rank}\ (X_2) = q \\ \Omega = [I_n \otimes \Sigma], & \Sigma : p \times p, \quad N = np, \quad K = kq. \end{cases}$$

$$Y = \begin{bmatrix} y_1' \\ \vdots \\ y_n' \end{bmatrix} : n \times p, \quad B = \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_k' \end{bmatrix} : k \times q \quad \text{and} \quad E = \begin{bmatrix} \varepsilon_1' \\ \vdots \\ \varepsilon_n' \end{bmatrix} : n \times p.$$

where

$$y' = (y_1', \cdots, y_n') \qquad \text{with} \qquad y_i : p \times 1,$$

$$\varepsilon' = (\varepsilon_1', \cdots, \varepsilon_n') \qquad \text{with} \qquad \varepsilon_i : p \times 1\ (i=1, \cdots, n),$$

$$\beta' = (\beta_1', \cdots, \beta_k') \qquad \text{with} \qquad \beta_j : q \times 1\ (j=1, \cdots, k).$$

Then the model (1.1) with (1.9) becomes

$$(1.10) \qquad Y = X_1 B X_2 + E, \qquad \mathrm{Cov}(E) = I_n \otimes \Sigma$$

This model is called a general multivariate regression model or GMANOVA model or sometimes growth curve model. Special forms of this model has been considered in various problems in association with growth curves, but it is Potthoff and Roy (1964) that gave the formulation (1.10). The problem of estimating the coefficient matrix B in the model (1.10) is treated by Rao (1965, 1967), Geisser (1970), Kariya (1983) etc. On the other hand, the problem of testing the general linear hypothesis in (1.6), i.e.,

$$(1.11) \qquad R_1 B R_2 = R_0, \quad R_1 : r_1 \times k, \quad R_2 : q \times r_2, \quad \mathrm{rank}(R_i) = r_i.$$

is treated by Khatri (1966), Gleser and Olkin (1970), Kariya (1978) etc. This testing problem is often called the GMANOVA problem and the hypothesis(1.11) shall be called a GMANOVA hypothesis. Some problems associated with the GMANOVA problem are considered by Hooper(1982, 1983), Marden(1984) etc. The analysis of the GMANOVA problem via invariance is a main theme of this book. The MANOVA problem in (C) is of course a special case of the GMANOVA problem with $X_2=I_p$ and $R_2=I_p$.

We usually assume normal distribution for the error term matrix E in each model:

$$E \sim N(0, I_n \otimes \Sigma) \quad \text{with} \quad \Sigma \in \mathcal{S}_+(p) \tag{1.12}$$

where $\mathcal{S}_+(p)$ denotes the set of $p \times p$ positive definite matrices. Here by the notation in (1.12), it is meant that the n rows of E independently follow the same normal distribution with mean O and covariance matrix Σ. More generally we make

Definition 1.1. Let U be an $n \times p$ random matrix and let $u_i : 1 \times p$ be the i-th row of U $(i=1, \cdots, n)$. Then the notation

$$U \sim N(\mu, A \otimes \Phi) \quad (A \in \mathcal{S}_+(n),\ \Phi \in \mathcal{S}_+(p))$$

means that the $np \times 1$ vector $(u_1, \cdots, u_n)'$ is normally distributed with mean $(\mu_1, \cdots, \mu_n)'$ and covariance matrix $A \otimes \Phi$, where μ_i is the i-th row of μ.

Lemma 1.1. When $U \sim N(\mu, A \otimes \Phi)$, then for any $B : m \times n$ and $C : p \times q$, $BUC \sim N(B\mu C,\ BAB' \otimes C'\Phi C)$.

1.2. *Examples of the GMANOVA problem.* Here we shall give some examples of the GMANOVA problem which are not special cases of the MANOVA problem. The following basic example is stated in terms of the growth of animals in Potthoff and Roy (1964) but we here describe it in terms of economic time series.

Example 1.1. Let $y(t)=(y_1(t), \cdots, y_p(t))'$ be a $p \times 1$ vector consisting of p economic variables, where $y(t)$ is observed at time $t=1, \cdots, n$, and $y(1), \cdots, y(n)$ are assumed to be independently distributed as normal distribution with mean $\mu(t)$ and covariance matrix $\Sigma \in \mathcal{S}_+(p)$:

$$y(t) \sim N(\mu(t), \Sigma) \quad (t=1, \cdots, n). \tag{1.13}$$

Further the i-th element $\mu_i(t)$ of $\mu(t)$ is assumed to obey a polynomial of $k-1$ degrees;

$$\mu_i(t)=\beta_{i0}+\beta_{i1}t+\cdots+\beta_{ik-1}t^{k-1} \quad (i=1, \cdots, p) \tag{1.14}$$

This is a model in which p economic time series $(y_1(t), \cdots, y_p(t))$ fluctuate interdependently at each time t and the mean growth process of each $y_i(t)$ is described by the polynomial (1.14). This model is written as a multivariate regression or MANOVA model;

$$Y=X_1B+E, \qquad E \sim N(0, I_n \otimes \Sigma) \tag{1.15}$$

where

$$Y=\begin{bmatrix} y(1)' \\ y(2)' \\ \vdots \\ y(n)' \end{bmatrix} : n \times p, \qquad X_1=\begin{bmatrix} 1 & 1 & 1^2 & \cdots & 1^{k-1} \\ 1 & 2 & 2^2 & \cdots & 2^{k-1} \\ \vdots & & & & \vdots \\ 1 & n & n^2 & \cdots & n^{k-1} \end{bmatrix} : n \times k \tag{1.16}$$

and

$$B=\begin{pmatrix} \beta_{10} & \beta_{20} & \cdots & \beta_{p0} \\ \beta_{11} & \beta_{21} & \cdots & \beta_{p1} \\ \vdots & & & \vdots \\ \beta_{1k-1} & \beta_{2k-1} & \cdots & \beta_{pk-1} \end{pmatrix} : k \times p. \tag{1.17}$$

In this model, the hypothesis that the p economic variables follow the same growth pattern completely or $\beta_{1j}=\beta_{2j}=\cdots=\beta_{pj}(j=0, 1, \cdots, k-1)$ is expressed as

$$BR_2=0 \tag{1.18}$$

where

$$(1.19)\qquad R_2=\begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \ddots & \vdots \\ 0 & -1 & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ \vdots & \cdots & \ddots & 1 \\ 0 & \cdots & 0 & -1 \end{pmatrix} : p\times(p-1)$$

Further, the hypothesis that the p economic variables follow the same growth pattern but they may be different in the original levels β_{i0}(at $t=0$) or the intercepts $(i=1,\cdots,p)$ is expressed as

$$(1.20)\qquad R_1BR_2=0 \qquad \text{with} \qquad R_1=[0, I_{k-1}] : (k-1)\times k.$$

This is the hypothesis that $\beta_{1j}=\beta_{2j}=\cdots=\beta_{pj}$ $(j=1, \cdots, k-1)$ but β_{i0} may be different.

Though the model (1.15) is a MANOVA model, the hypothesis (1.18) (or (1.20)) is a GMANOVA hypothesis different from the MANOVA hypothesis (1.8) as the coefficient matrix B is multiplied by the matrix R_2 from the right. Therefore the problem of testing (1.18) (or (1.20)) is not in the framework of the MANOVA problem, and it is regarded as a special case of the GMANOVA problem. That is, in this case, the structure of the hypothesis (1.18) (or (1.20)) makes the problem the GMANOVA problem though the model is a MANOVA model.

Example 1.2. Suppose there are groups of countries, where the first group consists of four Asian countries and the second group consists of three African countries. And it is judged that the countries in each group have taken the same process of growth. Here the problem is to test the hypothesis that the two groups have the same growth pattern based on the p economic variables of the same kind for each country;

$$y_{1j}(t)=(y_{1j1}(t),\ y_{1j2}(t),\dots,y_{1jp}(t))' \quad (j=1,\dots,4 : t=1,\dots,n)$$

and

$$y_{2i}(t)=(y_{2i1}(t),\ y_{2i2}(t),\dots,y_{2ip}(t))' \quad (i=1,2,3\ ;\ t=1,\dots,n).$$

Let

$$Y_1=\begin{bmatrix} y_{11}(1)' & y_{12}(1)' & y_{13}(1)' & y_{14}(1)' \\ y_{11}(2)' & y_{12}(2)' & y_{13}(2)' & y_{14}(2)' \\ \vdots & \vdots & \vdots & \vdots \\ y_{11}(n)' & y_{12}(n)' & y_{13}(n)' & y_{14}(n)' \end{bmatrix} : n\times 4p$$

and

$$Y_2=\begin{bmatrix} y_{21}(1)' & y_{22}(1)' & y_{23}(1)' \\ y_{21}(2)' & y_{22}(2)' & y_{23}(2)' \\ \vdots & \vdots & \vdots \\ y_{21}(n)' & y_{22}(n)' & y_{23}(n)' \end{bmatrix} : n\times 3p$$

Here Y_i is the matrix of all the p variables in all the countries belonging to the i-th group over $t=1,\dots,n$, where $i=1,2$. Since we have assumed that all the countries in each group had taken the same growth process, we have

$$E[y_{11}(t)]=\cdots=E[y_{14}(t)]\equiv\mu_1(t) \quad (t=1,\dots,n)$$

and

$$E[y_{21}(t)]=\cdots=E[y_{23}(t)]\equiv\mu_2(t) \quad (t=1,\dots,n).$$

Further we assume that $\mu_i(t)$ is approximated by a polynomial of degree k-1. Then similar to Example 1.1, define B_1 and B_2 for each group as in (1.17) and use the same X_1 in (1.16) to obtain

$$E(Y_1)=(X_1B_1, X_1B_1, X_1B_1, X_1B_1)=X_1B_1[I_p, I_p, I_p, I_p]$$

and

$$E(Y_2)=(X_1B_2, X_1B_2, X_1B_2)=X_1B_2[I_p, I_p, I_p].$$

Hence letting

(1.21) $\quad Y=[Y_1,Y_2] : n\times 7p, \quad B=[B_1,B_2] : k\times 2p \quad$ and

(1.22) $$X_2=\begin{bmatrix} I_p & I_p & I_p & I_p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_p & I_p & I_p \end{bmatrix} : 2p\times 7p,$$

our model is finally expressed as

(1.23) $$Y=X_1BX_2+E, \quad E \sim N(0, I_n\otimes[I\otimes\Sigma]),$$

where all the p variables in 7 countries are assumed to have the same covariance matrix Σ. The hypothesis to be tested is $B_1=B_2$ and so it is expressed as

(1.24) $$BR_2=0 \quad \text{with} \quad R_2=\begin{pmatrix} I \\ -I \end{pmatrix} : 2p\times p.$$

This is a GMANOVA problem with a covariance matrix of structure $I_7\otimes\Sigma$. Further, the hypothesis that the growth pattern between the two groups is the same except the initial levels or intercepts is of the form $R_1BR_2=0$ as in (1.20). It is noted that the model (1.23) is an expression of a model pooling cross-section data and time series data. The problems here will be treated in Chapter 4.

Examples 1.3. Let $x_i=(x_{1i}', x_{2i}')' : 2p\times 1$ be a random sample from $N(\mu, \Sigma)$ where x_{kj}: $p\times 1$ and $\mu=(\mu_1', \mu_2')'$ with $\mu_k : p\times 1$ $(k=1,2 : i=1,\dots,n)$. Then letting $Y'=[x_1,\dots,x_n]$ and $e=(1,\dots,1)' : n\times 1$

$$Y : n\times 2p \sim N(e\mu', I_n\otimes\Sigma) \quad \text{or}$$
$$Y=X_1B+E \text{ with } X_1=e,\ B=\mu' \text{ and } E \sim N(0, I_n\otimes\Sigma).$$

This is the MANOVA model, but the hypothesis $\mu_1=\mu_2$ is expressed as

$$BR_2=0 \quad \text{with} \quad R_2=\begin{bmatrix} I_p \\ -I_p \end{bmatrix}$$

Hence testing this hypothesis is not a special case of the MANOVA problem but a special case of the GMANOVA problem.

Example 1.4. Let

$$x_i=\begin{pmatrix} x_{1i} \\ x_{2i} \end{pmatrix}\begin{matrix} p \\ q \end{matrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma\right) \quad \text{and} \quad y_j=\begin{pmatrix} y_{1j} \\ y_{2j} \end{pmatrix}\begin{matrix} p \\ q \end{matrix} \sim N\left(\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}, \Sigma\right)$$

be random samples $(i=1,\dots,n;\ j=1,\dots,m)$, where x_i and y_j are independent. Further let

(1.25) $$Y=[x_1,\dots,x_n,y_1,\dots,y_m]' : (n+m)\times(p+q)$$

$$B=\begin{pmatrix} \mu_1' & \mu_2' \\ \eta_1' & \eta_2' \end{pmatrix} : 2\times(p+q), \quad \text{and} \quad X_1=\begin{pmatrix} e_1 & o \\ o & e_2 \end{pmatrix} : (n+m)\times 2$$

Then the model is expressed as

(1.26) $$Y=X_1B+E \quad \text{with} \quad E \sim N(0, I_{n+m}\otimes\Sigma)$$

and when the hypothesis $\mu_2=\eta_2$ is considered in this model, it is expressed as

$$R_1BR_2=0 \quad \text{with}$$

(1.27) $$R_1=(1,-1) : 1\times 2 \quad \text{and} \quad R_2=\begin{pmatrix} 0 \\ I_q \end{pmatrix} : (p+q)\times q$$

This is a special case of the GMANOVA problem.

1.3 Extensions of the GMANOVA probrem.

Example 1.5. In example 1.4, let us suppose that $\mu_2=\eta_2$ holds in the model (1.26) or (1.27) holds in the model, and suppose that we are interested in the problem of testing $\mu_1=\eta_1$ under this setting. In other words, our model is (1.26) with prior knowledge (1.27) on the coefficient matrix B and we wish to test $\mu_1=\eta_1$ or

$$R_3BR_4=0 \text{ where } R_3=(1,-1):1\times 2 \text{ and } R_4=\begin{pmatrix} I_p \\ 0 \end{pmatrix}. \tag{1.28}$$

This is the problem treated by Cochran and Bliss (1948), Rao (1949), Cochran (1964), Rao (1966), Subrahmanian and Subrahmanian (1971) etc., in association with discriminant analysis with covariates (see also Memon and Okamoto (1970), and Kshirsargar (1972) pp. 200–203). (See also Example 2.2.)

This example motivates the following extension of the GMANOVA problem. Let our model be given by

$$Y=X_1BX_2+E \quad \text{with} \quad R_1BR_2=R_0 \tag{1.29}$$

where X_i's and R_i's are the same as before (see (1.9) and (1.10)). Here $R_1BR_2=R_0$ is no longer the hypothesis to be tested but a part of the model. The hypothesis to be tested here is

$$R_3BR_4=R_5. \tag{1.30}$$

This extension of the GMANOVA problem has been made by Gleser and Olkin (1970) in terms of a canonical form. Recently it is treated by Banken (1984). We shall call the model (1.29) the extended GMANOVA model and the problem of testing (1.30) therein the extended GMANOVA problem. Example 1.5 and the following example due to Banken (1984) are special cases of hte extended GMANOVA problem.

Example 1.6. In onder to inveigate the effects of thyroxin and of thiouracil on the growth of young rats, 27 rats were randomly divided into three groups; 10 rats was in the first group, 7 in the second and 10 in the third. The first group was kept as a control. The second group was given thyroxin and the third group thiouracil. The weight of each rats was measured at the beginning of the experiment and then in four consecutive weeks. The data can be found in Box (1950). Let x_{ijt} be the weight of the j-th individual in the i-th group at the t-th week, $t=0,\dots,4$, $j=1,\dots,n_i$, $i=1,2,3$. The vectors $x_{ij}=(x_{ij0},\dots,x_{ij4})'s$ are assumed to be independently distributed as

$$x_{ij} \sim N(\mu_i, \Sigma) \qquad (\mu_i \in R^5,\ \Sigma \in \mathcal{S}_+(5))$$

where $\mu_i=(\mu_{i0},\dots,\mu_{i4})'$ and μ_{it} is assumed to be a polynomial of degree 2 in t, i.e.,

$$\mu_{it}=a_{i0}+a_{i1}t+a_{i2}t^2 \qquad (t=0,\dots,4).$$

Being randomly assigned to one of the three groups the expected weights of the rats should be equal at the beginning of the experiment:

$$a_{10}=a_{20}=a_{30}.$$

Here we want to test whether the expected growth curves are equal, i.e.,

$$H:\ \mu_1=\mu_2=\mu_3.$$

In matrix notation, the model is given by

$$Y=X_1BX_2+E, \quad E \sim N(0,\ I_{27}\otimes\Sigma)$$

with the constraint $R_1BR_2=0$, where

$$Y'=[x_{11},\dots,x_{110},x_{21},\dots,x_{27},x_{31},\dots,x_{310}],$$

$$X_1=\begin{pmatrix} e_{10} & 0 & 0 \\ 0 & e_7 & 0 \\ 0 & 0 & e_{10} \end{pmatrix}, \qquad X_2=\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 4 & 9 & 16 \end{pmatrix}$$

$$B=\begin{pmatrix} a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \\ a_{30} & a_{31} & a_{32} \end{pmatrix}, \quad R_1=\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}, \quad R_2=\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

and $e_i=(1,\dots,1)'\in R^i$. The null hypothesis is expressed as

$$H: R_3BR_4=0 \quad \text{with } R_3=\begin{pmatrix}1 & 0 & -1\\ 0 & 1 & -1\end{pmatrix} \text{ and } R_4=I.$$

Hence this problem is a special case of the extended GMANOVA problem stated above. These examples will be treated in Chapter 4.

The GMANOVA problem is further extended as follows:
The model is given by

$$(1.31) \qquad Y=X_1BX_2+E, \quad \mathrm{Cov}(E)=I\otimes\Omega$$
$$R_iBQ_i=H_i \ (i=1,\dots,m)$$

and the problem is to test

$$(1.32) \qquad R_{m+1}BQ_{m+1}=H_{m+1}$$

where R_i's, Q_i's, and H_i's are fixed matrices. We shall also call this problem an extended GMANOVA problem. An example of a model formulated in this way is the SUR model in (1.4) introduced in (B). There it has been remarked that the coefficient matrix $\tilde{B}$ of the form (1.5) cannot be put in a form as a single general linear restriction of the form $R_1BR_2=R_0$. But the coefficient matrix B in $Y=XB+E$ is expressed as a matrix B satisfying

$$(1.33) \quad R_{1j}BQ_{1j}=0 \ (j=2,\dots,p);\ R_{2j}BQ_{2j}=0 \ (j=1,3,\dots,p);$$
$$\dots\dots;\ R_{pj}BQ_{pj}=0 \ (j=1,\dots,p-1)$$

where

$$R_{1j}=[I_{k1},0,\dots,0], \quad Q_{1j}=(0,\dots,0,\overset{j}{1},0,\dots,0)' \ (j=2,\dots,p);\ \dots\dots;$$
$$R_{pj}=[0,\dots,0,I_{kp}], \quad Q_{pj}=(0,\dots,\overset{j}{1},0,\dots,0)' \ (j=1,\dots,p-1)$$

In particular, when $p=2$, the model is expressed as

$$(1.34) \qquad \begin{cases} Y=XB+E, & \mathrm{Cov}(E)=I\otimes\Sigma,\\ R_1BQ_1=0, & R_2BQ_2=0.\end{cases}$$

where

$$B=\begin{bmatrix}\beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\end{bmatrix}, \quad X=[X_1,X_2], \quad R_1=[I_{k1},0],$$
$$R_2=[0,I_{k2}], \quad Q_1=(0,1)'$$

and $Q_2=(1,0)'$. In fact, under (1.34) $\beta_{12}=0$ and $\beta_{21}=0$. Hence the SUR model (1.4) with (1.5) is a special case of the extended GMANOVA model in this sense, and so the problem of testing, say $\beta_{111}=0$ in $\beta_{11}=(\beta_{111},\dots,\beta_{11k1})'$, under this model is also a special case of the extended GMANOVA problem. In fact, $\beta_{111}=0$ is expressed as $R^*BQ^*=0$ with

$$R^*=(1,0,\dots,0) \quad \text{and} \quad Q^*=(1,0,\dots,0)'$$

However, without any conditions on X, R_{ij}'s and Q_{ij}'s, it is very difficult to treat such an extended GMANOVA problem since the restrictions are not "nested". This point is fully discussed in Chapters 4 and 6.

2. Testing on Some Covariance Structure

2.1. *OLSE and GLSE.* To state a second main problem in this book, we here consider some conditions on the identical equalities between the OLSE(ordinary least squares estimator) and the GLSE (generalized LSE) and between sample variances. Let us write again the model (1.1) as

$$(2.1) \qquad y=X\beta+\varepsilon, \quad E(\varepsilon)=0, \quad \mathrm{Cov}(\varepsilon)=\tau^2\Phi\equiv\Omega,$$

where X is an $N\times K$ matrix of rank K. An estimator of the form

$$(2.2) \qquad b(\hat{\Phi})=(X'\hat{\Phi}^{-1}X)^{-1}X'\hat{\Phi}^{-1}y$$

is called the GLSE where $\hat{\Phi}\in\mathcal{S}_+(N)$ is an estimator of Φ. In particular, $b(I)$ is called the OLSE and $b(\Phi)$ is sometimes called the GME (Gauss-Markov estimator) whether or not Φ is known. But we also often call $b(\Phi)$ GLSE when no confusion is caused. The following theorem states alternative conditions for the GME to be identically equal to the OLSE.

Theorem 2.1. A necessary and sufficient condition for $b(\Phi)\equiv b(I)$ is that one of the following (1), (2) and (3) holds.

(1) (Kruskal (1968)) The column space $L(X)$ of X is Φ-invariant.

(2) (Zyskind(1967)) $L(X)$ is spanned by some k latent vectors of Φ.

(3) (Rao(1967)) Let Z be an $N\times(N-K)$ matrix of rank $N-K$ such that $X'Z=0$. Then for some $\Gamma\in\mathcal{S}_+(K)$ and $\Delta\in\mathcal{S}_+(N-K)$, Φ is written as

$$\Phi=X\Gamma X'+Z\Delta Z'. \tag{2.3}$$

Proof of (3). (3) is a modified version of Rao (1967) due to Geisser (1970). Sufficiency is clear. To prove the necessity, suppose $b(\Phi)\equiv b(I)$. Then from $(X'\Phi^{-1}X)^{-1}X'\Phi^{-1}=(X'X)^{-1}X'$, we obtain

$$X'\Phi Z=0 \tag{2.4}$$

Since $H\equiv[X, Z]$ is of rank N, (2.4) implies

$$\Phi=(H')^{-1}H'\Phi HH^{-1}=(H')^{-1}\begin{bmatrix}X'\Phi X & 0\\ 0 & Z'\Phi Z\end{bmatrix}H^{-1}. \tag{2.5}$$

Here using $(H')^{-1}=[X(X'X)^{-1},\ Z(Z'Z)^{-1}]$ in (2.5) and setting

$$\Gamma=(X'X)^{-1}X'\Phi X(X'X)^{-1} \text{ and } \Delta=(Z'Z)^{-1}Z'\Phi Z(Z'Z)^{-1},$$

we obtain (2.3), completing the proof.

The structure of Φ in (2.3) is called Rao's covariance structure and the inverse of Φ is given by

$$\Phi^{-1}=X(X'X)^{-1}\Gamma^{-1}(X'X)^{-1}X'+Z(Z'Z)^{-1}\Delta^{-1}(Z'Z)^{-1}Z'. \tag{2.6}$$

We apply the above result to the GMANOVA model in (1.9) ;

$$Y=X_1BX_2+E,\qquad \mathrm{Cov}(E)=I_n\otimes\Sigma \tag{2.7}$$

where $Y: n\times p$, $X_1: n\times k$ of rank k, $X_2: l\times p$ of rank l and $\Sigma\in\mathcal{S}_+(p)$. Let

$$\hat{B}(\Sigma)=(X_1'X_1)^{-1}X_1'Y\Sigma^{-1}X_2'(X_2\Sigma^{-1}X_2')^{-1} \tag{2.8}$$

be the GME in the GMANOVA model (2.7) and let Z_2 be a $(p-l)\times p$ matrix such that $X_2Z_2'=0$ and $\mathrm{rank}(Z_2)=p-l$.

Corollary 2.1. A necessary and sufficient condition for $\hat{B}(\Sigma)=\hat{B}(I)$ is that for some $\Gamma\in\mathcal{S}_+(l)$ and $\Delta\in\mathcal{S}_+(p-l)$, Σ is of the form

$$\Sigma=X_2'\Gamma X_2+Z_2'\Delta Z_2. \tag{2.9}$$

Note that if $p=l$, (2.9) simply means $\Sigma\in\mathcal{S}_+(p)$ and $\hat{B}(\Sigma)=(X_1'X_1)^{-1}X_1'Y$, which is the GME in the MANOVA model.

In the model (2.1), unless Φ is of certain structure, Φ is not efficiently estimable based on y. For the number of unknown parameters (β, Φ, τ^2) is greater than the sample size N. This implies that in the model (2.1), the hypothesis that the covariance structure is of the form (2.3) will not be testable based on y. However, in the GMANOVA model (2.7), the hypothesis that Σ is of the form (2.9) is testable based on Y. This is a third main problem treated in this book. In the below, it will be stated more explicitly and some implications of the problem are given.

Secondly, we shall consider a condition on the identical equality between the OLS sample variance and the GLS sample variance. Let

$$s^2(\Phi)=[y-Xb(\Phi)]'\Phi^{-1}[y-Xb(\Phi)]/m \tag{2.10}$$

where $m=N$ or $N-K$. This is regarded as an estimate of τ^2 in (2.1) when Φ is known. We shall call $s^2(I)$ and $s^2(\Phi)$ the OLS sample variance and the GLS sample variance respectively.

Theorem 2.2. (Kariya (1980)) Let Φ_1 and Φ_2 be two covariance matrices. Then a necessary and sufficient condition for which $s^2(\Phi_1)=s^2(\Phi_2)$ holds for all y is that Φ_1 and Φ_2 are related as follows: For some $N\times N$ symmetric matrix H,

$$\Phi_2=N\Phi_1 N+H-NHN \tag{2.11}$$

or equivalently

$$\Phi_2=\Phi_1+H-NHN \tag{2.12}$$

where $N=I-X(X'X)^{-1}X'$. In particular, the class of Φ for which $s^2(\Phi)=s^2(I)$ holds for all y is the class of Φ of the form

$$\Phi=I+H-NHN \tag{2.13}$$

for some H.

Proof. Let Z be an $N\times(N-K)$ matrix such that $Z'X=0$, $ZZ'=N$ and $Z'Z=I_{N-K}$. Then as is easily shown

$$A^{-1}-A^{-1}X(X'A^{-1}X)^{-1}X'A^{-1}=Z(Z'AZ)^{-1}Z' \tag{2.14}$$

for any $A\in\mathcal{S}_+(N)$. Hence $s^2(\Phi_1)=s^2(\Phi_2)$ for all y is equivalent to

$$\Phi_1^{-1}-\Phi_1^{-1}X(X'\Phi_1^{-1}X)^{-1}X'\Phi_1^{-1}=\Phi_2^{-1}-\Phi_2^{-1}X(X'\Phi_2^{-1}X)^{-1}X'\Phi_2^{-1} \text{ or}$$

$$Z'\Phi_1 Z=Z'\Phi_2 Z.$$

Solving this equation based on the results in Rao and Mitra (1971, pages 24–25) yields (2.11) The converse is clear because $Z'\Phi_1 Z=Z'\Phi_2 Z$ from (2.11). Replacing H by $H+\Phi_1$ in (2.11) yields (2.12), while replacing H by $H-\Phi_1$ in (2.12) yields (2.11). This completes the proof.

Combining this result with Theorem 2.1 yields

Corollary 2.2. A necessary and sufficient condition for which $b(\Phi)\equiv b(I)$ and $s^2(\Phi)\equiv s^2(I)$ hold is that Φ is of the form

$$\Phi=X\Gamma X'+N \tag{2.15}$$

for some $\Gamma\in\mathcal{S}_+(K)$, where $b(\Phi)$ and $s^2(\Phi)$ are defined by (2.2) and (2.10) respectively.

Proof. Take Z to be the matrix used in Theorem 2.2, and observe that for this choice of Z, Theorem 2.1 holds. Then from (2.3), $Z'\Phi Z=\Delta$, while from (2.13), $Z'\Phi Z'=I_{N-K}$. Hence from the equality of Φ in (2.3) and (2.13), $\Delta=I_{N-K}$ follows. This implies (2.15) by (2.3). Sufficiency is clear from its form.

By this corollary; when Φ is of the form (2.15) in (2.1) even if Γ is unknown, $(b(\Phi), s^2(\Phi))$ for (β, τ^2) is identically equal to $(b(I), s^2(I))$. Conversely the class of Φ for which $(b(I), s^2(I))$ is as efficient as $(b(\Phi), s^2(\Phi))$ is given by the class of Φ of the form (2.15). Further, under normality $\varepsilon\sim N(0, \tau^2\Phi)$ and when Φ is of the form (2.15), the MLE (maximum likelihood estimator) for (β, τ^2) is $(b(I), s^2(I))$ for any $\Gamma\in\mathcal{S}_+(K)$.

To apply Theorem 2.2 and Corollary 2.2 to the GMANOVA model (2.7) with $\Sigma=\tau^2\Phi$, let

$$S(\Sigma) = (Y-X_1\hat{B}(\Sigma)X_2)'(Y-X_1\hat{B}(\Sigma)X_2)/n \tag{2.16}$$

and

$$v^2(\Phi)=\operatorname{tr}\Phi^{-1}S(\Phi)/np. \tag{2.17}$$

Corollary 2.3. (1) A necessary and sufficient condition for $v^2(\Phi)\equiv v^2(I)$ is that for some $H\in\mathcal{S}(p)$, Φ is of the form

$$\Phi=I+H-N_2HN_2 \quad \text{with} \quad N_2=I_p-X'_2(X_2X_2')^{-1}X_2 \tag{2.18}$$

(2) A necessary and sufficient condition for $\hat{B}(\Phi)\equiv\hat{B}(I)$ and $v^2(\Phi)\equiv v^2(I)$ is that for some $\Gamma\in\mathcal{S}_+(k)$, Φ is of the form

$$\Phi=X_2'\Gamma X_2+N_2. \tag{2.19}$$

Theorem 2.3. A necessary and sufficient condition for $S(\Sigma)\equiv S(I)$ is that Σ is of the form (2.9) or $\Sigma=X_2'\Gamma X_2+Z_2'\Delta Z_2$ for some $\Gamma\in\mathcal{S}_+(l)$ and $\Delta\in\mathcal{S}_+(p-l)$.

Proof. Let $N_1=I-X_1(X_1'X_1)^{-1}X_1'$ and $Q=I-\Sigma^{-1}X_2'(X_2\Sigma^{-1}X_2')^{-1}X_2$. Then from

$$Y-X_1B(\Sigma)X_2=N_1Y+(I-N_1)YQ,$$

we obtain

$$nS(\Sigma)=Y'N_1Y+Q'Y'(I-N_1)YQ. \tag{2.20}$$

Hence $S(\Sigma)\equiv S(I)$ for all Y implies

$$Q'Y'(I-N_1)YQ\equiv N_2Y'(I-N_1)YN_2 \quad \text{for all } Y. \tag{2.21}$$

Hence using $Q\Sigma^{-1}X_2'=0$, (2.21) implies $(I-N_1)YN_2\Sigma^{-1}X_2'\equiv 0$ for all Y, which in turn implies $N_2\Sigma^{-1}X_2=0$. This is nothing but the form of (2.4). Hence Σ must be of the form required. The converse is clear, completing the proof.

Combining this result with Corollary 2.1 yields

Corollary 2.4. A necessary and sufficient condition for $\hat{B}(\Sigma)\equiv B(I)$ and $S(\Sigma)\equiv S(I)$ is that Σ is of the form (2.9).

It is remarked that when $E\sim N(0,I_n\otimes\Sigma)$, the density of Y is expressed as

$$f(Y|B,\Sigma)=c|\Sigma|^{-np/2}\exp[-\frac{n}{2}\mathrm{tr}\Sigma^{-1}S(\Sigma)-\frac{1}{2}\mathrm{tr}\Sigma^{-1}X_2'(\hat{B}(\Sigma)-B)'$$
$$\times X_1'X_1(\hat{B}(\Sigma)-B)+\mathrm{tr}\Sigma^{-1}(Y-X_1\hat{B}(\Sigma)X_2)'X_1(\hat{B}(\Sigma)-B)X_2].$$

If Σ is of the form (2.9), by Corollary 2.4 $S(\Sigma)=S(I)$ and $\hat{B}(\Sigma)=\hat{B}(I)$ so that $(\hat{B}(I),S(I))$ is the MLE of (B,Σ). Otherwise the MLE of (B,Σ) is complicated (see Chapter 3).

Example 2.1 Consider the GMANOVA model (2.7) under the

following covariance structure

$$\Sigma=\sigma^2(1-\rho)I_p+\sigma^2\rho ee'$$

where $e=(1,\dots,1)'\in R^p$ and $-(p-1)^{-1}<\rho<1$. Then by Theorem 2.1 and Corollary 2.1 a necessary and sufficient condition for $\hat{B}(\Sigma)=\hat{B}(I)$ is that the column space $L(X_2')$ of X_2' is Σ-invariant or $\Sigma X_2'=X_2'A$ for some $A:l\times l$. Hence $\hat{B}(\Sigma)=\hat{B}(I)$ if and only if $ee'X_2'=X_2'A$ for some A, which holds if and only if either (1) $X_2e=0$ or (2) $L(X_2')$ contains e. In the case of (1), $e=Z_2'c$ for some $c\in R^l$ so that

(A) $\Sigma=X_2'[\sigma^2(1-\rho)(X_2X_2')^{-1}]X_2+Z_2'[\sigma^2(1-\rho)(Z_2Z_2')^{-1}+\sigma^2\rho cc']Z_2$,

while in the case of (2), $e=X_2'd$ for some $d\in R^l$ so that

(B) $\Sigma=X_2'[\sigma^2(1-\rho)(X_2X_2')^{-1}+\sigma^2\rho dd']X_2+Z_2'[\sigma^2(1-\rho)(Z_2Z_2')^{-1}]Z_2$.

Consequently by Theorem 2.3 under these covariance matrices, $S(\Sigma)\equiv S(I)$, where $S(\Sigma)$ is defined in (2.18).

2.2 *Choice between OLSE and GLSE.* Let $Y=X_1BX_2+E$ be the GMANOVA model in (2.7), where normality is assumed for E:

$$E\sim N(O,I_n\otimes\Sigma) \quad \text{with} \quad \Sigma\in\mathcal{S}_+(p) \tag{2.22}$$

By Corollary 2.1, the problem of testing the hypothesis that Σ belongs to the class

$$\mathcal{C}=\{\Sigma=X_2'\Gamma X_2+Z_2'\Delta Z_2,\ \Gamma\in\mathcal{S}_+(k),\ \Delta\in\mathcal{S}_+(l)\} \tag{2.23}$$

or the problem of testing

$$H:\Sigma\in\mathcal{C} \quad \text{versus} \quad K:\Sigma\notin\mathcal{C} \tag{2.24}$$

may be regarded as a problem of choice between the OLSE $\hat{B}(I)$ and a GLSE $\hat{B}(\Sigma)$, where Z_2 and $\hat{B}(\Sigma)$ are as before (see (2.8)). Here, if the hypothesis H is accepted, Σ is considered to belong to or to be close enough to the class $\mathscr{C}$ and the OLSE $\hat{B}(I)$ may be used rather than a GLSE $\hat{B}(\Sigma)$. It should be noted that this testing problem is testable in a growth curve model if $n>p$ and that it is regarded as a problem of choice between the OLSE and a GLSE if $l<p$. If $l=p$, $\hat{B}(I)=(X_1'X_1)^{-1}X_1'Y$ is the best linear unbiased estimator of B. Of course, when $K: \Sigma \notin \mathscr{C}$ is true, the covariance matrix of the OLSE is given by

$$(2.25)\qquad \mathrm{Var}(\hat{B}(I))=(X_1'X_1)^{-1}\otimes(X_2X_2')^{-1}X_2\Sigma X_2'(X_2X_2')^{-1},$$

while the covariance matrix of the Gauss-Markov estimator (GME) is given by

$$(2.26)\qquad \mathrm{Var}(\hat{B}(\Sigma))=(X_1'X_1)^{-1}\otimes(X_2\Sigma^{-1}X_2')^{-1}.$$

Hence the difference is

$$(2.27)\qquad \begin{aligned}&\mathrm{Var}(\hat{B}(I))-\mathrm{Var}(\hat{B}(\Sigma))\\ &=(X_1'X_1)^{-1}\otimes(X_2X_2')^{-1}X_2\Sigma Z_2'(Z_2\Sigma Z_2')^{-1}Z_2\Sigma X_2'(X_2X_2')^{-1},\end{aligned}$$

which is nonnegative definite, where (2.14) is used in the evaluation of the difference. Clearly this difference is zero if and only if $X_2\Sigma Z_2'=0$ or equivalently Σ belongs to the class $\mathscr{C}$. Therefore under $\Sigma \notin \mathscr{C}$, it is positive semi-definite. The relative efficiency of the OLSE is usually defined by

$$(2.28)\qquad \begin{aligned}\eta&=|\mathrm{Var}(\hat{B}(I))|/|\mathrm{Var}(\hat{B}(\Sigma))|\\ &=\{|X_2X_2'|^2/|X_2\Sigma^{-1}X_2'|\,|X_2\Sigma X_2|\}^k.\end{aligned}$$

From Bloomfield and Watson (1975) and Knott (1975), it follows that

$$(2.29)\qquad 1\geq\eta\geq\{\textstyle\prod_{i=1}^{t}[4\gamma_i\gamma_{p-i+1}/(\gamma_i+\gamma_{p-i+1})^2]\}^k \quad (t=\min(l,\ p-l))$$

where $\gamma_1\geq\cdots\geq\gamma_p>0$ are the characteristic roots of Σ. Therefore the lower bound for the efficiency η of the OLSE relative to the Gauss-Markov estimator $\hat{B}(\Sigma)$ depends on the dispersion of the roots $\{\gamma_i\}$

of Σ, and unless the variance $\Sigma(\gamma_i-\bar{\gamma}^2)/p$ is zero or $\gamma_1=\cdots=\gamma_p$, the lower bound cannot be 1. However, the bound is derived for which it holds for any X_2, and so for some X_2 the efficiency η can be close to 1. In fact, if $\Sigma\in\mathscr{C}$, $\eta=1$. This implies that the problem of a choice between the OLSE and a GLSE will be better to consider for a given X_2. The above testing problem concerns this problem and is treated in Chapter 5, where in deriving the LRT (likelihood ratio test) for the problem (2.24), the MLE (maximum likelihood estimator) is derived.

2.3. *Extension of the identity between OLSE and GLSE.* Here we extend the results in Theorem 2.1 and Corollary 2.1 to the case where on the regression coefficient, a linear restriction is imposed. Let

$$(2.30)\qquad y=X\beta+\varepsilon,\quad \mathrm{Cov}(\varepsilon)=\sigma^2\Phi,\quad X: N\times K,\quad \mathrm{rank}(X)=K$$

be a regression model with a linear restriction

$$(2.31)\qquad R\beta=r,\quad R: M\times K,\quad \mathrm{rank}(R)=M.$$

Then as is well known, when Φ is known, the best linear unbiased estimator (BLUE) or the Gauss Markov estimator is given by

$$(2.32)\qquad \bar{\beta}(\Phi)=b(\Phi)-H^{-1}R'(RH^{-1}R')^{-1}[Rb(\Phi)-r]$$

where $b(\Phi)$ is defined by (2.2) and

$$H=X'\Phi^{-1}X.$$

Here define

$$(2.33)\qquad Q=I-R'(RR')^{-1}R',\quad \gamma=Q\beta \quad\text{and}\quad \tilde{r}=R'(RR')^{-1}r.$$

Then $\beta=\gamma+\tilde{r}$ and the model in (2.30) is expressed as

$$(2.34)\qquad \tilde{y}=XQ\gamma+\varepsilon \quad\text{with}\quad \tilde{y}=y-X\tilde{r}.$$

Further, observing by the definition of the Penrose generalized inverse

$$(QHQ)^+=H^{-1}-H^{-1}R'(RH^{-1}R')^{-1}RH^{-1} \quad \text{and} \quad Q(QHQ)^+=(QHQ)^+,$$

the BLUE in (2.32) is expressed as

$$\tilde{\beta}(\Phi)=\tilde{\gamma}(\Phi)+\tilde{r} \quad \text{with} \quad \tilde{\gamma}(\Phi)=(QHQ)^+QX'\Phi^{-1}\tilde{y}. \tag{2.35}$$

and so the covariance matrix of $\tilde{\beta}(\Phi)$ is expressed as

$$\mathrm{Cov}(\tilde{\beta}(\Phi))=\mathrm{Cov}(\tilde{\gamma}(\Phi))=\sigma^2(QHQ)^+. \tag{2.36}$$

Now, since $\tilde{r}$ does not depend on Φ, from (2.35) a necessary and sufficient condition for $\tilde{\beta}(\Phi)\equiv\tilde{\beta}(I)$ is that

$$(QX'\Phi^{-1}XQ)^+QX'\Phi^{-1}=(QX'XQ)^+QX'.$$

Using $(QHQ)(QHQ)^+QX'\Phi^{-1}=QX'\Phi^{-1}$, this becomes

$$QX'\Phi^{-1}[I-XQ(QX'XQ)^+QX']=0.$$

Hence from Theorem 2.1 where no restriction exists, we obtain

Theorem 2.4. The OLSE under $R\beta=r$ is equal to the BLUE under $R\beta=r$ if and only if for some $\Gamma\in\mathcal{S}_+(M)$ and $\Delta\in\mathcal{S}_+(N-M)$

$$\Phi = XW\Gamma W'X' + Z\Delta Z' \tag{2.37}$$

where W is a $K\times(K-M)$ matrix of rank $K-M$ such that $WW'=Q$ and $W'W=I_{K-M}$, and Z is an $N\times(N-K+M)$ matrix of rank $N-K+M$ such that $Z'XW=0$.

We apply this result to the growth curve model

$$Y=X_1BX_2+E, \quad \mathrm{Cov}(E)=I_n\otimes\Sigma, \quad \Sigma\in\mathcal{S}_+(p) \tag{2.38}$$

with a linear restriction

$$R_1BR_2=R_0 \tag{2.39}$$

where $X_1 : n\times k$ of rank k, $X_2 : l\times p$ of rank l, $R_1 : r_1\times k$ of rank

r_1, and $R_2 : p\times r_2$ of rank r_2. As has been seen in 1.1, this model is regarded as a regression model in (2.30) with restriction (2.31) where $X=X_1\otimes X_2'$ and $R=R_1\otimes R_2'$. Hence, in our model here the BLUE in (2.32) is reduced to

$$\tilde{B}(\Sigma)=\hat{B}(\Sigma)-V_1[R_1\hat{B}(\Sigma)R_2-R_0]V_2(\Sigma) \tag{2.40}$$

where

$$\begin{aligned} &\hat{B}(\Sigma)=(X_1'X_1)^{-1}X_1'Y\Sigma^{-1}X_2'(X_2\Sigma^{-1}X_2')^{-1} \\ &V_1=(X_1'X_1)^{-1}R_1'[R_1(X_1'X_1)^{-1}R_1']^{-1} \quad \text{and} \\ &V_2(\Sigma)=[R_2'(X_2\Sigma^{-1}X_2')^{-1}R_2]^{-1}R_2'(X_2\Sigma^{-1}X_2')^{-1}. \end{aligned} \tag{2.41}$$

Now we let

$$Q=I_{lk}-R'(RR')^{-1}R' \quad \text{with} \quad R=R_1\otimes R_2'. \tag{2.42}$$

Corollary 2.5. $\tilde{B}(\Sigma)\equiv\tilde{B}(I)$ if and only if

$$I\otimes\Sigma=XW\Gamma W'X'+Z\Delta Z' \tag{2.43}$$

for some $\Gamma\in\mathcal{S}_+(r_1r_2)$ and $\Delta\in\mathcal{S}_+(lk-r_1r_2)$, where W and Z are defined for Q in (2.42) and $X=X_1\otimes X_2'$ in the same way as in Theorem 2.4.

It is noted that when the restriction $R_1BR_2=R_0$ is present and when $R_2\neq I$, the condition (2.43) is not expressed in terms of X_1 and X_2 only. But if the restriction is not present, the condition is reduced to the one in Corollary 2.1.

Though the covariance structure in (2.43) is a necessary and sufficient condition for the BLUE or GLSE $\tilde{B}(\Sigma)$ in (2.40) to be equal to the OLSE $\tilde{B}(I)$, it is difficult to put it in a form in which the covariance structure is testable. Hence we here consider the case $X_2=I$. In this case, the model (2.38) becomes the MANOVA model with a linear restriction (2.39);

$$(2.44)\qquad Y=X_1B+E,\quad \mathrm{Cov}(E)=I_n\otimes\Sigma,\quad R_1BR_2=R_0.$$

Then from (2.41), $\hat{B}(\Sigma)=(X_1'X_1)^{-1}X_1'Y\equiv\hat{B}(I)$, but $V_2(\Sigma)=V_2(I)$ does not necessarily hold unless $\mathrm{rank}(R_2)=r_2=p$ where $V_2(\Sigma)$ is given in (2.41). Therefore in the case $r_2<p$ the BLUE $\tilde{B}(\Sigma)$ in (2.40) is not identically equal to the OLSE $\tilde{B}(I)$ even under the model (2.44).

Theorem 2.5. In the model (2.44), $\tilde{B}(\Sigma)=\tilde{B}(I)$ if and only if

$$(2.45)\qquad \Sigma=R_2\Gamma R_2'+P_2\Delta P_2' \quad \text{for some } \Gamma\in\mathcal{S}_+(r_2) \text{ and } \Delta\in\mathcal{S}_+(p-r_2),$$

where P_2 is a $p\times(p-r_2)$ matrix of rank $p-r_2$ satisfying $P_2'P_2=I_{p-r_2}$ and $P_2'R_2=0$.

Proof. From (2.40) and $\hat{B}(\Sigma)\equiv\hat{B}(I)$, $\tilde{B}(\Sigma)=\tilde{B}(I)$ if and only if

$$V_1[R_1\hat{B}(I)R_2-R_0]V_2(\Sigma)=V_1[R_1\hat{B}(I)R_2-R_0]V_2(I).$$

By (2.41) with $R_1BR_2=R_0$ and $X_2=I$, this implies

$$R_1[\hat{B}(I)-B]\,[R_2(R_2'\Sigma R_2)^{-1}R_2'\Sigma-R_2(R_2'R_2)^{-1}R_2']=0.$$

Since $\hat{B}(I)-B$ spans $R^{k\times p}$, and since $\mathrm{rank}(R_1)=r_1\leq k$, this in turn implies

$$R_2(R_2'\Sigma R_2)^{-1}R_2'\Sigma=R_2(R_2'R_2)^{-1}R_2'.$$

In fact, $ABC=0$ for all B implies $A\otimes C'=0$ which implies $C=0$ unless $A=0$. Therefore in the same way as in the proof of Theorem 2.1, Σ is of the form (2.45). The converse is clear, proving the result.

The covariance structure (2.45) can be tested based on the model (2.44). The problem of testing Σ of the form (2.45) will be treated in Chapter 5.

Example 2.2. Let us consider Example 1.5, where the model is of the form (2.45) if the hypothesis (1.27) is a priori known, which is assumed in a classification model with covariates. Then by our theorem, the BLUE $\tilde{B}(\Sigma)$ under the restriction (1.27) is identically equal to the OLSE $\tilde{B}(I)$ if and only if Σ is of the form

$$\Sigma=\begin{pmatrix}0\\ I\end{pmatrix}\Gamma(0,I)+\begin{pmatrix}I\\ 0\end{pmatrix}\Delta(I,0)=\begin{pmatrix}\Delta & 0\\ 0 & \Gamma\end{pmatrix}\begin{matrix}p\\ q\end{matrix}\equiv\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$$

since $R_2=\begin{pmatrix}0\\ I\end{pmatrix}$, where $(p+q)$ and q in Example 1.4 correspond to p and r_2 here respectively. Hence $\tilde{B}(\Sigma)\equiv\tilde{B}(I)$ if and only if x_{1i} and x_{2i} (and y_{1j} and y_{2j}) are independent under normality or $\Sigma_{12}=0$. In this case, no additional efficiency in estimating μ_1 and η_1 can be expected by using the covariates x_{2i} and y_{2j} with $E(x_{2i})=E(y_{2j})$, and on the contrary using x_{2i} and y_{2j} under the independence will be less efficient than ignoring them. Hence before making a use of the covariates, it will be desirable to check if Σ is of the form above or if x_{2i} and y_{2j} are independent.

For references, discriminant problems under $\eta_1=\eta_2$ have been considered by Cochran and Bliss (1948), Rao (1949), Cochran (1964), Rao (1966), Memon and Okamoto (1970) etc. When Σ is known and $\Sigma_{12}\neq 0$, Cochran and Bliss (1948) constructed a discriminant function using x_{2i}'s and y_{2j}'s as well as x_{1i}'s and y_{1j}'s which is more efficient than the usual discriminant function using x_{1i}'s and y_{1j}'s. When Σ is unknown, they proposed a discriminant function in which Σ is replaced by an estimate. However when $\Sigma_{12}=0$ is close to zero, this discriminant function does not seem to be better than the usual one based on x_{1i}'s and y_{1j}'s. Of course, when $\Sigma_{12}=0$, it seems most reasonable to base discrimination solely on the basis of x_{1j}'s and y_{1j}'s. This motivates the problem of testing $\Sigma_{12}=0$ or if Σ is of the above form. This problem is treated in Exmaple 2.1 of Chapter 5.

2.4. *Case of SUR model.* It has been observed in 1.3 that an SUR model is an extended GMANOVA model. Here in an SUR model (1.4) with (1.5), a necessary and sufficient condition for the GLSE to be identically equal to the OLSE is derived. Rewrite the model as follows:

$$y = X\beta + \varepsilon, \qquad \mathrm{Cov}(\varepsilon) = \Sigma \otimes I_n \tag{2.46}$$

where

$$X = \begin{bmatrix} X_1 & & 0 \\ & \ddots & \\ 0 & & X_p \end{bmatrix} : n \times K, \qquad K = \Sigma k_i$$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} : np \times 1, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} : K \times 1 \quad \text{and} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix} : np \times 1.$$

Here the GLSE or the Gauss-Markov estimator is given by

$$b(\Sigma \otimes I) = (X'[\Sigma \otimes I]^{-1}X)^{-1}X'[\Sigma \otimes I]^{-1}y. \tag{2.47}$$

Then by Theorem 2.1, a necessary and sufficient condition for $b(\Sigma \otimes I) \equiv b(I \otimes I)$ is that for some $\Gamma \in \mathcal{S}_+(K)$ and $\Delta \in \mathcal{S}_+(np-K)$

$$\Sigma \otimes I = X\Gamma X' + Z\Delta Z', \tag{2.48}$$

where $Z : np \times (np-K)$ satisfies $Z'X = 0$ with $\mathrm{rank}(Z) = np-K$. Here we take

$$Z = \begin{bmatrix} Z_1 & & 0 \\ & \ddots & \\ 0 & & Z_p \end{bmatrix} \qquad \begin{matrix} \mathrm{rank}\ (Z_i) = n-k_i, \\ Z_i'X_i = 0 \end{matrix}$$

where $Z_i : n \times (n-k_i)$ $(i=1, \cdots, p)$. Further let $\Sigma = (\sigma_{ij})$,

$$\Gamma = (\Gamma_{ij}),\ \Gamma_{ij} : k_i \times k_j \quad \text{and} \quad \Delta = (\Delta_{ij}),\ \Delta_{ij} : (n-k_i) \times (n-k_j).$$

Then (2.48) is equivalent to

$$\sigma_{ij} I_n = X_i \Gamma_{ij} X_j' + Z_i \Delta_{ij} Z_j' \tag{2.49}$$

or equivalently

$$\Gamma_{ij} = \sigma_{ij}(X_i'X_i)^{-1}X_i'X_j(X_j'X_j)^{-1} \quad \text{and} \quad \Delta_{ij} = \sigma_{ij}(Z_i'Z_i)^{-1}Z_i'Z_j(Z_j'Z_j)^{-1}.$$

$(i, j=1, \cdots, p)$ This is a condition that σ_{ij} must satisfy for $b(\Sigma \otimes I) \equiv b(I \otimes I)$.

Next we consider a condition on X_i's for which $b(\Sigma \otimes I) \equiv b(I \otimes I)$ for all $\Sigma \in \mathcal{S}_+(p)$. Multiplying (2.49) by $(X_i'X_i)^{-1}X_i'$ from the left and by X_i from the right yields

$$\sigma_{ij} I_{k_i} = \Gamma_{ij} X_j' X_i.$$

This implies $k_j \geq k_i$, but from the symmetry of i and j it also implies $k_i \geq k_j$. Hence $k_i = k_j$ is necessary, under which

$$\Gamma_{ij} = \sigma_{ij}(X_j'X_i)^{-1}. \tag{2.50}$$

Further, pre-multiplying (2.49) by $M_i \equiv X_i(X_i'X_i)^{-1}X_i'$ and using (2.50) yields

$$\sigma_{ij} M_i = \sigma_{ij} X_i (X_j'X_i)^{-1} X_j'.$$

If $\sigma_{ij} \neq 0$, this implies

$$M_i = M_j,$$

under which $Z_i = Z_j$ can be taken.

Theorem 2.5. $b(\Sigma \otimes I) \equiv b(I \otimes I)$ for all $\Sigma \in \mathcal{S}_+(p)$ if and only if

$$M_1 = \cdots = M_p \tag{2.51}$$

Proof. The necessity is proved above, while the sufficiency follows by tracing back the proof of the necessity.

The condition (2.51) implies $k_1 = \cdots = k_p$. Of course, if $X_1 = \cdots = X_p$, (2.51) immediately follows. It is remarked that $M_1 = M_2$ if and only if $X_2 = X_1 A$ for some $A \in \mathcal{G}l(k_1)$.

3. Identifiability of Hypotheses.

3.1. *Definition.* Let

$$\mathcal{F}(\Theta) = \{f(x|\theta) | \theta \in \Theta\}$$

be a model or a class of pdf's (probability density functions) on R^n with respect to a σ-finite measure parametrized by θ, where Θ is a nonempty open set of R^k. For comparison, we give the usual definition of identifiability of $\mathcal{F}(\Theta)$.

Definition 3.1. A pdf $f(x|\theta_0)$ in $\mathcal{F}(\Theta)$ or simply $\theta_0 \in \Theta$ is said to be identifiable if for some $\theta_1 \in \Theta$

(3.1) $$f(x|\theta_0) = f(x|\theta_1) \quad \text{a.e. } x$$

implies $\theta_0 = \theta_1$. Further $\mathcal{F}(\Theta)$ or Θ is called identifiable if each $\theta \in \Theta$ is identifiable.

In any statistical problem, the identifiability of a model (or a class of pdf's under consideration) must be guaranteed in advance. In the below, assuming this identifiability for $\mathcal{F}(\Theta)$, we shall define the identifiability of a hypothesis testing problem. To treat the situation with nuissance parameter, let

(3.2) $$\mathcal{F}(\Theta) = \{f(x|\theta, \lambda),\ \theta \in \Theta,\ \lambda \in \Lambda\}$$

be a class of pdf's on R^n parametrized with (θ, λ) and consider a hypothesis

(3.3) $$H : \theta \in \Theta_0 \qquad (\Theta_0 \subset \Theta)$$

where $\Theta \times \Lambda$ is open in $R^k \times R^l$ and Θ_0 is a closed subset of Θ.

Definition 3.2. A hypothesis Θ_0 is said to be identifiable if for

some closed set $\Theta_1 \subset \Theta$

(3.4) $$\sup_{\theta \in \Theta_0} \sup_{\lambda \in \Lambda} f(x|\theta, \lambda) = \sup_{\theta \in \Theta_1} \sup_{\lambda \in \Lambda} f(x|\theta, \lambda) \quad \text{a.e.} x$$
implies $\Theta_0 = \Theta_1$.

If $\Lambda = \{\lambda_0\}$ (singleton) and if $\Theta_0 = \{\theta_0\}$ (simple hypothesis), this definition says that θ_0 is identifiable if $f(x|\theta_0, \lambda_0) = \sup_{\theta \in \Theta_1} f(x|\theta, \lambda_0)$ for some closed set Θ_1 implies $\Theta_1 = \{\theta_0\}$. Hence in this special case, Definition 3.2 is stronger than Definition 3.1. This fact may lead us to

Definition 3.3. A simple hypothesis $\theta_0 \in \Theta$ is said to be identifiable if for some $\theta_1 \in \Theta$.

(3.5) $$\sup_{\lambda \in \Lambda} f(x|\theta_0, \lambda) = \sup_{\lambda \in \Lambda} f(x|\theta_1, \lambda) \quad \text{a.e.} x$$

implies $\theta_0 = \theta_1$.

For simple hypotheses for Θ, we shall use this definition rather than Definition 3.2. Under this definition, if $\theta_0 \in \Theta$ is not identifiable, there exists $\theta_1 \neq \theta_0$ such that (3.5) holds. This implies that for testing θ_0 versus θ_1, the likelihood ratio is identically equal to 1 or

(3.6) $$\sup_{\lambda \in \Lambda} f(x|\theta_1, \lambda) / \sup_{\lambda \in \Lambda} f(x|\theta_0, \lambda) = 1 \quad \text{a.e.} x$$

It also implies that the problem of testing θ_0 versus $\Theta_2 (\subset \Theta)$ is not distinguishable from the problem of testing θ_1 versus Θ_2 from the likelihood ratio;

(3.7) $$\frac{\sup_{\theta \in \Theta_2} \sup_{\lambda \in \Lambda} f(x|\theta, \lambda)}{\sup_{\lambda \in \Lambda} f(x|\theta_0, \lambda)} = \frac{\sup_{\theta \in \Theta_2} \sup_{\lambda \in \Lambda} f(x|\theta, \lambda)}{\sup_{\lambda \in \Lambda} f(x|\theta_1, \lambda)}$$

On the other hand, Definition 3.2 may be associated with model selection. Let $\mathcal{F}(\Theta_i \times \Lambda)$ be a submodel of the model $\mathcal{F}(\Theta \times \Lambda)$ in (3.2), where Θ_i is a closed subset of $\Theta (i = 0, 1)$, and let $\hat{\theta}_i(x)$ and

$\lambda(x)$ be the MLE's (maximum likelihood estimators) for θ_i and λ under $\mathscr{F}(\Theta_i \times \Lambda)$ respectively. One may think of the selection of variables in regression. Then, according to our Definition 3.2, the two models $\mathscr{F}(\Theta_i \times \Lambda)$ $(i=0,1)$ are not identifiable if as in (3.4) the maximum likelihoods of the two models are identically equal;

$$(3.8) \qquad f(x|\hat{\theta}_0(x), \hat{\lambda}(x)) = f(x|\hat{\theta}_1(x), \hat{\lambda}(x)) \quad \text{a.e.} x$$

and if $\Theta_0 \neq \Theta_1$. Hence the concept of the identifiability in Definition 3.2 is also useful in the selection of models based on an information criterion. In particular, in AIC (Akaike's information criterion), the information for $\mathscr{F}(\Theta_i \times \Lambda)$ is defined as

$$(3.9) \quad I(\Theta_i \times \Lambda) = -2\log f(x|\hat{\theta}_i(x), \hat{\lambda}(x)) + 2p_i \ \ (p_i = \dim\Theta_i + \dim\Lambda).$$

If (3.8) holds and if $p_0 = p_1$, then the two models are regarded as unidentifiable, since $I(\Theta_0 \times \Lambda) = I(\Theta_1 \times \Lambda)$.

3.2 *Examples.*

Example 3.1. Let $y = X\beta + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2\Phi)$ be a normal regression model where X is an $n \times k$ fixed matrix of rank k and $\Phi \in \mathscr{S}_+(n)$ (the set of positive definite matrices). Here $\Lambda = \{(\beta, \sigma^2)\} | \beta \in R^k, \sigma^2 \in R\} = R^k \times R_+$ and $\Theta = \mathscr{S}_+(n)$. Then, the simple hypothesis $\theta = I$ in Θ is not identifiable by Definition 3.3 because by Corollary 2.2, for any θ_1 belonging to the class

$$\Theta_1 = \{\Phi | \Phi = X\Gamma X' + I - X(X'X)^{-1}X', \ \ \Gamma \in \mathscr{S}_+(k), |\Phi| = 1\} \ \ (\subset \Theta),$$

we obtain

$$\sup_{(\beta,\sigma^2)} f(x|\theta_0, (\beta, \sigma^2)) = \sup_{(\beta,\sigma^2)} f(x|\theta_1, (\beta, \sigma^2)).$$

Even by Definition 3.2, $\theta_0 = I$ is not identifiable because

$$\sup_{(\beta,\sigma^2)} f(y|\theta_0, (\beta, \sigma^2)) = \sup_{\theta \in \Theta_1} \sup_{(\theta,\sigma^2)} f(y|\theta, (\beta, \sigma^2)).$$

In fact, by Corollary 2.2, the MLE is the GLSE $b(\Phi)$ which is

identically equal to the OLSE $b(I)$, and the MLE of σ^2 is the sample variance $s^2(\Phi)$ in (2.10) which is identically equal to $s^2(I)$, so that for any $\Phi \in \Theta_1$

$$\begin{aligned}\sup_{(\beta,\sigma^2)} f(y|\Phi, (\beta, \sigma^2)) &= e^{-n/2}(2\pi s^2(I))^{-n/2}|\Phi|^{-1/2} \\ &= e^{-n/2}(2\pi s^2(I))^{-n/2} = \sup_{(\beta,\sigma^2)} f(y|I, (\beta, \sigma^2))\end{aligned}$$

If we do not require the condition $|\Phi| = 1$ in Θ_1, the likelihood ratio is

$$\begin{aligned}\sup_{(\beta,\sigma^2)} f(y|\Phi, (\beta, \sigma))/\sup_{(\beta,\sigma^2)} f(y|I, (\beta, \sigma^2)) \\ = [s^2(I)/s^2(\Phi)]^{n/2}|\Phi|^{-1/2} = |\Phi|^{-1/2}\end{aligned}$$

because $s^2(\Phi) = s^2(I)$. This is not necessarily one but a constant.

Example 3.2. In Example 3.1, assume $\varepsilon \sim N(0, \tau\Phi(\rho))$, where $\tau > 0$,

$$(3.10) \qquad \Phi(\rho) = I_n + \rho A, \quad \rho \in \{\rho | \Phi(\rho) \in \mathscr{S}_+(n)\}$$

and $A: n \times n$ is known. This contains exact or approximate models for serial correlation including the autoregressive model of the first order (see Anderson (1948) and Kadiyala (1970)). We here consider the problem

$$(3.11) \qquad H: \Phi(0) = I \quad \text{versus} \quad K: \Phi(\rho), \ \rho \neq 0 \ (\text{or } \rho > 0).$$

Then for a fixed ρ, the LRT rejects H when

$$\begin{aligned}\sup_{(\beta,\tau)} f(y|\beta, \tau, \Phi(\rho))/\sup_{(\beta,\tau)} f(y|\beta, \tau, I) \\ = |\Phi(\rho)|^{-1/2}[s^2(I)/s^2(\Phi(\rho))]^{n/2} > c.\end{aligned}$$

By Theorem 2.2, for a fixed ρ, H and K are observationally equivalent as hypotheses if and only if for some symmetric matrix B

$$(3.12) \qquad (I + \rho A)^{-1} = I + B - NBN$$

In general, for a given X, (3.12) does not hold exactly. But if it holds approximately, the likelihood ratio remains near $|\Phi(\rho)|^{-1/2}$

for any y. In such a case, the LRT is insensitive and close to the trivial test that rejects H when a tossed coin appears head, where the probability of the appearance is assumed to be equal to the significance level. The situation is the same for the case of the Durbin-Watson test, which can be regarded as an approximation to the LRT. Further, if the model (3.10) does not well approximate the process of data generation, there may exist observationally equivalent hypotheses for H or K.

As an example for which (3.12) holds, let us consider the intraclass covariance matrix

$$\Psi(\sigma^2, \lambda)=\sigma^2[(1-\lambda)I+\lambda ee'] \quad \text{where } e=(1, \cdots, 1)' \in R^n.$$

Then with $\tau=\sigma^2(1-\lambda)$, $\rho=\lambda/(n\lambda-\lambda+1)$ and $A=-ee'$, $\Psi(\sigma^2, \lambda)$ is expressed as $\tau\Phi(\rho)$ and the problem of testing $H: \lambda=0$ versus $K: \lambda\neq 0$ (or $\lambda>0$) is equivalent to (3.11), where $\rho<1/n$ from $-1/(n-1)<\lambda<1$. Because $\Phi(\rho)=I+[\rho/(1-n\rho)]ee'$, (3.12) holds with $B=[\rho/(1-n\rho)]ee'$ if $Ne=0$ or the column space $L(X)$ of X contains e, in particular if the regression model contains a constant term, regardless of the value of ρ, H and K are not identifiable as hypotheses. On the other hand, as has been seen, if $L(X)$ contains e, $b(\Phi(\rho))=b(I)$. Consequently in a regression model with intraclass covariance structure, if $L(X)$ contains e, $(b(I), s^2(I))$ is as efficient as $(b(\Psi(\sigma^2, \lambda)), s^2(\Psi(\sigma^2, \lambda)))$, and hypotheses concerning λ are not identifiable.

Let C be an $n\times n$ orthogonal matrix such that $C'AC=D=\text{diag}\{d_1, \cdots, d_n\}$. Then by pre- and post- multiplying (3.12) by C' and C respectively, we get $(I+\rho D)^{-1}=I+\Delta$, where $\Delta=C'(B-NBN)C=\text{diag}\{\delta_1, \cdots, \delta_n\}$. Hence (3.12) holds if and only if $\delta_i=(1+\rho d_i)^{-1}-1$ $(i=1, \cdots, n)$.

The following example is different from the context of 3.1, but it is closely related.

Example 3.3. Let $X_1, \cdots, X_n$ be a random sample from $N_p(\mu, \Sigma)$ or

$$X=[X_1, \cdots, X_n]': n\times p \sim N(e_n\mu', I_n\otimes\Sigma)$$

where $(\mu, \Sigma)\in\Theta=R^p\times\mathcal{S}_+(p)$ and $e_n=(1, \cdots, 1)'\in R^n$. In this model, let us consider the problem of testing independence

$$(3.13) \qquad (\mu, \Sigma)\in\Theta_0=\{(\mu, \Sigma)\in\Theta \mid \Sigma_{12}=0\} \quad \text{versus} \quad (\mu, \Sigma)\in\Theta.$$

where $\Sigma=(\Sigma_{ij})$ with $\Sigma_{ij}: p_i\times p_j$ $(i, j=1, 2)$. Then as is well known, the LRT is given by

$$L=\sup_{(\mu,\Sigma)\in\Theta} f(X\mid\mu, \Sigma)/\sup_{(\mu,\Sigma)\in\Theta_0} f(X\mid\mu, \Sigma)=|I-S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}|^{-n/2}$$

where S is the sample covariance matrix. On the other hand, take the restricted model $\{N_p(\mu, \Sigma) \mid (\mu, \Sigma)\in\tilde{\Theta}\}$ where $\tilde{\Theta}=\{(\mu, \Sigma)\in\Theta \mid \Sigma_{22}=I_{p_2}\}\subset\Theta$ and in this model consider the problem of testing independence

$$(3.14) \qquad (\mu, \Sigma)\in\tilde{\Theta}_0=\{(\mu, \Sigma)\in\tilde{\Theta} \mid \Sigma_{12}=0\} \quad \text{versus} \quad (\mu, \Sigma)\in\tilde{\Theta}.$$

Then it is shown in Chapter 5 that the LRT is also given by

$$\begin{aligned}\tilde{L}&=\sup_{(\mu,\Sigma)\in\tilde{\Theta}} f(X\mid\mu, \Sigma)/\sup_{(\mu,\Sigma)\in\tilde{\Theta}_0} f(X\mid\mu, \Sigma)\\ &=|I-S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}|^{-n/2}.\end{aligned}$$

Consequently the two problems (3.13) and (3.14) are not identifiable through the LR principle. In other words, the LRT statistic itself does not tell us which problem is tested.

Chapter 2

INVARIANCE APPROACH TO TESTING

1. A Review on Testing Theory.

1.1. *The Neyman-Pearson framework.* In the analyses of testing problems, we shall adopt a decison-theoretic approach established by Neyman, Pearson and Wald. For the details of this approach, the readers may refer to the classical but immortal book by Lehmann (1959). In this section a basic framework of the approach is reviewed with an emphasis on testing theory via invariance. Let

(1.1) $$\mathscr{F}(\Theta) = \{f(x|\theta) \mid \theta \in \Theta\} \quad (x \in \mathscr{X})$$

be a class of probability densities on a subset $\mathscr{X}$ of an Euclidean space with respect to a sigma-finite measure μ, where the functional form of f is known but f is parametrized by an unknown vector θ in a subset Θ of an Enclidean space. Of course $\mathscr{F}(\Theta)$ is assumed to be identifiable (see Section 3 of Chapter 1). Then a testing problem is described by two disjoint subsets of Θ, say Θ_0 and Θ_1, as follows:

(1.2) $$H : \theta \in \Theta_0 \quad \text{versus} \quad K : \theta \in \Theta_1$$

where H is considered a maintained or null hypothesis to be tested against the alternative hypothesis K. A decision to make here is

either to accept (not to reject) the null hypothesis H or to reject H relative to the alternative K based on an observation x. A decision function, which we shall call a test or a test function below, is a measurable function $\phi(x)$ from $\mathscr{X}$ into $[0,1]$, denoting the probability that H is rejected when x is observed. In particular, if $\phi(x)=1$, H is rejected. The average probability of a test for rejection under θ is defined by

(1.3) $$\pi(\phi, \theta) = E_\theta[\phi(x)] = \int_{\mathscr{X}} \phi(x) f(x|\theta) \mu(dx)$$

and this function is called the power function of a test ϕ when $\theta \in \Theta_1$ or when K is true, meaning the average probability of the correct decision under a test ϕ when K is true. As is well known, in the Neyman-Pearson theory, in the class of level α tests defined by

(1.4) $$\mathscr{D} = \{\phi \mid \phi \text{ is a test, } \sup_{\theta \in \Theta_0} \pi(\phi, \theta) \leq \alpha\} \quad (0 < \alpha < 1)$$

we wish to find a test which maximizes $\pi(\phi, \theta)$ under $\theta \in \Theta_1$ in some sense. Here the number α is called a significance level, and it corresponds to the maximum average probability of the incorrect decision under the test ϕ when H is true. The size of a test is defined to be the number $\sup_{\theta \in \Theta_0} \pi(\phi, \theta)$. In maximizing power, the following Generalized Neyman-Pearson Lemma is usually used.

Lemma 1.1. (See, e.g., Lehmann (1959) p. 83) Let $f_1, \ldots, f_{m+1}$ be real-valued functions defined on $\mathscr{X}$ and integrable with respect to μ, and suppose that for given constants $c_1, \ldots, c_m$, there exists a test function ϕ satisfying

(*) $$\int \phi f_i d\mu = c_i \quad (i=1, \ldots, m).$$

Let $\mathscr{C}$ be the class of tests for which (*) holds. Then there exists a test ϕ in $\mathscr{C}$ that maximizes

(**) $$\int \phi f_{m+1} d\mu$$

and if a test ϕ in $\mathscr{C}$ is of the form

$$(***)\qquad \phi(x)=\begin{cases}1 & \text{if } f_{m+1}(x)>\sum_{i=1}^{m}k_i f_i(x) \quad \text{a.e.}\\ 0 & \text{if } f_{m+1}(x)<\sum_{i=1}^{m}k_i f_i(x) \quad \text{a.e.}\end{cases}$$

for some constants $k_1, \ldots, k_m$, then it maximizes (**) in $\mathscr{C}$. If a test ϕ in $\mathscr{C}$ satisfies (***) with $k_i \geq 0 (i=1, \ldots, m)$, it maximizes (**) in the class of tests satisfying

$$\int \phi_0 f_i d\mu \leq c_i \qquad (i=1, \ldots, m).$$

If $m=1$ with $f_1(x)=f(x|\theta_0)$ and $f_2(x)=f(x|\theta_1)$ where $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$ are fixed, the above lemma is reduced to the Neyman and Pearson fundamental lemma.

However, except certain specific cases, there exists in general no test that maximizes the power $\pi(\phi, \theta)$ uniformly in $\theta \in \Theta_1$ in the class of level α tests $\mathscr{D}$. An approach to a solution to this indeterminacy is to restrict by a certain criterion the class $\mathscr{D}$ to a subclass of $\mathscr{D}$, in which we try to find a test which maximizes the power as much as possible or whose power function behaves well for the purpose of an analysis. Criteria to restrict $\mathscr{D}$ to a subclass of $\mathscr{D}$ may be categorized as follows :

1. Like invariance, associated with the decision-theoretic structure of a problem concerned, a natural structure is required for a test.
2. Like unbiasedness, similarity and minimaxity, an optimal property is required for the behavior of power function, and only a class of tests whose power functions satisfy this property is considered.

In addition, we may add the following two more criteria to restrict $\mathscr{D}$ to a subclass consisting of a single test or to derive a test.

3. Like the likelihood ratio (LR) principle, a derivation method is adopted which gives a unique test without reference to the power function.
4. Like the studentization of a should-be- or may-be-good estimator,

ad hoc or intuitive approaches are adopted.

Of course, these criteria are more or less mutually related. In fact, in most univariate cases and in some multivariate cases, these approaches often give a same test. However, the principle of unbiasedness or similarity is no longer effective in such incomplete models as GMANOVA model, SUR model etc. which we concern in this book, because of the lack of the completeness of the models.

For the completeness of a model an excellent exposition is found in Chapter 4 of Lehmann (1959). On the other hand, the invariance approach in 1 is often applicable to incomplete multivariate normal linear models. The likelihood ratio (LR) principle is also always applicable to any model whenever the form of the probability density is known though it often fails to give an explicit form of the test. The philosophy behind this LR principle differs from that of the Neyman-Pearson theory and so it might not be appropriate to compare the LRT (likelihood ratio test) to other optimal tests in terms of power. However, the LRT is always invariant (see Lehmann (1959) page 252 or Eaton (1983) page 263), hence it is naturally included in the analysis through invariance. The ad hoc approach in 4 is often difficult to apply in the case of incomplete models because a good estimator is difficult to find there. In addition, the studentization of a good estimator does not necessarily produce a good test (see Chapter 4). In some cases, approaches 3 and 4 are combined. For example, some parameters are first assumed to be known and a form of the MLE (maximum likelihood estimator) is obtained under this assumption. Then the MLE is studentized and some estimates are substituted for the unknown parameters which were assumed to be known. Alternatively, assuming some parameters are known, the LRT is derived and then estimates are substituted for the unknown parameters. However even in these cases, the tests thus obtained are mostly invariant. Based on these aspects, in this book we adopt the invariance criterion as our basic standpoint but we do not completely neglect the other criteria.

1.2. *Invariance.* For theories on invariance, Ferguson (1968) or Lehmann (1959) or Eaton (1983) will be served as basic references.

To describe the invariance principle in our testing theory, let

$$(1.5)\qquad \mathcal{P}(\Theta) = \{P_\theta \mid dP_\theta/d\mu = f(x\mid\theta),\ f(x\mid\theta)\in\mathcal{F}(\Theta)\}$$

where $\mathcal{F}(\Theta)$ is given by (1.1), and let $\mathcal{G}$ be a group of bijective and bimeasurable transformations from $\mathcal{X}$ onto $\mathcal{X}$ such that

$$(1.6)\qquad P_\theta\in\mathcal{P}(\Theta) \Rightarrow gP_\theta\in\mathcal{P}(\Theta) \quad \text{for any } g\in\mathcal{G}$$

where $gP_\theta(A)=P_\theta(g^{-1}(A))$ for a Borel set A of $\mathcal{X}$. When a given model $\mathcal{P}(\Theta)$ in (1.5) or $\mathcal{F}(\Theta)$ in (1.1) admits such a group $\mathcal{G}$, it is said to be invariant under the group $\mathcal{G}$ or $\mathcal{G}$ is said to preserve model $\mathcal{P}(\Theta)$. The relation (1.6) with the identifiability of $\mathcal{F}(\Theta)$ means that there exists a parameter θ' in Θ such that $gP_\theta=P_{\theta'}$ and the correspondence between θ and θ' is designated by $\theta'=\bar{g}\theta$. Then $\bar{\mathcal{G}}=\{\bar{g}\mid g\in G\}$ forms a group as a homomorphic image of $\mathcal{G}$. In this situation, if the induced group $\bar{\mathcal{G}}$ leaves the testing problem (1.2) invariant in the sense that

$$(1.7)\qquad \bar{g}(\Theta_0)=\Theta_0 \quad\text{and}\quad \bar{g}(\Theta_1)=\Theta_1 \quad \text{for any } \bar{g}\in\bar{\mathcal{G}},$$

it is said that the testing problem is left invariant under the group $\mathcal{G}$. The condition (1.7) means that $\mathcal{G}$ preserves the hypotheses or $\mathcal{G}$ preserves $\mathcal{P}(\Theta_i)$'s. Now when a testing problem is left invariant under a group $\mathcal{G}$, we require a test function $\phi(x)$ to be invariant;

$$(1.8)\qquad \phi(gx)=\phi(x) \quad \text{for all } g\in\mathcal{G} \text{ and } x\in\mathcal{X},$$

and restrict our attention to the class of invariant level α tests, say $\mathcal{D}_I$. The requirement (1.8) says that any point in the orbit of x defined by

$$\mathcal{G}x=\{gx\in\mathcal{X}\mid g\in\mathcal{G}\}$$

is considered equivalent to x for making a decision on the problem. That is, we reject the null hypothesis with the same probability $\phi(x)$ for any observation falling in $\mathcal{G}x$ Hence if $\phi(x_1)\neq\phi(x_2)$, x_1 and x_2 are on different orbits or $\mathcal{G}x_1\neq\mathcal{G}x_2$, A measurable function $T(x)$ from $\mathcal{X}$ onto a measurable space $\mathcal{T}$ satisfying

(a) $T(gx)=T(x)$ for all $g\in\mathcal{G}$ and

(b) $T(x_1)=T(x_2)$ implies $x_1=gx_2$ for some $g\in\mathcal{G}$

is called a maximal invariant. Condition (a) says that a maximal invariant is constant on orbits whereas condition (b) says that each orbit gets a different value, that is, T distinguishes orbits. Therefore an invariant test $\phi(x)$ is expressed as

$$(1.9)\qquad \phi(x)=\psi(T(x)) \quad \text{for some test } \psi \text{ in } \mathcal{T}.$$

Also a maximal invariant parameter $\tau(\theta)$ is defined to be a map τ from Θ onto a space Υ such that

$$\tau(\bar{g}\theta)=\tau(\theta) \quad \text{for all } \bar{g}\in\bar{\mathcal{G}} \text{ and}$$
$$\tau(\theta_1)=\tau(\theta_2) \quad \text{implies} \quad \theta_1=\bar{g}\theta_2 \quad \text{for some } \bar{g}\in\bar{\mathcal{G}}.$$

Here the measurability of τ is not necessarily required though it is usually measurable. Under this situation, the following lemma is well known and easy to prove (see, e.g., Lehmann (1959) page 220).

Lemma 1.2. The distribution of a maximal invariant $T(x)$ under θ depends on θ only through a maximal invariant parameter $\tau(\theta)$. Hence for any $\phi\in\mathcal{D}_I$, the power function $\pi(\phi,\theta)=E_\theta[\phi(x)]$ depends on θ only through $\tau(\theta)$.

With this lemma, the problem of testing $H:\theta\in\Theta_0$ versus $K:\theta\in\Theta_1$ in (1.2) is now reduced by invariance as follows. Let $\mathcal{D}_I^T$ be the class of level α tests on the range space $\mathcal{T}$ of $T(x)$ and let

$$\mathcal{P}^T = \{P_\tau^T \mid \tau=\tau(\theta),\ P_\tau^T = P_\theta \circ T^{-1},\ P_\theta \in \mathcal{P}(\Theta)\} \tag{1.10}$$

be the class of probability distributions induced by $T=T(x)$ through $\mathcal{P}(\Theta)$ in (1.5). Then the problem is regarded as

$$H' : \tau\in\tau(\Theta_0) \quad \text{versus} \quad K' : \tau\in\tau(\Theta_1) \tag{1.11}$$

and in this reduced problem, we wish to maximize in some sense the power function

$$\tilde{\pi}(\psi,\ \tau) = E_\tau[\psi(T)] = E_\theta[\psi(T(x))] = \pi(\psi\circ T,\ \theta)$$

with respect to ψ in $\mathcal{D}_I^T$. In this maximization, the (Generalized) Neyman-Pearson Lemma is applied to the class $\mathcal{P}^T$ in (1.10). Or we may stay with the original problem and try to directly maximize the power $\pi(\phi, \theta)$ in the class of invariant tests $\mathcal{D}_I$. A test ϕ^* in $\mathcal{D}_I$ (or ψ^* in $\mathcal{D}_I^T$ with $\phi^*=\psi^*\circ T$ a.e.) is called a UMPI(uniformly most powerful invariant) test if for any $\phi\in\mathcal{D}_I$

$$\pi(\phi^*,\ \theta) \ge \pi(\phi,\ \theta) \quad \text{for all } \theta\in\Theta_1,$$

and a test ϕ^* in $\mathcal{D}_I$ (or ψ^* in $\mathcal{D}_I^T$ with $\phi^*=\psi^*\circ T$ a.e.) is called an LBI (locally best invariant) test if for any $\phi\in\mathcal{D}_I$ there exists an open neighborhood Θ_1 of Θ_0 such that

$$\pi(\phi^*,\ \theta) \ge \pi(\phi,\ \theta) \quad \text{for all } \theta\in\Theta_1-\Theta_0.$$

Of course, even in the class $\mathcal{D}_I$, there exist in general no UMPI test and no LBI test. But an LBI tests exist in many multivariate testing problems under normality assumption.

To summarize, the following procedures are usually taken for analysis of an invariant testing problem.

(1) To find a group which leaves the given problem invariant.

(2) To choose a convenient maximal invariant under the group found.

(3) To derive the distribution of the maximal invariant chosen.

(4) To derive an optimal (UMPI or LBI etc.) test based on the distribution.

(5) To derive or approximate the null distribution of the optimal test derived.

In addition, one may also check the performance of the power function. For this purpose, one may consider the following things.

(6) To derive the nonnull distribution of the test and check the behavior of the power function.

(7) To investigate whether the test is minimax or locally minimax. A test ϕ^* is called minimax in $\mathcal{D}$ with respect to Δ if

$$\inf_{\theta\in\Delta}\pi(\phi^*,\ \theta) = \sup_{\phi\in\mathcal{D}}\inf_{\theta\in\Delta}\pi(\phi,\ \theta)$$

where Δ is a suitable subset of Θ_1. The definition of local minimaxity is given in Section 6.

(8) To examine whether the power function is monotonically increasing as the parameter goes further from the null hypothesis. This property is often called the monotonicity property of a power function.

(9) To check whether the test is admissible in $\mathcal{D}_I$ or even in $\mathcal{D}$. A test ϕ^* of level α in $\mathcal{D}_I$ (or $\mathcal{D}$) is called admissible if there exists no test ϕ of size α in $\mathcal{D}_I$ (or $\mathcal{D}$) such that

$$\pi(\phi,\ \theta) \ge \pi(\phi^*,\ \theta) \quad \text{for all} \quad \theta\in\Theta_1 \quad \text{and}$$
$$\pi(\phi,\ \theta) \le \pi(\phi^*,\ \theta) \quad \text{for all} \quad \theta\in\Theta_0$$

with strict inequality for at least one $\theta\in\Theta_0\cup\Theta_1$.

(10) To investigate if the test is robust. In other words, one may question if, when the model is enlarged, the null distribution of the test remains the same and the optimality of the test is preserved. Those properties are often called null robustness and optimality robustness respectively.

Practically speaking, once the procedures from (1) to (5) are performed, the problem is basically solved and the test can be applied to practical problems. In particular, if a UMPI test is found, the procedures from (6) to (9) can be saved, provided the invariance principle is accepted. This is because the UMPI test dominates any other test in $\mathscr{D}_I$ in power. However, the UMPI test does not always exist and then one may look for a test with such a local optimality as LBI property. In fact, though it depends on the interest of analysis, the LBI property will be an important alternative to the UMPI property in the sense that when the parameter is far distant from the null hypothesis, most of "reasonable" tests can detect easily that the null hypothesis is false.

1.3. *Significance probability.* The Neyman-Pearson testing theory is frequently criticized for the arbitrariness of choice of significance level. However, once an optimal test is obtained with critical region of the form $u=u(x)>c$, then the test statistic u would be considered a measure of showing how strongly the data contradicts (or supports) the hypothesis. It is not quite so not only because the scale of u is arbitrary but because it depends on the probaility law of x. A correct measure is the significance probability or critical probability given by

$$\hat{\alpha}(u(x_0))=\sup_{\theta\in\Theta_0} P_\theta(u(x)>u(x_0)) \tag{1.12}$$

where x_0 is an observation. As Lehmann (1959) stated, it is good and important practice to determine not only whether the hypothesis is accepted or rejected at the given significance level, but also to determine the smallest significance level $\hat{\alpha}(u(x_0))$ in (1.12) at which the hypothesis would be rejected for the given observation. This number gives an idea of how strongly the data contradicts or supports the hypothesis, and enables those who use the statistical result to reach a verdict based on the significance level of their choice.

1.4. The following lemmas are sometimes used in this book.

Lemma 1.3. Let $f(x|\theta)=c(\theta)\exp[\Sigma_{j=1}^k \theta_j t_j(x)]$ be a density on $\mathscr{X}$ with respect to a σ-finite measure μ where $\theta\in\Theta$ and Θ is open in R^k. Then for any integrable function $\phi(x)$, the derivatives of all orders with respect to the θ_j's of $h(\theta)=\int\phi(x)f(x|\theta)\ d\mu$ can be computed under the integral sign.

Lemma 1.4. Suppose a group $\mathscr{G}$ acting on $\mathscr{X}$ is generated by two subgroups $\mathscr{H}$ and $\mathscr{K}$ and let $s(x)$ be a maximal invariant under $\mathscr{H}$ satisfying the condition :

$$s(x)=s(x') \quad \text{implies} \quad s(kx)=s(kx') \quad \text{for any } k\in\mathscr{K}.$$

Let $\mathscr{Y}$ be the range space of s and let $\mathscr{K}^*=\{k^*|k^*y=s(kx)$ for $y=s(x)$, $k\in\mathscr{K}$, $x\in\mathscr{X}\}$. Then $t(s(x))$ is a maximal invariant under $\mathscr{G}$ where t is a maximal invariant under $\mathscr{K}^*$.

These lemmas are found, e.g., in Lehmann (1959).

2. Maximality of a Group Leaving a Problem Invariant.

2.1. *Normality preserving transformation.* As has been summarized in Section 1, the first step to analyze a testing problem by invariance is to find a group of transformations leaving the problem invariant. There according to the invariance principle, it will be desirable to choose a maximal group if possible. The problem of the maximality has never been explicitly questioned even in a specific problem until Banken (1984), where a group leaving the GMANOVA problem invariant is shown to be maximal in the general affine linear group. Of course, such maximality should be in general questioned in the group under the action of which a given model is preserved. Then a maximal subgroup of the maximal group preserving null and alternative hypotheses concerned in the model should be taken. In multivariate testing problems, it is often the case that a normal linear model is adopted and a subgroup of the general affine linear group is chosen as a group preserving the normal model. In this section, we consider the maximality of such a group preserving a normal model. Let R^{np} denote the set of all $n\times p$ matrices and let

$$(2.1)\qquad \mathcal{P}(\mathcal{M}\times\mathcal{S}_+(p))=\{N(M, I_n\otimes\Sigma)\,|\,(M,\Sigma)\in\mathcal{M}\times\mathcal{S}_+(p)\}$$

be a model of np-dimmensional normal distributions with mean $M\in\mathcal{M}$ and covariance matrix of the form $I_n\otimes\Sigma$ where $\mathcal{M}\subset R^{np}$ is a set of $n\times p$ matrices (see Section 1 of Chapter 1 for the notation $N(M,I_n\otimes\Sigma)$). This is a typical model we often encounter in multi-variate testing problems. Since a transformation h in a group preserving $\mathcal{P}(\mathcal{M}\times\mathcal{S}_+(p))$ (i.e., $P\in\mathcal{P}(\mathcal{M}\times\mathcal{S}_+(p))$ implies $hP=Ph^{-1}\in\mathcal{P}(\mathcal{M}\times\mathcal{S}_+(p))$) should beb ijective and Lebesgue bimeasurable, let $\mathcal{H}$ be the group of all bijective and Lebesgue bimeasurable transformations from R^{np} onto R^{np} and let $\mathcal{H}^*$ be the group of all homeomorphisms in $\mathcal{H}$. Then $\mathcal{H}^*$ is a subgroup of $\mathcal{H}$. The following theorem charac-terizes a transformation h in $\mathcal{H}$ which preserves the model (2.1).

Theorem 2.1. (Nabeya and Kariya (1984)) Let $h\in\mathcal{H}$, and let $M^{(k)}\in R^{np}$ $(k=0,\ldots,n)$ satisfy

(2.2) for some $c\in R^p$, $M^{(k)}c$'s are not contained in a hyperplane of R^n.

Suppose that for any $M=M^{(k)}$ $(k=0,\ldots,n)$ and for any $\Sigma\in\mathcal{S}_+(p)$, there exists $(\Theta^{(k)}, \Psi^{(k)})\in R^{np}\times\mathcal{S}_+(p)$ such that $h(X)\sim N(\Theta^{(k)}, I_n\otimes\Psi^{(k)})$ when $X\sim N(M^{(k)}, I_n\otimes\Sigma)$. Then there exist constant matrices $P\in\mathcal{O}(n)$, $A\in\mathcal{G}l(p)$ and $H\in R^{np}$ such that

$$(2.3)\qquad h(x)=PxA+H \qquad \text{for a.e. } x\in R^{np}.$$

The condition (2.2) in Theorem 2.1 is always satisfied for the model (2.1) if the space $\mathcal{M}$ of M is R^{np}. However, if $\mathcal{M}=\{\begin{pmatrix}M_1\\0\end{pmatrix}\in$

$R^{np}\,|\,M_1\in R^{mp}\}$, a transformation h in $\mathcal{H}$ which is not of the form (2.3) can preserve a normal model as is shown in the following example.

Example 2.1. Let $n=2$ and $p=1$ and let

$$\mathcal{P}(0\times\mathcal{S}_+(1))=\{N(0,\ I_2\sigma^2)\,|\,\sigma^2>0\}$$

Then h defined by

$$h\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}\right)=\begin{pmatrix}\cos(x_1^2+x_2^2) & -\sin(x_1^2+x_2^2)\\ \sin(x_1^2+x_2^2) & \cos(x_1^2+x_2^2)\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}$$

preserves $\mathcal{P}(0\times\mathcal{S}_+(1))$ and it is easy to see $h\in\mathcal{H}$. Let $\mathcal{H}^*$ be the set of h in $\mathcal{H}$ preserving $\mathcal{P}(0\times\mathcal{S}_+(1))$. Then $\mathcal{H}^*$ is easily shown to be a subgroup of $\mathcal{H}$.

This implies that when $\mathcal{M}=\{\begin{pmatrix}M_1\\0\end{pmatrix}\in R^{np}\,|\,M_1\in R^{mp}\}$ with $n\geq 2$, the subgroup of $\mathcal{H}^*$ defined by

$$\mathcal{H}_0=\{h\,|\,h(x)=PxA+H \text{ for some } (P,A,H)\in\mathcal{O}(n)\times\mathcal{G}l(p)\times R^{np}\},$$

is not a maximal subgroup of $\mathcal{H}^*$ for preservation of model $\mathcal{P}(\mathcal{M}\times\mathcal{S}_+(p))$ in (2.1). For example, let

$$(2.4)\qquad X=\begin{pmatrix}X_1\\X_2\end{pmatrix}\sim N\left(\begin{pmatrix}M_1\\0\end{pmatrix}, I_n\otimes\Sigma\right) \quad\text{with}\quad (M_1,\Sigma)\in R^{mp}\times\mathcal{S}_+(p)$$

be a canonical form of the MANOVA model. Then the group $\mathcal{H}_0=\mathcal{O}(m)\times\mathcal{O}(n-m)\times\mathcal{G}l(p)\times R^{mp}$ acting on X by

$$\begin{pmatrix}X_1\\X_2\end{pmatrix}\to\begin{pmatrix}\Gamma_1 & 0\\ 0 & \Gamma_2\end{pmatrix}\begin{pmatrix}X_2\\X_2\end{pmatrix}A+\begin{pmatrix}F_1\\0\end{pmatrix}$$

is not a maximal group of $\mathcal{H}^*$ preserving (2.4) where $(\Gamma_1,\Gamma_2,A,F_1)\in\mathcal{H}_0$. In fact, the model in Example 2.1 is a special case of this model with $m=0$, $p=1$ and $n=2$. However if we reduce the model (2.4) by sufficiency to

$$(2.5)\qquad \begin{cases} X_1 \sim N(M_1, I_m \otimes \Sigma), \quad X_2'X_2 \sim W_p(\Sigma, n-m) \\ X_1 \text{ and } X_2'X_2 \text{ are independent,} \end{cases}$$

where $(M_1, \Sigma) \in R^{mp} \times \mathcal{S}_+(p)$ and $W_p(\Sigma, n-m)$ denotes the Wishart distribution with mean $(n-m)\Sigma$ and degrees of freedom $n-m$, then we obtain that the group $\mathscr{H}_1 = \mathcal{O}(n) \times \mathscr{G}l(p) \times R^{mp}$ as a group preserving (2.5) where $\mathscr{H}_1$ acts on $(X_1, X_2'X_2)$ by

$$h(X_1, X_2'X_2) = (PX_1A + F_1, A'X_2'X_2A) \quad \text{for } h = (P, A, F_1) \in \mathscr{H}_1,$$

is maximal in $\tilde{\mathscr{H}}^*$. Here $\tilde{\mathscr{H}}$ denotes the group of all bijective and bimeasurable transformation from $R^{mp} \times \mathcal{S}_+(p)$ onto itself and $\tilde{\mathscr{H}}^*$ denotes the subgroup of homeomorphisms in $\tilde{\mathscr{H}}$.

Theorem 2.2. (Nabeya and Kariya (1984)). Let $h \in \tilde{\mathscr{H}}$ and $n-m \geq p$, and assume (2.2) for the space of M_1. Suppose that for any $M_1 = M_1^{(k)} (k=0, 1, \cdots, n)$ and for any $\Sigma \in \mathcal{S}_+(p)$, there exists $(\Theta^{(k)}, \Psi^{(k)}) \in R^{mp} \times \mathcal{S}_+(p)$ such that when $X_1 \sim N(M_1^{(k)}, I_m \otimes \Sigma)$, $S \sim W_p(\Sigma, n-m)$ and X_1 and S are independent, then $h_1(X_1, S) \sim N(\Theta^{(k)}, I_m \otimes \Sigma)$, $h_2(X_1, S) \sim W_p(\Sigma, n-m)$ and $h_1(X_1, S)$ and $h_2(X_1, S)$ are independent, where $h(X_1, S) = (h_1(X_1, S), h_2(X_1, S)) \in R^{mp} \times \mathcal{S}_+(p)$. Then there exists $(P, A, F_1) \in \mathscr{H}_1$ such that

$$h(X_1, S) = (PX_1A + F_1, A'SA) \quad \text{a.e } (X_1, S) \in R^{mp} \times \mathcal{S}_+(p).$$

This theorem implies the maximality of $\mathscr{H}_1$ in $\tilde{\mathscr{H}}^*$. It is noted that the condition (2.2) holds if the space of Θ_1 contains the space $\{\Theta_1 \in R^{mp} | \Theta_1 = (\theta, 0), \theta \in R^m\}$. Hence it holds in particular if the space of Θ_1 is R^{mp}, which is the case of the MANOVA problem. On the other hand, even if we start with the group $\mathscr{H}_0$ which is not maximal in $\mathscr{H}^*$, we eventually get the model (2.5) via Lemma 1.4 because $X_2'X_2$ is a maximal invariant under the subgroup $\mathcal{O}(n-m)$ of $\mathscr{H}_0$ acting upon X_2 by $X_2 \to P_2X_2$. Therefore, though $\mathscr{H}_0$ is not maximal in $\mathscr{H}^*$ as a group preserving the model (2.4), the invariance reduces the model in (2.5) upon which the group $\mathscr{H}_1$ acts and $\mathscr{H}_1$ is maximal in $\tilde{\mathscr{H}}^*$.

3. Distribution of a Maximal Invariant.

3.1. *Representation theorem.* As has been stated in section 1, when a testing problem (1.2) is left invariant under a group $\mathscr{G}$, it is a usual procedure to choose a convenient maximal invariant (step(2)) and then derive the distribution of the maximal invariant (step(3)). However, sometimes maximal invariants are too complicated to analytically treat, which is the case of GMANOVA problem. In addition, it is not always necessary to derive an explicit form of the distribution. Indeed, when we are interested in locally optimal tests, we simply need a local form of the distribution in the neighborhood of the null hypothesis (see Section 4). In such a case and also in other cases, what is known as a representation of the probability ratio of a maximal invariant is very useful. There are several versions of the representation concerning conditions on the group leaving a problem invariant and the sample sapce acted upon by the group, but all of them give the same expression:

$$(3.1)\qquad R(T(x)) \equiv \frac{dP^T_{\theta_1}}{dP^T_{\theta}}(T(x)) = \frac{\int_{\mathscr{G}} f(gx|\theta_1)\chi(g)\nu(dg)}{\int_{\mathscr{G}} f(gx|\theta_0)\chi(g)\nu(dg)}$$

Here $T(x)$ is a maximal invariant, P^T_θ is the distribution of $T(x)$ under θ, $f(x|\theta)$ is a density with respect to a relatively invariant measure μ, $\mathscr{G}$ is a locally compact group, ν is a left invariant measure on $\mathscr{G}$ and $\chi(g)$ is a multiplier function of μ (i.e. $\mu(gA) = \chi(g)\mu(A)$ for $g \in \mathscr{G}$ and a Borel set A), which is the inverse of the Jacobian $x \to gx$ in the case that $\mathscr{G}$ is a matrix group and sample space $\mathscr{X}$ is a subset of an Euclidean space. A few sets of precise conditions for which (3.1) holds are stated below. The representation

(3.1) says that the Radon Nikodym derivative of the distribution of a maximal invariant under θ_1 with respect to that under θ_0 is, when evaluated at $T=T(x)$, given by the ratio of the integrals of the densities under θ_1 and θ_0 over the group $\mathscr{G}$. In testing problems, θ_1 and θ_0 are respectively chosen from the alternative hypothesis space Θ_1 and the null hypothesis space Θ_0 so that (3.1) represents the likelihood ratio of $T(x)$ under θ_1 and θ_0. Hence if the null distribution $P^T_{\theta_0}$ of a maximal invariant T is the same for all $\theta_0 \in \Theta_0$, which is true when Θ_0 is simple, a test with critical region $R(T(x))>c$ is by the Neyman Pearson Lemma MPI (most powerful invariant) test for testing Θ_0 versus the fixed θ_1 in Θ_1. If the test happens to be independent of θ_1, it is UMPI for testing Θ_0 versus Θ_1. In case that a locally optimal test is interested in, the ratio in (3.1) is expanded in the neighborhood of Θ_0 and the integral is evaluated locally. In these procedures, it is not necessary to choose an explicit form of a maximal invariant $T(x)$ and step (3) can be skipped.

3.2. *References.* Historically it was Stein (1956) who first gave the expression (3.1), but he did not explicitly state the conditions for which it is valid. Stein's representation was applied by Giri (1965) and Schwartz (1967) etc. In a line with Stein, Schwartz (1966) gave a set of conditions for the validity, but his conditions are rather complicated. His result is introduced in Farrell (1976). Wijsman (1966, 1967) took a differential-geometric approach and gave a sufficient condition for (3.1) by using the concept of Cartan $\mathscr{G}$ space. Koehn (1970) generalized some results of Wijsman (1967), while Bondar (1976) considered conditions for (3.1) through a topological argument. Recently taking Bourbaki's approach and using a quotient measure, Andersson (1982) established some results concerning (3.1) in terms of proper action of the group. Wijsman

(1983) studied on the properness of the action in applied problems. Further Wijsman (1984) also studied on the global cross-section for factorization of measures and applied it to the representation of the distribution of a maximal invariant.

3.3. *Andersson's approach.* Based on Bourbacki's approach, Andersson (1982), using quotient measures, obtains the representation of densities of maximal invariants. His framework is relatively simpler than Wijsman's (1967), as will be seen below. Let $\mathscr{G}$ be a locally compact sigma-compact Hausdorff topological group and let $\mathscr{X}$ be a locally compact sigma-compact Hausdorff topological space. Suppose $\mathscr{G}$ acts topologically on $\mathscr{X}$ (i.e., the map $\mathscr{G}\times\mathscr{X}\ni(g,x)\to gx \in\mathscr{X}$ is continuous). A left (right) invariant measure on $\mathscr{G}$ is denoted by ν_l(ν_r respectively), and let $\mathscr{X}/\mathscr{G}$ be the quotient space $\mathscr{X}/\mathscr{G}\equiv\mathscr{Q}$ with quotient topology. In Andersson (1982), the natural projection π from $\mathscr{X}$ onto the quotient space $\mathscr{Q}$ is taken as a maximal invariant. In fact $\pi(x)=\mathscr{G}x$ for $x\in\mathscr{X}$ is a maximal invariant. Further let $f_i(x)$ $(i=1,2)$ be two probability densities on $\mathscr{X}$ with respect to a relatively invariant measure μ with multiplier χ (i.e., $\mu(gA)=\chi(g)\mu(A)$ where $g\in\mathscr{G}$ and A is a measurable subset of $\mathscr{X}$). Under this set-up, a sufficient condition for the representation (3.1) to hold with $T=\pi$ is provided by the notion of a proper action:

Definition 3.1. Consider the map K of $\mathscr{G}\times\mathscr{X}$ into $\mathscr{X}\times\mathscr{X}$ given by $K(g,x)=(gx,x)$. Then the action of $\mathscr{G}$ on $\mathscr{X}$ is said to be proper if $K^{-1}(C)$ is compact for each compact subset $C\subset\mathscr{X}\times\mathscr{X}$, and the space $\mathscr{X}$ is called a proper $\mathscr{G}$-space if the action of $\mathscr{G}$ is proper.

Theorem 3.1. (Andersson (1982)) Suppose $\mathscr{G}$ acts properly on $\mathscr{X}$. Then the probability ratio of the maximal invariant π under f_i $(i=1,2)$ is given by

$$R(\pi(x)) \equiv \frac{dP_2^\pi}{dP_1^\pi}(\pi(x)) = \frac{\int_{\mathscr{G}} f_2(gx)\chi(g)\nu_l(dg)}{\int_{\mathscr{G}} f_1(gx)\chi(g)\nu_l(dg)} \quad \text{a.e.}\,(P_1^\pi). \tag{3.2}$$

A proof of this theorem is sketched in Appendix.

The Definition 3.1 of the properness on the action of $\mathscr{G}$ is not so easy to handle as it stands. To give a set of equivalent conditions, define for $A, B \subset \mathscr{X}$,

$$((A, B)) = \{g \in \mathscr{G} \mid gA \cap B \neq \varnothing\}. \tag{3.3}$$

Definition 3.2. $\mathscr{X}$ is called a Cartan $\mathscr{G}$-space if for each $x \in \mathscr{X}$, there exists a neighborhood V of x such that $((V, V))$ has compact closure.

Lemma 3.1. Under the present set-up, the following conditions are equivalent.

(1) $\mathscr{G}$ acts properly on $\mathscr{X}$.

(2) For any x and y in $\mathscr{X}$, there exist neighborhoods V_x and V_y of x and y respectively such that $((V_x, V_y))$ has compact closure.

(3) $\mathscr{X}$ is a Cartan $\mathscr{G}$-space and $\mathscr{X}/\mathscr{G}$ is Hausdorff.

(4) For $K \subset \mathscr{X}$ compact, $((K, K))$ has compact closure.

Proof. The equivalence between (1) and (2) follows from Bourbaki (1966) III§4.4, Proposition 7, while the equivalence among (2), (3) and (4) follows from Palais (1961) Theorem 1.2.9.

As a matter of a fact, the sigma compactness of $\mathscr{G}$ and $\mathscr{X}$ assumed in our set up is not necessary to obtain the result in Lemma 3.1. It is also remarked that if A and B are compact in $\mathscr{X}$, $((A, B))$ is closed. The following result is useful in application.

Lemma 3.2. (Palais (1961) 1.3.3) Let $\mathscr{Y}$ be a locally compact space acted upon topologically by a locally compact group $\mathscr{G}$. Then if $\mathscr{X}$

is a proper (Cartan) $\mathscr{G}$-space, so is $\mathscr{X} \times \mathscr{Y}$.

Wijsman (1983) develops a useful tool for proving the action of $\mathscr{G}$ on $\mathscr{X}$ proper.

3.4. *Wijsman's approach.* Let $\mathscr{X}$ be a nonempty open subset of R^n, $\mathscr{G}$ a Lie subgroup of the general linear group $\mathscr{Gl}(n)$ acting linearly on $\mathscr{X}$, and ν Lebesgue measure on $\mathscr{X}$.

Theorem 3.2. (Wijsman (1967)). Suppose $\mathscr{X}$ is a Cartan $\mathscr{G}$-space. Then (3.1) holds with $\chi(g) = |g'g|^{1/2}$ for $g \in \mathscr{G}$.

The proof requires some knowledge on differentiable geometry. It is remarked that Wijsman (1972) extended the result to the case where $\mathscr{G}$ is a Lie subgroup of the general affine linear group.

Though the set up for $\mathscr{X}$ and $\mathscr{G}$ in Theorem 3.2 is more restrictive than that in Theorem 3.1, as is shown in Lemma 3.1, the condition for $\mathscr{X}$ to be a Cartan $\mathscr{G}$-space is weaker and easier to check than the condition for $\mathscr{X}$ to be a proper $\mathscr{G}$-space. In Chapter 3, Wijsman's theorem is used, while in Chapter 5 Andersson's theorem is used for comparison.

To state a condition for Cartan $\mathscr{G}$-space, let $\mathscr{G}$ be a locally compact group and $\mathscr{X}$ a completely regular Hausdorff space. When $\mathscr{G}$ acts freely on $\mathscr{X}$ (i.e., $g \neq e$ implies $gx \neq x$ for all $x \in \mathscr{X}$), $\mathscr{X}$ is said to be a $\mathscr{G}$ principal bundle. If $\mathscr{X}$ is a $\mathscr{G}$ principle bundle, for each (x_1, x_2) there is a unique element $f(x_1, x_2) \in \mathscr{R}$ such that $x_2 = f(x_1, x_2)x_1$ where

$$\mathscr{R} = \{(x_1, x_2) \in \mathscr{X} \times \mathscr{X} \mid x_1 = gx_2 \text{ for some } g \in \mathscr{G}\}.$$

Hence if f is continuous from $\mathscr{R}$ into $\mathscr{G}$, $\mathscr{X}$ is called a Cartan principal bundle.

Lemma 3.3. (Palais (1961) 1.1.3) A $\mathscr{G}$-principal bundle $\mathscr{X}$ is a

Cartan principal bundle if and only if it is a Cartan $\mathscr{G}$-space.

As an example, let us consider the MANOVA problem. Let $X: n\times p \sim N(\binom{\Theta}{0}, I_n\otimes\Sigma)$, where $X=(X_i)$ with $X_i: n_i\times p$ $(i=1,2,3)$, $\Theta=(\Theta_j):(n_1+n_2)\times p$ with $\Theta_j: n_j\times p$, $(j=1,2)$ and $n_1+n_2+n_3=n$ with $n_3\geq p$. The MANOVA problem tests $\Theta_2=0$ and the group $\tilde{\mathscr{G}}=\mathcal{O}(n_1)\times\mathcal{O}(n_2)\times\mathcal{O}(n_3)\times\mathscr{G}l(p)\times R^{n_1 p}$ acts on the space of X by

$$(X_1, X_2, X_3) \to (\Gamma_1X_1A'+F,\ \Gamma_2X_2A',\ \Gamma_3X_3A')$$

where $(\Gamma_1, \Gamma_2, \Gamma_3, A, F)\in\tilde{\mathscr{G}}$. It follows from Lemma 1.4 that a maximal invariant is a function of X_2 and X_3. Hence letting

$$\mathscr{X}_i=\{x\in R^{n_i\times p}\mid \operatorname{rank}(x)=\min(n_i, p)\}\ (i=2,3),$$

it suffices to show that $\mathscr{X}_2\times\mathscr{X}_3$ is a Cartan $\mathcal{O}(n_2)\times\mathcal{O}(n_3)\times\mathscr{G}l(p)$ space because $R^{n_2p}\times R^{n_3p}-\mathscr{X}_2\times\mathscr{X}_3$ has measure 0. Further, since $\mathcal{O}(n_2)\times\mathcal{O}(n_3)$ is compact, by Definition 3.2 it suffices to show that $\mathscr{X}_2\times\mathscr{X}_3$ is a Cartan $\mathscr{G}l(p)$ space. But by Lemma 3.2 and 3.3, it suffices to show that $X_3A'=X_3$ for $X_3\in\mathscr{X}_3$ implies $A=I$, which follows from the definition of $\mathscr{X}_3$ and $n_3\geq p$. Therefore by Theorem 3.2, the probability ratio of a maximal invariant is given by (3.1) where $\mathscr{G}=\mathcal{O}(n_2)\times\mathcal{O}(n_3)\times\mathscr{G}l(p)$, $f(X_2, X_3|\theta)$ is the density of (X_2, X_3), $\theta_0=(0, \Sigma)$, $\theta_1=(\Theta_2, \Sigma)$ and $\chi(g)=|AA'|^{(n_2+n_3)/2}$ is the inverse of Jacobian of $(X_2, X_3)\to g(X_2, X_3)$ for $g\in\mathscr{G}$.

4. Locally Best Invariant Test

4.1. *Derivation of an LBI test.* In this section, a basic idea for deriving an LBI (locally best invariant) test is provided. As was done in Section 1, let $\mathscr{X}$ be a nonempty open subset of R^n and let $\mathscr{G}$ be a closed subgroup of an affine linear group acting on the left of $\mathscr{X}$. To apply Wijsman's representation theorem described in Section 3, it is assumed that $\mathscr{X}$ is a Cartan $\mathscr{G}$-space. Let $\mathscr{F}(\Theta)$

denote a class of probability densities with respect to Lebesgue measure on $\mathscr{X}$ such that each density in $\mathscr{F}(\Theta)$ is of the form

$$f(x|\theta)=\beta(\theta)q(\psi(x:\theta)),\quad \theta\in\Theta \tag{4.1}$$

where β, q and ψ are known functions, $\psi:\mathscr{X}\times\Theta\to R_+$ is a measurable function for each fixed θ, q is a fixed integrable function from R_+ into $[0,\infty)$ and Θ is a nonempty open subset of R^p. In a normal model, $q(z)=\exp(-z)$ and $\psi(x:\theta)$ is quadratic in x. Suppose the group $\mathscr{G}$ leaves invariant the problem

$$H:\theta=\theta_0 \quad \text{versus} \quad K:\theta\neq\theta_0.$$

Then by Theorem 3.1, the ratio of the distributions of a maximal invariant $T=t(x)$ is given by

$$R\equiv(dP^T_{\theta_1}/dP^T_{\theta_0})(t(x))=H(x|\theta_1)/H(x|\theta_0) \tag{4.2}$$

with

$$H(x|\theta)=\int_{\mathscr{G}}f(gx|\theta)\chi(g)\nu(dg) \tag{4.3}$$

where $\theta_1\in\Theta_1$, P^T_θ is the distribution of T under θ, $\chi(g)$ is the inverse of the Jacobian of transformation $x\to gx$ and ν is an invariant measure on $\mathscr{G}$. Here assuming that q and β are continuously twice differentiable, we expand the integrand in the numerator of the ratio in (4.2) as

$$\begin{aligned} f(x|\theta_1)=(\beta(\theta_0)+o_1)\{&q(\psi(x:\theta_0)) \\ &+q^{(1)}(\psi(x:\theta_0))[\psi(x:\theta_1)-\psi(x:\theta_0)] \\ &+\frac{1}{2}q^{(2)}(\psi^*(x:\theta_0,\theta_1))[\psi(x:\theta_1)-\psi(x:\theta_0)]^2\} \end{aligned} \tag{4.4}$$

where $\beta(\theta_1)=\beta(\theta_0)+o_1$ with $o_1=o(||\theta_1-\theta_0||)$, $\psi^*(x:\theta_1,\theta_0)=c\psi(x:\theta_1)+(1-c)\psi(x:\theta_0)$ for some $0\leq c\leq 1$, and $q^{(i)}=\partial^i q/\partial z^i$. Here $||a||$ denotes the Euclidean norm. Then with $D=H(x|\theta_0)$, the ratio R in (4.2) is expressed as

$$1+\int_{\mathcal{G}} q^{(1)}(\psi(gx,\theta_0))[\psi(gx:\theta_1)-\psi(gx:\theta_0)]\chi(g)\nu(dg)/D+M \tag{4.5}$$

where $M \equiv M(x:\theta_1,\theta_0)$ is a remainder term. Here if we can show that the second term is expressed as $\gamma(\theta_1,\theta_0)s(x)$ with $\gamma(\theta_1,\theta_0)=O(||\theta_1-\theta_0||)$ and that for any invariant test function $\phi(x)$ of size α

$$\int \phi(x)M(x:\theta_1,\theta_0)dP^T_{\theta_0}=o(||\theta_1-\theta_0||) \quad \text{uniformly in } \phi, \tag{4.6}$$

then from (4.5) the power function of ϕ is given by

$$\pi(\phi,\theta_1)=\alpha+E_{\theta_0}[\phi(x)\gamma(\theta_1,\theta_0)s(x)]+o(||\theta_1-\theta_0||) \tag{4.7}$$

Hence, by the Generalized Neyman-Pearson Lemma stated in Section 1, the test based on $s(x)$ is LBI. Sometimes the second term in the right side of (4.4) vanishes, in which case higher order derivatives of q need to be considered. In this manner, most LBI tests can be derived. In fact, our LBI tests in this book are basically derived in this way (see, e.g., Kariya and Sinha (1984)).

5. Distributions of Test Statistics

5.1. *Null Distribution.* Once an optimal test is derived, the null distribution of the test statistic is needed to obtain critical points. But as has been mentioned in Section 1, what we really need in reporting results will be significance probabilities. In most multivariate testing problems, it is however difficult to derive the exact null distributions of those optimal tests or to compute the significance probabilities. Even if the exact null distributions are derived, they are often too complicated to use without the aid of a computer. In fact, they often involves matrix variate hypergeometric functions or Zonal polynomials. Here by derivation, it is usually meant that the distributions are expressed as a form where no incomputed integrals are involved however complicated the form may be. But sometimes, a simple integral form suffices for a computational

purpose. In this sense, a common computational program package will be desired to be developed to compute the significance probabilities for normal models. To illustrate a typical case, let $u=u(x)$ be a test statistic with critical region $u>c$ and let $f(x)$ be the density of $N(0, I_n \otimes I_p)$ under a null hypothesis where $x: n\times p$. In fact, the null distributions of most tests treated in normal models do not depend on unknown parameters or the tests are similar, so that we may assume $N(0, I_n\otimes I_p)$. In particular, if a group leaves a problem invariant and if the induced group acts transitively on the parameter space under null hypothesis, then the null distribution of any invariant test is free from unknown parameters (see also Section 9). In such a case, $N(0, I_n\otimes I_p)$ can be assumed. Then the null distribution of $u=u(x)$ is obtained by

$$F(u_0)=P(u(x)\le u_0)=\int_{u(x)\le u_0} f(x)dx$$

Hence for a given observation x_0, what we need is to evaluate the numerical value of this integral with $u_0=u(x_0)$ through a computer so that we obtain the significance probability $1-F(u_0)$. Of couse, sometimes it may be useful to obtain an explicit form of $F(u_0)$ in terms of e.g., a power series, but it is not always required for a computational purpose.

On the other hand, in the case of comparatively large samples, the null distributions are often approximated by their asymptotically expanded expressions up to certain orders of the sample sizes. The derivations are often made by first expanding a test statistic stochastically (or in the concept of in-probability) up to a certain order, then evaluating the characteristic function of the expanded statistic and finally inverting the characteristic function. However, the question of the validity of this approach is often left open in articles so that the expansion is formal. On this problem, Bhattacharya and Ghosh (1978) and Feller (1966) may be good references, (see also Kariya and Maekawa (1981) and Fujikoshi (1984)). In this book, the validity of our expansions is also left open.

5.2. *Nonnull distribution.* The exact nonnull distributions of optimal tests in multivariate testing problems are still more difficult to derive and even if they are obtained, the expressions are often intractable to deduce some results on the behavior of the power functions. On the other hand, to obtain asymptotic expansions of the nonnull distribution, it is usually required to confine the alternative hypotheses to a contour in which they come closer to the null hypotheses in certain orders of sample size n as n gets larger. This is because the tests considered there are consistent or the power functions go to 1 as n increases under a fixed parameter point. In this sense, the information we can collect through the nonnull asymptotic distributions is about the local behaviors of the power functions in the neighborhood of the null hypotheses. If a test is LBI (locally best invariant), it should have a good property in the neighborhood of a null hypothesis though the asymptotic expansion of the nonnull distribution of the LBI test may be interesting to make a comparison with those of other tests.

6. Local Minimaxity

6.1. *Minimaxity.* For testing problem (1.2), a test ϕ^* in the class of level α tests $\mathcal{D}$ is called minimax in $\mathcal{D}$ with respect to a contour Δ if

$$\inf_{\theta\in\Delta}\pi(\phi^*,\ \theta)=\sup_{\phi\in\mathcal{D}}\inf_{\theta\in\Delta}\pi(\phi,\ \theta) \tag{6.1}$$

where Δ is a subset of the space Θ_1 of the alternative hypothesis. Since $\mathcal{D}$ includes the trivial test $\phi_0\equiv\alpha$, the right side of (6.1) is not less than α. Hence if the power function $\pi(\phi,\ \theta)$ is continuous for each $\phi\in\mathcal{D}$ (which is guaranteed if the model (1.1) is an

exponential family (see Lemma 1.3)), and if $\bar{\Theta}_0\cap\bar{\Theta}_1\neq\phi$, any test is minimax with respect to $\Delta=\Theta_1$, where $\bar{\Theta}_i$ denotes the closure of Θ_i $(i=0,1)$. This is why the minimaxity is restricted to a contour Δ of Θ_1. When the given problem remains invariant under a group $\mathcal{G}$, it is often the case that Δ is chosen to be an orbit defined by a maximal invariant parameter $\tau=\tau(\theta)$;

$$\Delta_\tau=\{\theta\in\Theta_1\,|\,\tau=\tau(\theta)\}, \tag{6.2}$$

and that it is asked whether a UMPI test is minimax in $\mathcal{D}$ (not in $\mathcal{D}^I$) with respect to Δ_τ for each τ. A typical way to show minimaxity is to use the well known proposition that a (an extended) Bayes rule with constant risk is minimax (see e.g., Ferguson (1967) page 91). Of course, the power function of an invariant test is constant on each orbit Δ_τ in (6.2) since it depends on θ only through $\tau(\theta)$. Hence if an invariant test is minimax with respect to Δ_τ for all τ, it must be UMPI by the definition of (6.1). On the other hand, according to the Hunt-Stein theorem, if a minimax test with respect to Δ_τ for each τ exists, it can be found in the class of invariant tests of level α. To state the Hunt-Stein theorem, let $\mathcal{G}$ be a locally compact and sigma-compact group with Borel sigma-field $\mathcal{L}$ and suppose the problem (1.2) under the model $\mathcal{F}(\Theta)$ in (1.1) is left invariant under $\mathcal{G}$. Further assume the map $(g,x)\to gx$ is jointly measurable.

Theorem 6.1. (Hunt-Stein Theorem) Suppose the following condition holds :

(HS) There exists a sequence of probability measures $\{\nu_n\}$ on $(\mathcal{G},\mathcal{L})$ such that for any $g\in\mathcal{G}$ and $B\in\mathcal{L}$

$$\lim_{n\to\infty}|\nu_n(Bg)-\nu_n(B)|=0. \tag{6.3}$$

Then there exists an almost invariant minimax test of level α, where a measurable function f on $(\mathcal{X},\mathcal{B},\mu)$ is almost invariant if

$f(gx)=f(x)$ a.e. (μ).

A proof is found in Lehmann (1959) p336. Bonder and Milones (1981) looked into the various conditions for which the theorem holds. In particular, it is shown that the condition (HS) is equivalent to the existence of a summing sequence :

(S) There is a sequence $\{B_n\}$ of compact subsets of $\mathscr{G}$ with $B_n \subset B_{n+1}$, $0<\nu(B_n)<\infty$ and $\mathscr{G}=\cup B_n$ such that $\nu(B_n g \cup B_n)/\nu(B_n) \to 1$ uniformly in g on compact subsets of $\mathscr{G}$ where ν is a right invariant measure on $\mathscr{G}$.

Also it is shown there that (HS) is implied by Stein's condition, which in turn is implied by the solvability of $\mathscr{G}$ (see also Kudo (1955) or Kiefer (1957) for the latter implication).

A remark is that one can get an invariant test from an almost invariant test by the following lemma.

Lemma 6.1. (See Lehmann (1959) page 225). Assume (1) $(\mathscr{X},\mathscr{B},\mu)$ is a sigma-finite measure space, (2) a group $\mathscr{I}$ with a sigma-field $\mathscr{C}$ acts on $\mathscr{X}$, (3) the map $(g,x)\to gx$ is measurable and (4) there exists a sigma-finite measure ν on $\mathscr{I}$ such that for any $g\in\mathscr{I}$ and $B\in\mathscr{C}$, $Bg\in\mathscr{C}$ and if $\nu(C)=0$ for some $C\in\mathscr{I}$, then $\nu(Cg)=0$ for any $g\in\mathscr{I}$. Then for any almost invariant function $f(x)$ on $\mathscr{X}$, there exists an invariant function $h(x)$ such that $h(x)=f(x)$ a.e.

In applying the Hunt-Stein theorem to a testing problem, the condition (HS) is rather restrictive. In particular, the general linear group $\mathscr{G}l(n)$ $(n\geq 2)$ does not satisfy it (see Lehmann (1959) Section 8.4, Example 9). But in fact, it is pointed out in Bondar and Milones (1981) that the converse of the theorem holds when $\mathscr{G}$ is an almost connected group. Usually if a group $\mathscr{G}$ leaving a given problem invariant does not satisfy (HS) condition, we take a solvable

subgroup $\mathscr{G}_0$ of $\mathscr{G}$ and try to show the minimaxity of a given (UMPI) test in the class of $\mathscr{G}_0$-invariant tests. However, even in doing so, it is in general difficult to verify the minimaxity. The minimaxity of Hotelling's T^2 test was questioned by Giri, Kiefer and Stein (1963) and it was shown for a very special case. It was Salavaevskii (1968) that established it for a general case with surprisingly voluminous computations.

6.2. *Local minimaxity.* In most multivariate testing problems, no UMPI test exists so that no invariant minimax test with respect to all the contours (6.2) defined by a maximal invariant parameter τ exists. But as has been pointed out, there often exist LBI tests in those problems, to which the concept of local minimaxity corresponds. In fact, Giri and Kiefer (1964) created the concept of local minimaxity as follows. Let $(\mathscr{X},\mathscr{B})$ be a Borel measurable space where $\mathscr{X}\subset R^n$ and let $p(\cdot:\varDelta,\xi)$ be a density with respect to a σ-finite measure μ where $\varDelta$ is a real parameter, ξ is nuisance parameter and the range of ξ may depend on $\varDelta$. Consider a testing problem $H_0:\varDelta=0$ versus $\varDelta=\lambda>0$.

Definition 6.1. A test ϕ^* is locally minimax of level $\alpha(0<\alpha<1)$ for testing $H_0:\varDelta=0$ versus $\varDelta=\lambda$ as $\lambda\to 0$ if

$$\lim_{\lambda\to 0}\frac{\inf_\xi \pi(\phi^*,\ (\lambda,\xi))-\alpha}{\sup_{\phi_\lambda\in\mathscr{D}}\inf_\xi \pi(\phi_\lambda,\ (\lambda,\xi))-\alpha}=1 \qquad (6.4)$$

where $\pi(\phi,\ (\lambda,\xi))=E[\phi\,|\,\lambda,\xi]$ is the power function of ϕ and $\mathscr{D}$ is the class of level α tests.

Giri and Kiefer (1964) gave a sufficient condition for a given test to be locally minimax, which is stated in Chapter 3. In invariant testing problems, $\varDelta$ will be a function of a maximal invariant parameter and ξ will belong to a contour defined by $\varDelta=\lambda$, and in showing an LBI test to be locally minimax with respect to this contour, the class $\mathcal{D}$ in (6.5) is replaced by the class of level α tests invariant under a group satisfying the (HS) condition in Theorem 6.1.

7. Monotonicity

7.1. *Monotonicity.* We consider a monotonicity property that power increases as the parameter goes far from the null hypothesis. In an invariant problem, as has been stated in Section 1, the power function (1.3) of an invariant test ψ is a function of a maximal invariant parameter $\tau=\tau(\theta)$:

$$\bar{\pi}(\psi,\ \tau)=E_\theta[\psi(T(x))] \qquad \theta\in\Theta_0\cup\Theta_1.$$

Here τ is assumed to be $\tau=(\tau_1,\dots,\tau_p)'$ with $\tau_1\geq\dots\geq\tau_p\geq 0$ and to satisfy $\tau(\Theta_0)=0$, which is a typical case in multivariate testing problems. Then ψ is said to be of monotonicity if $\bar{\pi}(\psi,\ \tau)$ is an increasing function of each τ_i for fixed τ_j's $(j\neq i)$, where τ_i moves over the interval $[\tau_{i-1},\ \tau_{i+1}]$. The monotonicity of ψ of course implies the unbiasedness of ψ :

$$\bar{\pi}(\psi,\ \tau)\geq\alpha \quad \text{for all } \tau\in\tau(\Theta_1) \text{ (or } \theta\in\Theta_1).$$

In situations where there exist UMP or UMPI tests, the power functions are usually functions of a single parameter and of monotonicity, which is often implied by the monotone likelihood ratio property of the densities (see Lehmann (1959) Chapter 3). On the unbiasedness of invariant tests, Perlman and Olkin's paper (1980) will be helpful as a reference which includes a survey as well as new developments in this area. As is pointed out in their paper, the unbiasedness and monotonicity properties of the LBI (locally best invariant) test in the MANOVA problem has remained an open question. This implies that the unbiasedness of the LBI test in the

GMANOVA problem is a harder problem to settle. The difficulty may be related to the drawback of the LBI tests stated in Section 4 of Chapter 3.

To show the monotonicity property of an invariant test, the following lemma is often used (which is an application of Anderson's Theorem (1955) in Das Gupta, Anderson and Mudholkar (1965)).

Lemma 7.1. Let $x_1,\dots,x_s$ and Y be independent random vectors and matrix where $x_i : p\times 1\sim N(\lambda_i\mu_i,\Sigma_i)$ $(i=1,\dots,s)$ and Y be any random matrix. Suppose a set C (regarded as an acceptance region) in the space of $(x_1,\dots,x_s,Y)$ is convex and symmetric in x_i for x_j's and Y being fixed $(j\neq i)$. Then the probability $P(C^c)$ is increasing in each λ_i.

8. Admissibility of Tests.

8.1. *Conventional approaches.* As has been stated in Section 1, a test ϕ^* in the class $\mathcal{D}_\alpha$ of level α tests is called admissible if there exists no test ϕ in $\mathcal{D}_\alpha$ (or $\mathcal{D}_{I\alpha}$) such that $\pi(\phi,\ \theta)\geq\pi(\phi^*,\ \theta)$ for all $\theta\in\Theta_1$ and $\pi(\phi,\ \theta)\leq\pi(\phi^*,\ \theta)$ for all $\theta\in\Theta_0$ with strict inequality for at least one $\theta\in\Theta_0\cup\Theta_1$. As methods for showing this property for a given test ϕ^*, Kiefer (1966) listed the following ;

(a) the Bayes method, of showing that a procedure is unique Bayes with respect to some proper prior distribution,

(b) the method which uses exponential type, and

(c) the method based on local properties.

Making use of the method (c), it follows that the unique most powerful (MP) test for testing $\theta\in\Theta_0$ versus $\theta=\theta_1\in\Theta_1$ where θ_1 is fixed is admissible because no other tests dominate the MP test at $\theta=\theta_1$. In particular, the unique UMPI test and the unique LBI test are admissible in $\mathcal{D}_{I\alpha}$. Thus the LBI tests we derive in the

GMANOVA problem, an extended GMANOVA problem, the problem of testing on Rao's covariance structure etc., are all admissible. However in these problems, the admissibility or inadmissibility of the LRT's and other related tests is hard to establish by the methods (a), (b) and (c). In fact, for the method (a), it is very difficult to find in our problems the prior distributions that yield the tests concerned. In Kiefer and Schwartz (1965) and Schwartz (1967) showed by this method the admissibility of a lot of tests often encountered in applications. A difference between the problems treated there and our problems treated in this book is that our model possibly reduced by sufficiency is still an incomplete model which allows an ancillary statistic.

The method (b) is developed by Stein (1956) and is applied to many problems such as the MANOVA problem, the problems of testing independence etc. by many authors (e.g., see Schwartz (1967a)).

In an exponential type model of completeness, Matthes and Truax (1967) obtained the minimal complete class of tests consisting of all admissible tests. Here a subclass $\mathscr{C}_{I\alpha}$ of $\mathscr{D}_{I\alpha}$ is said to be essentially complete if for any test ϕ in $\mathscr{D}_{I\alpha}$ there exists a test ϕ^* in $\mathscr{C}_{I\alpha}$ such that $\pi(\phi^*, \theta) \geq \pi(\phi, \theta)$ for all $\theta \in \Theta_1$. When the inequality is strict for at least one $\theta \in \Theta_1$, $\mathscr{C}_{I\alpha}$ is called a complete class. Hence a test which does not belong to a complete class is not admissible.

Even in incomplete models, the method (b) is applicable. In the problem of testing on means with covariates, which is an extended GMANOVA problem, using Theorem 5.8 of Wald (1950) that the closure of proper Bayes tests under weak* limits is an essentially complete class of tests, Marden and Perlman (1980) characterized the weak* limit of sequences of Bayes tests and obtained necessary and sufficient conditions for admissibility. Applications of the

results yields the inadmissibility of well-known tests in the problem such as the studentized test based on an efficient estimator and the LRT with a restriction on the significance level. Using the same approach, Marden (1981, 1982, 1983) extended the results to the GMANOVA problems, the problem of testing independence with incomplete data treated by Eaton and Kariya (1983), and the problem of testing a hypothesis against a multivariate one-sided alternative.

9. Robustness

9.1. *Null robustness.* Robustness properties of a test may be studied from the two sides: (1) Robustness under null hypothesis and (2) robustness under alternative hypothesis. We first consider (1). By (1) it is usually meant that the critical point of a size α test computed under an assumed model, say $\mathscr{F}(\Theta) = \{f(x|\theta) \mid \theta \in \Theta\}$ in (1.1), is stable against a departure from the model. That is, even if the model $\mathscr{F}(\Theta)$ in (1.1) is enlarged to

$$\mathscr{H}(\Theta) = \{h(x|\theta) \mid \theta \in \Theta,\ h \in \mathscr{H}\},$$

(1) means that the critical point computed under the null hypothesis of the given density $f(x|\theta)$ defining $\mathscr{F}(\Theta)$ is effective or approximately effeictive for all h in $\mathscr{H}(\Theta)$, where $\mathscr{H}$ is a class of densities having the same parameter space as f and containing f. But here we call a test null robust in $\mathscr{H}(\Theta)$ if the null distribution derived under $\mathscr{F}(\Theta)$ in (1.1) remains the same for all h in $\mathscr{H}(\Theta)$. Null robustness in this sense has been considered in a class of elliptically contoured distributions or left orthogonally invariant distributions by many authors (Dempster (1969), Dawid (1977), Kariya and Eaton (1977), Chmieleswki (1980), Kariya (1981), Jensen and Good (1981), Eaton and Kariya (1984) etc.).

Here we introduce the results due to Kariya (1981). Let $\mathcal{O}(n)$ and

$\mathcal{S}_+(p)$ denote the set of $n\times n$ orthogonal matrices and the set of $p\times p$ positive definite matrices respectively as before. For an $n\times p$ random matrix X, let $\mathcal{L}(X)$ denote the distribution of X. We shall call X left $\mathcal{O}(n)$-invariant about M if $\mathcal{L}(Q(X-M))=\mathcal{L}(X-M)$ for all $Q\in\mathcal{O}(n)$. Also, we shall call $\mathcal{L}(X)$ elliptically symmetric about M with scale matrix $\Sigma\in\mathcal{S}_+(p)$ if $\mathcal{L}(gy)=\mathcal{L}(y)$ for all $g\in\mathcal{O}(np)$, where $y=(y_1,\dots,y_n)'$, y_i is the i-th row of $Y=(X-M)\Sigma^{-1/2}$. Let $\mathcal{X}=\{X: n\times p \mid \mathrm{rank}(X)=p\}$ and throughout this section $n\geq p$ is assumed. Further, let $\mathcal{F}_L(M)$ and $\mathcal{F}_E(M,I_n\otimes\Sigma)$ denote respectively the class of np-dimensional left $\mathcal{O}(n)$-invariant distributions about M such that $P(X-M\in\mathcal{X})=1$ and the class of np-dimensional elliptically symmetric distributions about M with scale matrix $\Sigma\in\mathcal{S}_+(p)$ such that $P(X-M\in\mathcal{X})=1$. Clearly

$$\mathcal{F}_E(M, I_n\otimes\Sigma)\subseteq\mathcal{F}_L(M) \quad \text{for all } M: n\times p \text{ and } \Sigma\in\mathcal{S}_+(p).$$

If $\mathcal{L}(X)\in\mathcal{F}_E(M,I_n\otimes\Sigma)$ has a density, it is expressed as

$$(9.1) \qquad f(X\mid M,\Sigma)=|\Sigma|^{-n/2}q(\mathrm{tr}\,\Sigma^{-1}(X-M)'(X-M)),$$

where $q:[0,\infty)\to[0,\infty)$ and if $\mathcal{L}(X)\in\mathcal{F}_L(M)$ has a density, it is expressed as

$$(9.2) \qquad f(X\mid M)=\psi((X-M)'(X-M)),$$

where $\psi:\mathcal{S}_+(p)\to[0,\infty)$. A left $\mathcal{O}(n)$-invariant distribution which is not elliptically symmetric is the matrix variate t-distribution, whose density is given by

$$f_0(X)=c\,|I_p+(X-M)'(X-M)|^{-(m+n)/2},$$

where c is a normalizing constant. Frazer and Ng (1980) treated a multivariate regression model with this distribution for the error term.

In most multivariate hypothesis testing problems, $\mathcal{L}(X)$ is assumed to be normal:

$$(9.3) \qquad \mathcal{L}(X)=N(M,\Omega) \qquad M: n\times p,\ \Omega\in\mathcal{S}_+(np).$$

Here $E(X)=M$ and $\mathrm{Cov}(x)=\Omega$, where $x=(x_1,\dots,x_n)'$ and x_i is the i-th row of X. After transformation, many null hypotheses can be stated as

$$(9.4) \qquad H: (M,\Sigma)\in\Theta_0\times\Lambda_0,$$

where in (9.3), $\Omega=I_n\otimes\Sigma$, $\Sigma\in\mathcal{S}_+(p)$, and $\Theta_0\times\Lambda_0\subset R^{np}\times\mathcal{S}_+(p)$. It is noted that when $\Omega=I_n\otimes\Sigma$ in (9.3), the rows of X are independent with common covariance matrix Σ. Here we assume

$$(9.5) \qquad 0\in\Theta_0, \text{ or } \Theta_0 \text{ contains } 0 \text{ in } R^{np}.$$

Usually, this assumption is satisfied; if necessary, replace X by $X-M_0$ and Θ_0 by Θ_0-M_0, where M_0 is a fixed point in Θ_0. Typical problems of the form (9.4) with (9.5) are the MANOVA, GMANOVA problems, the problems of testing independence or equality of covariance matrices or sphericity. In these problems, except for some special cases, there exist no UMP (uniformly most powerful) tests and usually many tests are proposed in each problem. A feature that these tests have in common is similarity, which is often implied by invariance. In fact, under the null hypotheses in invariant problems, the groups leaving the problems invariant often act transitively on the parameter space $\Theta_0\times\Lambda_0$ so that the null distribution of these tests do not depend on $(M,\Sigma)\in\Theta_0\times\Lambda_0$, and the tests become similar. Now, let us consider the uniqueness of the null distributions in $\mathcal{F}_L$ and $\mathcal{F}_E$, where

$$\mathcal{F}_L=\cup\{\mathcal{F}_L(M)\mid M\in\Theta_0\}$$

and

$$\mathcal{F}_E=\cup\{\mathcal{F}_E(M,I_n\otimes\Sigma)\mid(M,\Sigma)\in\Theta_0\times\Lambda_0\}.$$

Let $\mathcal{Z}=\{Z\in\mathcal{X}\mid Z'Z=I_p\}$. Let $\mathcal{G}l(p)$ denote the group of $p\times p$ nonsingular matrices and let $\mathcal{GU}(p)$ (or $\mathcal{GT}(p)$) denote the group of

$p\times p$ nonsingular upper (or lower) triangular matrices with positive diagonal elements.

Below, $t(X)$ denotes a test statistic for the problem (9.4) with (9.5).

Theorem 9.1 A necessary and sufficient condition for $\mathcal{L}(t(X))$ to remain the same for all $\mathcal{L}(X)\in\mathcal{F}_L$ is that when $\mathcal{L}(X)=N(M, I_n\otimes\Sigma)$, the following conditions (i) and (ii) hold:

(i) $\mathcal{L}\{t(X-M)\}=\mathcal{L}\{t(X)\}$ for all $M\in\Theta_0$ and all $\Sigma\in\mathcal{S}_+(p)$.

(ii) $\mathcal{L}\{t(X)\}=\mathcal{L}\{t(Z)\}$ for $M=0$ and all $\Sigma\in\mathcal{S}_+(p)$, where $X=ZA$ with $Z\in\mathcal{Z}$ and $A\in\mathcal{S}_+(p)$.

Corollary 9.1 The null distributuion of $t(X)$ is unique in $\mathcal{F}_L$ if the following conditions (i)′ and (ii)′ hold:

(i)′ $t(X-M)=t(X)$ for all $M\in\Theta_0$.

(ii)′ $t(XC)=t(X)$ for all $C\in\mathcal{S}_+(p)$ or for all $C\in\mathcal{GU}(p)$ or for all $C\in\mathcal{GT}(p)$.

Corollary 9.2 If (i)′ and (ii)′ hold, $\mathcal{L}\{t(X)\}$ is unique in $\mathcal{F}_E$.

To apply Corollary 1.1 to a specific test, conditions (i)′ and (ii)′ need to be verified. Usually (i)′ is satisfied, but (ii)′ is not in most problems. In invariant problems, if the groups leaving the problems invariant contain as a subgroup $\mathcal{S}_+(p)$ or $\mathcal{GU}(p)$ or $\mathcal{GT}(p)$ acting on X by $X\to XC$, then the condition (ii)′ is clearly satisfied. The MANOVA and GMANOVA problems are typical examples which satisfy (ii)′.

Next, we consider the uniqueness of the null distribution of a test $t(X)$ in $\mathcal{F}_E$. Let $\mathcal{U}=\{U\in\mathcal{X}\mid \mathrm{tr}\,U'U=1\}$ and $||X||=(\mathrm{tr}X'X)^{1/2}$.

Theorem 9.2 A necessary and sufficient condition for $\mathcal{L}(t(X))$ to remain the same for all $\mathcal{L}(X)\in\mathcal{F}_E$ is that when $\mathcal{L}(X)=N(M, I_n\otimes\Sigma)$,

the following conditions (iii) and (iv) hold:

(iii) $\mathcal{L}\{t((X-M)\Sigma^{-1/2})\}=\mathcal{L}\{t(X)\}$ for all $(M,\Sigma)\in\Theta_0\times\Lambda_0$.

(iv) $\mathcal{L}\{t(X)\}=\mathcal{L}\{t(X/||X||)\}$ for $M=0$, $\Sigma=\sigma^2I$ and $\sigma^2>0$.

Corollary 9.3 The (null) distribution of $t(X)$ is unique in $\mathcal{F}_E$ if the following conditions (iii)′ and (iv)′ hold:

(iii)′ $t((X-M)\Sigma^{-1/2})=t(X)$ for all $(M,\Sigma)\in\Theta_0\times\Lambda_0$.

(iv)′ $t(\alpha X)=t(X)$ for all $\alpha>0$.

Chapter 3

THE GENERAL MANOVA PROBLEM

1. Problem and Summary

1.1. *Problem.* The purpose of this chapter is to study the testing problem in a general MANOVA model or a growth curve model in a systematic way, which is described in Section 1 of Chapter 1. That is, the model considered here is

$$Y = X_1 B X_2 + E, \quad E \sim N(0, I \otimes \Omega) \tag{1.1}$$

and under this model the following testing problem will be studied;

$$H : X_3 B X_4 = X_0 \quad \text{versus} \quad K : X_3 B X_4 \neq X_0. \tag{1.2}$$

We called this problem the general MANOVA problem and often abbreviate it as the GMANOVA problem. Our analysis into the problem is based on the invariance principle and only fully invariant tests will be paid a prime attention to. For the general theory of invariance, the readers may be referred to Lehmann (1959), Ferguson (1967) and Eaton (1983). In association to the necessity for our analysis, some selective topics of the theory are reviewed in Chapter 2. Especially, the representation theorem for the probability ratio of distributions of a maximal invariant, which is due

to Wijsman (1967), plays an important role in the analysis. Andersson's version (1983) is also available, which is used in Chapter 5.

1.2. *References.* In contrast to the usual MANOVA problem (see Chapter 1), very little work had been done on the GMANOVA problem using invariance before Kariya (1978). References to work on this model prior to that of Potthoff and Roy (1964) and Rao (1965) can be found in Rao (1967) and Gleser and Olkin (1970) and the papers referred to below. Khatri (1966) derived the LRT (likelihood ratio test) for the above problem by transforming the model and using a conditional argument, and he also proposed some other related tests based on similarities between this problem and the usual MANOVA problem. Using Khatri's approach, some related problems in the model were studied by Krishnaiah (1966). In another direction, Gleser and Olkin (1970) reduced the problem to a canonical form and using the invariance principle, derived the LRT. Kiefer and Schwartz (1965) briefly treated the problem in a special case and proposed a nonivariant Bayes test. In the same case Stein (1966) proposed a conditional test based on the principle of conditionality. Fujikoshi (1973) proved the unconditional monotonicity of the power functions of the LRT and the tests proposed by Khatri (1966). In the paper (1973), he also derived the nonnull asymptotic distributions of these tests. Bayesian Analyses are made by Geisser (1970, 1980), Lee and Geisser (1972, 1975), Fearn (1975).

Kariya (1978), using an invariance-sufficiency argument, established an essentially complete class theorem for fully invariant tests, derived the LBI (locally best invariant) test and showed that it is locally minimax. Kariya and Kanazawa (1978) derived the null distribution of the LBI test for a special case. The admissibility of tests in the GMANOVA problem is questioned by Marden and

Perlman (1980) and Marden (1983), and complete classes are described. In particular, the LRT is shown to be inadmissible (see Section 12). In the group of affine transformations, Banken (1983) proved the maximality of the invariant group Gleser and Olkin (1970) and Kariya (1978) used to analyze the problem, while Nabeya and Kariya (1984) showed it in the group of all homeomorphisms. Further, Banken (1984) made an extension of the GMANOVA problem and reduced it to the GMANOVA problem via invariance. On the other hand, Hooper (1982, 1983) considered the problem of simultaneous interval estimation in the GMANOVA model.

In applications, Ware and Bowden (1977) have applied the model to a circadian rhythm analysis and Zerbe and Jones (1980) to a time series analysis.

1.3. *Summary.* In Section 2, a canonical form of the problem (1.2) under the model (1.1) is obtained and two groups, say $\mathscr{H}$ and $\mathscr{G}$, leaving the canonical problem invariant are chosen where $\mathscr{H}$ is a subgroup of $\mathscr{G}$. Then a test invariant under $\mathscr{G}$ is characterized through a maximal invariant under $\mathscr{H}$, which consists of 4 random matrices (T_1, T_2, T_3, T_4). This is because an analytically tractable maximal invariant under $\mathscr{G}$ is difficult to find. In Section 3, we establish the maximality of the group $\mathscr{G}$ as a group leaving the problem invariant.

In Section 4, by sufficiency approach it is shown that the class of $\mathscr{G}$-invariant tests based on (T_1, T_2) alone forms an essentially complete class among all the $\mathscr{G}$-invariant tests. Under this result, the conditional problem given T_2 becomes exactly the same as the MANOVA problem. Here it is noted that the marginal distribution of T_2 does not depend on unknown parameters, but T_2 is a part of a sufficient statistic so that T_2 is an ancillary statistic. Some distributional results are prepared and the LRT is derived.

In Section 5, the essentially complete class in Section 4 is again

obtained by invariance approach, where Wijsman's theorem for the representation of the probability ratio of a maximal invariant is used. This section is also a preparation for Sections 6–9. In Section 6, the distributions of maximal invariants are discussed.

In Section 7, the LRT and its related tests are considered. In a special case the LRT is shown to be UMPI (uniformly most powerful invariant) in the class of conditional $\mathscr{G}$-invariant level α tests given T_2, though it is not UMPI in the class of all the $\mathscr{G}$-invariant level α tests. Some additional properties of the LRT are investigated. In Section 8, a unique LBI (locally best invariant) test is derived. Hence it is locally uniformally better than any other $\mathscr{G}$-invariant test including the LRT and so it is admissible in the class of $\mathscr{G}$-invariant tests. In Section 9, the LBI test is shown to be locally minimax. In Section 10, the exact distribution of the LBI test are derived for a certain case and the asymptotic null distribution up to n^{-1} is derived in a general case.

In Section 11, the monotonicity property of the tests are discussed and in Section 12, the admissibility and the robustness of the tests are briefly treated.

Recall that $\mathcal{O}(n)$, $\mathscr{G}l(p)$ and $\mathscr{S}_+(p)$ denote respectively the group of $n\times n$ orthogonal matrices, the group of $p\times p$ nonsingular matrices and the set of $p\times p$ positive definite matrices. For $A\in\mathscr{S}_+(p)$, $A^{1/2}$ denotes the symmetric square root of A, *i.e.*, $(A^{1/2})^2=A$ unless otherwise stated.

2. A Canonical Form and Invariance

2.1. *Canonical form.* To analyze the problem stated in Section 1, we restrict our attention to the class of tests which are invariant under a group leaving the problem invariant. As in Gleser and Olkin (1970), we reduce the problem to a canonical form and derive a maximal invariant under a particular group leaving the problem invariant. However as Gleser and Olkin pointed out, there is a bigger group leaving the problem invariant although its maximal invariant is rather complicated. In Gleser and Olkin (1970), invariance is utilized only to derive the LRT, so their attention is restricted to a smaller group under which an analytically tractable maximal invariant is available. On the other hand, since our concern is in a possible reduction of the problem through invariance, our attention is focused on the class of tests invariant under the bigger group. We shall call the bigger group full group and a test invariant under the full group a fully invariant test. However the smaller group is utilized to describe the class of fully invariant tests since an analytically tractable maximal invariant under the full group is not available.

The model under consideration is

$$Y : n\times p \sim N(X_1BX_2, I_n\otimes\Omega), \tag{2.1}$$

where

$$\begin{aligned} &X_1 : n\times(n_1+n_2), && \mathrm{rank}(X_1)=n_1+n_2,\\ &X_2 : (p_1+p_2)\times p, && \mathrm{rank}(X_2)=p_1+p_2,\\ &p=p_1+p_2+p_3, && n=n_1+n_2+n_3,\\ &B : (n_1+n_2)\times(p_1+p_2), && \text{and } \Omega\in\mathcal{S}_+(p). \end{aligned}$$

And the problem is to test the GMANOVA hypothesis

$$H_0 : X_3BX_4=X_0 \quad\text{versus}\quad H_1 : X_3BX_4\neq X_0, \tag{2.2}$$

where

$$\begin{aligned} &X_3 : n_1\times(n_1+n_1), && \mathrm{rank}(X_3)=n_1,\\ &X_4 : (p_1+p_2)\times p_2, && \mathrm{rank}(X_4)=p_2. \end{aligned}$$

For a canonical reduction, we use

Lemma 2.1. For any $m\times n$ matrix A with $\mathrm{rank}(A)=m$, there exist matrices $E\in\mathcal{G}l(m)$ and $P\in\mathcal{O}(n)$ such that $A=E(I_m, 0)P$.

By this lemma we write X_1 and X_2 as

$$X_1=P_1\begin{pmatrix} I\\ 0\end{pmatrix}E_1 \ ;\ P_1\in\mathcal{O}(n),\quad E_1\in\mathcal{G}l(n_1+n_2) \quad\text{and}$$

$$X_2=E_2(I, 0)P_2\ ;\ E_2\in\mathcal{G}l(p_1+p_2),\quad P_2\in\mathcal{O}(p).$$

Let

$$Y^*=P_1'YP_2',\quad B^*=E_1BE_2,\quad \Omega^*=P_2'\Omega P_2. \tag{2.3}$$

So $Y^*\sim N(\begin{pmatrix} B^* & 0\\ 0 & 0\end{pmatrix}, I_n\otimes\Omega^*)$, where $B^* : (n_1+n_2)\times(p_1+p_2)$. On the other hand, for the hypothesis we write X_3BX_4 as

$$X_3BX_4=X_3E_1^{-1}E_1BE_2E_2^{-1}X_4=X_3E_1^{-1}B^*E_2^{-1}X_4. \tag{2.4}$$

And apply the above lemma to $X_3E_1^{-1}$ and $E_2^{-1}X_4$ to get

$$X_3E_1^{-1}=E_3(I, 0)P_3\ ;\ E_3\in\mathcal{G}l(n_1),\quad P_3\in\mathcal{O}(n_1+n_2) \quad\text{and}$$

$$E_2^{-1}X_4=P_4\begin{pmatrix} 0\\ I\end{pmatrix}E_4 \ ;\ P_4\in\mathcal{O}(p_1+p_2),\quad E_4\in\mathcal{G}l(p_2).$$

Hence from (2.4), the null hypothesis is equivalent to

$$H_0 : (I_{n_1}, 0)\,\Theta\begin{pmatrix} 0\\ I_{p_2}\end{pmatrix}=E_3^{-1}X_0E_4^{-1}, \tag{2.5}$$

where $\Theta=P_3B^*P_4 : (n_1+n_2)\times(p_1+p_2)$. Let $n_3=n-(n_1+n_2)$ and

$$\tilde{\Theta}\equiv\begin{pmatrix}\bar{\Theta}\\ 0\end{pmatrix}\equiv\begin{pmatrix}\Theta & 0\\ 0 & 0\end{pmatrix}\equiv\begin{array}{c} \begin{array}{ccc} p_1 & p_2 & p_3 \end{array}\\ \begin{pmatrix}\Theta_{11} & \Theta_{12} & \Theta_{13}\\ \Theta_{21} & \Theta_{22} & \Theta_{23}\\ 0 & 0 & 0\end{pmatrix}\end{array}\begin{array}{c} \\ n_1\\ n_2\\ n_3\end{array} \tag{2.6}$$

$$\Theta_{13}\equiv0 \text{ and } \Theta_{23}\equiv0$$

where $\bar{\Theta}=(\Theta, 0) : (n_1+n_2)\times p$. Then the hypothesis H_0 in (2.5) is equivalent to $H_0 : \Theta_{12}=E_3^{-1}X_0E_4^{-1}$.

Here without loss of generality, we assume $X_0=0$ so that it is expressed as $H_0 : \Theta_{12}=0$. Now define Z and Σ respectively by

$$Z=\begin{pmatrix} P_3 & 0 \\ 0 & I \end{pmatrix} Y^* \begin{pmatrix} P_4 & 0 \\ 0 & I \end{pmatrix} : n\times p, \text{ and } \Sigma=\begin{pmatrix} P_4' & 0 \\ 0 & I \end{pmatrix} \Omega^* \begin{pmatrix} P_4 & 0 \\ 0 & I \end{pmatrix} : p\times p.$$

Then our canonically reduced model is $Z\sim N(\tilde{\Theta}, I_n\otimes\Sigma)$ or

$$(2.7)\quad Z=\begin{array}{c} \begin{array}{ccc} p_1 & p_2 & p_3 \end{array} \\ \begin{pmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \\ Z_{31} & Z_{32} & Z_{33} \end{pmatrix} \end{array} \begin{array}{c} \\ n_1 \\ n_2 \\ n_3 \end{array} \sim N\left(\begin{pmatrix} \Theta_{11} & \Theta_{12} & 0 \\ \Theta_{21} & \Theta_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; I_n\otimes\begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix}\right)$$

where $\Sigma_{ij} : p_i\times p_j$ and the problem is to test

$$(2.8)\qquad H_0 : \Theta_{12}=0 \quad \text{versus} \quad H_1 : \Theta_{12}\neq 0.$$

This is a canonical form of the problem. Here we may further reduce the model (2.7) by sufficiency and get

$$\bar{Z}=\begin{pmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \end{pmatrix} \sim N(\bar{\Theta}, I_{n_1+n_2}\otimes\Sigma)$$

$$(2.9)\qquad V\equiv(V_{ij})=(Z_{3i}'Z_{3j}) : p\times p \sim W_p(\Sigma, n_3) \text{ and}$$

$\bar{Z}$ and V are independent,

where $\bar{\Theta}=(\Theta, 0)$ as in (2.6) and $W_p(\Sigma, n_3)$ denotes the Wishart distribution with mean $n_3\Sigma$ and degrees of freedom n_3. The density function with respect to the Lebesgue measure is

$$f_{(z,v)}(\bar{z}, v|\bar{\Theta}, \Sigma)=c|\Sigma|^{-\frac{m}{2}}\operatorname{etr}(-\frac{1}{2}\Sigma^{-1}(\bar{z}-\bar{\Theta})'(\bar{z}-\bar{\Theta}))$$
$$\times|\Sigma|^{-\frac{n}{2}}|v|^{\frac{n-p-1}{2}}\operatorname{etr}(-\frac{1}{2}\Sigma^{-1}v).$$

Clearly this is an exponential family but since we have the extra information $\Theta_{13}=0$, $\Theta_{23}=0$ on the mean $\bar{\Theta}$ of $\bar{Z}$, the sufficient statistic $(\bar{Z}, V)$ is not complete for $(\tilde{\Theta}, \Sigma)$ or (Θ, Σ) and our model is a "curved" exponential family. In terms of the natural parameters

of the exponential family, the above density can be written as

$$f_{(z,v)}(\bar{z}, v|\Delta, \eta)=\beta(\Delta, \eta)h(v)\exp[-\frac{1}{2}\operatorname{tr}\Delta(\bar{z}'\bar{z}+v)+\operatorname{tr}\bar{z}'\eta],$$

where $\Delta\equiv\Sigma^{-1}$ and $\eta\equiv\bar{\Theta}\Sigma^{-1}$. It is worth noting that neither the information $\Theta_{13}=0$, $\Theta_{23}=0$ on $\bar{\Theta}$ nor the hypothesis $\Theta_{12}=0$ is linear in the natural parameter (Δ, η) in the density. This nonlinearity of the model and the hypothesis makes the analysis of the problem difficult. We also note that when $p_1=p_3=0$, the problem is nothing but the usual MANOVA problem. In this chapter we mainly concern the case $p_3\neq 0$ where we have the extra information on $\tilde{\Theta}$ or $(\tilde{\Theta}, \Sigma)$. From the viewpoint of a curved model, the problem of estimating a GMANOVA model is considered in Kariya (1983).

Before the problem is analyzed, let us make a correspondence between the canonical variable Z and the original variable Y. Let $\bar{X}_1$ and $\bar{X}_2$ be respectively $n\times n_3$ and $p_3\times p$ matrices such that

$$\operatorname{rank}([X_1, \bar{X}_1])=n, \quad \bar{X}_1'X_1=0, \quad \operatorname{rank}\left(\begin{bmatrix} X_2 \\ \bar{X}_2 \end{bmatrix}\right)=p \text{ and } X_2\bar{X}_2'=0.$$

Also let $\bar{X}_3$ and $\bar{X}_4$ be respectively $n_2\times(n_1+n_2)$ and $(p_1+p_2)\times p_1$ matrices such that

$$\operatorname{rank}\left(\begin{bmatrix} X_3 \\ \bar{X}_3 \end{bmatrix}\right)=n_1+n_2, \quad X_3\bar{X}_3'=0,$$
$$\operatorname{rank}([\bar{X}_4, X_4])=p_1+p_2 \text{ and } \bar{X}_4'X_4=0.$$

Then for $P_i's$ and $E_i's$ in (2.3) we can take

$$E_1=(X_1'X_1)^{1/2}, \qquad E_2=(X_2X_2')^{1/2}$$

$$P_1=[X_1(X_1'X_1)^{-1/2}, \bar{X}_1(\bar{X}_1'\bar{X}_1)^{-1/2}] \text{ and } P_2=\begin{pmatrix} (X_2X_2')^{-1/2}X_2 \\ (\bar{X}_2\bar{X}_2')^{-1/2}\bar{X}_2 \end{pmatrix}$$

and for P_3, P_4, E_3 and E_4,

$$E_3=(X_3E_1^{-2}X_3')^{1/2}, \quad E_4=(X_4'E_2^{-2}X_4)^{1/2}$$

$$P_3=\begin{pmatrix}(X_3E_1^{-2}X_3')^{-1/2}X_3E_1^{-1}\\ (\overline{X}_3E_1^2\overline{X}_3')^{-1/2}\overline{X}_3E_1\end{pmatrix} \quad \text{and}$$

$$P_4=[E_2\overline{X}_4(\overline{X}_4'E_2^2\overline{X}_4)^{-1/2},\quad E_2^{-1}X_4(X_4'E_2^{-2}X_4)^{-1/2}].$$

Using these expressions, from (2.3) and (2.7) Z_{ij}'s are written in the original term. In particular,

$$Z_{12}=[X_3(X_1'X_1)^{-1}X_3']^{-1/2}X_3\hat{B}_0X_4[X_4'(X_2X_2')^{-1}X_4]^{-1/2}$$

$$\hat{B}_0=(X_1'X_1)^{-1}X_1'YX_2'(X_2X_2')^{-1},$$

$$Z_{13}=[X_3(X_1'X_1)^{-1}X_3']^{-1/2}X_3(X_1'X_1)^{-1}X_1'Y\overline{X}_2'(\overline{X}_2\overline{X}_2')^{-1/2}, \tag{2.10}$$

$$Z_{32}=(\overline{X}_1'\overline{X}_1)^{-1/2}X_1'YX_2'(X_2X_2')^{-1}X_4[X_4'(X_2X_2')X^4]^{-1/2}, \text{ and}$$

$$Z_{33}=(\overline{X}_1'\overline{X}_1)^{-1/2}\overline{X}_1'Y\overline{X}_2'(\overline{X}_2\overline{X}_2')^{-1/2}$$

It is noted that $\overline{X}_1(\overline{X}_1'\overline{X}_1)^{-1}\overline{X}_1'=I_n-X_1(X_1'X_1)^{-1}X_1'$ and $\overline{X}_2'(\overline{X}_2\overline{X}_2')^{-1}\overline{X}_2=I_p-X_2'(X_2X_2')^{-1}X_2$.

2.2. *Reduction via invariance.* Now to analyze the problem via invariance, we consider the groups which leave invariant the problem under the model (2.7). Let

$$\mathscr{A}=\left\{\begin{pmatrix}A_{11}&A_{12}&A_{13}\\0&A_{22}&A_{23}\\0&0&A_{33}\end{pmatrix}\in\mathscr{G}l(p)\,|\,A_{ii}\in\mathscr{G}l(p_i)\ \ (i=1,2,3)\right\} \tag{2.11}$$

$$\mathscr{F}=\left\{F:n\times p\,|\,F=\begin{array}{c}\begin{array}{ccc}p_1&p_2&p_3\end{array}\\ \begin{pmatrix}F_{11}&0&0\\F_{21}&F_{22}&0\\0&0&0\end{pmatrix}\end{array}\begin{array}{c}\\ n_1\\ n_2\\ n_3\end{array}\right\} \quad \text{and} \tag{2.12}$$

$$\mathscr{P}=\left\{\begin{pmatrix}P_1&0&0\\0&P_2&0\\0&0&P_3\end{pmatrix}\in\mathscr{O}(n)\,|\,P_i\in\mathscr{O}(n_i)\right\}=\mathscr{O}(n_1)\times\mathscr{O}(n_2)\times\mathscr{O}(n_3) \tag{2.13}$$

And consider the group

$$\mathscr{G}=\mathscr{P}\times\mathscr{A}\times\mathscr{F}=\mathscr{O}(n_1)\times\mathscr{O}(n_2)\times\mathscr{O}(n_3)\times\mathscr{A}\times\mathscr{F} \tag{2.14}$$

with group operation $(P_2,A_2,F_2)\cdot(P_1,A_1,F_1)=(P_2P_1,\ A_2A_1,\ P_2F_1A_2'+F_1)$ for $(P_i,A_i,F_i)\in\mathscr{G}$ (i=1,2). Then $\mathscr{G}$ leaves the problem invariant by the left action

$$g(Z)=PZA'+F \quad \text{for } g=(P,A,F)\in\mathscr{G}. \tag{2.15}$$

In fact, the group acting on the parameter space is $\overline{\mathscr{G}}=\mathscr{G}$ with the left action

$$\bar{g}(\Theta,\Sigma)=(P\Theta A'+F,\ A\Sigma A') \quad \text{for } \bar{g}=(P,A,F)\in\overline{\mathscr{G}}. \tag{2.16}$$

And it is easy to see that the distribution of $g(Z)$ under $\bar{g}(\Theta,\Sigma)$ is the same as that of Z under (Θ,Σ) and that $\overline{\mathscr{G}}\equiv\mathscr{G}$ preserves the hypothesis. Hence the problem is left invariant under $\mathscr{G}$. We shall call this group the full group. Since it is difficult to find an analytically tractable maximal invariant under $\mathscr{G}$, we consider a smaller group $\mathscr{H}=\mathscr{O}(n_3)\times\mathscr{A}\times\mathscr{F}$, which is isomorphic to the subgroup $\{I_{n_1}\}\times\{I_{n_2}\}\times\mathscr{O}(n_3)\times\mathscr{A}\times\mathscr{F}$ of the group $\mathscr{G}$. Hence $\mathscr{H}$ leaves the problem invariant under the action where $h=\bar{h}=(A,\ F)\in\mathscr{H}\equiv\overline{\mathscr{H}}$.

$$h(Z)=PZA'+F \quad \text{and} \quad \bar{h}(\Theta,\Sigma)=(P\Theta A'+F,\ A\Sigma A') \tag{2.17}$$

where $h=\bar{h}=(P_3,\ A,\ F)\in\mathscr{H}\equiv\overline{\mathscr{H}}$ with $P=\mathrm{DIAG}\{I_{n_1},\ I_{n_2},\ P_3\}$. Here $\mathrm{DIAG}\{A,B,C\}$ denotes the block diagonal matrix with A, B and C as diagonal blocks. Note that under the subgroup $\mathscr{O}(p_3)$ acting on Z by $Z\to PZ$, the sufficient statistic $(\overline{Z},V)$ in (2.9) is a maximal invariant. This implies a maxinal invariant under $\mathscr{H}$ is a function of $(\overline{Z},V)$ and that under the full group $\mathscr{G}$, a maximal invariant is also a function of $(\overline{Z},V)$ since $\mathscr{G}$-invariance implies $\mathscr{H}$-invariance. In this sense, the model (2.9) reduced by sufficiency can be regarded as the model reduced by invariance, namely by the group $\mathscr{O}(n_3)$. This point will be again referred later. And a maximal invariant under $\mathscr{H}$ is given by

Proposition 2.1. (Gleser and Olkin (1970)) A maximal invariant under the group $\mathscr{H}$ is $s(Z)=\{s_1(Z),\ s_2(Z)\}$ where

$$(2.18)\qquad \begin{aligned} s_1(Z)&=(Z_{12},Z_{13})\begin{pmatrix} V_{22} & V_{23} \\ V_{32} & V_{33}\end{pmatrix}^{-1}(Z_{12},Z_{13})' \\ s_2(Z)&=\begin{pmatrix} Z_{13} \\ Z_{23}\end{pmatrix}V_{33}^{-1}\begin{pmatrix} Z_{13} \\ Z_{23}\end{pmatrix}'. \end{aligned}$$

A proof is given in Appendix.

We remark that Gleser and Olkin (1970) imposed the conditions $n_1+n_2\leq p_3$ and $n_1\leq p_2$ in the statement of the proposition. But this is unnecessary.

Proposition 2.2. (Gleser and Olkin (1970)) (a) Uner $\bar{\mathscr{H}}=\mathscr{H}$, a maximal invariant parameter is

$$\Upsilon\equiv\Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}' \quad \text{where} \quad \Sigma_{22.3}=\Sigma_{22}-\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32}.$$

(b) Under $\bar{\mathscr{G}}=\mathscr{G}$, a maximal invariant parameter is the set of the ordered characteristic roots of $\Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}'$; with $\delta_i=\mathrm{ch}_i(\Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}')$

$$\delta=(\delta_1,\ldots,\delta_{n1})', \quad \delta_1\geq\ldots\geq\delta_{n_1}\geq 0$$

where $\mathrm{ch}_i(A)$ denotes the i-th largest characteristic root of A.

From these propositons and from the general theory of invariance, the power function of an invariant test under $\mathscr{H}$ is a function of $\Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}'$ and the power function of an invariant test under $\mathscr{G}$ is a function of δ (Lemma 1.2 of Chapter 2).

Next we describe the class of fully invariant tests, that is, tests invariant uder $\mathscr{G}$. Let $\mathscr{D}(\mathscr{H})$ be the class of all invariant tests of level α under $\mathscr{H}$ and let $\mathscr{D}(\mathscr{G})$ be the class of all invariant tests of level α under $\mathscr{G}$. Instead of the form of the maximal invariant $s(Z)$ derived in Proposition 2.1, we choose an alternative form which is in one-to-one correspondence with $s(Z)$. From (2.18),

$$s_1(Z)=(Z_{12}-Z_{13}V_{33}^{-1}V_{32})V_{22.3}^{-1}(Z_{12}-Z_{13}V_{33}^{-1}V_{32})'+Z_{13}V_{33}^{-1}Z_{13}',$$

and

$$s_2(Z)=\begin{pmatrix} Z_{13}V_{33}^{-1}Z_{13}' & Z_{13}V_{33}^{-1}Z_{23}' \\ Z_{23}V_{33}^{-1}Z_{13}' & Z_{23}V_{33}^{-1}Z_{23}'\end{pmatrix}, \quad \text{where } V_{22.3}^{-1}=V_{22}-V_{23}V_{33}^{-1}V_{32}.$$

A convenient choice of a maximal invariant under $\mathscr{H}$ is $t(Z)=\{t_1(Z),\ t_2(Z),\ t_3(Z),\ t_4(Z)\}$, where

$$(2.19)\qquad T_1\equiv t_1(Z)=xV_{22.3}^{-1}x' : n_1\times n_1 \quad \text{with}$$

$$(2.20)\qquad x=(I+T_2)^{-1/2}(Z_{12}-Z_{13}V_{33}^{-1}V_{32}),$$

$$(2.21)\qquad T_2\equiv t_2(Z)=Z_{13}V_{33}^{-1}Z_{13}' : n_1\times n_1$$

$$(2.22)\qquad T_3\equiv t_3(Z)=Z_{23}V_{33}^{-1}Z_{23}' : n_2\times n_2 \quad \text{and}$$

$$(2.23)\qquad T_4\equiv t_4(Z)=Z_{13}V_{33}^{-1}Z_{23}' : n_1\times n_2.$$

Here $(I+T_2)^{-1/2}\in\mathscr{S}_+(n_1)$. With this choice, any $\mathscr{H}$-invariant test is a measurable function of $t(Z)$.

Next, in terms of $t(Z)$ we describe the class of $\mathscr{G}$-invariant tests. Since $\mathscr{H}$ can be regarded as a subgroup of $\mathscr{G}$, $\mathscr{D}(\mathscr{G})\subset\mathscr{D}(\mathscr{H})$, that is, any $\mathscr{G}$-invariant test is $\mathscr{H}$-invariant. Consequently a $\mathscr{G}$-invariant test is a function of the maximal invariant $t(Z)$. Therefore for any test ϕ in $\mathscr{D}(\mathscr{G})$, we can write

$$(2.24)\qquad \phi(Z)=\phi_0(t_1(Z),\ t_2(Z),\ t_3(Z),\ t_4(Z))$$

for some ϕ_0 defined on the space of (T_1,T_2,T_3,T_4).

Lemma 2.3. Suppose a group G satisfies $G=HK$ for two subgroups K,H of G. Suppose G acts on a set $\mathscr{X}$ and let $\tau:\mathscr{X}\to\mathscr{I}$ be a maximal invariant under H where $\mathscr{I}$ is the range of τ. Suppose $K\to\bar{K}$ is a homomorphism of K onto a group $\bar{K}$ such that $\tau(k(x))=\bar{k}\tau(x)$ for all $k\in K$ where $\bar{k}$ is the homomorphic image of k. Then a function ψ defined on $\mathscr{X}$ is G-invariant if and only if (1) ψ is H-invariant, say $\psi(x)=\tilde{\psi}(\tau(x))$ and (2) $\tilde{\psi}(\bar{k}t)=\tilde{\psi}(t)$ for all $k\in K$ and all $t\in\mathscr{I}$.

Proof. Suppose ψ is G-invariant. Then ψ is H-invariant, so $\psi(x)=\tilde{\psi}(\tau(x))$ for some $\tilde{\psi}$ defined on $\mathcal{I}$. Also $\psi(kx)=\tilde{\psi}(\tau(kx))=\tilde{\psi}(\bar{k}\,\tau(x))=\psi(x)=\tilde{\psi}(\tau(x))$, hence $\tilde{\psi}(\bar{k}t)=\tilde{\psi}(t)$ for all $t\in\mathcal{I}$. The converse is clear since for any $g\in G$, $g=hk$ for some $h\in H$ and $k\in K$ and $\tau(gx)=\tau(hkx)=\tau(kx)=\bar{k}\tau(x)$.

Proposition 2.3. The class of $\mathcal{G}$-invariant tests $\mathcal{D}(\mathcal{G})$ can be specified as the set of all tests ϕ in $\mathcal{D}(\mathcal{H})$ such that whenever ϕ is expressed in the form of (2.24), ϕ_0 satisfies

$$(2.25)\qquad \phi_0(P_1t_1P_1',\ P_1t_2P_1',\ P_2t_3P_2',\ P_1t_4P_2') = \phi_0(t_1, t_2, t_3, t_4)$$

Proof. Let $H=\{I_m\}\times\{I_{m_2}\}\times\mathcal{H}=\{I\}\times\{I\}\times\mathcal{O}(n_3)\times\mathcal{A}\times\mathcal{F}$, $K=\mathcal{O}(n_1)\times\mathcal{O}(n_2)\times\{I_{n_3}\}\times\{I_p\}\times\{0\}$ and $\bar{K}=\tilde{\mathcal{O}}$ in Lemma 2.3. Then all the assumptions in the lemma are satisfied. Note

$$\bar{k}t(Z) = \{P_1t_1(Z)P_1',\ P_1t_2(Z)P_1',\ P_2t_3(Z)P_2',\ P_1t_4(Z)P_2'\}$$

for $\bar{k}=\begin{pmatrix}P_1 & 0\\ 0 & P_2\end{pmatrix}\in\mathcal{O}(n_1)\times\mathcal{O}(n_2)$. Hence the result follows.

The proposition also follows from Lemma 1.4 in Chapter 2.

By this proposition, an $\mathcal{H}$-invariant test is a measurable function from the range space $\mathcal{T}_1\times\mathcal{T}_2\times\mathcal{T}_3\times\mathcal{T}_4$ of $T\equiv(T_1,T_2,T_3,T_4)\equiv t(Z)$ into $[0,1]$, where $\mathcal{T}_1=\mathcal{T}_2=R^{n_1(n_1+1)/2}$, $\mathcal{T}_3=R^{n_2(n_2+1)/2}$ and $\mathcal{T}_4=R^{n_1n_2}$, while a $\mathcal{G}$-invariant test is characterized as an $\mathcal{H}$-invariant test satisfying (2.25). In addition, in terms of a maximal invariant parameter $\delta=(\delta_i)=\{\mathrm{ch}_i(\Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}')\}$ in Proposition 2.2, the hypothesis in (2.8) is stated as

$$(2.26)\qquad H_0:\delta_1=0 \quad\text{versus}\quad H_1:\delta_1>0,$$

which is the reduced problem via $\mathcal{G}$-invariance. This formulation is sometimes used without reference.

Finally to make a correspondence between $(x, V_{22.3}, T_2)$ and the original terms, let $N_1=I-X_1(X_1'X_1)^{-1}X_1'$,

$$(2.27)\qquad A_1 = [X_3(X_1'X_1)^{-1}X_3']^{-1/2}X_3(X_1'X_1)^{-1}Y_1' \quad\text{and}$$

$$A_2 = X_2'(X_2X_2')^{-1}X_4[X_4'(X_2X_2')^{-1}X_4]^{-1/2}.$$

Then by using (2.10), we obtain

Lemma 2.3. (1) $V_{22\cdot3}=[X_4'(X_2X_2')^{-1}X_4]^{-1/2}X_4'[X_2(Y'N_1Y)^{-1}X_2']^{-1}X_4 \times[X_4'(X_2X_2')^{-1}X_4]^{-1/2}$.

(2) $Z_{12}-Z_{13}V_{33}^{-1}V_{32} = A_1Y(Y'N_1Y)^{-1}X_2'[X_2(Y'N_1Y)^{-1}X_2']^{-1}X_2A_2$.

(3) $T_2 = A_1Y\{(Y'N_1Y)^{-1} -(Y'N_1Y)^{-1}X_2'[X_2(Y'N_1Y)^{-1}X_2']^{-1}X_2(Y'N_1Y)^{-1}\}Y'A_1'$.

Proof. The proof is straightforward, and omitted.

From this lemma, $T_1 = xV_{22.3}^{-1}x'$ is expressed as

$$(2.28)\qquad T_1 = (I+T_2)^{-1/2}A_1Y(Y'N_1Y)^{-1}X_2'[X_2(Y'N_1Y)^{-1}X_2']^{-1}X_4 \times\{X_4'[X_2(Y'N_1Y)^{-1}X_2']^{-1}X_4\}^{-1}X_4'[X_2(Y'N_1Y)^{-1}X_2']^{-1} \times X_2(Y'N_1Y)^{-1}Y'A_1'(I+T_2)^{-1/2}.$$

In particular, if $X_2=I$, then $T_2=0$ and

$$(2.29)\qquad T_1=A_1YX_4[X_4'(Y'N_1Y)X_4]^{-1}X_4'Y'A_1'.$$

3. Maximality of the Group.

3.1. *Banken's result.* In section 2, the group $\mathcal{G}$ in (2.14) has been called the full group relative to our problem. In this section, this terminology will be justified based on Banken (1983). That is, it will be shown that the group $\mathcal{G}$ is maximal among the affine linear groups which leave our problem invariant. First it is noted that any affine linear transformation $f(x)=Bx+c$ from R^{np} into R^{np} is written as

$$\bar{f}(x)=\Sigma_{i=1}^{r}P_iXQ_i'+C \quad\text{with}\quad B=\Sigma_{i=1}^{r}P_i\otimes Q_i$$

for some $P_i: n\times n$, $Q_i: p\times p$ and r, where X and C are $n\times p$ matrices with $X'=[x_1,\dots,x_n]$ and $C'=[c_1,\dots,c_n]$, $x=(x_1',\dots,x_n')'$ and $c=(c_1',\dots,c_n')'$. The next theorem is due to Banken (1983).

Theorem 3.1. (Banken (1983)) Suppose $X: n\times p \sim N(\mu, I_n \otimes \Sigma)$ with $\mu \in R^{np}$ and $\Sigma \in \mathcal{S}_+(p)$. Then the maximal group of affine linear transformations which leaves the model invariant is $\mathcal{K}=\mathcal{O}(n)\times \mathcal{Gl}(p)\times R^{np}$.

Now we shall show the maximality of our group. The following result is due to Banken (1983) but the proof is different.

Theorem 3.2. When $Z: n\times p \sim N(\tilde{\Theta}, I_n \otimes \Sigma)$ with $\tilde{\Theta}$ in (2.6) as in (2.7), the maximal affine linear group that leaves the problem (2.8) invariant is $\mathcal{G}=\mathcal{P}\times\mathcal{A}\times\mathcal{F}$ in (2.14).

Proof. Let $f(Z)=\Sigma_{j=1}^{r} P_j Z Q_j' + C$ be an affine transformation. For f to preserve the covariance structure of the form $I\otimes\Sigma$, it follows from the proof of Theorem 3.1 that $B \equiv \Sigma_{j=1}^{r} P_j \otimes Q_j = P\otimes A$ for some $P\in\mathcal{O}(n)$ and $A\in\mathcal{Gl}(p)$ so that $f(Z)=PZA'+C$.

Now for this f to preserve the mean structure of $\tilde{\Theta}$ and the hypothesis $\Theta_{12}=0$, $P\tilde{\Theta}A'+C\in\mathcal{F}$ for any $\tilde{\Theta}$ of the form (2.6) with $\Theta_{12}=0$, where $\mathcal{F}$ is given in (2.12). From the case $\tilde{\Theta}=0$, $C\in\mathcal{F}$ follows.

To show $P\in\mathcal{P}$ and $A\in\mathcal{A}$, write $P=(P_{ij})$ with $P_{ij}: n_i\times n_j$ and $A'=E=(E_{ij})$ with $E_{ij}: p_i\times p_j$ and let $(P\tilde{\Theta}E)_{ij}: n_i\times p_j$ denote the (i,j) block matrix of $P\tilde{\Theta}E$. From $(P\tilde{\Theta}E)_{1i}=0$ $(i=2,3)$, $(P\tilde{\Theta}E)_{22}=0$ and $(P\tilde{\Theta}E)_{3j}=0$ $(j=1,2,3)$,

$$(3.1)\quad \begin{cases} \text{(a)}\ P_{11}\Theta_{11}E_{1i}+P_{12}\Theta_{21}E_{1i}+P_{12}\Theta_{22}E_{2i}=0 & (i=2,3) \\ \text{(b)}\ P_{21}\Theta_{11}E_{13}+P_{22}\Theta_{21}E_{13}+P_{22}\Theta_{22}E_{23}=0 & \\ \text{(c)}\ P_{31}\Theta_{11}E_{1j}+P_{32}\Theta_{21}E_{1j}+P_{32}\Theta_{22}E_{2j}=0 & (j=1,2,3) \end{cases}$$

Here set $\Theta_{21}=0$ and $\Theta_{22}=0$ to get $E_{13}'\Theta_{11}'(P_{11}', P_{21}', P_{31}')=0$. Multiplying this by $(P_{11}', P_{21}', P_{31}')'$ from the right and using $\Sigma_{i=1}^{3} P_{i1}'P_{i1}=I$ produces $E_{13}'\Theta_{11}'=0$ for any Θ_{11}, implying $E_{13}=0$. Similarly setting $\Theta_{11}=0$ and $\Theta_{21}=0$ and using the same argument, we obtain $E_{23}=0$. Since $E\in$

$\mathcal{Gl}(p)$, $E_{13}=0$ and $E_{23}=0$ implies the nonsingularity of $H\equiv(E_{ij})$ $(i,j=1,2)$. And under $E_{13}=0$ and $E_{23}=0$ the remaining equations are (a) with $i=2$ and (c) with $j=1,2$. Next, setting $\Theta_{11}=0$, from (c) with $j=1,2$, we obtain $P_{32}(\Theta_{21}, \Theta_{22})H=0$ for any Θ_{21} and Θ_{22}. Hence $P_{32}=0$ follows from $|H|\neq 0$. Further, setting $\Theta_{21}=0$ and $\Theta_{22}=0$, from (c) with $j=1,2$ we obtain $P_{31}\Theta_{11}(E_{11}, E_{12})=0$ for any Θ_{11}. Since (E_{11}, E_{12}) is of full rank by $|H|\neq 0$, $P_{31}=0$ follows. Since $P\in\mathcal{O}(n)$, $P_{31}=0$ and $P_{32}=0$ implies $Q=(P_{ij})$ $(i,j=1,2)\in\mathcal{O}(n_1+n_2)$ and so $P_{13}=0$ and $P_{23}=0$. Now the remaining equation is only (a) with $j=2$ in (3.1). There setting $\Theta_{11}=0$,

$$(3.2)\qquad P_{12}(\Theta_{21}, \Theta_{22})K=0 \quad \text{for any } \Theta_{21} \text{ and } \Theta_{22},$$

where $K=[E_{12}', E_{22}']': (p_1+p_2)\times p_2$. Since K is of full rank, for some $F_1\in\mathcal{O}(p_1+p_2)$ and $F_2\in\mathcal{Gl}(p_2)$, $K=F_1\begin{bmatrix} I \\ 0\end{bmatrix}F_2$. Hence (3.4) becomes $P_{12}(\Theta_{21}, \Theta_{22})F_1\begin{bmatrix} I \\ 0\end{bmatrix}=0$ for any Θ_{21} and Θ_{22}, from which $P_{12}=0$ can be concluded. This implies $P_{21}=0$ and so $P=\mathrm{DIAG}\{P_{11}, P_{22}, P_{33}\}\in\mathcal{P}$. Finally from the remaining equation $P_{11}\Theta_{11}E_{12}=0$ and from $P_{11}\in\mathcal{O}(n_1)$, $E_{12}=0$ follows. Therefore $E=(E_{ij})$ is block lower triangular so that $E'=A\in\mathcal{A}$, completing the proof.

Corollary 3.1. In the MANOVA problem, the group $\mathcal{O}(n_1)\times\mathcal{O}(n_2)\times\mathcal{O}(n_3)\times\mathcal{Gl}(p_2)\times R^{np_2}$ is the maximal affine linear group that leaves the problem invariant.

Proof. This follows from the fact that the MANOVA with $p_1=p_3=0$.

The following theorem now follows from Theorem 2.2 in Chapter 2.

Theorem 3.3. The group $\mathcal{G}=\mathcal{P}\times\mathcal{A}\times\mathcal{F}$ in (2.14) is essentially maximal in the group of all homeomorphisms as a subgroup leaving the problem (2.8) invariant.

Proof. A subgroup $\mathcal{O}(n_3)$ in $\mathcal{P}\equiv\mathcal{O}(n_1)\times\mathcal{O}(n_2)\times\mathcal{O}(n_3)$ reduces the model (2.7) to the sufficiently reduced model (2.9) in which $\bar{Z}$ and V are independent and $(\bar{Z}, V)$ is sufficient. Hence by Theorem 2.2 in Chapter 2 and the argument thereafter, the group $\tilde{\mathcal{G}}=\mathcal{O}(n_1)\times\mathcal{O}(n_2)\times\mathcal{G}l(p)\times R^{(n_1+n_2)p}$ is maximal in the group of all homeomorphisms, say $\mathcal{K}$, as a subgroup leaving the model (2.9) invariant. Therefore it follows from Theorem 3.2 that $\mathcal{G}^*=\mathcal{O}(n_1)\times\mathcal{O}(n_2)\times\mathcal{A}\times\mathcal{F}$ is maximal in $\mathcal{K}$ as a subgroup leaving the problem (2.8) invariant. Since Lemma 1.4 in Chapter 2 justifies the stepwise reduction of the model (2.7) to (2.9) by $\mathcal{O}(n_3)$, $\mathcal{G}=\mathcal{G}^*\times\mathcal{O}(n_3)$ is maximal in $\mathcal{K}$ as a subgroup leaving the problem invariant. This completes the proof.

4. Essentially Complete Classes Via Sufficiency.

4.1. *Essentially complete classes.* In section 2, it was shown that a maximal invariant under $\mathcal{H}$ consists of 4 statistics $T_i (i=1, \cdots, 4)$, where $T=(T_1, T_2, T_3, T_4)$ is given in (2.19)-(2.23), and that a $\mathcal{G}$-invariant test is characterized as an $\mathcal{H}$-invariant test satisfying $\phi(P_1T_1P_1', P_1T_2P_2', P_2T_3P_2', P_1T_4P_2')=\phi(T_1, T_2, T_3, T_4)$. In this section, we shall show that for any $\mathcal{G}$-invariant test ψ of level α based on T, there exists a $\mathcal{G}$-invariant level α test ϕ of based on (T_1, T_2) only such that the power of ϕ is the same as that of ψ. In other words, the class of $\mathcal{G}$-invariant tests of level α based on (T_1, T_2) only is shown to be an essentially complete class in the class of all $\mathcal{G}$-invariant tests of level α, $\mathcal{D}(\mathcal{G})$. This implies that without loss of generality, (T_3, T_4) can be discarded from our consideration. We establish the result by a sufficiency approach and by an invariance approach. Let $\mathcal{C}(\mathcal{H})$ and $\mathcal{C}(\mathcal{G})$ be respectively the class of $\mathcal{H}$-invariant tests of level α based on (T_1, T_2) only and the class of $\mathcal{G}$-invariant tests of level α based on (T_1, T_2) only.

Theorem 4.1. (1) $\mathcal{C}(\mathcal{H})$ forms an essentially complete class in $\mathcal{D}(\mathcal{H})$.
(2) $\mathcal{C}(\mathcal{G})$ forms an essentially complete class in $\mathcal{D}(\mathcal{G})$.
(3) Under $H_0: \boldsymbol{\Theta}_{12}=0$, T_1 and T_2 are independent.

To prove Theorem 4.1, the distributional properties of the statistics T_i $(i=1, \cdots, 4)$ in (2.19)-(2.23) are examined.

Definition 4.1. Let U_1, U_2 and U_3 be random matrices. Then U_1 and U_2 are said to be conditionally independent given U_3 if

$$P(U_1\in A_1,\ U_2\in A_2 \mid U_3)=\prod_{i=1}^{2} P(U_i\in A_i \mid U_3),$$

for Borel sets A_i's on the spaces of U_i's $(i=1, 2)$, where $P(\cdot \mid U_3)$'s are versions of conditional distributions given U_3.

Lemma 4.1. (1) Given (Z_{13}, V_{33}), $x\sim N((I+T_2)^{-1/2}\Theta_{12}, I\otimes\Sigma_{22.3})$. Hence the conditional distribution of x given (Z_{13}, V_{33}) depends on (Z_{13}, V_{33}) only through T_2.
(2) $V_{22.3}\sim W_{p_2}(\Sigma_{22.3}, n_3-p_3)$, and $V_{22.3}$ is independent of (x, T_2, T_3, T_4).
(3) Given (Z_{13}, V_{33}), T_1 and (T_3, T_4) are conditionally independent.
(4) The joint distribution of (T_2, T_3, T_4) does not depend on the maximal invariant parameter $\Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}'$.

Proof. First note
(a) Z_{12} given $Z_{13}\sim N(\Theta_{12}+Z_{13}\Sigma_{33}^{-1}\Sigma_{32},\ I_{n_1}\otimes\Sigma_{22.3})$
(b) $Z_{13}\sim N(0,\ I_{n_1}\otimes\Sigma_{33})$
(c) $V_{33}^{-1/2}V_{32}$ given $V_{33}\sim N(V_{33}^{-1/2}\Sigma_{33}^{-1}\Sigma_{32},\ I\otimes\Sigma_{22.3})$
(d) $V_{33}\sim W_{p_3}(\Sigma_{33}, n_3)$ and $V_{22.3}\sim W_{p_2}(\Sigma_{22.3}, n_3-p_3)$
(e) $V_{22.3}$ is independent of (V_{32}, V_{23}).

Therefore it follows that
(f) for given (Z_{13}, V_{33}), Z_{12}, $V_{33}^{-1/2}V_{32}$ and $Z_{13}V_{33}^{-1}Z_{23}'$ are mutually

conditionally independent,

(g) Z_{13} and V_{33} are independent, and

(h) $V_{22.3}$ is independent of all other statistics.

Consequently (1) follows from (a) and (c), (2) follows from (h), (3) follows from (a) and the definition of conditional independence, and (4) follows from the forms of T_i's, for example, write T_3 as

$$T_3 = Z_{23}\Sigma_{33}^{-1/2}(\Sigma_{33}^{-1/2}V_{33}\Sigma_{33}^{-1/2})^{-1}\Sigma_{33}^{-1/2}Z'_{23}.$$

This completes the proof.

Using (1) in Lemma 4.1, Theorem 4.1 (3) follows, while Theorem 4.1 (1) follows from

Lemma 4.2. The statistic (T_1, T_2) is sufficient for the distributions of T.

Proof. First it is shown that T_1 and (T_3, T_4) are conditionally independent given T_2. From (2) in Lemma 4.2, it suffices to show that x and (T_3, T_4) are conditionally independent given T_2. This follows easily from (1) and (3) in Lemma 4.1. Second using the conditional independence of T_1 and (T_3, T_4) given T_2, the conditional distribution of (T_3, T_4) given (T_1, T_2) is the same as that of (T_3, T_4) given T_2. In fact let $A \subset \mathcal{T}_1, B \subset \mathcal{T}_2$ and $C \subset \mathcal{T}_3 \times \mathcal{T}_4$ be measurable sets and let I_F denote the indicator function of a set F. Then

$$\int_{[T_1\in A]\cap[T_2\in B]} P((T_3, T_4)\in C \mid T_1, T_2)dP$$

$$= \int_{[T_1\in A]\cap[T_2\in B]} I_C(T_3, T_4)dP$$

$$= \int_{[T_2\in B]} I_A(T_1) I_C(T_3, T_4)dP$$

$$= \int_{[T_2\in B]} P(T_1\in A, (T_3, T_4)\in C \mid T_2)dP$$

$$= \int_{[T_2\in B]} P(T_1\in A \mid T_2)P((T_3, T_4)\in C \mid T_2)dP$$

(from (1) in Lemma 4.1.)

$$= \int_{[T_2\in B]} I_A(T_1)P((T_3, T_4)\in C \mid T_2)dP$$

(from the definition of conditional expectation)

$$= \int_{[T_1\in A]\cap[T_2\in B]} P((T_3, T_4)\in C \mid T_2)dP$$

Therefore $P((T_3, T_4)\in C \mid T_1, T_2) = P((T_3, T_4)\in C \mid T_2)$ a.e. (T_1, T_2). Finally, since the conditional distribution of (T_3, T_4) given T_2 is paramenterfree by (4) in Lemma 4.1, the result follows.

Now we shall prove Theorem 4.1 (2). First we state a theorem due to Hall, Wijsman and Ghosh (1965). Let $(\mathcal{X}, \mathcal{A}, P)$ where $P\in\mathcal{P}$ be a probability space and let G be a group of measurable bijective transformations of $\mathcal{X}$ such that for each $g\in G$, $g\mathcal{A}=\mathcal{A}$ and $gP\in\mathcal{P}$, where $gP(A)=P(g^{-1}(A))$ for $A\in\mathcal{A}$. Let $\mathcal{A}_I=\{A\in\mathcal{A} \mid gA=A$ for all $g\in\mathcal{G}\}$ be the invariant sub-sigma-field of $\mathcal{A}$ and let $\mathcal{A}_S$ be the sufficient sub-sigma-field of $\mathcal{A}$. An $\mathcal{A}$-measurable function f is said to be almost invariant if for each $g\in G$, $f(gx)=f(x)$ a.e. $(\mathcal{P})$.

Lemma 4.3. (Hall, Wijsman and Ghosh (1965)) Suppose (i) for each $g\in G$, $g\mathcal{A}_S=\mathcal{A}_S$ and (ii) if a function f is $\mathcal{A}_S$-measurable and almost invariant, there exists a $\mathcal{A}_S\cap\mathcal{A}_I$-measurable function $\bar{f}$ such that $f=\bar{f}$ a.e. $(\mathcal{P})$. Then $\mathcal{A}_{SI}=\mathcal{A}_S\cap\mathcal{A}_I$ is a sufficient sigma-field of $\mathcal{A}_I$.

We apply this to our problem. Take the Borel field of the space $\mathcal{T}_1\times\mathcal{T}_2\times\mathcal{T}_3\times\mathcal{T}_4$ of T for $\mathcal{A}$ in Lemma 4.3, take the Borel field of the space $\mathcal{T}_1\times\mathcal{T}_2$ of (T_1, T_2) for $\mathcal{A}_S$ above since (T_1, T_2) is sufficient, and take $\mathcal{O}(n_1)\times\mathcal{O}(n_2)$ for G. The action of G on T is $(P_1T_1P'_1, P_1T_2P'_1, P_2T_3P'_2, P_1T_4P'_2)$ for $(P_1, P_2)\in G$. Then the condition (i) follows. The condition (ii) follows from Lemma 6.1 in Chapter 2

(Lehmann (1959) Chapter 6 Theorem 4). Another proof is given in Appendix.

4.2. *The conditional MANOVA problem and the LRT.* From Theorem 4.1 and Lemma 4.1, our problem is now reduced to the following; in the model

$$(4.1)\quad \begin{aligned} &x \text{ given } T_2 \sim N((I+T_2)^{-1/2}\Theta_{12}, I\otimes\Sigma_{22.3}),\\ &V_{22.3} \sim W_{p_2}(\Sigma_{22.3}, n_3-p_3), \quad T_2=Z_{13}V_{33}^{-1}Z_{13}' \text{ with}\\ &Z_{13} \sim N(0, I_{n_1}\otimes\Sigma_{33}) \text{ and } V_{33} \sim W_{p_3}(\Sigma_{33}, n_3), \text{ and}\\ &(x, T_2) \text{ and } V_{22.3} \text{ are independent}\end{aligned}$$

our problem is to test $\Theta_{12}=0$ versus $\Theta_{12}\neq 0$. Hence the conditional problem given T_2 is exactly the same as the MANOVA problem. And so all the results in the MANOVA problem are effective in this conditional problem. In addition even if we take (T_3, T_4) into account, the joint density of $(x, V_{22.3}, T_2, T_3, T_4)$ is by Lemma 4.1 given by

$$(4.2)\quad f_1(x\,|\,(I+T_2)^{-1/2}\Theta_{12}, \Sigma_{22.3})f_2(V_{22.3}\,|\,\Sigma_{22.3})dx\,dV_{22.3}\,dP^{(T_2,T_3,T_4)}$$

where f_1 and f_2 are respectively the densities of x and $V_{22.3}$ with respect to the Lebesgue measure dx and $dV_{22.3}$ in (4.1) and $P^{(T_2,T_3,T_4)}$ is the joint distribution of (T_2, T_3, T_4). Since $P^{(T_2,T_3,T_4)}$ does not depend on unknown parameters, the LRT (likelihood ratio test) for testing $\Theta_{12}=0$ versus $\Theta_{12}\neq 0$ is given by

$$\frac{\max_{(\Theta_{12},\Sigma_{22.3})} f_1(x\,|\,(I+T_2)^{-1/2}\Theta_{12}, \Sigma_{22.3})f_2(V_{22.3}\,|\,\Sigma_{22.3})}{\max_{\Sigma_{22.3}} f_1(x\,|\,0, \Sigma_{22.3})f_2(V_{22.3}\,|\,\Sigma_{22.3})}$$

This implies that the LRT in the conditional MANOVA problem is equal to the LRT in the GMANOVA problem. Therefore the LRT in the GMANOVA problem is given by

$$(4.3)\qquad |I+T_1| > c$$

(see e.g., Anderson (1958)). In Chapter 7, the LRT and its related

tests are considered.

5. Essentially Complete Class Via Invariance.

5.1 *Reduced problem.* In this section, we take an invariance approach to the proof of Theorem 4.1 (2) : The class $\mathscr{C}(\mathscr{G})$ of $\mathscr{G}$-invariant tests based on (T_1, T_2) only forms an essentailly complete class in the class $\mathscr{D}(\mathscr{G})$ of all $\mathscr{G}$-invariant tests based on $T=(T_1, T_2, T_3, T_4)$.

The approach here is taken in a wider framework by Kariya and Sinha (1983) and is based on the representation theorem on the ratio of distributions of a maximal invariant due to Wijsman (1967) (see Section 3 of Chapter 2). From Proposition 2.1, a maximal invariant under $\mathscr{G}$ depends on Z only through

$$(5.1)\qquad Z_0 \equiv (U_1, U_2, U_3) \equiv ((Z_{12}, Z_{13}), Z_{23}, (Z_{32}, Z_{33}))$$

and a maximal invariant parameter under $\bar{\mathscr{G}}$ depends on (Θ, Σ) only through $\xi\xi' \equiv \Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}' \equiv \Upsilon$ with $\xi=\Theta_{12}\Sigma_{22.3}^{-1/2}$. Hence the power function of a $\mathscr{G}$-invariant test depends on (Θ, Σ) only through ξ (in fact, by Proposition 2.2 it depends on (Θ, Σ) only through the characteristic roots of $\xi\xi'$), and so without loss of generality, we can set

$$(5.2)\qquad \Theta_{11}=0,\ \Theta_{21}=0,\ \Theta_{22}=0,\ \Sigma=I \text{ and } \Theta_{12}=\xi.$$

Then the marginal density of Z_0 in (5.1) becomes

$$(5.3)\qquad \begin{aligned} f(Z_0\,|\,\xi) = c\exp[&-\tfrac{1}{2}\mathrm{tr}(U_1-\xi^*)(U_1-\xi^*)'\\ &-\tfrac{1}{2}\mathrm{tr}\,U_2U_2'-\tfrac{1}{2}\mathrm{tr}\,U_3U_3'],\end{aligned}$$

where $\xi^*=(\xi, 0) : n_1\times(p_2+p_3)$. Under this density, the group $\mathscr{G}$ acting on the left of Z is also reduced to the subgroup $\mathscr{G}_0=\mathscr{O}(n_1)\times\mathscr{O}(n_2)\times\mathscr{O}(n_3)\times\mathscr{A}_0$ acting on the left of Z_0 by

$$(5.4)\qquad gZ_0 = g(U_1, U_2, U_3) = (P_1U_1B',\ P_2Z_{23}A_{33}',\ P_3U_3B')$$

where $g=(P_1, P_2, P_3, B)\in\mathscr{G}_0$ and

$$(5.5)\qquad \mathscr{A}_0 = \{B=\begin{pmatrix} A_{22} & A_{23} \\ 0 & A_{33}\end{pmatrix} \mid A_{ii}\in\mathscr{G}l(p_i),\ i=2,3\}.$$

Therefore, from the viewpoint of invariance, our problem is to test

$$(5.6)\qquad H_0 : \xi=0 \quad\text{versus}\quad H_1 : \xi\neq 0$$

under the density (5.3) of Z_0 in (5.1) with group $\mathscr{G}_0$ acting on the left of Z_0 by (5.4). To be precise, we have to apply a stepwise invariance reduction; first reduce the original problem by the group $\mathscr{A}_1\times\mathscr{F}$ where $\mathscr{F}$ is defined by (2.12) and $\mathscr{A}_1=\{A\in\mathscr{A}\mid A_{11}\in\mathscr{G}l(p_1), A_{23}=0, A_{22}=I, A_{33}=I\}$ with $\mathscr{A}$ in (2.11), and then the second reduction is made by the group $\mathscr{G}_0$. This stepwise reduction is justified by Lemma 1.4 in Chapter 2, as is the case of Proposition 2.3.

5.2. *Probability ratio of a maximal invariant.* To apply to the reduced problem (5.6) Wijsman's theorem on the ratio of distributions of a maximal invariant (see in Section 3 of Chapter 2), let $K=K(Z_0)$ be a maximal invariant under $\mathscr{G}_0$ acting on the left of Z_0 by (5.4) and let P_ξ^K denote the distribution of K under ξ. Further define

$$(5.7)\quad \chi(g) = |BB'|^{(n_1+n_3)/2}|A_{33}A_{33}'|^{n_2/2} = |A_{22}A_{22}'|^{(n_1+n_3)/2}|A_{33}A_{33}'|^{n/2}$$

$$(5.8)\quad \mu_0(dB) = (|A_{22}A_{22}'|^{-(p_2+p_3)/2}dA_{22})(|A_{33}A_{33}'|^{-p_3/2}dA_{33})dA_{23}$$

$$(5.9)\qquad \mu(dB) = \chi(g)\mu_0(dB),\quad M=n_1+n_3-p_3 \quad\text{and}$$

$$(5.10)\qquad H(Z_0\mid\xi) = \int_{\mathscr{G}_0} f(gZ_0\mid\xi)\nu(dP)\mu(dB)$$

where $g=(P_1, P_2, P_3, B)\in\mathscr{G}_0$, $\nu(dP)=\nu_1(dP_1)\nu_2(dP_2)\nu_3(dP_3)$ with $P=(P_1, P_2, P_3)$ and ν_i's are the invariant probability measures over $\mathcal{O}(p_i)$'s $(i=1,2,3)$.

Lemma 5.1. The Radon-Nikodym derivative of P_ξ^K with respect

to P_0^K at $K=K(Z_0)$ is given by

$$(5.11)\qquad R_\xi \equiv (dP_\xi^K/dP_0^K)(K(Z_0)) = H(Z_0\mid\xi)/H(Z_0\mid 0)$$

Proof. Wijsman's theorem requires to show that the space of Z_0 is a Cartan $\mathscr{G}_0$-space. To show this, by Lemma 3.2 in Chapter 2, it suffices to show that the space of U_3, say $\mathscr{U}_3$, in (5.1) is a Cartan $\mathscr{A}_0$-space under transformation $U_3\to U_3B'$. Here we regard the space $\mathscr{U}_3$ as the set of $n_3\times(p_2+p_3)$ matrices of rank p_2+p_3 (because of $n_3\geq p$) by eliminating a set of Lebesgue measure zero. Then $U_3B'=U_3$ with $U_3\in\mathscr{U}_3$ implies $B=I$. That is, no element B in $\mathscr{A}_0$ except $B=I$ leaves U_3 fixed. Therefore by Lemma 3.3 in Chapter 2, $\mathscr{U}_2$ is a Cartan $\mathscr{A}_0$-space. Hence Wijsman's theorem is now applicable to our problem in (5.6) under the density (5.3). Also based on its definition, the space of Z_0 can be shown to be a Cartan $\mathscr{G}_0$-space. Next, we can identify $\chi(g)$ in (5.7) with the inverse of the Jacobian of the transformation (5.4) and $\mu_0(dB)$ in (5.8) with a left invariant measure on $\mathscr{A}_0$ in (5.5). Therefore, Wijsman's theorem states that the probability ratio $(dP_\xi^K/dP_0^K)(K(Z_0))$ is given by (5.11). This completes the proof.

It is noted that the left invariant measure (5.8) on $\mathscr{A}_0$ is not equal to a right invariant measure on $\mathscr{A}_0$. By this reason, the left action of $\mathscr{G}_0$ on the space of Z_0 in (5.4) must be distinguished from the right action: $\tilde{g}(Z_0)=(P_1'U_1B, P_2'U_2A_{33}, P_3'U_3B)$.

We shall evaluate (5.11). Let $C=(C_{ij})\in\mathscr{A}_0$ be a matrix such that

$$(5.12)\qquad C(U_1'U_1+U_3'U_3)C' = I_{p_2+p_3},$$

and define

$$(5.13)\qquad (W_2, W_3) \equiv U_1C' = (Z_{12}C_{22}'+Z_{13}C_{23}', Z_{13}C_{33}') \quad\text{and}$$

$$(5.14)\qquad \Delta \equiv \operatorname{tr}\xi\xi' = \operatorname{tr}\Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}'.$$

Lemma 5.2. The ratio R_ξ in (5.11) is evaluated as

$$(5.15)\quad R_\xi = \int_{\mathcal{O}(n_1)\times\mathcal{A}_0} \exp[-\tfrac{1}{2}\mathrm{tr}BB' + \mathrm{tr}\xi' P_1(W_2A'_{22} + W_3A'_{23}) - \tfrac{1}{2}\Delta]\mu(dB)\nu(dP_1) \Big/ \int \exp[-\tfrac{1}{2}\mathrm{tr}BB']\mu(dB)$$

Proof. Write (5.10) with (5.3) as

$$(5.16)\quad H(Z_0|\xi) = c_0 \int_{\mathcal{G}_0} \exp[-\tfrac{1}{2}\mathrm{tr}(P_1U_1B' - \xi^*)'(P_1U_1B' - \xi^*) - \tfrac{1}{2}\mathrm{tr}A_{33}Z'_{23}Z_{23}A'_{33} - \tfrac{1}{2}\mathrm{tr}BU'_3U_3B']\mu(dB)\nu(dP)$$

Here transforming B into BC with C satisfying (5.12), the inside of [ ] in (5.16) becomes

$$-\tfrac{1}{2}[\mathrm{tr}BB' - 2\mathrm{tr}\xi' P_1(W_2A'_{22} + W_3A'_{23}) + \mathrm{tr}A_{33}C_{33}Z'_{23}Z_{23}C'_{33}A'_{33} + \Delta]$$

Further, transforming A_{33} into $A_{33}(I + C_{33}Z'_{23}Z_{23}C'_{33})^{-1/2}$ yields the inside of [ ] in the integrand of the numerator in (5.15), while under $\xi=0$, it gives the inside of [ ] in the denominator. Some multiplicative constants coming out by these transformations are cancelled out between the numerator and the denominator. In addition, using $\nu_i(\mathcal{O}(n_i))=1$ for $i=2,3$ yields the result.

5.3. *Essentially complete class.* Since $R_\xi = (dP^K_\xi/dP^K_0)(K(Z_0))$ is the density of $K=K(Z_0)$ with respect to P^K_0, since C in (5.12) does not depend on Z_{23} so that R_ξ in (5.15) does not depend on Z_{23} and since P^K_0 is independent of the parameter ξ by (5.3), $(U_1, U_3) = ((Z_{12}, Z_{13}), (Z_{32}, Z_{33}))$ is sufficient for the family $\{dP^K_\xi = R_\xi dP^K_0 | \xi \in R^{n_1 p_2}\}$. Though we used the parametrization $\xi = \Theta_{12}\Sigma^{-1/2}_{22.3}$, P^K_ξ in fact depends on ξ only through the characteristic roots of $\xi\xi'$. We summarize this fact as

Theorem 5.1. Let $(U_1, U_3) = ((Z_{12}, Z_{13}),\ (Z_{32}, Z_{33}))$. Then (1) a maximal invariant (T_1, T_2) under $\mathcal{K}_1 \equiv \mathcal{A}_0$ acting on the left of (U_1, U_3) by $(U_1, U_3) \to (U_1B', U_3B')$ for $B \in \mathcal{K}_1 \equiv \mathcal{A}_0$ is sufficient for the distributions of (T_1, T_2, T_3, T_4).

(2) A maximal invariant under $\mathcal{G}_1 \equiv \mathcal{O}(n_1)\times\mathcal{O}(n_3)\times\mathcal{A}_0$ acting on the left of (U_1, U_3) by $(U_1, U_3) \to (P_1U_1B', P_3U_3B')$ for $(P_1, P_3, B) \in \mathcal{G}_1$ is sufficient for the distributions of a maximal invariant under $\mathcal{G}$ acting on the left of Z as in (2.15).

Proof. (2) is clear by the above argument, while the above proof for R_ξ is effective for the group $\mathcal{K}_1 = \mathcal{A}_0$ without the subgroup $\mathcal{O}(n_1)\times\mathcal{O}(n_2)\times\mathcal{O}(n_3)$.

This theorem implies that the class $\mathcal{C}(\mathcal{G})$ of $\mathcal{G}$-invariant test based on (T_1, T_2) forms an essentially complete class in the class $\mathcal{D}(\mathcal{G})$ of $\mathcal{G}$-invariant tests based on (T_1, T_2, T_3, T_4). Because (T_1, T_2) is a maximal invariant under the subgroup $\mathcal{K}_1 = \mathcal{A}_0$ of $\mathcal{G}_1$ as in Theorem 5.1 (1).

Based on this theorem, our original problem may be restated from the viewpoint of invariance as follows: with (U_1, U_3) test $H_0 : \xi = 0$ versus $H_1 : \xi \neq 0$ under the density

$$(5.17)\quad f(U_1, U_3|\xi) = c\exp[-\tfrac{1}{2}\mathrm{tr}(U_1 - \xi^*)'(U_1 - \xi^*) - \tfrac{1}{2}\mathrm{tr}U'_3U_3]$$

where the group $\mathcal{G}_1 = \mathcal{O}(n_1)\times\mathcal{O}(n_3)\times\mathcal{A}_0$. acts on the left of (U_1, U_3) by $(U_1, U_3) \to (P_1U_1B', P_3U_3B')$ for $(P_1, P_3, B) \in \mathcal{G}_1$. Or the problem may be also stated as follows; test $H_0 : \delta_1 = 0$ versus $H_1 : \delta_1 > 0$ under the density of a maximal invariant under $\mathcal{G}_1$, where δ_i is the i-th largest characteristic root of $\xi\xi' = \Theta_{12}\Sigma^{-1}_{22.3}\Theta'_{12}$. Let $J = J(U_1, U_3)$ be a maximal invariant under $\mathcal{G}_1$ and let P^J_ξ be the distribution of J under ξ. Then in the same way as above, we obtain

Lemma 5.3. $(dP^J_\xi/dP^J_0)(J(U_1, U_3)) = R_\xi$ where R_ξ is given by (5.15).

In Marden (1983), the GMANOVA model is regarded as $U_1 \sim N(\xi^*, I_{n_1}\otimes\Sigma^*)$ and $U_3 \sim N(0, I_{n_3}\otimes\Sigma^*)$ from an invariance point of view, where $\Sigma^* = (\Sigma_{ij})$ $(i, j = 2, 3)$.

6. Distributions of Maximal Invariants.

6.1 *Distributions of maximal invariants* As has been stated, a usual invariance approach to a testing problem is to first find a group leaving the problem invariant, then find a maximal invariant, third derive the distribution of the maximal invariant, fourth derive an optimal test based on the distribution and finally consider the null distribution of the test statistic. In addition, the nonnull distribution is often considered to investigate the behavior of the power function. In our case, a maximal invariant under $\mathscr{G}$ is characterized through a maximal invariant under $\mathscr{H}$. However, without this characterization, it could be handled through an application of Wijsman's theorem in the same way as in Section 5, though we have not taken this approach. In Section 5, we observed that a maximal invariant under the group $\mathscr{G}_1=\mathcal{O}(n_1)\times\mathcal{O}(n_3)\times\mathscr{A}_0$ with $\mathscr{A}_0$ in (5.5) which acts on the left of $(U_1, U_3)\equiv((\boldsymbol{Z}_{12}, \boldsymbol{Z}_{13}), (\boldsymbol{Z}_{32}, \boldsymbol{Z}_{33}))$ by $(U_1, U_3)\rightarrow(P_1U_1B', P_3U_3B')$ for $g=(P_1, P_3, B)\in\mathscr{G}_1$ is sufficient for the distributions of a maximal invariant under $\mathscr{G}$ acting on the left of $\boldsymbol{Z}$ in (2.15). In this section, we consider some distributional aspects of a maximal invariant under $\mathscr{G}_1$ acting on (U_1, U_3).

Let $J=J(U_1, U_3)$ be a maximal invariant under $\mathscr{G}_1$ and let P_ξ^J be the distribution of J under $\xi=\Theta_{12}\Sigma_{22.3}^{-1/2}$. Then by Lemma 5.3, $(dP_\xi^J/dP_0^J)(J(U_1, U_3))=R_\xi$ with R_ξ in (5.15). Therefore, once the null distribution dP_0^J of J is obtained, the nonnull distribution dP_ξ^J of J is obtained as $dP_\xi^J=R_\xi dP_0^J$. In addition, in deriving an optimal test based on dP_ξ^J, it is not necessary to derive the null distribution dP_0^J, because R_ξ is the density of J with respect to dP_0^J being free from ξ. From this viewpoint, what is required is to evaluate R_ξ in (5.15). Though we do not do this completely, we here slightly modify it for a future use. Let $D=\int_{\mathscr{A}_0}\exp[-\frac{1}{2}\mathrm{tr}BB']\mu(dB)$ and let

$$\text{(6.1)}\qquad \begin{cases} Q(B) = W_2A_{22}'+W_3A_{23}', \\ h(B) = \exp[-\frac{1}{2}\mathrm{tr}BB']/D \end{cases}$$

where $(\boldsymbol{W}_2, \boldsymbol{W}_3)$ is defined in (5.13). Then R_ξ in (5.15) is evaluated as

$$\text{(6.2)}\qquad R_\xi = \exp(-\frac{1}{2}\Delta)\int_{\mathscr{A}_0}\int_{\mathcal{O}(n_1)}\exp(\mathrm{tr}\xi'P_1Q(B))\nu_1(dP_1)\,h(B)\mu(dB)$$

Here usually $\exp(\mathrm{tr}\xi'P_1Q(B))$ is expanded as $\Sigma[\mathrm{tr}\xi'P_1Q(B)]^j/j!$ and the integration of each term over $\mathcal{O}(n_1)\times\mathscr{A}_0$ is evaluated formally by using Zonal polynomials or matrix-variate hypergeometric functions. As a special case, when $n_1=1$,

$$\begin{aligned}\int_{\mathcal{O}(n_1)}&\exp(\mathrm{tr}\xi'P_1Q(B))\nu_1(dP_1)\\ &= [\exp(Q(B)\xi') + \exp(-Q(B)\xi')]/2\\ &= \textstyle\sum_{j=0}^{\infty}[Q(B)\xi']^{2j}/(2j)!.\end{aligned}$$

Integrating this over $\mathscr{A}_0$ term by term, we can identify R_ξ with the ratio $f(t_1|\lambda)/f_1(t_1|0)$, where $f_1(t_1|\lambda)$ is the density of nonnormalized noncentral F-distribution with d.f. p_2 and $n_3-p_2-p_3+1$ and noncentrality $\lambda=\gamma/2(1+t_2)$, and $\gamma=\xi\xi'$: 1×1. However, this fact is easily obtained by using the distributional results in Lemma 4.1. In fact, $(n_3-p_2-p_3+1)T_1/p_2$ given $T_2=t_2$ is distributed as noncentral F distribution with d.f. p_2 and $n_3-p_2-p_3+1$ and noncentrality parameter λ as is in the case of Hotelling T^2-statistic. Note that $T_1=\boldsymbol{x}V_{22.3}^{-1}\boldsymbol{x}'$ where $\boldsymbol{x}\sim N((1+t_2)^{-1/2}\Theta_{12}, \Sigma_{22.3})$ and $V_{22.3}\sim W_{p_2}(\Sigma_{22.3}, n_3-p_3)$.

When $n_1>p_2$, T_1 cannot have a density with respect to the Lebesgue measure on the coordinate space $\mathscr{T}_1$ of T_1. Even in this case, R_ξ in (6.1) is the density of a maximal invariant $\boldsymbol{J}=\boldsymbol{J}(U_1, U_3)$ under $\mathscr{G}_1$ with respect to P_0^J.

6.2. *Case* $n_1 \leq p_2$. A book by Johnson and Kotz (1972) serves as a general reference here.

Definition 6.1. A central multivariate Beta distribution $B_m\left(\frac{a}{2}, \frac{b}{2}\right)$ is defined to be the distribution with density

$$(6.3) \quad f(t) = c_0 |t|^{(a-m-1)/2} |I-t|^{(b-m-1)}, \quad t \in \mathcal{S}_+(m),\ I-t \in \mathcal{S}_+(m)$$

where $\min(a,b) \geq m$, a, b integers and

$$(6.4) \quad c_0 = \prod_{i=1}^{m} \Gamma\left(\frac{a+b-i+1}{2}\right) \Big/ \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma\left(\frac{a-i+1}{2}\right) \Gamma\left(\frac{b-i+1}{2}\right).$$

Definition 6.2. A multivariate F-distribution $F_m(2a, 2b)$ is defined to be the distribution with density

$$(6.5) \qquad f(t) = c_0 |t|^{a-\frac{1}{2}(m+1)} |I+t|^{-(a+b)}, \quad t \in \mathcal{S}_+(m)$$

where $2a$ and $2b$ are integers and $a, b > (m-1)/2$.

Lemma 6.1. (1) When $T \sim F_m(2a, 2b)$, $(I+T^{-1})^{-1} \sim B_m(a, b)$ and $(I+T)^{-1} \sim Be_m(b, a)$.
(2) When $U: m \times p \sim N(0, I_n \otimes \Sigma)$ and $S \sim W_p(\Sigma, n)$ with $n \geq p \geq m$ are independent, then $T = US^{-1}U' \sim F_m(p, m+n-p)$.

Theorem 6.1. (1) When $n_1 \leq p_2$, p_3 and $n_3 \geq p$, under the null hypothesis $\Theta_{12} = 0$, T_1 and T_2 are independently distributed as

$$T_1 \sim F_{n_1}(p_2, n_1+n_3-p_2-p_3) \quad \text{and} \quad T_2 \sim F_{n_1}(p_3, n_1+n_3-p_3).$$

(2) When $n_1 \leq p_2$, the nonnull conditional density of T_1 given $T_2 = t_2$ is given by

$$(6.6) \quad f_1(t_1 | t_2 : \Upsilon) = c |t_1|^{(p_2-n_1-1)/2} |I+t_1|^{-(n_1+n_3-p_3)/2}$$
$$\times \exp(-\tfrac{1}{2}\mathrm{tr}\Delta)\, {}_1F_1\left(\frac{n_1+n_3-p_3}{2} ; \frac{p_2}{2} : \frac{1}{2}(I+t_1^{-1})^{-1}\right)$$

where ${}_1F_1\ (a ; b : A)$ is a matrix-variate hypergeometric function

(see, e.g. James (1964)),

$$(6.7) \qquad \Delta = (I+t_2)^{-1/2} \Upsilon (I+t_2)^{-1/2}, \quad \text{and} \quad \Upsilon = \Theta_{12} \Sigma_{22.3}^{-1} \Theta_{12}'.$$

This distribution with density (6.6) is called noncentral multivariate F-distribution. Hence when $n_1 \leq \min(p_2, p_3)$, the joint density of (T_1, T_2) is given by

$$f(t_1, t_2 | \Upsilon) = f_1(t_1 | t_2 : \Upsilon) f_2(t_2)$$

where $f_2(t)$ is the density of $F_{n_1}(p_3 ; n_1+n_3-p_3)$.

These results are well known and found, e.g., in James (1964), Olkin and Rubin (1964) and Johnson and Kotz (1972).

It is noted that when $n_1 \leq \min(p_2, p_3)$, the density of a maximal invariant under the transformation $(T_1, T_2) \rightarrow (P_1 T_1 P_1', P_1 T_2 P_1')$ ($P_1 \in \mathcal{O}(n_1)$) is given by

$$(6.8) \qquad \int_{\mathcal{O}(n_1)} f(P_1 t_1 P_1', P_1 t_2 P_1' | \Upsilon) \nu_1(dP_1)$$

where ν_1 is the invariant probability measure (see e.g., James (1964), and see also Chapter 2). Of course, (6.8) is equivalent to $R_t\, dP_0^J$ with R_t in (6.2) provided $n_1 \leq \min(p_2, p_3)$.

Lemma 6.2. (1) (Mitra (1970)) When $U \sim Be_m\left(\frac{a}{2}, \frac{b}{2}\right)$, $\alpha' U \alpha / \alpha' \alpha \sim Be_1\left(\frac{a}{2}, \frac{b}{2}\right)$, where $\alpha \in R^m$ and $\alpha \neq 0$.
(2) When $U \sim Be_m\left(\frac{a}{2}, \frac{b}{2}\right)$, $E(U) = (a/(a+b))I$.

Proof. (1) See Mitra (1970). (2) Since the distribution of U is the same as that of PUP' with $P \in \mathcal{O}(m)$, we obtain $E(U) = \alpha I$ for some α. Taking $\alpha = (1, 0, \ldots, 0)'$ in (1) yields $\alpha = a/(a+b)$.

7. The Likelihood Ratio Test (LRT) And the Related Tests.

7.1. *LRT*. In our problem Khatri (1966) derived the LRT through a conditional augument in the original model. On the other hand, by using invariance, Gleser and Olkin (1970) derived it in the canonical model in Section 2. We also gave a derivation in 4.2. Here we study the LRT with the related tests proposed by Khatri (1966) for some special cases.

It is recalled that a reduced version of our model is given by

$$(7.1)\qquad \begin{cases} x \text{ given } T_2 \sim N((I+T_2)^{-1/2}\Theta_{12},\ I_{n_1}\otimes\Sigma_{22.3}), \\ V_{22.3} \sim W_{p_2}(\Sigma_{22.3},\ n_3-p_3) \\ T_2 = Z_{13}V_{33}^{-1}Z_{13}' \quad \text{with} \\ Z_{13} \sim N(O, I_{n_1}\otimes\Sigma_{33}) \text{ and } V_{33} \sim W_{p_3}(\Sigma_{33}, n_3) \end{cases}$$

where (x, T_2) and $V_{22.3}$ are independent. In this model, the problem is to test $H_0 : \Theta_{12}=0$. Then as has been seen in 4.2 or from Gleser and Olkin (1970), the LRT is given by

$$(7.2)\qquad \mathscr{K}_4 : \frac{|V_{22.3}|}{|V_{22.3}+x'x|} = \frac{1}{|I+xV_{22.3}^{-1}x'|} = \frac{1}{|I+T_1|} < k_4.$$

Hence in our terms, the LRT is a $\mathscr{G}$-invariant test based on T_1 alone since the cut-off point k_4 is independent of T_2 (see below). From this fact we may pose a question whether the class of $\mathscr{G}$-invariant tests based on T_1 alone forms an essentially complete class. The answer is negative as will be shown. Then one may ask whether the LRT is admissible in $\mathscr{C}(\mathscr{G})$ the class of $\mathscr{G}$-invariant level α tests based on (T_1, T_2). The question on the admissibility of the LRT will be negatively answered in Section 11. However the similarity of our problem to the usual MANOVA problem makes us conclude, without proof, that the LRT is conditionally admissible given $T_2=t_2$ in the class of conditional level α tests (see Schwartz (1967)). In fact, given $T_2=t_2$, our problem is exactly the same as the MANOVA problem and the LRT in (7.2) is exactly the same as the LRT in the conditional MANOVA problem (see 4.2). Based

on the analogy, Khatri (1966) proposed the following tests as in the MANOVA problem:

Roy's maximal root test: $\mathscr{K}_1$

$$(7.3)\qquad \mathscr{K}_1 : \mathrm{ch}_1(T_1) = \mathrm{ch}_1(xV_{22.3}^{-1}x') > k_1,$$

Lawley-Hotelling's trace test: $\mathscr{K}_2$

$$(7.4)\qquad \mathscr{K}_2 : \mathrm{tr}\,T_1 = \mathrm{tr}\,xV_{22.3}^{-1}x' > k_2,$$

Pillai's trace test: $\mathscr{K}_3$

$$(7.5)\qquad \mathscr{K}_3 : \mathrm{tr}\,T_1(I+T_1)^{-1} = \mathrm{tr}\,x(x'x+V_{22.3})^{-1}x' > k_3.$$

From Theorem 4.1 (3), under $H: \Theta_{12}=0$, T_1 and T_2 are independent and so the critical points k_i's $(i=1,\dots,4)$ can be determined independently of T_2. Therefore these are unconditional $\mathscr{G}$-invariant tests based on T_1 only. To investigate some properties of these tests, we define

$$(7.6)\qquad \mathscr{C}_0(\mathscr{G}) = \{\phi\in\mathscr{C}(\mathscr{G})\,|\,\phi \text{ is a test based on } T_1 \text{ alone}\}$$

and

$$(7.7)\qquad \mathscr{C}_1(\mathscr{G}) = \{\phi\in\mathscr{C}(\mathscr{G})\,|\,E_0[\phi(T_1,T_2)\,|\,T_2]\le\alpha \text{ a.e.}(T_2)\},$$

where $E_0[\,\cdot\,|T_2]$ denotes the conditional expectation of $\cdot$ given T_2. Clearly

$$\mathscr{C}_0(\mathscr{G})\subset\mathscr{C}_1(\mathscr{G})\subset\mathscr{C}(\mathscr{G}).$$

The null distributions of the tests $\mathscr{K}_i$'s do not depend on T_2 and they are the same as those of the corresponding tests in the MANOVA problem. Hence the tables for percentage points or the asymptotic null distibutions in the MANOVA problem can be utilized for the determination of k_i's or the significance probabilities.

7.2. *UMP property.* When $n_1=1$, the LRT is shown to be UMP in $\mathscr{C}_1(\mathscr{G})$. When $n_1=1$, all the four tests $\mathscr{K}_i$'s are identical. Let

$$(7.8) \quad W_1 = (n_3-p_2-p_3+1)T_1/p_2 \quad \text{and} \quad W_2 = (n_3-p_3+1)T_2/p_3,$$

and let $h(w_1,w_2|\gamma)$ be the density of (W_1,W_2) where $\gamma=\Theta_{12}\Sigma_{22\cdot3}^{-1}\Theta_{12}'$: 1×1. Then by Theorem 6.1, h is expressed as

$$(7.9) \quad h(w_1,w_2|\gamma) = h_1(w_1|\lambda)h_2(w_2) \quad \text{with} \quad \lambda=\gamma/2(1+t_2),$$

where the conditional density $h_1(w_1|\lambda)$ of W_1 given W_2 is the density of noncentral F-distribution with d.f. (degrees of freedom) p_2 and $n_3-p_2-p_3+1$ and noncentrality λ, and $h_2(w_2)$ is the density of F-distribution with d.f. p_3 and n_3-p_3+1. Since $h_1(w_1|\lambda)$ has a monotone likelihood ratio (MLR) property in λ, hence in γ for given t_2 or w_2, for testing $\gamma=0$ vs $\gamma>0$, a test ϕ with the critical region

$$(7.10) \qquad w_1 > c(t_2) \quad \text{or} \quad t_1 > c(t_2)$$

is UMP(uniformly most powerful) (Lehmann (1959) Chapter 3). But from the independence of T_1 and T_2 under $H_0: \Theta_{12}=0$, the cut-off point $c(t_2)$ in (7.10) can be determined freely from t_2, and test ϕ with critical region (7.10) with $c(t_2)\equiv c$ is the LRT. Further, since the conditional power function $\pi(\phi,\lambda|t_2)=E_\gamma[\phi|T_2=t_2]$ is strictly increasing in $\lambda=\gamma/2(1+t_2)$ for each t_2 by the MLR property and since the distribution of W_2 or T_2 gives a positive measure to any nonempty open set, the power function $\pi(\phi,\gamma)$ is strictly increasing in γ. Hence we obtain a generalized version of Hotelling's T^2-problem.

Theorem 7.1. When $n_1=1$, the LRT $\mathscr{K}_4 (=\mathscr{K}_1=\mathscr{K}_2=\mathscr{K}_3)$ is UMPI (UMP invariant) in $\mathscr{C}_1(\mathscr{G})$ and the power function is strictly increasing in γ. In particular, $\mathscr{K}_4$ is UMPI in $\mathscr{C}_0(\mathscr{G})$.

This result is also proved in Giri (1962).

It should be noted that, since the proof is based on a conditional argument, the class for which the UMPI property holds is not $\mathscr{C}(\mathscr{G})$ but $\mathscr{C}_1(\mathscr{G})$ in (5.6). In the class $\mathscr{C}(\mathscr{G})$, no UMPI test exists even

when $n_1=1$. In fact, by Neyman-Pearson Lemma, testing $\gamma=0$ versus $\gamma=\gamma_0$, a MP test in $\mathscr{C}(\mathscr{G})$ is given by

$$h(w_1,w_2|\gamma_0)/h(w_1,w_2|0) > c.$$

This unconditional critical region is expressed as

$$(7.11) \qquad \Sigma_{j=0}^{\infty} b_j\left(\frac{\gamma_0}{1+t_2}\right)^j\left(\frac{t_1}{1+t_1}\right)^j \exp\left(-\frac{\gamma_0}{2(1+t_2)}\right) > c$$

for certain $b_j>0$ and the critical region in (7.11) depends on γ_0. Hence this MP test is different from the LRT and it dominates in power at $\gamma=\gamma_0$ the LRT that is UMPI in $\mathscr{C}_1(\mathscr{G})$. Therefore the LRT is not UMPI in $\mathscr{C}(\mathscr{G})$ and no UMPI test exists in $\mathscr{C}(\mathscr{G})$.

7.3. *Distribution of the LRT.* As has been remarked, our conditional problem is nothing but the MANOVA problem. Hence using a distributional result in the MANOVA problem, we obtain without any condition on n_1

Proposition 7.1. Under the null hypothesis $\Theta_{12}=0$, the LRT statistic l in (7.2) has the same distribution as $\hat{l}=\prod_{i=1}^{n_1} L_i$, where $L_i\sim Be((n_3-p_2-p_3+i)/2,\ p_2/2)$ and L_i's are independent. Further, under the alternative hypothesis $\Theta_{12}\neq0$ with rank $(\Theta_{12})=1$, the LRT statistic l has the same distribution as $\hat{l}=\prod_{i=1}^{n_1} L_i$, where L_i's are independent, $L_i\sim Be((n_3-p_2-p_3+i)/2,\ p_2/2)$ $(i=1,\dots,n_1-1)$, L_{n_1} given $T_2=t_2\sim Be((n_3-p_2-p_3+n_1)/2,\ p_2/2 : \lambda(t_2,\Upsilon))$, noncentral beta with noncentral parameter $\lambda(t_2,\Upsilon)=\mathrm{tr}(I+t_2)^{-1/2}\Upsilon(I+t_2)^{-1/2}$, and $\lambda(T_2,\Upsilon)\sim \mathrm{tr}\Upsilon Be((n_1+n_3-p_3)/2,\ p_3/2)$ with $\Upsilon=\Theta_{12}\Sigma_{22\cdot3}^{-1}\Theta_{12}'$.

Proof. For the null case and the conditional part in the nonnull case, we can refer to Anderson (1958) or Eaton (1983). To show the last part, write $\lambda(T_2,\Upsilon)=\beta'(I+T_2)^{-1}\beta$ with $\Upsilon=\beta\beta'$ and $\beta: n_1\times1$ since $\mathrm{rank}(\Theta_{12})=1$. Then from Lemma 6.1 and Theorem 6.1, $(I+T_2)^{-1}\sim Be_{n_1}((n_1+n_3-p_3)/2,\ p_3/2)$. Hence applying Lemma 6.2

yields the result.

Proposition 7.2. Assume rank $(\Theta_{12})=1$. Then the distribution of the LRT statistic l has a strict monotone likelihood ratio (MLR) property in$-\mathrm{tr}\Upsilon$ and so the power function is strictly increasing in $\mathrm{tr}\Upsilon$.

Proof. Given t_2, the distribution of l has a strict MLR property in$-\lambda(t_2,\Upsilon)$ (see Das Gupta and Perlman (1972)). Let $h(l|\lambda)$ be the density of l given $\lambda=\lambda(t_2,\Upsilon)$ and let $g(\lambda|\mathrm{tr}\Upsilon)$ be the density of λ. Then the density of l is given by

$$f(l|\mathrm{tr}\Upsilon) = \int_0^{\mathrm{tr}\Upsilon} h(l|\lambda)g(\lambda|\mathrm{tr}\Upsilon)d\lambda.$$

Since $g(\lambda|\mathrm{tr}\Upsilon)$ is strictly totally positive of order 2 as a function of λ and $\mathrm{tr}\Upsilon$ (see Lehmann (1959)), $f(l|\mathrm{tr}\Upsilon)$ has a strict MLR property in$-\mathrm{tr}\Upsilon$. The second part follows from the first part (see Lehmann (1959)), completing the proof.

In the original term, T_1 is given by (2.29). The asymptotic null distributions of the tests $\mathscr{K}_1\sim\mathscr{K}_4$ are given in Section 10 and some other properties of these tests are considered in Sections 11 and 12 where the inadmissibility of the tests is shown based on Marden (1983).

7.4. *Some other tests.* When the admissibility of tests in the GMANOVA problem is questioned, Marden (1983) included the following invariant tests in his consideration;

i) $\mathrm{ch}_1(T_0) > c$, ii) $\mathrm{tr}T_0 > c$

iii) $\mathrm{tr}T_0(I+T_0)^{-1} > c$ and iv) $|I+T_0| > c$

where

$$T_0 = (Z_{12},Z_{13})\begin{pmatrix} V_{22} & V_{23} \\ V_{32} & V_{33}\end{pmatrix}^{-1}(Z_{12},Z_{13})'$$

These tests are appropriate for testing $(\Theta_{12},\Theta_{13})=0$ versus $(\Theta_{12},\Theta_{13})\neq 0$ without assuming $\Theta_{13}=0$ and they are $\mathscr{G}$-invariant. He also considered some other tests which are not $\mathscr{G}$-invariant. The generalized Bayes test based on

$$|I+T_2|^{p_2}|I+T_1|^{n_1+n_3-p_3} > c$$

is also considered and expected to be close in power to the LRT.

7.5. *The case* $X_2=I$. It is often the case that a model we consider in applications is not a GMANOVA model but a MANOVA model;

$$Y = X_1B+E,\quad E\sim N(O,\ I\otimes\Omega) \tag{7.12}$$

while a hypothesis is formulated as a GMANOVA hypothesis:

$$X_3BX_4 = X_0, \tag{7.13}$$

where the notation is the same as before. Then in terms of the canonical form, the model is expressed as

$$Z = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \\ Z_{31} & Z_{32}\end{pmatrix} \sim N(\begin{pmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \\ 0 & 0\end{pmatrix},\quad I_n\otimes\Sigma), \tag{7.14}$$

and the problem is to test $H:\Theta_{12}=0$. Though this is not the so-called MANOVA problem, it is essentially the MANOVA problem from an invariance point of view. In fact, via invariance it is reduced to the problem of testing $\Theta_{12}=0$ in the model $(Z'_{12},Z'_{22},Z'_{32})'\sim N((\Theta'_{12},\Theta'_{22},0')',\ I_n\otimes\Sigma_{22})$ which is a MANOVA problem. Therefore, the above four tests are most commonly used and the Pillai test is LBI (see Schwartz (1967)). Further, we obtain

Proposition 7.3. For testing (7.13) in the MANOVA model (7.12), when $\min(n_1, p_2)=1$, all the four tests are equivalent and they are UMPI in $\mathscr{C}(\mathscr{G})$.

In the GMANOVA problem with $n_1>1$, and $\mathrm{rank}(X_2)=p_1+p_2<p$, even when $p_2=1$, none of the four tests are even *conditionally* UMPI though they are again equivalent. This is a different aspect.

In the original term, when $X_2=I$, T_1 is expressed as (2.30) :

$$T_1 = [X_3(X_1'X_1)^{-1}X_3']^{-1/2}YX_4[X_4'Y'N_1YX_4]^{-1} \times X_4'Y'[X_3(X_1'X_1)^{-1}X_3']^{-1/2} \qquad (7.15)$$

where $N_1 = I-X_1(X_1'X_1)^{-1}X_1'$. It is noted that $T_2=0$ if $X_2=I$.

7.6. *Examples.* The case $n_1=1$ corresponds to the case that the order of the matrix X_3 in the restriction $X_3BX_4=X_0$ is $1\times(1+n_2)$. In the following examples, the LRT is UMPI in the class $\mathcal{C}_1(\mathcal{G})$ because $n_1=1$. However, it is not UMPI in $\mathcal{C}(\mathcal{G})$ unless $X_2=I$, and there is a locally best invariant test in the class, which is derived in the next section. But we may use the LRT for its distributional convenience and the optimal property we have seen.

Example 7.1. (Potthoff and Roy (1964)). Suppose we have n_1+1 groups of animals with m_j animals in the j-th group, and with each group being subjected to a different treatment. Animals in all groups are measured at the same q points in time and the q observations on a given animal are assumed to be multivariate normal with unknown covariance matrix Ω. The growth curve associated with the j-th group at time t is assumed to be

$$\xi_{j,0}+\xi_{j,s}t+\cdots+\xi_{j,s}t^s \ (j=1,\cdots,n_1+1).$$

Then the model can be expressed as $Y\sim N(E(Y),\ I\otimes\Omega)$, where

$$E(Y) = E\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n_1+1}\end{bmatrix} = \begin{bmatrix} e_{m_1} & & & 0 \\ & e_{m_2} & & \\ & & \ddots & \\ 0 & & & e_{m_{n_1+1}}\end{bmatrix}\begin{bmatrix} \xi^1 \\ \xi^2 \\ \vdots \\ \xi^{n_1+1}\end{bmatrix}\begin{bmatrix} 1 & \cdots\cdots & 1 \\ t_1 & \cdots\cdots & t_q \\ t_1^2 & \cdots\cdots & t_q^2 \\ & \vdots & \\ t_1^s & \cdots\cdots & t_q^s\end{bmatrix}.$$

$Y=(Y_{lk}^{(j)})$ $(l=1,\cdots,m_j : k=1,\cdots,q)$, $e_{m_j}=(1,\cdots,1) : m_j\times 1$, and $\xi^j=(\xi_{j,0},\xi_{j,1},\cdots,\xi_{j,s}) : 1\times(s+1)$ $(j=1,\cdots,n_1+1)$. Here suppose we wish to test the hypothesis that there is no difference among the (n_1+1) groups. Then the hypothesis can be written as $X_3BX_4=X_0$, where

$$X_3=[I_{n_1}, e_{n_1}] : n_1\times(n_1+1),\quad X_4=I_{s+1},\quad X_0=0 \quad\text{and}\quad B=\begin{bmatrix}\xi^1 \\ \vdots \\ \xi^{n_1+1}\end{bmatrix}.$$

Therefore in the case $n_1=1$, we have two groups of animals and test the hypothesis that the two groups are equal.

Example 7.2. Let $y(t)=(y_1(t),\cdots,y_p(t))'$ be a $p\times 1$ vector consisting of p economic variables, where $y(t)$ is observed at time $t=1,\cdots,n$ and $y(t)$'s obey $N(\mu(t),\Sigma)$ independently. Here $\mu_i(t)=\beta_{i0}+\beta_{i1}t$ $(i=1,\cdots,p)$. Then as has been seen in Example 1.1, the model is expressed as $Y=X_1B+E$ with $E\sim N(I_n\otimes\Sigma)$, where

$$Y=\begin{pmatrix} y(1)' \\ \vdots \\ y(n)'\end{pmatrix} : n\times p,\quad X_1=\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n\end{pmatrix} : n\times 2 \quad\text{and}\quad B=\begin{pmatrix}\beta_{10} & \beta_{20}\cdots\beta_{p0} \\ \beta_{11} & \beta_{21}\cdots\beta_{p1}\end{pmatrix} : 2\times p$$

And the hypothesis that the growth rates β_{i1}'s in $\mu_i(t)=\beta_{i0}+\beta_{i1}t$ are all equal or $\beta_{11}=\beta_{21}=\cdots=\beta_{p1}$ expressed as

$$X_3BX_4=0 \text{ with } X_3=[0\ 1] \text{ and } X_4=\begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ -1 & 1 & \cdot & \cdot & \cdot \\ 0 & -1 & \cdot & \cdot & 0 \\ \cdot & 0 & \cdot & \cdot & 1 \\ 0 & \cdot & 0 & \cdot & -1\end{pmatrix} : p\times(p-1).$$

Hence when the means $\mu_i(t)$ are linear in t, the problem corresponds to the case $n_1=1$. In this problem, because $X_2=I$, the LRT is UMPI in the whole class $\mathcal{C}(\mathcal{G})$.

Exmaple 7.3. From Example 1.3 in Chapter 1, the problem of testing $\mu_1=\mu_2$ in $N_{2p}(\mu, \Sigma)$ based on a random sample where $\mu=(\mu_1', \mu_2')' : 2p\times 1$ is regarded as a GMANOVA problem with $n_1=1$ and $X_2=I$. Hence the the LRT is UMPI in $\mathcal{C}(\mathcal{G})$.

Example 7.4. From Example 1.4 in Chapter 1 the problem of testing $\mu_2=\eta_2$ based on random samples from $N(\mu, \Sigma)$ and $N(\eta, \Sigma)$, where $\mu=(\mu_1', \mu_2')' : (p+q)\times 1$ and $\eta=(\eta_1', \eta_2')' : (p+q)\times 1$, is regarded as a GMANOVA problem with $n_1=1$ and $X_2=I$. Therefore the LRT is UMPI in $\mathcal{C}(\mathcal{G})$.

8. Locally Best Invariant (LBI) Test.

8.1. *LBI tests.* In this section, we derive a unique LBI test in $\mathcal{C}_\alpha(\mathcal{G})$, the class of $\mathcal{G}$-invariant *size* α tests in $\mathcal{C}(\mathcal{G})$, the class of $\mathcal{G}$-invariant level α tests based on (T_1, T_2) (see Sections 4 and 5) and a unique LBI test in $\mathcal{C}_{0\alpha}(\mathcal{G})$, the class of $\mathcal{G}$-invariant *size* α tests based on T_1 alone.

Let

$$\Delta \equiv \Delta(\delta) = \Sigma_{i=1}^{n_1}\delta_i = \operatorname{tr}\Theta_{12}\Sigma_{22\cdot 3}^{-1}\Theta_{12}'$$

where $\delta=(\delta_1, \cdots, \delta_{n_1})$ is the vector of the ordered characteristic roots of $\Theta_{12}\Sigma_{22\cdot 3}^{-1}\Theta_{12}'$. The following theorems give our main results here.

Theorem 8.1. There is a $\Delta_0>0$ such that on the set $\{\delta | \Delta(\delta)<\Delta_0\}$, the power function of any test ϕ in $\mathcal{C}_\alpha(\mathcal{G})$ is evaluated as

$$\pi(\phi, \delta) = \alpha + B(\phi)\Delta + o(\Delta) \tag{8.1}$$

where $o(\Delta)$ is uniform in $\phi\in\mathcal{C}_\alpha(\mathcal{G})$,

$$B(\phi) = a_1E_0\{\phi(T_1, T_2)\operatorname{tr}[(I+T_2)^{-1}[a_0T_1(I+T_1)^{-1}-I] + n_1]\}, \tag{8.2}$$

$$a_0=(n_1+n_3-p_3)/p_2, \quad a_1=(2n_1)^{-1} \tag{8.3}$$

and E_0 denotes the expectation under $H_0 : \Theta_{12}=0$. That is,

$$\lim_{\Delta\to 0} \sup_\phi |[\pi(\phi, \delta)-\alpha-B(\phi)\Delta]/\Delta| = 0.$$

Further, the test with the critical region

$$\mathcal{K}_5 : \operatorname{tr}(I+T_2)^{-1}[a_0T_1(I+T_1)^{-1}-I] > k_5$$

is the unique LBI test in $\mathcal{C}_\alpha(\mathcal{G})$ and so it is admissible in $\mathcal{C}_\alpha(\mathcal{G})$.

The proof of this theorem is given later.

Theorem 8.2. Pillai's test $\mathcal{K}_3 : \operatorname{tr}T_1(I+T_1)^{-1} > k_3$ is the unique LBI test in $\mathcal{C}_{0\alpha}(\mathcal{G})$.

Proof. Under the null hypothesis H_0, T_1 and T_2 are independent. Hence by Theorem 8.1, for $\phi\in\mathcal{C}_{0\alpha}(\mathcal{G})$

$$B(\phi) = a_1\operatorname{tr}E_0(I+T_2)^{-1}\{a_0E_0[\phi(T_1)T_1(I+T_1)^{-1}-I\}.$$

But from $T_2=Z_{13}V_{33}^{-1}Z_{13}'$ with $Z_{13}\sim N(0, I\otimes\Sigma_{33})$, the distribution of T_2 is the same as that of QT_2Q' for each $Q\in\mathcal{O}(n_1)$, which implies $E_0(I+T_2)^{-1}=cI$ for some $c>0$ (see Lemma 6.2). Consequently,

$$B(\phi) = a_1a_0cE_0[\phi(T_1)\operatorname{tr}T_1(I+T_1)^{-1}] - a_1a_0cn_1.$$

Applying the Generalized Neyman-Pearson Lemma (see Section 1 of Chapter 2) and maximizing $B(\phi)$ yields the test $\mathcal{K}_3$. This completes the proof.

Theorem 8.2 also follows from Schwartz (1967a). In fact, since given $T_2=t_2$, the conditional problem is the MANOVA problem and since $\operatorname{tr}(I+T_2)^{-1}\le n_1$, the result follows.

In the case $p_3=0$ where T_2 vanishes or $X_2=I$ the LBI test $\mathcal{K}_5$ in Theorem 8.1 is reduced to Pillai's test $\mathcal{K}_3$, and the case $p_1=p_3=0$ where our problem is the MANOVA problem, both $\mathcal{K}_3$ and $\mathcal{K}_5$ are naturally reduced to Pillai's test in the MANOVA problem.

In the investigation of the admissiblity of the GMANOVA tests (see Section 12), Marden (1983) says as follows. The LBI test $\mathcal{K}_5$

has the drawback, shared by the Pillai test $\mathscr{K}_3$, that there may be a sequence $\{\delta^{(n)}\}$ of parameter points for which $\delta_1^{(n)} \to \infty$ but the power does not approach one. The numerical work shows that this drawback can be serious for small n, but lessens as n increases. In comparing the LRT to the LBI test, it appears that the former is better when $\delta=(\delta_1, 0, \cdots, 0)'$ and δ_1 is large, and the latter is better when $\delta=(\delta_1, \cdots, \delta_1)'$ and δ_1 is small. For smaller n the differences are more pronouced. A more detailed study would be needed to pin down the relative advantages of the two tests. Marden (1983) in fact carried out a Monte Carlo Study to see the difference in power of the two tests (and some other tests) for various values of Δ for $\alpha=0.05$. Each difference was calculated using 100 pseudo-observations. The table below exhibits the maximum the LRT beats the LBI test and the maximum the LBI test beats the LRT test, where each maximum was taken over between 10 and 15 parameter points. It also contains the maximum power the LBI test was observed to have.

Maxima (standard error)

(n_3, n_1, p_2, p_3)	Form of δ	LRT beats LBI	LBI beats LRT	Power of LBI
(5, 2, 2, 2)	$(\delta_1, 0)$	.34 (.05)	.17 (.05)	.59
	$(\delta_1, \delta_1/10)$	.21 (.05)	.14 (.05)	.78
	$(\delta_1, \delta_1/2)$	.01 (.02)	.30 (.05)	.98
	(δ_1, δ_1)	.009 (.002)	.220 (.005)	.97
(10, 2, 5, 2)	$(\delta_1, 0)$	.324 (.005)	.040 (.004)	.69
	$(\delta_1, \delta_1/10)$	.28 (.05)	.05 (.04)	.86
	$(\delta_1, \delta_1/2)$	.030 (.002)	.087 (.004)	.986
	(δ_1, δ_1)	.021 (.002)	.084 (.004)	.985
(20, 2, 5, 5)	$(\delta_1, 0, 0, 0, 0)$	.075 (.003)	.009 (.003)	.976
	$(\delta_1, \delta_1, 0, 0, 0)$	.02 (.02)	.06 (.03)	.998
	$(\delta_1, \delta_1, \cdots, \delta_1)$	.001 (.001)	.046 (.003)	1
(20, 5, 2, 2)	$(\delta_1, 0)$	.116 (.003)	.002 (.002)	.985
	(δ_1, δ_1)	.000 (.000)	.030 (.003)	.999

8.2. *Evaluation of probability ratio.* To prove Theorem 8.1, we first review our reduced problem in Section 5. Let

$$(8.4) \qquad (U_1, U_3) \equiv ((Z_{12}, Z_{13}), \ (Z_{32}, Z_{33}))$$

and let $\mathscr{G}_1=\mathcal{O}(n_1)\times\mathcal{O}(n_3)\times\mathscr{A}_0$ with

$$\mathscr{A}_0 = \{B=\begin{pmatrix} A_{22} & A_{23} \\ 0 & A_{33}\end{pmatrix} \mid |A_{ii}| \neq 0 \ (i=1,2)\}$$

and the group acting on the left of (U_1, U_3) by

$$(8.5) \qquad g(U_1, U_3) = (P_1U_1B', P_3U_3B') \quad \text{for } g=(P_1, P_3, B)\in\mathscr{G}_1.$$

Then we have shown in Section 5 that the class $\mathcal{C}(\mathscr{G}_1)$ of $\mathscr{G}_1$-invariant level α tests based on (U_1, U_3) forms an essentially complete class in $\mathscr{D}(\mathscr{G})$. Noting that (T_1, T_2) is a function of (U_1, U_3), the class $\mathcal{C}(\mathscr{G})$ of $\mathscr{G}$-invariant level α tests based on (T_1, T_2) is equal to the class $\mathcal{C}(\mathscr{G}_1)$ when each test ϕ in $\mathcal{C}(\mathscr{G}_1)$ is expressed as a test on the coordinate space $\mathscr{T}_1\times\mathscr{T}_2$ of (T_1, T_2). Hence an LBI test in $\mathcal{C}_\alpha(\mathscr{G}_1)$ is an LBI test in $\mathcal{C}_\alpha(\mathscr{G})$ and vice versa. Further, let $J=J(U_1, U_3)$ be a maximal invariant under $\mathscr{G}_1$ acting on (U_1, U_3) as in (8.2), and let P_ξ^J be the distribution of J under $\xi=\Theta_{12}\Sigma_{22.3}^{-1/2}$. Then the ratio $dP_\xi^J/dP_0^J\equiv R_\xi$ is given by

$$(8.6) \qquad R_\xi = \exp(-\Delta/2)\int_{\mathscr{A}_0}\int_{\mathcal{O}(n_1)}\exp[\operatorname{tr}\xi' P_1(W_2A_{22}'+W_3A_{23}')] \times\nu_1(dP_1)h(B)\mu(dB)$$

where

$$(8.7) \quad h(B) = \exp(-\frac{1}{2}\operatorname{tr}BB')/D \text{ with } D=\int_{\mathscr{A}_0}\exp(-\frac{1}{2}\operatorname{tr}BB')\mu(dB)$$

$$(8.8) \qquad \mu(dB) = |A_{22}A_{22}'|^{(M-p_2)/2}|A_{33}A_{33}'|^{(n-p_3)/2}dA_{22}dA_{33}dA_{23}$$

$$M=n_1+n_3-p_3,$$

$$(8.9) \qquad (W_2, W_3)\equiv U_1C'=(Z_{12}C_{22}'+Z_{13}C_{23}', \ Z_{13}C_{33}') : n_1\times(p_2+p_3)$$

(8.10) $\quad C(U_1'U_1+U_3'U_3)C' = I_{p_2+p_3}$ and $C=(C_{ij})\in\mathscr{A}_0$

(see Lemma 5.3 and ((6.2)).

Now we shall evaluate R_t around $\Delta(\delta)=0$ as follows.

Lemma 8.1. The ratio R_t in (8.4) is evaluated as

(8.11) $\quad R_t = 1 + a_1\{\mathrm{tr}(I+T_2)^{-1}[a_0T_1(I+T_1)^{-1}-I]+n_1\}\Delta + o(\mathrm{tr}\xi\xi'),$

where $o(\mathrm{tr}\xi\xi')$ is uniform in (T_1, T_2) or (U_1, U_3).

To prove this, expand the integrand in (8.6) as $1+K+\frac{1}{2}K^2+o(K^2)$ where

(8.12) $\quad K= \mathrm{tr}\xi'P_1W_2A_{22}' + \mathrm{tr}\xi'P_1W_3A_{23}' \equiv K_1+K_2,$ say

First, note that the integral of $(\mathrm{tr}P_1Q)^k$ over $\mathcal{O}(n_1)$ with respect to $\nu_1(dP_1)$ is zero for k odd and

Lemma 8.2. (James (1964) (22) and (23) with k=1. See also Appendix D.) Let A and B be $p\times p$ matrices and let ν be the invariant probability measure on $\mathcal{O}(p)$. Then

(8.13) $$\int_{\mathcal{O}(p)}\mathrm{tr}APBP'\nu(dP) = \mathrm{tr}A\mathrm{tr}B/p \quad \text{and}$$

(8.14) $$\int_{\mathcal{O}(p)}(\mathrm{tr}AP)^2\nu(dP) = \mathrm{tr}AA'/p.$$

Hence the integral of K in (8.12) over $\mathcal{O}(n_1)$ is zero. We evaluate the integration of $K^2=K_1^2+2K_1K_2+K_2^2$ term by term. First, from Lemma 8.2, the integration of K_1^2 over $\mathcal{O}(n_1)$ is equal to $\mathrm{tr}\,W_2A_{22}'\xi'\xi A_{22}W_2'/n_1$, and the integration of K_2^2 over $\mathcal{O}(n_1)$ is $\mathrm{tr}\,W_3A_{33}'\xi'\xi A_{33}W_3'/n_1$. On the other hand, the intergration of K_1K_2 over $\mathscr{A}_0$ is zero because K_1K_2 is an odd function of A_{23} and because $A_{23}\sim N(0, I\otimes I)$ independently of A_{22} and A_{33} from the form of $h(B)\mu(dB)$ in (8.7) and (8.8). Hence

(8.15) $$\iint_{\mathcal{O}(n_1)\times\mathscr{A}_0}K^2\nu_1(dP_1)h(B)\mu(dB)$$
$$=\int_{\mathscr{A}_0}[\mathrm{tr}\,W_2'W_2A_{22}'\xi'\xi A_{22}+\mathrm{tr}\,W_3'W_3A_{23}'\xi'\xi A_{23}]h(B)\mu(dB)/n_1.$$

To evaluate the integration of $\mathrm{tr}\,W_2'W_2A_{22}'\xi'\xi A_{22}$ over $\mathscr{A}_0$, note that from (8.7) and (8.8), the marginal density of A_{22} is

$$h_2(A_{22}) = c|A_{22}A_{22}'|^{M/2}\exp(-\frac{1}{2}\mathrm{tr}A_{22}A_{22}')\mu_2(dA_{22}) \quad \text{with}$$
$$\mu_2(dA_{22}) = |A_{22}A_{22}'|^{-p_2/2}dA_{22}.$$

Lemma 8.3. Under $A=BC$ where $(B,C)\in G_T^+(p)\times\mathcal{O}(p)$, $\mathscr{G}l(p)$ is homemorphic to $G_T^+(p)\times\mathcal{O}(p)$ and an invariant measure $|AA'|^{-p/2}dA$ on $\mathscr{G}l(p)$ is factored as $\lambda(dB)\tau(dC)$, where $G_T^+(p)$ denotes the set of $p\times p$ lower triangular matrices with positive diagonal elements, $\lambda(dB)=(\prod b_{ii}^{-i})dB$ with $B\in G_T^+(p)$ is an invariant measure on $G_T^+(p)$ and $\tau(dC)$ is the invariant probability measure on $\mathcal{O}(p)$.

Proof. See, e.g., Wijsman (1967) page 398 or Eaton (1983) page 213.

Based on this lemma, let $\mu_2(dA_{22})=\lambda_2(dE)\tau_2(dQ)$ with $A_{22}=EQ$, where $\tau(dQ)$ is the invariant probability measure on $\mathcal{O}(p_2)$ with $Q\in\mathcal{O}(p_2)$ and $\lambda_2(dE)=(\prod e_{ii}^{-i})dE$ with $E=(e_{ij})\in G_T^+(p_2)$. Under this decomposition, the integration of $\mathrm{tr}\,W_2'W_2Q'E'\xi'\xi EQ$ in (8.15) over $\mathcal{O}(p_2)$ is $\mathrm{tr}\,W_2'W_2\mathrm{tr}EE'\xi'\xi/p_2$ by Lemma 8.2 and the integration of this term over $G_T^+(p_2)$ with respect to $c|EE'|^{M/2}\exp(-\frac{1}{2}\mathrm{tr}EE')\tau_2(dE)$ is $M\mathrm{tr}\,W_2'W_2\mathrm{tr}\xi'\xi/p_2$ since EE' is distributed as $W_{p_2}(I, M)$. Therefore

(8.16) $$\int_{\mathscr{A}_0}(\mathrm{tr}\,W_2'W_2A_{22}'\xi'\xi A_{22})h(B)\mu(dB) = (M/p_2)\Delta\mathrm{tr}\,W_2'W_2.$$

On the other hand, since the marginal distribution of A_{23} is $N(0, I_{p_2}\otimes I_{p_3})$, the integration of $\mathrm{tr}\,W_3'W_3A_{23}'\xi'\xi A_{23}$ over $\mathscr{A}_0$ is simply $\mathrm{tr}\,W_3'W_3\mathrm{tr}\xi'\xi$. Therefore from (8.15) and (8.16),

$$\frac{1}{2}\iint_{O(n_1)\times\mathcal{A}_0} K^2\nu(dP_1)h(B)\mu(dB) = (2n_1)^{-1}\Delta[(M/p_2)\mathrm{tr}\,W_2'W_2 + \mathrm{tr}\,W_3'W_3]. \tag{8.17}$$

Here we identify $\mathrm{tr}\,W_i'W_i$ by

Lemma 8.4. (1) $\mathrm{tr}\,W_2'W_2 = \mathrm{tr}\,x(x'x+V_{22.3})^{-1}x'(I+T_2)^{-1}$
$= \mathrm{tr}\,T_1(I+T_1)^{-1}(I+T_2)^{-1}$

(2) $\mathrm{tr}\,W_3'W_3 = n_1 - \mathrm{tr}(I+T_2)^{-1}$.

The proof of this lemma is given in Appendix *E*.

This lemma with (8.17) gives the first two terms of R_ξ in (8.11). To complete the proof Lemma 8.1, we need to show the integration of the term $o(K^2)$ over $\mathcal{A}_0\times\mathcal{O}(n_1)$ is of order $o(\mathrm{tr}\xi\xi')$ uniformly in (U_1, U_2) or (T_1, T_2). Using Schwarz's inequality in (8.12),

$$\begin{aligned} K^2 &\le \mathrm{tr}\xi\xi'\mathrm{tr}\,W_2'W_2A_{22}'A_{22} + 2(\mathrm{tr}\xi\xi')[\mathrm{tr}\,W_2'W_2A_{22}'A_{22} \\ &\quad\times \mathrm{tr}\,W_3'W_3A_{33}'A_{33}]^{1/2} + \mathrm{tr}\xi\xi'\mathrm{tr}\,W_3'W_3A_{33}'A_{33} \\ &\le \mathrm{tr}\xi\xi'(p_2+p_3)[\mathrm{tr}A_{22}A_{22}'+\mathrm{tr}A_{33}A_{33}' \\ &\quad + 2(\mathrm{tr}A_{22}A_{22}'\mathrm{tr}A_{33}A_{33}')^{1/2}] \equiv \mathrm{tr}\xi\xi'\psi(A_{22}, A_{33}) \end{aligned} \tag{8.18}$$

The second inequality follows from $\mathrm{tr}AB \le \mathrm{tr}A\mathrm{tr}B$ for $A, B\in\mathcal{S}_+(p)$ and $\mathrm{tr}\,W_iW_i' \le p_2+p_3$ $(i=2, 3)$, which follows from (8.9) and (8.10). Therefore

$$\iint_{O(n_1)\times\mathcal{A}_0} o(K^2)\nu_1(dP_1)h(B)\mu(dB) = o(\mathrm{tr}\xi\xi')\int_{\mathcal{A}_0}\psi(A_{22}, A_{33})h(B)\mu(dB) = o(\mathrm{tr}\xi\xi'), \tag{8.19}$$

which completes the proof of Lemma 8.1.

8.3. *Proof of Theorem* 8.1. Let $\phi(T_1, T_2)\in\mathcal{C}_\alpha(\mathcal{G})=\mathcal{C}_\alpha(\mathcal{G}_1)$, the power function of ϕ is for $\Delta=\mathrm{tr}\xi\xi'$ near zero

$$\pi(\phi, \delta) = \int\phi(T_1, T_2)R_\xi P(d(T_1, T_2)\,|\,\xi=0) = \alpha+B(\phi)\Delta+o(\Delta)$$

where R_ξ is given by Lemma 8.1. Applying the Generalized Neyman-Pearson Lemma (see Section 1 of Chapter 2) and maximizing $B(\phi)$ with respect to $\phi\in\mathcal{C}_\alpha(\mathcal{G})$ yields the unique LBI test $\mathcal{K}_5$ given in Theorem 8.1. This completes the proof.

9. Local Minimaxity of the LBI Test.

9.1. *Definition of local minimaxity.* Once an optimal test is obtained, our next concern is to derive or approximate the null distribution. But for continuation of the argument, we first consider minimaxity. Generally speaking, the minimaxity property of a test is hard to establish as has been discussed in Section 6 of Chapter 2. In Hotelling's T^2-problem, which is a special case of the MANOVA problem, Salaevskii (1968) proved the minimax property of Hotelling's T^2-test (see Chapter 2 for the minimaxity of a test). In this section, we prove that the LBI test is locally minimax in the sense of Giri and Kiefer (1964) (see also Schwartz (1967) and Giri (1977)). In Giri and Kiefer (1964) Lemma 1 states the conditions under which a given test is locally minimax as follows. Let $(\mathcal{X},\mathcal{B})$ be a measurable space where $\mathcal{X}\subset R^l$ and $\mathcal{B}$ is a Borel σ-field. Let $p(\cdot : \Delta,\xi)$ be a density with respect to a σ-finite measure μ where Δ is a real parameter, ξ is nuissance parameter and the range of ξ may depend on Δ. Consider a testing problem $H_0 : \Delta=0$ versus $H_1 : \Delta=\lambda$ $(\lambda>0)$.

Definition 9.1. A test ϕ^* is said to be locally minimax of level α $(0<\alpha<1)$ for testing $H_0 : \Delta=0$ versus $\Delta=\lambda$ as $\lambda\to 0$ if

$$\lim_{\lambda\to 0}\frac{\inf_\xi \pi(\phi^*, (\lambda,\xi))-\alpha}{\sup_{\phi_\lambda\in\mathcal{Q}_\alpha}\inf_\xi \pi(\phi_\lambda, (\lambda,\xi))-\alpha} = 1 \tag{9.1}$$

where $\pi(\phi,(\lambda,\xi))=E[\phi\,|\,\lambda,\xi]$ is the power function of ϕ and $\mathcal{Q}_\alpha$ is the class of tests of level α.

Lemma 9.1. Under the following three assumptions, the test ϕ^* is locally minimax for testing $H_0 : \Delta=0$ versus $\Delta=\lambda$ as $\lambda\to 0$.

Assumption 1. There exists a statistic $U(x)$ such that U is bounded and positive and has a continuous distribution function for each (Δ, ξ), which is equicontinuous in (Δ, ξ) for $\Delta<\Delta_0$, and such that $\phi^*(x)=1$ if $U(x)>c$ and $\phi^*(x)=0$ otherwise.

Assumption 2. $E(\phi^*|0,\xi) = \alpha$ and $E(\phi^*|\lambda,\xi) = \alpha + h(\lambda) + g(\lambda,\xi)$, where $g(\lambda,\xi)=o(h(\lambda))$ uniformly in ξ, $h(\lambda)>0$ for $\lambda>0$ and $h(\lambda)=O(\lambda)$.

Assumption 3. There exist probability measures $\eta_{0,\lambda}$ and $\eta_{1,\lambda}$ on the sets $\{\Delta=0\}$ and $\{\Delta=\lambda\}$ respectively for which

$$\frac{\int p(x:\lambda,\xi)\eta_{1,\lambda}(d\xi)}{\int p(x:0,\xi)\eta_{0,\lambda}(d\xi)} = 1 + h(\lambda)[k(\lambda)+U(x)r(\lambda)] + B(x,\lambda),$$

where $0<c_1<r(\lambda)<c_2<\infty$ for λ sufficiently small, $k(\lambda)=O(1)$ and $B(x,\lambda)=o(h(\lambda))$ uniformly in x.

9.2. *Local minimaxity of the LBI test.* The definition of local minimaxity in (9.1) is a natural extension of the definition of minimaxity

$$\inf_\xi \pi(\phi^*, (\lambda,\xi)) = \sup_{\phi_\lambda}\inf_\xi \pi(\phi_\lambda, (\lambda,\xi)) \tag{9.2}$$

where ξ is regarded as an original parameter and λ is a real-valued function of ξ so that for given λ, $\lambda=\lambda(\xi)$ gives a restriction on the parameter space of ξ. One of our reduced problems is to test H_0 : $\Theta_{12}=0$ in the model

$$\begin{pmatrix} Z_{12} & Z_{13} \\ Z_{32} & Z_{33} \end{pmatrix} \sim N\left(\begin{pmatrix} \Theta_{12} & 0 \\ 0 & 0 \end{pmatrix}, I_{n_1+n_3}\otimes\begin{pmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} \end{pmatrix}\right) \tag{9.3}$$

Here ξ may be regarded as $(\Theta_{12},\Sigma_{22},\Sigma_{23},\Sigma_{33})$ and λ will be taken to

be

$$\lambda=\lambda(\Theta_{12},\Sigma_{22},\Sigma_{23},\Sigma_{33}) = \operatorname{tr}\Theta_{12}\Sigma_{22\cdot 3}^{-1}\Theta_{12}'. \tag{9.4}$$

For a given λ, it may be possible to find a minimax test ϕ_λ^* which satisfies (9.2) on the contour

$$C_\lambda = \{(\Theta_{12},\Sigma_{22},\Sigma_{23},\Sigma_{33})\mid \lambda(\Theta_{12},\Sigma_{22},\Sigma_{23},\Sigma_{33})=\lambda\} \tag{9.5}$$

or to show that a given test is minimax on the contour C_λ. However, not only it is difficult but also ϕ_λ^* may depend on λ. In the definition of local minimaxity, λ is let to go to zero. But $\pi(\phi,(\lambda,\xi))$ is continuous in (Θ,Σ) and $\pi(\phi,(0,\xi))=\alpha$ under the null hypothesis $\Theta_{12}=0$, since $\Theta_{12}=0$ if and only if $\lambda=0$. Hence (9.2) is reduced to a trivial relation as $\lambda\to 0$. This is why in (9.1) α is subtracted from the both sides of (9.2) and the ratio of the two terms is taken before $\lambda\to 0$. Understanding this point, we prove

Theorem 9.1. The LBI test $\mathscr{K}_5$ is locally minimax with respect to the contour C_λ in (9.5) as $\lambda\to 0$ under the model (9.3).

It should be noted that the class of tests Q_α in which we claim the local minimaxity of $\mathscr{K}_5$ is not restricted to the class of $\mathscr{G}$ invariant tests of size α. It is the class of all tests of size α for testing $\Theta_{12}=0$ in the model (9.3). Further, this theorem simply shows the local minimax of $\mathscr{K}_5$ on the contour C_λ. In the MANOVA problem, Schwartz (1967) defined a local family of contours containing such a contour as C_λ. His argument is applicable to our case to prove the local minimaxity of $\mathscr{K}_5$ for a wider class of contours.

It is also remarked that if any $\mathscr{G}$-invariant test is to be locally minimax with respect to C_λ, it must be the one which is LBI. In fact, if there exists a locally minimax $\mathscr{G}$-invariant test of size α, we can restrict the class Q_α to the class of $\mathscr{G}$-invariant tests of size α, $\mathscr{C}_\alpha(\mathscr{G})$, (see Section 6 of Chapter 2) and by Theorem 8.1, for a $\mathscr{G}$-invariant test ϕ of size α, the power function of ϕ is given by

$$\pi(\phi,\ \delta) = \alpha + \Delta B(\phi) + o(\Delta)$$

with $\Delta=\lambda$ in (9.4). Hence in (9.1)

$$\sup_\phi \inf_\xi \pi(\phi,\ \delta) - \alpha = \sup_\phi[\Delta B(\phi)+o(\Delta)]$$

so that

$$\lim_{\Delta\to 0}[\Delta B(\phi^*)+o(\Delta)]/\sup_\phi[\Delta B(\phi)+o(\Delta)] = B(\phi^*)/\sup_\phi B(\phi).$$

For the right side to be equal to 1, $B(\phi^*)=\sup_\phi B(\phi)$, implying ϕ^* is LBI. Hence the Pillai test in our problem is not locally minimax though the Pillai test in the MANOVA problem is locally minimax.

9.3. *Proof of Theorem* 9.1. A crucial point in the proof is to apply the Hunt-Stein theorem (see Section 6 of Chapter 2) to reduce the class Q_α of tests of size α to a class of tests of size α which are invariant under a group. In fact, the Hunt-Stein theorem says that if a group is solvable, there exists a minimax test which is invariant under the group. Hence the class Q_α in the definition (9.1) can be replaced by

$$\sup_{\phi\in I_\alpha}\inf_\xi \pi(\phi,\ (\lambda,\xi)) - \alpha$$

where I_α is the class of invariant tests. However, the group we have been adopting does not satisfy the condition in the Hunt-Stein theorem, because

$$\mathscr{A}_0 = \{B=(A_{ij})\ (i,j=2,3)\,|\,A_{32}=0,\ |A_{ii}|\neq 0\ (i=1,2)\}$$

in $\mathscr{G}_1$ of Section 8 is not solvable. For this reason, in this section we take group $\mathscr{L}=\mathscr{O}(n_1)\times\mathscr{O}(n_3)\times\mathscr{B}$ which is a subgroup of $\mathscr{G}_1=\mathscr{O}(n_1)\times\mathscr{O}(n_3)\times\mathscr{A}_0$ in Section 8 where $\mathscr{B}$ is the solvable group

(9.6) $$\mathscr{B} = \left\{B=\begin{pmatrix} B_{22} & B_{23} \\ 0 & B_{33}\end{pmatrix}\,\middle|\,B_{ii}\in\mathscr{GU}(p_i)\ (i=2,3)\right\}.$$

Here $\mathscr{GU}(p_i)$ is the group of $p_i\times p_i$ upper triangular matrices. Note $B\in\mathscr{GU}(p_2+p_3)$ and the group $\mathscr{L}$ satisfies the condition in the Hunt-Stein theorem. This group $\mathscr{L}$ leaves the problem $\Theta_{12}=0$ invariant in the model (9.3) under the action

(9.7) $$g(U_1, U_3) = (P_1U_1B',\ P_3U_3B') \quad \text{for } g=(P_1,P_3,B)\in\mathscr{L}$$

where as before

(9.8) $$(U_1, U_3)=((Z_{12},Z_{13}),\ (Z_{32},Z_{33})).$$

Since $\mathscr{L}$ is a subgroup of $\mathscr{G}_1$, any $\mathscr{G}_1$ invariant test is $\mathscr{L}$-invariant.

Lemma 9,2. Under the group $\mathscr{L}$, a maximal invariant parameter is $\Sigma_{22.3}^{-1/2\prime}\Theta_{12}'\Theta_{12}\Sigma_{22.3}^{-1/2}$, where $\Sigma_{22.3}^{-1/2}\in\mathscr{GU}(p_2)$.

Proof. Similar to the proof of Lemma 2.1 and so omitted.

As a consequence of this lemma, without loss of generality we can replace $(\Theta_{12},\Sigma_{22},\Sigma_{23},\Sigma_{33})$ by $(\xi,I,0,I)$ with $\xi=\Theta_{12}\Sigma_{22.3}^{-1/2}$ where $\Sigma_{22.3}^{1/2}\in\mathscr{GU}(p_2)$, whenever we consider the distribution of a maximal invariant $K=K(U_1,U_2)$ under $\mathscr{L}$. Therefore formally the argument here is parallel to the argument in Section 8 and the probability ratio $R_\xi=dP_\xi^K/dP_0^K$ here is also given by

(9.9) $$R_\xi = \exp(-\Delta/2)\int_{\mathscr{B}}\int_{\mathscr{O}(n_1)}\exp[\operatorname{tr}\xi'P_1(W_2B_{22}'+W_3B_{23}')] \times\nu_1(dP_1)h(B)\beta(dB)$$

where

(9.10) $$h(B) = \exp(-\frac{1}{2}\operatorname{tr}BB')/D \quad \text{with}$$

$$D=\int_{\mathscr{B}}\exp(-\frac{1}{2}\operatorname{tr}BB')\beta(dB),$$

$$(9.11)\quad \beta(dB)=|B_{22}B'_{22}|^{M/2}\tau_2(dB_{22})\,|B_{33}B'_{33}|^{(n_1+n_3)/2}\,\tau_3(dB_{33})dB_{23}$$

$$\tau_j(dB_{jj})=\left(\prod_{i=1}^{p_j} b(j)_{ii}^{p_j-i+1}\right)dB_{jj}\ \ (j=2,3),$$

$$M=n_1+n_3-p_3,$$

$$(9.12)\quad (W_2,\ W_3)\equiv U_1C'=(Z_{12}C'_{22}+Z_{13}C'_{32},\ Z_{13}C'_{33}),$$

$$(9.13)\quad C(U'_1U_1+U'_3U_3)C'=I\ \text{ and }\ C=(C_{ij})\in\mathcal{B}.$$

Here $b(j)_{ik}$ is the (i,k) element of B_{jj}. Note $\tau_j(dB_{jj})$ is an invariant measure on $\mathcal{GU}(p_j)$ $(j=2,3)$.

Under this preparation, we verify Assumptions 1, 2 and 3 required in Lemma 9.1. To check Assumption 1, let

$$(9.14)\quad U=\operatorname{tr}(I+T_2)^{-1}[b_1T_1(I+T_1)^{-1}-b_2I]+1$$

where $b_1=(n_1+n_3-p_3)/(2n_1p_2)$ and $b_2=1/2n_1$. Then the LBI test $\mathcal{K}_5$ is described by $\phi^*(U)$, where $\phi^*=1$ if $U>c_\alpha$ and $\phi^*=0$ otherwise, and U is bounded and positive. And with $\Delta=\operatorname{tr}\xi\xi'$ and $\xi=\Theta_{12}\Sigma_{22.3}^{-1/2}$, the distribution of U is continuous for each (Δ,ξ). Further, by (8.1), for $\Delta<$some Δ_0, the distribution function of U is given by

$$(9.15)\quad F^U(c|\Delta,\xi)=1-E[\phi_c|\Delta,\xi]=1-\alpha(c)-B(\phi_c)\Delta+o_c(\Delta)$$

where $\phi_c(U)=1$ if $U>c$ and $\phi_c(U)=0$ otherwise, and $\alpha(c)=E[\phi_c|0,0]=E_0[\phi_c]$. Here the remainder term $o_c(\Delta)$ is defined by $E[\phi_c(U)\bar{o}(K^2)|0,0]$, where $\bar{o}(K^2)$ is given by the left side of (8.19). From (8.18) and (8.19), $|\bar{o}(K^2)|\le a\Delta$ for some $a>0$. Hence for $c'>c$ and for $\Delta<\Delta_0$

$$|o_{c'}(\Delta)-o_c(\Delta)|\le a\Delta|E_0(\phi_c)-E_0(\phi_{c'})|$$

and so

$$|F^U(c'|\Delta,\xi)-F^U(c|\Delta,\xi)|\le|\alpha(c)-\alpha(c')|$$
$$+|B(\phi_c)-B(\phi_{c'})|\Delta+|\alpha(c)-\alpha(c')|a\Delta,$$

Since $\alpha(c)$ and $B(\phi_c)$ do not depend on (Δ,ξ) and since these are

uniformly continuous, $F^U(\cdot\,|\Delta,\xi)$ is equicontinuous for $\Delta<$some Δ_0. Therefore Assumption 1 is verified.

To check Assumption 2, set $\phi_\alpha\equiv\alpha$ in (8.1) to obtain $B(\phi_\alpha)=0$ from $\pi(\phi_\alpha,\Delta)=\alpha$. Since the LBI test ϕ^* is a unique maximizer of $B(\phi)$ with respect to $\phi\in\mathcal{C}_\alpha(\mathcal{G})$ for $0<\alpha<1$, we obtain $B(\phi^*)>0=B(\phi_\alpha)$. Further, since by (8.1)

$$E[\phi^*|\Delta,\xi]=\alpha+B(\phi^*)\Delta+o(\Delta),$$

setting $\Delta=\lambda$, $h(\lambda)=B(\phi^*)\lambda$ and $g(\lambda,\xi)=o(\operatorname{tr}\xi\xi')=o(\Delta)$, Assumption 2 follows. Finally we verify Assumption 3. Under the group $\mathcal{L}$, the probability ratio of the distributions of a maximal invariant $K=K(U_1,U_3)$ under $\xi=\Theta_{12}\Sigma_{22.3}^{-1/2}$ is given by R_ξ in (9.9). Of course, R_ξ is the density of K with respect to P_0^K.

Lemma 9.3. The ratio R_ξ is evaluated as

$$(9.16)\quad R_\xi=1-\frac{\Delta}{2n_1}\operatorname{tr}(I+T_2)^{-1}+I(W_2,\xi)+o(\Delta)$$

where $o(\Delta)$ is uniform in W_1 and W_2, with $\tau_2(dB_{22})$ in (9.11),

$$(9.17)\quad I(W_2,\xi)=\int_{\mathcal{GU}(p_2)}\frac{1}{2n_1}\operatorname{tr}W_2B'_{22}\xi'\xi B_{22}W'_2|B_{22}B'_{22}|^{M/2}$$
$$\times\exp\left(-\frac{1}{2}\operatorname{tr}B_{22}B'_{22}\right)\tau_2(dB_{22})/D_0$$

$$D_0=\int_{\mathcal{GU}(p_2)}|B_{22}B'_{22}|^{M/2}\exp\left(-\frac{1}{2}\operatorname{tr}B_{22}B'_{22}\right)\tau_2(dB_{22}).$$

Proof. Expanding the integrand of R_ξ in (9.9) as $1+J+\frac{1}{2}J^2+o(J^2)$ where

$$(9.18)\quad J=\operatorname{tr}\xi'P_1W_2B'_{22}+\operatorname{tr}\xi'P_1W_3B'_{23}\equiv J_1+J_2,\ \text{say},$$

the integration of J over $\mathcal{O}(n_1)$ is zero by the same reason as in the case of R_ξ in Section 8. Further, in the same way as in (8.18), the integration of the term $o(J^2)$ is shown to be $o(\Delta)$ uniformly in (U_1,U_3). We evaluate the integration of J^2. From Lemma 8.3, the integration of J_1^2 over $\mathcal{O}(n_1)$ is $\operatorname{tr}W_2B'_{22}\xi'\xi B_{22}W'_2/n_1$, while the integration of J_2^2 over $\mathcal{O}(n_1)\times\mathcal{B}$ yields $\operatorname{tr} W_3W_3'\operatorname{tr}\xi\xi'/n_1$ which is equal to $\Delta[n_1-\operatorname{tr}(I+T_2)^{-1}]/n_1$ by Lemma 8.4, since $B_{23}\sim N(0, I\otimes I)$. And the integration of J_1J_2 is zero. Hence

$$R_\xi = \exp(-\Delta/2)\{1+\frac{1}{2n_1}\Delta[n_1-\operatorname{tr}(I+T_2)^{-1} + \int_{\mathcal{B}} \operatorname{tr} W_2B_{22}'\xi'\xi B_{22}W_2'h(B)\beta(dB)]+o(\Delta)\}$$

Since $\exp(-\Delta/2)=1-\Delta/2+o(\Delta)$ and since $h(B)$ is the density of B with respect to $\beta(dB)$ in (9.11), using the marginal density of B_{22}, we obtain the result.

In Section 8, the integral in (8.17) was over $\mathcal{G}l(p_2)$ instead of $\mathcal{GU}(p_2)$ (see (8.15) or (8.16)) and hence resulted in $(2n_1p_2)^{-1}M\Delta\operatorname{tr} W_2W_2'$ for arbitrary ξ. This is not true now, but if we can show that there is a special value of ξ for which the same result holds, then for $\lambda <$ some λ_0.

(9.19) $$R_\xi = 1+\lambda[U-1]+o(\Delta)$$

follows from (9.16) and the definition of U in (9.14). Therefore putting the probability measure η on that particular ξ, Assumption 3 is verified with $h(\lambda)=B(\phi^*)\lambda$, $k(\lambda)=-B(\phi^*)^{-1}$, $r(\lambda)=B(\phi^*)^{-1}$ and $B(x,\lambda)$ being the last term in (9.16). We show the existence of ξ such that

$$I(W_2, \xi) = (2n_1p_2)^{-1}M\Delta\operatorname{tr} W_2W_2'$$

for any W_2 where $\Delta=\operatorname{tr}\xi\xi'$. For any matrices A and Γ, define

(9.20) $$r(A,\Gamma)=D_0^{-1}\int_{\mathcal{GU}(p_2)}(\operatorname{tr}\Gamma S'AS)|S'S|^{M/2}\exp(-\frac{1}{2}\operatorname{tr}S'S)\tau_2(dS).$$

where for simplicity $B_{22}=S'$. Let $\Gamma=(\gamma_{ij})$, $A=(a_{ij})$ and $S=(s_{ij})$. Then $\operatorname{tr}\Gamma SAS'=\sum_{ijkl}\gamma_{ki}s_{ij}a_{jl}s_{kl}$ but the integration of $s_{ij}s_{kl}$ gives 0 unless $i=k$ and $j=l$. Thus only $\sum_{i\geq j}\gamma_{ij}a_{jj}s_{ij}^2$ remains in the integration. Define for $i\geq j$

(9.21) $$c_{ij} = D_0^{-1}\int s_{ij}^2|SS'|^{M/2}\exp(-\frac{1}{2}\operatorname{tr}SS')\tau_2(dS) \ (>0).$$

Let C be the lower triangular matrix with elements c_{ij} and let $u'=(u_1,\dots,u_{p_2})$ with $u_i=\gamma_{ii}$. Then

$$r(A,\Gamma) = \sum_{i\geq j}u_ia_{jj}c_{ij} = \sum_j(\sum_{i\geq j}u_ic_{ij})a_{jj}.$$

In order that this be proportional to $\operatorname{tr}A$, $\sum_{i\geq j}u_ic_{ij}=\text{const}=c_0$, say, is necessary and sufficient. That is,

$$u'C = c_0e'$$

is necessary and sufficient, where $e'=(1,\dots,1)$, and so $u'=c_0e'C^{-1}$ since $|C|\neq 0$. With this choice for u, $r(A,\Gamma)=c_0\operatorname{tr}A$ and in particular $r(I,\Gamma)=c_0\operatorname{tr}I=c_0p_2$. But as in (8.17), $r(I,\Gamma)=M\operatorname{tr}\Gamma$. Therefore $c_0=Mp_2^{-1}\operatorname{tr}\Gamma$ and finally $r(A,\Gamma)=Mp_2^{-1}\operatorname{tr}A\operatorname{tr}\Gamma$ provided the i-th diagonal element of Γ is proportional to the i-th element of $e'C^{-1}$ for all $i=1,\dots,p_2$. It is noted that $c_{ii}=n^2+2n$ and $c_{ij}=n$ $(i\neq j)$ so that all the elements of $e'C^{-1}$ are positive. Taking $\Gamma=\xi'\xi$ and $A=W_2'W_2$ proves our claim.

Thus we have verified all the assumptions required in Lemma 9.1 under the action of the group $\mathcal{L}$. By virtue of the Hunt-Stein theorem, we obtain Theorem 9.1.

10. Distribution of the LBI Test.

10.1. *Exact distribution.* When $n_1=1$, we first derive the exact distribution of the LBI test as in Kariya and Kanazawa (1978). When $n_1=1$, let

(10.1) $W_1=(n_3-p_3-p_2+1)T_1/p_2$ and $W_2=(n_3-p_3+1)T_2/p_3$.

Then as has been seen in Chapter 7, the conditional distribution of W_1 given W_2 is a noncentral F distribution with degrees of freedom $(p_2, n_3-p_3-p_2+1)$ and noncentral parameter $\lambda(t_2,\gamma)=\gamma/2(1+t_2)$ where $\gamma=\Theta_{12}\Sigma_{22\cdot3}^{-1}\Theta_{12}'$. Using this fact, the distribution of

(10.2) $$S_1=T_1/(1+T_1) \quad \text{and} \quad S_2=1/(1+T_2)$$

is given by

(10.3) $$\begin{cases} S_1 \text{ given } S_2 \sim Be(p_2/2,\ r/2{:}s_2\gamma/2) \\ S_2 \qquad\qquad \sim Be(q/2,\ p_3/2) \end{cases}$$

where $Be(a,b{:}\tau)$ denotes noncentral beta distribution with degrees of freedom (a,b) and noncentral parameter τ, $Be(a,b)\equiv Be(a,b{:}0)$,

(10.4) $$r=n_3-p_3-p_2+1 \quad \text{and} \quad q=n_3-p_3+1.$$

Let $b(z{:}\alpha,\beta)=[B(\alpha,\beta)]^{-1}z^{\alpha-1}(1-z)^{\beta-1}$, $P_0(j{:}\mu)=\mu^j e^{-\mu}/j!$ and $\tau=\lambda/2$. Then the joint distribution of (S_1,S_2) is given by

(10.5) $$h(s_1,s_2{:}\tau)=\Sigma_{k=0}^{\infty}P_0(k{:}\tau s_2)b(s_1{:}\ \tfrac{1}{2}p_2+k,\ \tfrac{1}{2}r)b(s_2{:}\ \tfrac{1}{2}q,\ \tfrac{1}{2}p_3)$$

where $0<s_1,s_2<1$ (see e.g., Johnson and Kotz (1970)). Since the LBI test statistic is

(10.6) $$U=[a_0T_1(1+T_1)^{-1}-1]/(1+T_2)=[a_0S_1-1]S_2$$

with $a_0=(1+n_3-p_3)/p_2$, the nonnull distribution function of U is given by

(10.7) $$Q(x{:}\tau)=P((a_0S_1-1)S_2\le x).$$

To evaluate this, we distinguish the following four cases: (1) $x\ge a_0-1$, (2) $0\le x<a_0-1$, (3) $-1\le x<0$ and (4) $x<-1$. It is noted that $a_0>1$ and

$$P_0(k{:}\tau s_2)=\textstyle\sum_{j=0}^{\infty}e^{2\tau}p_0(k{:}\tau)P_0(j{:}\tau)(-1)^j s_2^{k+j}.$$

The following results are easily obtained by integrating (10.5) over each region

Case (1) $x\ge a_0-1: Q(x:\tau)=1$ and so $Q(x:0)=1$.

Case (2) $0\le x<a_0-1: Q(x:\tau)=K_1(x:\tau)+K_2(x:\tau)$, where

$$K_1(x:\tau)=\textstyle\sum_k\sum_j e^{2\tau}P_0(k:\tau)P_0(j:\tau)(-1)^j$$
$$\times\Big[B\Big(\frac{q}{2}+k-j,\ \frac{p_3}{2}\Big)/B\Big(\frac{q}{2},\ \frac{p_3}{2}\Big)\Big]I\Big(\frac{x+1}{a_0}:\frac{p_3}{2}+k,\ \frac{r}{2}\Big)$$

$$K_2(x:\tau)=\textstyle\sum_k\sum_j\sum_\alpha\sum_\beta e^{2\tau}P_0(k:\tau)P_0(j:\tau)(-1)^{j+\alpha}(\alpha+l)^{-1}x^{\alpha+l}$$
$$\times a_0^{-(l+\alpha+\beta)}\binom{p_3/2-1}{\alpha}\binom{\alpha+\beta+l-1}{\beta}\Big[B\Big(l,\ \frac{p_3}{2}\Big)B\Big(\frac{p_2}{2}+k,\ \frac{r}{2}\Big)\Big]^{-1}$$
$$\times J\Big(\frac{a_0}{x+1}:j+\alpha+\beta+1,\ \frac{r}{2}\Big)$$

where $l=q/2+k+j$,

$$I(z:\alpha,\beta)=\int_0^z b(t:\alpha,\beta)dt \quad\text{and}\quad J(z:\alpha,\beta)=\int_1^z t^{\alpha-1}(t-1)^{\beta-1}dt.$$

Hence the null distribution is given by

$$Q(x:0)=I\Big(\frac{x+1}{a_0}:\frac{p_2}{2},\ \frac{r}{2}\Big)$$
$$+\textstyle\sum_\alpha\sum_\beta\binom{p_3/2-1}{\alpha}\binom{\alpha+\beta+q/2-1}{\beta}(-1)^\alpha\Big(\alpha+\frac{q}{2}\Big)^{-1}x^{\alpha+q/2}a_0^{-q/2-\alpha-\beta}$$
$$\times\Big[B\Big(\frac{q}{2},\ \frac{p_3}{2}\Big)B\Big(\frac{p_2}{2},\ \frac{r}{2}\Big)\Big]^{-1}J\Big(\frac{a_0}{x+1}:\alpha+\beta+1,\ \frac{r}{2}\Big)$$

Case (3) $-1\le x<0: Q(x:\tau)=K_1(x:\tau)+K_3(x:\tau)$, where

$$K_3(x:\tau)=\textstyle\sum_k\sum_j\sum_\alpha\sum_\beta e^{2\tau}P_0(k:\tau)P_0(j:\tau)(-1)^{j+\alpha}(\alpha+l)^{-1}(-x)^{\alpha+l}a_0^\beta$$
$$\times\binom{p_3/2-1}{\alpha}\binom{\alpha+\beta+l-1}{\beta}\Big[B\Big(\frac{p_2}{2}+k+\beta,\ \frac{r}{2}\Big)/B\Big(\frac{q}{2},\frac{p_3}{2}\Big)B\Big(\frac{p_2}{2}+k,\frac{r}{2}\Big)\Big]$$
$$\times I\Big(\frac{x+1}{a_0}:\frac{p_2}{2}+k+\beta,\ \frac{r}{2}\Big).$$

Hence the null distribution is given by

$$Q(x:0)=I\Big(\frac{x+1}{a_0}:\frac{p_2}{2},\ \frac{r}{2}\Big)$$
$$+\textstyle\sum_\alpha\sum_\beta(-1)^\alpha\Big(\alpha+\frac{q}{2}\Big)^{-1}(-x)^{\alpha+q_3/2}a_0^\beta\binom{p_3/2-1}{\alpha}\binom{\alpha+\beta+q/2-1}{\beta}$$
$$\times\Big[B\Big(\frac{p_2}{2}+\beta,\ \frac{r}{2}\Big)/B\Big(\frac{p_3}{2},\ \frac{q}{2}\Big)B\Big(\frac{p_2}{2},\ \frac{r}{2}\Big)\Big]I\Big(\frac{x+1}{a_0}:\frac{p_2}{2}+\beta,\ \frac{r}{2}\Big)$$

Case (4) $x<-1 : Q(x:\tau)=0$ and so $Q(x:0)=0$.

For a given α, the cut-off point k_5 is determined by $Q(k_5:0)=1-\alpha$ and the power function of the LBI test is given by $1-Q(k_5:\tau)$ with $\tau=\gamma/2$. On the other hand, the significance probability for an observed value u_0 of U is $1-Q(u_0:0)$.

10.2. *Limiting distributions.* In a general case, let us first observe that when $n_3\to\infty$, the limiting null distribution of the LBI test statistic

$$(10.8)\qquad U_5\equiv \mathrm{tr}(I+T_2)^{-1}[bT_1(I+T_1)^{-1}-p_2I]+n_1p_2-n_1p_2p_3/b$$

is χ^2-distribution with degrees of freedom n_1p_2, i.e., $\chi^2(n_1p_2)$, where $b=n_1+n_3-p_3$. The addition of the correction term $n_1p_1-n_1p_2p_3/b$ is justified below. We first assume $\Sigma=I$ without loss of generality as U_5 is invariant. Then since $\mathrm{plim}_{n_3\to\infty}[V_{33}/n_3]=I$ and since $Z_{13}\sim N(0,I_{n_1}\otimes I_{p_3})$,

$$(10.9)\qquad \mathrm{plim}_{n_3\to\infty}T_2=\mathrm{plim}_{n_3\to\infty}\left(\frac{Z_{13}}{\sqrt{n_3}}\right)\left(\frac{V_{33}}{n_3}\right)^{-1}\left(\frac{Z_{13}}{\sqrt{n_3}}\right)'=0$$

Hence, as $Z_{12}\sim N(0,I_{n_1}\otimes I_{p_2})$ under the null hypothesis and as $Z_{13}\sim N(0,I_{n_1}\otimes I_{p_3})$,

$$(10.10)\qquad \mathrm{plim}_{n_3\to\infty}n_3T_1=\mathrm{plim}_{n_3\to\infty}(I+T_2)^{-1/2}\left[Z_{12}-Z_{13}\left(\frac{V_{33}}{n_3}\right)^{-1}\left(\frac{V_{32}}{n_3}\right)\right]$$
$$\times\left(\frac{V_{22.3}}{n_3}\right)^{-1}\left[Z_{12}-Z_{13}\left(\frac{V_{33}}{n_3}\right)^{-1}\left(\frac{V_{32}}{n_3}\right)\right](I+T_2)^{-1/2}=Z_{12}Z_{12}'$$

Therefore, from $\lim\,(b/n_3)=1$,

$$(10.11)\qquad \mathrm{plim}_{n_3\to\infty}U_5=\mathrm{tr}Z_{12}Z_{12}'$$

which is distributed as $\chi^2(n_1p_2)$.

Under the alternative hypothesis, $Z_{12}\sim N(\Theta_{12},I\otimes\Sigma_{22})$ where Θ_{12} and Σ_{22} depend on n (see Section 2). Here we assume

$$\lim_{n\to\infty}\Theta/\sqrt{n}=\Phi\quad\text{and}\quad \lim_{n\to\infty}\Sigma=\bar{\Sigma}.$$

Note $\bar{\Sigma}\in\mathcal{S}_+(p)$. This assumption is equivalent to saying that $\lim X_1'X_1/n$ is a positive definite constant matrix in the original term. By a similar argument, the limiting distribution of bT_1 under the alternative is the same as that of $\tilde{X}\tilde{X}'$, where $\tilde{X}\sim N(\sqrt{b}\Phi_{12}\bar{\Sigma}_{22.3}^{-1/2}, I_{n_1}\otimes I_{p_2})$. Therefore the limiting behavior of the distribution of U_5 is the same as that of $\tilde{U}_5\equiv\mathrm{tr}\tilde{X}\tilde{X}'$. Since $\mathrm{tr}\tilde{X}\tilde{X}'\sim\chi^2(n_1p_1 : b\mathrm{tr}\Phi_{12}\bar{\Sigma}_{22.3}^{-1}\Phi_{12}')$,

$$\frac{U_5-\mathrm{tr}\Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}'}{[2n_1p_2+4\mathrm{tr}\Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}']^{1/2}}\to N(0,1)\quad\text{as } n \text{ or } b\to\infty.$$

For a given level α, the critical point k_n for U_5 is chosen so that $\alpha=P(U_5>k_n|\Theta_{12}=0)$ for each n. Hence $\alpha=\lim_{n\to\infty}P(U_5>k_n|\Theta_{12}=0)=P(\chi^2_{n,p}>\chi^2_\alpha)$ or $k_n\to\chi^2_\alpha$. Therefore

$$\lim P(U_5>k_n|\Theta_{12}\neq0)=\lim P(\chi^2(n_1p_2 : b\mathrm{tr}\Phi_{12}\bar{\Sigma}_{22.3}^{-1}\Phi_{12}')>\chi^2_\alpha|\Phi_{12}\neq0)=1,$$

implying the LBI test is consistent.

10.3. *Asymptotic null distributions.* For the Lawley-Hotelling test, the Pillai test and the LRT, Fujikoshi (1973) derived the asymptotic expansions of the nonnull distributions up to order n^{-2} in the cases of $\Theta_{12}=O(1)$ and $\Theta_{12}=O(n^{1/2})$. We here simply concern the asymptotic expansions of these tests and the LBI test. The following results are based on Fujikoshi's expressions with $\Theta_{12}=0$ for the former three tests (see also Fujikoshi (1970)). For notation, let

$$f=n_1p_2,\quad \gamma=n_1+p_2+1,\quad \beta=n_1^2+p_2^2-5$$
$$l_0=(3f-8)\gamma^2+4\gamma+4(f+2),\quad l_1=-12f\gamma^2,\quad l_2=6(3f+8)\gamma^2$$
$$l_3=-4[(3f+16)\gamma^2+4\gamma+4(f+2)]\text{ and } l_4=3[(f+8)\gamma^2+4\gamma+4(f+2)].$$

Further, let $G_\alpha(x)$ denotes the distribution function of $\chi^2(\alpha)$.

(1) The Lawley-Hotelling test statistic: $U_2\equiv m_2\mathrm{tr}T_1$ where $m_2=n_3-p_2-p_3-1$.

$$P(U_2\le x)=G_f(x)+\frac{f\gamma}{4m_2}[G_f(x)-2G_{f+2}(x)+G_{f+4}(x)]$$

$$+\frac{f}{96m_2^2}\sum_{\alpha=0}^{4} l_\alpha G_{f+\alpha}(x)+O(m_2^{-3})$$

(2) The Pillai test statistic: $U_3 \equiv m_3 \mathrm{tr} T_1(I+T_1)^{-1}$ where $m_3=n_1+n_3-p_3=b$.

$$(10.12)\quad P(U_3 \leq x)=G_f(x)+\frac{f\gamma}{4m_3}[-G_f(x)+2G_{f+2}(x)-G_{f+4}(x)] + \frac{f}{96m_3^2}\sum_{\alpha=0}^{4} l_\alpha G_{f+\alpha}(x)+O(m_3^{-3})$$

(3) The LRT Statistic : $U_4 \equiv m_4 \log|I+T_1|$ with $m_4=n_3-p_3+(n_1-p_2-1)/2$.

$$P(U_4 \leq x)=G_f(x)+\frac{2f\beta}{96m_4^2}[-G_f(x)+G_{f+4}(x)]+O(m_4^{-3})$$

The choice of the correction factors m_2 and m_3 in U_2 and U_3 is due to Fujikoshi (1973 a), while m_4 is known as the Bartlett correction factor (see also Sugiura and Fujikoshi ((1969)).

Now using the distribution of the Pillai test statistic U_3 above, we derive the asymptotic distribution of the LBI test statistic U_5 in (10.8) up to order n^{-1} or b^{-1}. There the correction factor $n_1p_2-n_1p_2p_3/b$ simplifies the final expression of the asymptotic distribution. Our result is as follows.

Theorem 10.1. The asymptotic null distribution of the LBI test statistic U_5 up to order b^{-1} is given by

$$(10.13)\quad P(U_5 \leq x)=G_f(x)+\frac{f\gamma}{4b}[-G_f(x)+2G_{f+2}(x)-G_{f+4}(x)] + \frac{fp_3}{2b}[G_f(x)-G_{f+2}(x)]+O(b^{-2})$$

where $b=n_1+n_3-p_2$ and $G_f(x)$ is the distribution function of $\chi^2(f)$ with $f=n_1p_2$.

The proof is given in 10.4 below.

It is noted that the first two terms in the right hand side of

(10.13) form the asymptotic null distribution of U_3 up to order b^{-1}. Hence the difference is the term $fp_3[G_f(x)-G_{f+2}(x)]/2b$ (up to order b^{-1}). If $p_3=0$ where T_2 vanishes, the result is naturally reduced to the case of U_3.

It should be remarked that the expansion is formal and the validity is not checked here. On this point, the readers may refer to Bhattacharya and Ghosh (1978) for a general theory, Kariya and Maekawa (1981) for regression and Fujikoshi (1984) for a multivariate case.

10.4. *Proof of Theorem* 10.1. Without loss of generality, $\Sigma=I$ is assumed. Then (10.9) implies

$$T_2=\frac{1}{b}Z_{13}Z'_{13}+o_p(b^{-1}) \quad \text{and}$$

$$(I+T_2)^{-1}=I-T_2+T_2^2(I+T_2)^{-1}=I-\frac{1}{b}Z_{13}Z'_{13}+o_p(b^{-1})$$

as T_2 is nonnegative definite. Hence from (10.8),

$$U_5=\mathrm{tr}(I-\frac{1}{b}Z_{13}Z'_{13}+o_p(b^{-1}))[bT_1(I+T_1)^{-1}-p_2I]+n_1p_2-n_1p_2p_3/b$$
$$=U_3-\frac{1}{b}\mathrm{tr}Z_{13}Z'_{13}[bT_1(I+T_1)^{-1}]+(p_2/b)\mathrm{tr}Z_{13}Z'_{13}-n_1p_2p_3/b+o_p(b^{-1})$$

where $U_3=b\mathrm{tr}T_1(I+T_1)^{-1}$ as before. Using the independence between T_1 and Z_{13} under the null hypothesis, the characteristic function of U_5 is given by

$$(10.14)\quad \psi_5(t)=E\{\exp(itU_3)[1-\frac{it}{b}\mathrm{tr}Z_{13}Z'_{13}(bT_1(I+T_1)^{-1}) + it(p_2/b)\mathrm{tr}Z_{13}Z'_{13}-itn_1p_2p_3/b+o_p(b^{-1})]\}$$
$$=\psi_3(t)-t(p_2/b)E[\exp(itU_3)iU_3]+o_p(b^{-1})$$
$$=\psi_3(t)-t(p_2/b)\frac{\partial}{\partial t}\psi_3(t)+o_p(b^{-1})$$

where $\psi_3(t)=E[\exp(itU_3)]$ and $E(Z_{13}Z'_{13})=p_2I_{n_1}$ was used. On the other hand since the characteristic function of $G_f(x)$ is $(1-2it)^{-f/2}$, the expression (10.12) implies

$$\psi_3(t) = (1-2it)^{-f/2} + \frac{f\gamma}{4b}[-(1-2it)^{-f/2} +2(1-2it)^{-1-f/2}-(1-2it)^{-2-f/2}] + o(b^{-1})$$

Therefore

$$-t(p_2/b)\frac{\partial}{\partial t}\psi_3(t) = (p_2 f/2b)(-2it)(1-2it)^{-1-f/2} + o(b^{-1})$$
$$=(p_2 f/2b)[(1-2it)^{-f/2}-(1-2it)^{-1-f/2}] + o(b^{-1}).$$

Substituting these into (10.14) and inverting it yields the desired result.

11. Monotonicity of the LRT and LBI Test.

11.1. *Monotonicity of the LRT.* In the MANOVA problem, the monotone property of the power functions of Roy's test, Lawley-Hotelling's test and the LRT, which corresponds to those $\mathscr{K}_1, \mathscr{K}_3$ and $\mathscr{K}_4$ defined in Section 7 in our problem, are shown by Das Gupta, Anderson and Mudholkar (1964) based on Lemma 7.1 in Chapter 2. However the monotonicity of the power function of Pillai's test even in the MANOVA problem remains unsolved, though Perlman (1978) proved it in the case that the critical point k_3 is less than 1 or the significance level is large. This implies that the monotonicity of the power function of the LBI test in our problem is harder to establish.

In the GMANOVA problem, Khatri (1966) proved the monotonicity of the conditional power functions of Roy's test $\mathscr{K}_1$, Lawley-Hotelling's test $\mathscr{K}_2$, and the LRT $\mathscr{K}_4$ given T_2, while Fujikoshi (1973) proved it unconditionally. Here using the results in the MANOVA problem we shall see it in a similar way as in Fujikoshi (1973). In Section 4, it has been seen that a reduced version of our model is given by

$$(11.1)\qquad \begin{cases} x \text{ given } T_2 \sim N((I+T_2)^{-1/2}\Theta_{12},\ I_{n_1}\otimes\Sigma_{22.3}), \\ V_{22.3} \sim W_p(\Sigma_{22.3},\ n_3-p_3) \end{cases}$$

where (x, T_2) and $V_{22.3}$ are independent and the marginal distribution of T_2 does not depend on unknown parameters. Since the power function of a $\mathscr{G}$-invariant test is a function of the characteristic roots $\delta_1\geq\ldots\geq\delta_{n_1}\geq 0$ of $\Upsilon=\Theta_{12}\Sigma_{22.3}^{-1}\Theta_{12}'$, as far as the power function is concerned, we can assume without loss of generality that $\Upsilon=\mathrm{diag}\{\delta_1,\ldots,\delta_{n_1}\}$, $\Sigma_{22.3}=I$ and $\Theta_{12}=\mathrm{diag}\{\sqrt{\delta_1},\ldots,\sqrt{\delta_{n_1}}\} : n_1\times p_2$. And since the problem of testing $\Theta_{12}=0$ versus $\Theta_{12}\neq 0$ in the conditional model is exactly the same as the MANOVA problem, it follows that the conditional power function of a $\mathscr{G}$-invariant test based on $T_1=xV_{22.3}^{-1}x'$ alone is a function of the characteristic roots $D(T_2,\Upsilon) =\mathrm{diag}\{d_1(T_2,\Upsilon)\geq \ldots \geq d_{n_1}(T_2,\Upsilon)\geq 0\}$ of

$$(11.2)\qquad H(T_2,\Upsilon) = (I+T_2)^{-1/2}\Upsilon(I+T_2)^{-1/2}$$

(see, e.g., Anderson (1958)). Here it is easily seen from (11.2) that $D(T_2,\Upsilon_1)-D(T_2,\Upsilon_2)$ is nonnegative definite whenever $\Upsilon_1-\Upsilon_2$ is nonnegative definite or $\delta_i^{(1)}\geq\delta_i^{(2)}$ for all i. Therefore if the conditional power function $\pi(\phi,D\,|\,T_2)$ of a $\mathscr{G}$-invariant test ϕ based on T_1 alone is increasing in $d_j(T_2,\Upsilon)$ for each j, then

$$\pi(\phi,\ D(T_2,\Upsilon_1)\,|\,T_2) \geq \pi(\phi,\ D(T_2,\Upsilon_2)\,|\,T_2) \quad (\text{a.e. } T_2)$$

if $\delta_i^{(1)}\geq\delta_i^{(2)}$ for all i. Taking the expectation with respect to T_2 yields the unconditional monotonicity of the power function of ϕ or $\pi(\phi,\Upsilon_1) \geq\pi(\phi,\Upsilon_2)$. This implies that the tests $\mathscr{K}_1$, $\mathscr{K}_2$ and $\mathscr{K}_4$ are of monotonicity property in power because they enjoy the conditional monotonicity we summarize these as

Proposition 11.1. If the power function of a $\mathscr{G}$-invariant test ϕ based on T_1 alone has the conditional monotonicity in $d_i(T_2,\Upsilon)$, it is monotonically increasing in each δ_i. In particular, the power functions of the tests $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_4$ are increasing in each δ_i so that the tests are unbiased, and the power function of Pillai's $\mathcal{K}_3$ is increasing in δ_i if the cut-off point k_3 is less than 1.

The last statement follows from the former part and Perlamn (1973).

11.2. *Monotonicity of the LBI test.* As has been remarked, it is difficult to prove the monotonicity property for the LBI test. Following the same argument as in Perlman (1973), Kariya (1975) obtained the following result though the proof is omitted.

Proposition 11.2. If the significant point k_5 of the LBI test ϕ_5 with critical region

$$a_0 \operatorname{tr}(I+T_2)^{-1}T_1(I+T_1)^{-1} - \operatorname{tr}(I+T_2)^{-1} > k_5$$

satisfies $-n_1 < k_5 \leq -n_1+1$, the power function $\pi(\phi_5, c\delta)$ increases monotonically in c for each fixed $\delta=(\delta_1,\dots,\delta_{n_1})$ where $a_0=(n_1+n_3-p_3)/p_2$ and $c \geq 0$. Hence under the condition, the LBI test is unbiased.

However the condition $-n_1 < k_5 \leq -n_1+1$ implies large values of significance level and so it will not be practical.

12. Admissibility and Robustness

12.1. *Marden's result.* Using the same arguments as in Marden and Perlman (1980), the admissibility of the tests in the GMANOVA problem was studied by Marden (1980, 1983) where Wald's result is used that the set of proper Bayes tests and their weak* limits forms an essentially complete class. To state his result (1980), let

$$T_0 = (I+T_2)^{1/2}(I+T_1)(I+T_2)^{1/2} - I$$

$$(12.1) \qquad = (Z_{12}, Z_{13})\begin{pmatrix} V_{22} & V_{23} \\ V_{32} & V_{33} \end{pmatrix}^{-1}(Z_{12}, Z_{13})',$$

$$\tilde{S} = (I+T_2)^{-1/2}(I+T_1)^{-1/2}T_1(I+T_1)^{-1/2}(I+T_2)^{-1/2} \quad \text{and}$$

$$\tilde{R} = (I+T_0)^{-1/2}T_0(I+T_0)^{-1/2}$$

It is noted that

$$(12.2) \qquad 0 \leq R < I, \quad \operatorname{rank}(R)=q,$$
$$0 \leq \tilde{S} < I, \quad \operatorname{rank}(\tilde{S})=t, \quad 0 < I-\tilde{R}+\tilde{S} < I,$$

where $q=\min(p_2+p_3, n_1)$, $t=\min(n_1, p_2)$, and $A>B$ $(A \geq B)$ means $A-B$ is positive (semi-) definite. Since (T_1, T_2) is homeomorphic to $(\tilde{S}, \tilde{R})$, any $\mathcal{G}$-invariant test is a function of $(\tilde{S}, \tilde{R})$, upon which $\mathcal{G}$ acts by $(\tilde{S}, \tilde{R}) \to (\Gamma\tilde{S}\Gamma', \Gamma\tilde{R}\Gamma')$ where $\Gamma \in \mathcal{O}(n_1)$. Here our problem is to test $\delta=0$ versus $\delta \neq 0$ based on $(\tilde{S}, \tilde{R})$, where $\delta=(\delta_1,\dots,\delta_{n_1})$ and δ_i is the i-th largest characteristic root of $\Upsilon=\Theta_{12}\Sigma_{22\cdot 3}^{-1}\Theta_{12}'$. For each T_0 or $\tilde{R}$, choose $\Gamma_R \in \mathcal{O}(n_1)$ continuously such that

$$(12.3) \qquad R = \Gamma_R \tilde{R} \Gamma_R' = \operatorname{diag}\{r_1,\dots,r_{n_1}\}$$

where r_i is the i-th largest characteristic root of $\tilde{R}$, and let

$$(12.4) \qquad S=\Gamma_R \tilde{S} \Gamma_R', \quad r=(r_1,\dots,r_q) \quad \text{and}$$

$$(12.5) \qquad \mathcal{R}^q = \{a \in R^q \mid a=(a_i),\ 1>a_1> \dots >a_q>0\}.$$

Then $r \in \mathcal{R}^q$ a.e. $(Z_{12}, Z_{13}, \tilde{V})$ where $\tilde{V}=\begin{pmatrix} V_{22} & V_{23} \\ V_{32} & V_{33} \end{pmatrix}$.

Definition 12.1. A set $E \subset \mathcal{R}^q$ is said to be nonincreasing with respect to weak submajorization (abbreviated as w-nonincreasing) if

$$(12.6) \qquad r \in E,\ r' \in \mathcal{R}^q \text{ and } r' \leq_w r \text{ implies } r' \in E,$$

where $\leq_w$ is the partial ordering with $\mathcal{R}^q$ given by

$$r' \leq_w r \quad \text{if} \quad r_1' \leq r_1,\ r_1'+r_2' \leq r_1+r_2,\ \dots,\ \textstyle\sum_{i=1}^q r_i' \leq \sum_{i=1}^q r_i$$

and $r' <_w r$ means (12.6) with the inequalities strict (see Marshall and Olkin (1979)).

Define the class $\mathcal{E}$ to consist of all closed (in $\mathcal{R}^q$), convex and w-nonincreasing subsets of $\mathcal{R}^q$, and let

$$(12.7)\quad \begin{cases} \Psi = \{\delta\in R^{n_1} \mid \delta=(\delta_i),\ \delta_1\geq \dots \geq\delta_{n_1}\geq 0\}, \\ \Psi_0 = \{\delta\in\psi \mid \Delta(\delta)\leq 1\} \ \text{ and } \ \Psi_1=\{\delta\in\psi \mid \Delta(\delta)\geq 1\}, \end{cases}$$

where $\Delta=\Delta(\delta)=\sum_{i=1}^{n_1}\delta_i$. Further for any finite measure π^0 on Ψ_0 and locally finite measure on π^1 on Ψ_1 define

$$(12.8)\quad d(Z_{12},Z_{13},\tilde{V} : \pi^0,\pi^1) = \int_{\Psi_0}[(R_\delta-1)/\Delta]\pi^0(d\delta)+\int_{\Psi_1} R_\delta\pi^1(d\delta).$$

where the local finiteness of π^1 means that π^1 (A)$<\infty$ for any compact set A of Ψ_1. Here R_δ is equal to R_ξ in (8.6) with ξ replaced by

$$(12.9)\qquad \xi_0\equiv \operatorname{diag}\{\sqrt{\delta_1},\dots,\sqrt{\delta_{n_1}}\} : n_1\times p_2.$$

In fact, $R_\xi=R_{\xi_0}$ because of the invariance of ν_1 and $h\mu$ in (8.6) under $\xi\to P\xi Q$ with $(P, Q)\in\mathcal{O}(n_1)\times\mathcal{O}(p_2)$.

Now denote by $\mathcal{B}$ the class of $\mathcal{G}$-invariant tests ϕ of the form

$$(12.10)\quad \phi(x) = \begin{cases} 1 & \text{if } r\notin E \\ 1 & \text{if } d(Z_{12},Z_{13},\tilde{V} : \pi^0,\pi^1)>c \quad \text{a.e. } (Z_{12},Z_{13},\tilde{V}) \\ 0 & \text{otherwise} \end{cases}$$

where $E\in\mathcal{E}$, $|c|<\infty$, and for $r\in E^\circ$(=interior of E), $|d(Z_{12},Z_{13},\tilde{V} : \pi^0\pi^1)|<\infty$. Then we obtain

Theorem 12.1. (Marden (1980 or 1983)) All tests in $\mathcal{B}$ are admissible among $\mathcal{G}$-invariant tests. If $p_2\geq n_1$, $\mathcal{B}$ is the minimal complete class of invariant tests.

It is noted that the LBI test $\mathcal{K}_5$ is admissible in $\mathcal{D}_I$, the class of $\mathcal{G}$-invariant tests, because it is a unique best test in $\mathcal{D}_I$ around $\delta=0$. But by taking point mass at $\delta=0$ as π^0 and $\pi^1=0$ in (12.8), the admissibility also follows from the unique Bayes-ness.

Next to consider the admissibility of $\mathcal{G}$-invariant tests based on T_1 alone, let

$$(12.11)\qquad \phi = \begin{cases} 1 & \text{if } \operatorname{ch}(T_1)=(\operatorname{ch}_1(T_1),\dots,\operatorname{ch}_t(T_1))\notin A \\ 0 & \text{otherwise} \end{cases}$$

where $\operatorname{ch}_i(T_1)$ denotes the i-th largest characteristic root of T_1 and $A\subseteq\mathcal{R}^t$ is decreasing in the sense that $a=(a_i)\in A$, $a^*=(a_i^*)\in\mathcal{R}^t$ and $a_i^*\leq a_i$ for all i imply $a^*\in A$. The LRT $\mathcal{K}_4$, Roy's test $\mathcal{K}_1$, Lawley-Hotelling's test $\mathcal{K}_2$ and Pillai's test $\mathcal{K}_3$ are all of this form.

Theorem 12.2. (1) (Marden and Perlman (1980)) Let $n_1=1$. A size α test ϕ of the form (12.11) is admissible among $\mathcal{G}$-invariant tests if and only if $\alpha\leq\alpha^*$, where $\alpha^*=P((1+n_3-p_3)T_1/p_2>1)$.
(2) (Marden (1983)) Let $n_1>1$. No nontrivial test of the form (3.1) is not in $\mathcal{B}$. Hence if $p_2\geq n_1>1$, any such test is inadmissible.

The result (2) implies that all the tests $\mathcal{K}_i$'s (i=1,...,4) are inadmissible when $p_2\geq n_1>1$.

12.2. *Robustness.* In the framework described in Section 9 of Chapter 2, the robustness of the tests in the GMANOVA problem was studied by Kariya (1981) and Kariya and Sinha (1984). Recall that $\mathcal{F}_L(M)$ denotes the class of $n\times p$ dimmensional left $\mathcal{O}(n)$-invariant distributions about $M\in R^{n\times p}$ such that $P(\mathcal{X})=1$ for $P\in\mathcal{F}_L(M)$ and $\mathcal{F}_E(M, I_n\otimes\Sigma)$ denotes the class of $n\times p$ dimmensional elliptically symmetric distributions about M with scale matrix $\Sigma\in\mathcal{S}_+(p)$ such that $P(\mathcal{X})=1$ for $P\in\mathcal{F}_E(M, I_n\otimes\Sigma)$, where $\mathcal{X}=\{X : n\times p \mid \operatorname{rank}(X)=p\}$. Clearly

$$\mathcal{F}_E(M, I_n\otimes\Sigma)\subset\mathcal{F}_L(M) \quad \text{for all } M : n\times p \text{ and } \Sigma\in\mathcal{S}_+(p)$$

If $P\in\mathcal{F}_E(M, I_n\otimes\Sigma)$ has a density, it is expressed as

$$(12.12)\quad f(X|M,\Sigma) = |\Sigma|^{-n/2}q(\operatorname{tr}\Sigma^{-1}(X-M)'(X-M)) \quad (X\in\mathcal{X})$$

where $q : [0,\infty) \to [0,\infty)$, while if $P \in \mathcal{F}_L(M)$ has a density, it is expressed as

$$f(X|M) = q((X-M)'(X-M)) \tag{12.13}$$

where $q : \mathcal{S}_+(p) \to [0,\infty)$. Of course, $\mathcal{F}_E(M, I_n \otimes \Sigma)$ contains $N(M, I_n \otimes \Sigma)$.

Theorem 12.3. (1) The LBI test, the LRT, Roy's test, Lawley-Hotelling's test and Pillai's test are null robust or their null distributions remain the same for all $P \in \cup \{\mathcal{F}_E(\Theta) | \Theta$ is of the form (2.6)}.

(2) (Kariya and Sinha (1984)) The LBI test under normality is still LBI in the class of distributions whose densities are of the form (12.12) and satisfy

$$\int_{\mathcal{A}} (\mathrm{tr} AA')^{i/2} |\bar{q}^{(i)}(\mathrm{tr} AA')| \mu(dA) < \infty \quad (i=1,2,3)$$

$$\bar{q}^{(3)}(x) \le 0 \text{ and } \bar{q}^{(3)} \text{ is nondecreasing,}$$

where $\bar{q}^{(i)}(x) = d^i \bar{q}(x)/dx^i$, $\bar{q}$ is the marginal density of $(Z_{12}, Z_{13}, Z_{23}, Z_{32}, Z_{33})$ under $\Theta_{12}=0$ and $\Sigma=I$, $\mathcal{A}=\{A=(A_{ij})\ i,j=2,3 | \bar{A} \in \mathcal{G}l(p_2+p_3),\ A_{32}=0\}$, and μ is an invariant measure on $\mathcal{A}$.

Proof. (1) follows from Corollary 9.1 in Chapter 2. In fact, the conditions (i)' and (ii)' are easily checked since our group $\mathcal{G}$ contains as a subgroup $\mathcal{G}\mathcal{U}(p)$ acting on Z by $Z \to ZC'$ where $C \in \mathcal{G}\mathcal{U}(p)$ and $\mathcal{G}\mathcal{U}(p)$ denotes the group of $p \times p$ nonsingular upper triangular matrices. The proof of (2) is omitted here.

Chapter 4

EXTENDED GMANOVA PROBLEMS

1. Introduction

1.1. *Summary.* In Chapter 1, a GMANOVA model with general linear restrictions on the coefficient matrix was called an extended GMANOVA model and the problem of testing a general linear restriction in an extended GMANOVA model was called an extended GMANOVA problem. A typical example of an extended GMANOVA model is an SUR model (see Section 1 of Chapter 1) and so the problem of testing on the coefficient vector of, say, the first equation in the SUR model is an extended GMANOVA problem. Other examples are also found in Section 1 of Chapter 1. In this Chapter, we concern some extended GMANOVA problems.

In Section 2, a systematic approach to an extended GMANOVA problem is taken under an assumption on the general linear restrictions and regression matrices. This assumption makes it possible to have a similar canonical form as we obtained in Chapter 3 for the GMANOVA problem, so that a group of a similar form as in the GMANOVA problem leaves the canonical form invariant. In fact, Banken (1984) gave an LBI (locally best invariant) test in the canonical form by using a correspondence between the extended GMANOVA prbblem and the GMANOVA problem while the LRT (likelihood ratio test) had been derived in the canonical form by Gleser and Olkin (1970). Some examples listed in Chapter 1 which satisfy the assumption required are treated and the LBI tests are given.

In Section 3, a problem of testing on means with incomplete data is shown to be an extended GMANOVA problem and the problem is treated from our point of view, though it can be handled in a usual way. In this case, the assumption made in Section 2 is not satisfied but a UMPI (uniformly most powerful invariant) test exists.

In Section 4, in a two equation SUR model we treat the problem of testing the hypothesis that a part of the coefficient vector of the first equation is zero. Of course, if we ignore the second equation which is correlated with the first equation, the usual F-test is UMPI as is well known. But taking into account the second equation, no LBI test in general exists, much less UMPI test. Hence the F-test is no longer UMPI in the SUR model. However, in a particular case that the regression matrix of the first equation is a submatrix of that of the second equation, it is shown to be UMPI. A reduction of the problem to a canonical form is made independently of the way adopted in Section 2 because the assumption made in Section 2 is not satisfied.

In Section 5, the GMANOVA problem is treated under some covariance structure. Hence it is not an extended GMANOVA problem, but the argument is rather similar to the arguments in this chapter. First, under intra-class covariance structure, the LRT (likelihood ratio test) is shown to be UMPI if a condition on the regression matrix is satisfied and an example in Chapter 1 is analyzed. Second, under Rao's covariance structure, a GMANOVA problem is shown to be a MANOVA problem. Finally, in the growth curve model treated in Example 1.2 of Chapter 3 where the covariance matrix is of certain structure, the GMANOVA problem there is also shown to be a MANOVA problem.

2. An extended GMANOVA Problem.

2.1. *Canonical form.* In this section, we shall treat the extended GMANOVA problem described in Section 2 of Chapter 1. Let

$$Y = X_1 B X_2 + E, \quad E \sim N(0, I_n \otimes \Omega) \tag{2.1}$$

be a GMANOVA model where X_1 is an $n \times k$ matrix of rank k and X_2 is a $q \times p$ matrix of rank q. In the extended GMANOVA problem, it is assumed that the following prior information is available on the coefficient B:

$$X_3 B X_4 = X_0 \tag{2.2}$$

where X_3 is an $m_3 \times k$ known matrix of rank m_3 and X_4 is a $q \times r_4$ known matrix of rank r_4. And the problem is to test the hypothesis

$$H: X_5 B X_6 = X_7 \tag{2.3}$$

where X_5 is an $m_5 \times k$ known matrix of rank m_5 and X_6 is a $q \times r_6$ known matrix of rank r_6. Without loss of generality, we may assume $X_0 = 0$ and $X_7 = 0$. Since a general treatment of the problem (2.3) under the restriction (2.2) is difficult, we here make a crucial assumption for the analysis below.

Assumption 2.1. The matrices X_3, X_5, X_4 and X_6 satisfy

$$M_3 M_5 = M_5 M_3 \quad \text{and} \quad M_4 M_6 = M_6 M_4$$

where

$$M_i = (X_1'X_1)^{-1/2} X_i' [X_i (X_1'X_1)^{-1} X_i']^{-1} X_i (X_1'X_1)^{-1/2} \quad (i=3, 5) \quad \text{and}$$
$$M_j = (X_2 X_2')^{-1/2} X_j [X_j' (X_2 X_2')^{-1} X_j]^{-1} X_j' (X_2 X_2')^{-1/2} \quad (j=4, 6)$$

This assumption implies that there exists an orthogonal matrix which diagonalizes M_i and M_{i+2} simultaneously ($i=3,4$). Banken (1984) associated this kind of assumption with the estimability of (2.2) and (2.3) though his assumption is slightly different.

Now to derive a canonical form of the problem, we simply deal with the cases (I) $M_3M_5=M_5$ and $M_4M_6=M_4$, (II) $M_3M_5=M_5$ and $M_4M_6=0$, (III) $M_3M_5=M_3$ and $M_4M_6=M_6$ and (IV) $M_3M_5=0$ and $M_4M_6=M_6$. A more general case is dealt with in a similar manner and in fact briefly treated later.

Case (I) : $M_3M_5=M_5$ and $M_4M_6=M_4$. This assumption means that the row space of $X_5(X_1'X_1)^{-1/2}$ is included in the row space of $X_3(X_1'X_1)^{-1/2}$ and the column space of $(X_2X_2')^{-1/2}X_4$ is included in the column space of $(X_2X_2')^{-1/2}X_6$. In other words, X_5 and X_4 are respectively nested to X_3 and X_6 relative to $(X_1'X_1)^{-1}$ and $(X_2X_2')^{-1}$. Clearly $m_5 \leq m_3$ and $r_4 \leq r_6$ are implied. Let

$$X_1 = P_1\begin{bmatrix} I_k \\ 0 \end{bmatrix}F_1, \quad P_1 \in \mathcal{O}(n),\ F_1 \in \mathcal{G}l(k) \tag{2.4}$$
$$X_2 = F_2[I_q\ \ 0]P_2, \quad F_2 \in \mathcal{G}l(q),\ P_2 \in \mathcal{O}(p)$$

Then, as has been observed in Section 2 of Chapter 3,

$$Y^* = P_1' Y P_2' = \begin{array}{c} \begin{pmatrix} Y_1^* \\ Y_2^* \end{pmatrix} \end{array}\!\!\begin{array}{l} k \\ n-k \end{array} \sim N\left(\begin{pmatrix} B^* & 0 \\ 0 & 0 \end{pmatrix},\ I_n \otimes \Omega^*\right) \tag{2.5}$$

where $B^*=F_1BF_2$ and $\Omega^*=P_2'\Omega P_2$. From

$$X_iBX_{i+1} = X_iF_1^{-1}B^*F_2^{-1}X_{i+1} \quad (i=3,5)$$

Under the assumption, there exist $P_0 \in \mathcal{O}(k)$, $Q_0 \in \mathcal{O}(q)$, $F_i \in \mathcal{G}l(m_i)$ and $F_{i+1} \in \mathcal{G}l(r_{i+1})$ ($i=3,5$) such that

$$X_3F_1^{-1} = F_3(0, I_{m_3})P_0, \quad X_5F_1^{-1} = F_5(\underset{k-m_3}{0},\ \underset{m_3-m_5}{0},\ I_{m_5})P_0$$

$$F_2^{-1}X_4 = Q_0\begin{pmatrix} 0 \\ I_{r_4} \end{pmatrix}F_4, \quad F_2^{-1}X_6 = Q_0\begin{pmatrix} 0 \\ I_{r_6} \end{pmatrix}F_6 \tag{2.6}$$

Here letting $\Theta = P_0B^*Q_0$, from (2.5) and (2.6)

$$Z \equiv \begin{pmatrix} P_0 & 0 \\ 0 & I \end{pmatrix} Y^* \begin{pmatrix} Q_0 & 0 \\ 0 & I \end{pmatrix} \sim N\left(\begin{pmatrix} \Theta & 0 \\ 0 & 0 \end{pmatrix},\ I_n \otimes \Sigma\right). \tag{2.7}$$

Define

$$\tilde{\Theta} \equiv \begin{pmatrix} \Theta & 0 \\ 0 & 0 \end{pmatrix} = \begin{array}{c} \begin{array}{cccc} p_1 & p_2 & p_3 & p_4 \end{array} \\ \begin{pmatrix} \Theta_{11} & \Theta_{12} & \Theta_{13} & 0 \\ \Theta_{21} & \Theta_{22} & \Theta_{23} & 0 \\ \Theta_{31} & \Theta_{32} & \Theta_{33} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{l} \\ n_1 \\ n_2 \\ n_3 \\ n_4 \end{array} \tag{2.8}$$

where $n_1=k-m_3$, $n_2=m_3-m_5$, $n_3=m_5$ and $n_4=n-k$ so that $\Sigma n_i=n$, and $p_1=q-r_6$, $p_2=r_4$, $p_3=r_6-r_4$ and $p_4=p-q$ so that $\Sigma p_i=p$. Then from (2.6) the prior restriction (2.2) becomes

$$\Theta_{23}=0 \quad \text{and} \quad \Theta_{33}=0 \tag{2.9}$$

and the hypothesis (2.3) becomes $H: \Theta_{32}=0$ and $\Theta_{33}=0$. Hence from (2.9), we test

$$H: \Theta_{32}=0 \tag{2.10}$$

Case (II) : $M_3M_5=M_5$ and $M_4M_6=0$. Then in a similar way as in the case (I), there exists $P_0 \in \mathcal{O}(k)$ and $Q_0 \in \mathcal{O}(q)$ such that (2.6) holds with $F_2^{-1}X_6=Q_0\binom{0}{I}F_6$ replaced by

$$F_2^{-1}X_6 = Q_0GF_6 \quad \text{with} \quad G = \begin{pmatrix} 0 \\ I \\ 0 \end{pmatrix}\begin{array}{l} q-r_4-r_6 \\ r_6 \\ r_4 \end{array}$$

In this case, we also obtain (2.7) and (2.8) where n_i's are the same as before, but $p_1=q-r_4-r_6$, $p_2=r_4$, $p_3=r_6$ and $p_3=p-q$. And the prior restriction (2.2) remains unchanged as in (2.9) but the hypothesis (2.3) simply becomes

$$H: \Theta_{32}=0 \tag{2.10}$$

Therefore in the both cases the hypothesis to be tested is the hypothesis H in (2.10) in the model (2.7) with (2.8) and (2.9). Gleser and Olkin (1970) directly extended the canonical form of the GMANOVA problem to this canonical form (not via (2.2) and (2.3)) and derived the LRT (likelihood ratio test)

$$\frac{\left|I+(Z_{33}Z_{34})\begin{pmatrix} V_{33} & V_{34} \\ V_{43} & V_{44} \end{pmatrix}^{-1}\begin{pmatrix} Z'_{33} \\ Z'_{34} \end{pmatrix}\right|}{\left|I+(Z_{32}Z_{33}Z_{34})\begin{pmatrix} V_{22} & V_{23} & V_{24} \\ V_{32} & V_{33} & V_{34} \\ V_{42} & V_{43} & V_{44} \end{pmatrix}^{-1}\begin{pmatrix} Z'_{32} \\ Z'_{33} \\ Z'_{34} \end{pmatrix}\right|} < c \tag{2.11}$$

where $n-k\geq p$ is assumed and

$$Z=\begin{array}{c} \begin{array}{cccc} p_1 & p_2 & p_3 & p_4 \end{array} \\ \begin{pmatrix} Z_{11} & Z_{12} & Z_{13} & Z_{14} \\ Z_{21} & Z_{22} & Z_{23} & Z_{24} \\ Z_{31} & Z_{32} & Z_{33} & Z_{34} \\ Z_{41} & Z_{43} & Z_{43} & Z_{44} \end{pmatrix} \end{array} \begin{array}{c} \\ n_1 \\ n_2 \\ n_3 \\ n_4 \end{array}. \tag{2.12}$$

$$V\equiv(V_{ij})=(Z'_{i4}Z_{4j}) : p\times p \quad (i,j=1,\dots,4)$$

2.2. *Reduction via invariance and LBI test.* Following Banken (1984), let

$$\mathcal{P} = \{Q=\begin{pmatrix} Q_1 & & 0 \\ & \ddots & \\ 0 & & Q_4 \end{pmatrix} | Q_i\in\mathcal{O}(n_i),\ i=1,\dots,4\}$$

$$\mathcal{A} = \{A=(A_{ij})\in\mathcal{G}l(p) \,|\, A_{ij} : p_i\times p_j,\ A_{ij}=0 \text{ for } i>j,\ i,j=1,\dots,4\}. \tag{2.13}$$

and

$$\mathcal{F} = \{F=(F_{ij}) : n\times p \,|\, F_{ij} : n_i\times p_j,\ F_{ij}=0 \text{ for } i+j=k,\ k=5,\dots,8\}.$$

Then the group $\mathcal{G}=\mathcal{P}\times\mathcal{A}\times\mathcal{F}$ leaves the problem invariant under the action

$$g(Z) = PZA' + F \quad \text{and} \quad g(\tilde{\Theta},\Sigma) = (P\tilde{\Theta}A'+F,\ A\Sigma A')$$

where $g=(P,A,F)\in\mathcal{G}$. Then as has been shown in the case of the GMANOVA problem, it is shown that the group is maximal in the general affine linear group of homomorphisms leaving the problem invariant (see Section 3 of Chapter 2). Define the subgroup of $\mathcal{G}$

$$\mathcal{H} = \{g\in\mathcal{G} \,|\, g=(P,A,F),\ P=I\} = \{I\}\times\mathcal{A}\times\mathcal{F} \tag{2.14}$$

The following lemma is similar to Lemma 2.1.

Lemma 2.1. (1) A maximal invariant statistic under $\mathcal{H}$ is $S=(S_1,S_2,S_3)$ with

$$S_1=s_1(Z) = (Z_{32},Z_{33},Z_{34})\begin{pmatrix} V_{22} & V_{23} & V_{24} \\ V_{32} & V_{33} & V_{34} \\ V_{42} & V_{43} & V_{44} \end{pmatrix}^{-1}(Z_{32},Z_{33},Z_{34})'$$

$$S_2=s_2(Z) = \begin{pmatrix} Z_{23} & Z_{24} \\ Z_{33} & Z_{34} \end{pmatrix}\begin{pmatrix} V_{33} & V_{34} \\ V_{43} & V_{44} \end{pmatrix}^{-1}\begin{pmatrix} Z_{23} & Z_{24} \\ Z_{33} & Z_{34} \end{pmatrix}' \quad \text{and}$$

$$S_3=s_3(Z) = \begin{pmatrix} Z_{14} \\ Z_{24} \\ Z_{34} \end{pmatrix} V_{44}^{-1} \begin{pmatrix} Z_{14} \\ Z_{24} \\ Z_{34} \end{pmatrix}'.$$

(2) A maximal invariant parameter under $\mathcal{H}$ is $\Upsilon=\Theta_{32}\Sigma_{22.34}^{-1}\Theta'_{32}$ with

$$\Sigma_{22.34}=\Sigma_{22}-(\Sigma_{23}\,\Sigma_{24})\begin{pmatrix} \Sigma_{33} & \Sigma_{34} \\ \Sigma_{43} & \Sigma_{44} \end{pmatrix}^{-1}\begin{pmatrix} \Sigma_{32} \\ \Sigma_{42} \end{pmatrix}. \tag{2.15}$$

Here define

$$\begin{cases} T_1=t_1(Z)=x\,V_{22.34}^{-1}x' \\ T_2=t_2(Z)=(0,I_{n_3})S_2(0,I_{n_3})' \end{cases} \tag{2.16}$$

where

$$(2.17)\quad x=(I+T_2)^{-1/2}\left[Z_{32}-(Z_{33}, Z_{34})\begin{bmatrix} V_{33} & V_{34} \\ V_{43} & V_{44} \end{bmatrix}^{-1}(V_{23}, V_{24})'\right]$$

and $V_{22.34}$ is defined in the same way as $\Sigma_{22.34}$. Then analogous to Lemma 4.2 of Section 4 in Chapter 3, we obtain

Theorem 2.1. (1) The statistic $T=(T_1, T_2)$ is sufficient for the family of distributions of S.
(2) Conditional on T_2, $x \sim N((I+T_2)^{-1/2}\Theta_{32}, I\otimes\Sigma_{22.34})$.

Under this theorem, the rest of the argument is completely similar to those of the GMANOVA problem. In particular, a $\mathcal{G}$-invariant test is described as a $\mathcal{H}$-invariant test satisfying

$$(2.18)\quad \phi(t_1(gZ), t_2(gZ)) = \phi(P_3t_1(Z)P_3', P_3t_2(Z)P_3')$$

for $g\in\mathcal{G}$ and $P_3\in\mathcal{O}(n_3)$, and the LRT in (2.11) is written as

$$(2.19)\quad |I+T_1| > c$$

and the LBI test is given by

$$(2.20)\quad W_5 \equiv \mathrm{tr}(I+T_2)^{-1}[bT_1(I+T_1)^{-1}-p_2I] > c$$

where $b=n_3+n_4-p_3-p_4$

To obtain the null distribution of W_5, let

$$(2.21)\quad U_5 \equiv W_5 + n_3p_2 - n_3p_2(p_3+p_4)/b$$

Then the null distribution of U_5 is by Theorem 10.1 in Chapter 3 given by

$$(2.22)\quad P(U_5\leq x) = G_f(x) + \frac{f\gamma}{4b}[-G_f(x)+2G_{f+2}(x)-G_{f+4}(x)] + \frac{f(p_3+p_4)}{2b}[G_f(x)-G_{f+2}(x)] + O(b^{-2}),$$

where $f=n_3p_2$, $\gamma=n_3+p_2+1$ and $G_f(x)$ denotes the χ^2-distribution

function with degrees of freedom f.

We first remark that when $n_3=1$ in the above cases (I) and (II), the test with critical region $T_1>c$ is equivalent to the LRT and it is UMPI (uniformly most powerful invariant) in the class of conditional level α tests (given T_2). Further $(1+n_4-p_2-p_3-p_4)T_1/p_2$ is distributed as $F(p_2, 1+n_4-p_2-p_3-p_4 : \lambda)$ with $\lambda=\gamma/2(1+T_2)$, where $F(\alpha,\beta:\lambda)$ denotes F-distribution with degrees of freedom α and β and noncentral parameter λ. This result is parallel to the result in Section 7 of Chapter 3. An example of the case $n_3=1$ is found in Example 2.2 below. Secondly, in Section 7 of Chapter 3, it is observed that the GMANOVA problem with $X_2=I_p$ is essentially the MANOVA problem even if rank $(X_4)=r_4<q=p$. However, in the extended GMANOVA problem, this is not the case. This is because in the latter problem, the prior information $\Theta_{33}=0$ is available (see the cases (I) and (II)). Hence in the case $p_2=1$, the LRT is no longer UMPI, which is in contrast to the GMANOVA case with $p_2=1$. Thirdly, the local minimaxity of the LBI test also follows from Theorem 9.1 in Chapter 3.

Case (III) $M_3M_5=M_3$ and $M_4M_6=M_6$. This case is parallel to the case (I) and in the same way as before, the prior restriction (2.2) becomes in (2.8)

$$(2.23)\quad \Theta_{32}=0 \quad\text{and}\quad \Theta_{33}=0$$

and the hypothesis (2.3) becomes $\Theta_{23}=0$ and $\Theta_{33}=0$ or

$$(2.24)\quad H: \Theta_{23}=0$$

where $n_1=k-m_5$, $n_2=m_5-m_3$, $n_3=m_3$, $n_4=n-k$, $p_1=q-r_4$, $p_2=r_6$, $p_3=r_4-r_6$ and $p_4=p-q$. Hence in this case, the problem is to test (2.24) in the model (2.7) with (2.8) and (2.23), Or from an invariance point of view, it is equivalent to the problem of testing $\Theta_{23}=0$ in the reduced model

$$\begin{pmatrix} Z_{22} & Z_{23} & Z_{24} \\ \tilde{Z}_{32} & \tilde{Z}_{33} & \tilde{Z}_{34} \end{pmatrix} \sim N\left(\begin{pmatrix} \Theta_{22} & \Theta_{23} & 0 \\ 0 & 0 & 0 \end{pmatrix}, I_n \otimes \tilde{\Sigma}\right), \quad \tilde{\Sigma} = (\Sigma_{ij}) \ (i, j = 2, 3, 4)$$

where $\tilde{Z}_{i3} = \begin{pmatrix} Z_{i3} \\ Z_{i4} \end{pmatrix}$ $(i=2,3,4)$ and Z_{ij}'s are given in (2.12). This is a reduced form of the GMANOVA problem treated in Chapter 3.

Case (IV) $M_3M_5=0$ and $M_4M_6=M_6$. This case is parallel to the case (II), and the prior restriction (2.2) becomes $\Theta_{32}=0$ and $\Theta_{33}=0$ and the hypothesis becomes $\Theta_{23}=0$, where $n_1=k-m_3-m_5$, $n_2=m_5$, $n_3=m_3$, $n_4=n-k$, $p_1=q-r_4$, $p_2=r_6$, $p_3=r_4-r_6$ and $p_4=p-q$. Hence this case is equivalent to the case (III).

It is noted that in the case of $M_3M_5=M_5$ and $M_4M_6=M_6$, the prior restriction (2.2) contains the hypothesis (2.3) so that no testing problem occurs.

2.3. *Examples.* The following example is a repetition and continuation of Example 1.5 in Chapter 1.

Example 2.1. (Banken (1984)). In a biological investigation the effects of thyroxin and thiouracil on the growth of young rats were studied. For this purpose, 27 rats were randomly assigned to three groups; 10 to the first and third groups and 7 to the second group. The first group was kept as a control, while the second and the third were given thyroxin and thiouracil respectively. The weight of each animal was measured at the beginning of the experiment and then in four consecutive weeks. The data can be found in Box (1950). Let x_{ijt} be the weight of the j-th individual in the t-th week, where $t=0,\ldots,4$, $j=1,\ldots,n_i$, and $i=1,2,3$ with $n_1=10$, $n_2=7$, and $n_3=10$. The vectors $x_{ij}=(x_{ij0},\ldots,x_{ij4})'$ are assumed to be independently normally distributed

$$x_{ij} \sim N(\mu_i, \Sigma) \quad (\mu_i \in R^5,\ \Sigma \in \mathcal{S}_+(5))$$

where $\mu_i=(\mu_{i0},\ldots,\mu_{i4})'$ and $\mu_{it}=a_{i0}+a_{i1}t+a_{i2}t^2$ $(t=0,\ldots,4)$.

As the rats were randomly assigned, the expected weight of the rats should be equal at the beginning of the experiment:

$$a_{10}=a_{20}=a_{30}. \tag{2.25}$$

And we want to test if the expected growth curves are equal, i.e.,

$$H: \mu_1=\mu_2=\mu_3 \quad \text{or} \quad a_{1j}=a_{2j}=a_{3j} \ (j=0,1,2) \tag{2.26}$$

Example 1.5. in Chapter 1 shows that the model is written as

$$Y = X_1BX_2 + E, \quad E \sim N(0,\ I_{27} \otimes \Sigma),$$

where $B: 3\times 3$, the prior knowledge (2.25) is written as $X_3BX_4=0$ with

$$X_3 = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \quad \text{and} \quad X_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

and the hypothesis (2.24) is written as $X_5BX_6=0$ with $X_5=X_3$ and $X_6=I$. Therefore, clearly the Assumption in the Case (I) is fulfilled. In this case, $n_1=1$, $n_2=0$, $n_3=2$, $n_4=24$, $p_1=0$, $p_2=2$, $p_3=1$ and $p_4=2$, and the values of T_1 and T_2 in (2.16) is computed as

$$T_1 = \begin{bmatrix} 1.484880 & 0.535907 \\ 0.535907 & 0.315512 \end{bmatrix}$$

$$T_2 = \begin{bmatrix} 0.0120777 & 0.0360276 \\ 0.0360276 & 0.0364151 \end{bmatrix}$$

Banken (1984) reported that the hypothesis is rejected at level 1% when the LRT, Lawley-Hotelling's trace test or Roy's maximum trace test are used, and that the LBI test also rejected the hypothesis at 1% level where the null distribution of the LBI test is simulated with 1,000 random samples. One may use the asymptotic null distribution up to order n^{-1} of the LBI test statistic U_5 in (2.22), U_5 is computed as $U_5=17.61$ for this data and so from

(2.22) with $b=23$, $f=4$ and $r=5$

$$P(U_5\leq 17) = -0.0019586+O(0.0018903)$$

where the tables in Abramowitz and Stegun (1975) are used for χ^2-distribution. Though the negative value of $P(U_5\leq 17)$ up to order n^{-1} is not necessarily appropriate, the result will support Banken's experiment as $O(b^{-2})=O(0.0018903)$.

Example 2.2. (See Examples 1.4 and 1.5 in Chapter 1) Let $x_i=(x_{1i}', x_{2i}')' : (p+q)\times 1$ and $y_j=(y_{1j}', y_{2j}')' : (p+q)\times 1$ be random sample from $N(\mu,\Sigma)$ and $N(\eta,\Sigma)$ respectively ($i=1,\dots,n$; $j=1,\dots m$), where $\mu=(\mu_1', \mu_2)'$ with $\mu_2 : q\times 1$ and $\eta=(\eta_1', \eta_2')'$ with $\eta_2 : q\times 1$. Here let

$$Y'=(x_1,\dots,x_n,y_1,\dots,y_m) : (p+q)\times(n+m),$$

$$X_1=\begin{pmatrix} e_1 & 0 \\ 0 & e_2 \end{pmatrix} : (n+m)\times 2 \quad \text{and} \quad B=\begin{pmatrix} \mu_1' & \mu_2' \\ \eta_1' & \eta_2' \end{pmatrix} : 2\times(p+q)$$

Then the model is expressed as

$$Y = X_1B + E, \quad E\sim N(0, I_{n+m}\otimes\Sigma). \tag{2.27}$$

Assume that we know $\mu_2=\eta_2$ or

(2.28) $X_3BX_4=0$ with $X_3=(1, -1) : 1\times 2$ and $X_4=\begin{pmatrix} 0 \\ I_q \end{pmatrix} : (p+q)\times q$.

Under this condition, we wish to test

(2.29) $X_5BX_6=0$ with $X_5=(1, -1) : 1\times 2$ and $X_6=\begin{pmatrix} I_p \\ 0 \end{pmatrix} : (p+q)\times p$.

As has been stated in Section 1 of Chapter 1, this is the problem treated by Cochran and Bliss (1948), Rao (1949), Kariya and Kanazawa (1978) etc., in association with discriminant analysis with covariates, and the above formulation shows that the problem is not a GMANOVA problem but an extended GMANOVA problem. Further it is easy to see that the Assumption in the Case (II) is

satisfied, since $M_3=M_5$ holds from $X_3=X_5$ and $M_4M_6=0$ holds from $X_2=I$. Hence the test based on W_5 in (2.20) is LBI. As has been remarked, though $X_2=I$, there exists no UMPI test even when $n_3=1$ or $p_2=1$. The LRT is conditional UMPI when $n_3=1$.

The admissiblity of the tests in this problem is studied by Marden and Perlman (1980), and it is shown that the LRT is not admissible if the significance level is large and that the test based on the studentization of an efficient estimator is inadmissible.

2.4. *General case.* In 2.2 above, four special cases in which Assumption 2.1 is satisfied were treated. Even for a general case, the procedure of the reduction is similar. In fact, by Assumption 2.1, in place of (2.6) there exist $P_0\in\mathcal{O}(k)$ and $Q_0\in\mathcal{O}(q)$ such that

$$X_3F_1^{-1} = F_3(0,I_{m_3})P_0, \quad X_5F_1^{-1} = F_5(\overset{k-s}{0}, \overset{m_5}{I}, \overset{s-m_5}{0})P_0$$

$$F_2^{-1}X_4 = Q_0\begin{pmatrix} 0 \\ I_{r_4} \end{pmatrix}F_4, \quad F_2^{-1}X_6=Q_0HF_6 \quad \text{with} \quad H=\begin{array}{c} \begin{pmatrix} 0 \\ I \\ 0 \end{pmatrix} \end{array}\begin{array}{l} q-t \\ r_6 \\ t-r_6 \end{array}$$

These expressions lead to the following decomposition of $\tilde{\Theta}$:

$$\tilde{\Theta}=\begin{array}{c} \begin{array}{ccccc} p_1 & p_2 & p_3 & p_4 & p_5 \end{array} \\ \begin{pmatrix} \Theta_{11} & \Theta_{12} & \Theta_{13} & \Theta_{14} & 0 \\ \Theta_{21} & \Theta_{22} & \Theta_{23} & \Theta_{24} & 0 \\ \Theta_{31} & \Theta_{32} & 0 & 0 & 0 \\ \Theta_{41} & \Theta_{42} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}\begin{array}{l} \\ n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \end{array} \tag{2.30}$$

and the model is $Z\sim N(\tilde{\Theta},I_n\otimes\Sigma)$. Here the prior restriction (2.2) is equivalent to $\Theta_{ij}=0$ $(i,j=3,4)$, which is incorporated in $\tilde{\Theta}$, while the hypothesis (2.3) is equivalent to

$$H: \Theta_{22}=0,\ \Theta_{23}=0,\ \Theta_{32}=0 \text{ and } \Theta_{33}=0,$$

where

$$\begin{cases} n_1=k-s, \ n_2=s-m_3, \ n_3=m_3+m_5-s, \ n_4=s-m_5, \ n_5=n-k \\ p_1=q-t, \ p_2=t-r_4, \ p_3=r_4+r_6-t, \ p_4=t-r_6, \ p_5=p-q. \end{cases}$$

If n_i(or p_j)≤ 0, set it to be zero. The above four cases we have treated correspond to the situation where either $s=m_3$ or $s=m_5$ and either $t=r_4$ or r_6, which is equivalent to the situation where $s\leq \max(m_3,m_5)$ and $t\leq \max(r_4,r_6)$. Hence the situation we have not treated is the one where $s>\max(m_3,m_5)$ and/or $t>\max(r_4,r_6)$. These cases are left out here because of the difficulty.

3. Testing on Means With Incomplete Data.

3.1. *Problem.* In this section, we shall transform the problem of testing on means of normal population with missing data into the extended GMANOVA problem treated in Section 2. The model considered here is

$$(3.1) \qquad \begin{aligned} \tilde{Z} &\sim N(e_n\mu',\ I_n\otimes\Sigma) \\ \tilde{W}_2 &\sim N(e_{m_2}\mu_2',\ I_{m_2}\otimes\Sigma_{22}) \end{aligned}$$

where $\tilde{Z}: n\times p$ and $\tilde{W}_2: m_2\times p_2$ are independent, $e_j=(1,\ldots,1)'\in R^j$, $p_1+p_2=p$ and $\mu=(\mu_1',\mu_2')': p\times 1$ with $\mu_2: p_2\times 1$. Under this model, the following problem is considered;

$$(3.2) \qquad H: \mu_2=0 \quad \text{versus} \quad K: \mu_2\neq 0.$$

Under the model (3.1), we have n p-dimmensional observation vectors which are random samples from $N_p(\mu,\Sigma)$ and m_2 p_2-dimmensional observation vectors which are random samples from $N(\mu_2,\Sigma_{22})$.

In the last decade, methods for statistical inference with missing or extra data have been extensively investigated. The articles by Hartly and Hocking (1971), and Kariya, Krishnaiah and Rao (1983) provide overviews of the subject and extensive bibliographies (see also Section 2 of Chapter 5). Most methods are however

proposed by intuitive or ad hoc approaches and not much attention is paid to their optimal properties. Some common features of previous work in this field are as follows: (1) normal models are assumed, (2) some simple patterns for missing or extra data are assumed and (3) repeated samples from a specific pattern are assumed.

For the problem (3.2) under consideration, Bhargava (1962, 1975) derived the LRT (likelihood ratio test) and Eaton and Kariya (1983) showed that it is UMPI. In this section, we show that the problem is a particular case of the extended GMANOVA problem and the same result is obtained. Incidentally, for the problem of testing $\mu=0$ in the model (3.1), Bhargava (1975) derived the LRT, and Morrison and Bhoj (1973) investigated the behavior of the power of the LRT numerically. In Cohen (1977), it is shown that in the case of $p_1=p_2=1$, Hotelling's T^2-test ignoring the extra data $\tilde{W}_2$ is admissible, while in Eaton and Kariya (1983) it is shown that there exists no LBI (locally best invariant) test. In addition, for the problem of testing $\mu_1=0$ in the model (3.1), it is shown in Eaton and Kariya (1983) that there exists no LBI test, contrary to the problem (3.2). In the case of $p_1=p_2=1$, Khatri, Bhargava and Shah (1974), Little (1976) and Sarkar (1979) treated the problem. Our purpose here is to treat the problem (3.2) in association with the extended GMANOVA problem.

Now to incorporate this problem into the framework of an extended GMANOVA problem, let $P\in\mathcal{O}(n)$ and $Q\in\mathcal{O}(m_2)$ such that $Pe_n=(\sqrt{n},0,\ldots,0)'$ and $Qe_{m_2}=(\sqrt{m_2},0,\ldots,0)'$, and define

$$(3.3) \qquad Z = P\tilde{Z} = \begin{array}{cc} p_1 & p_2 \\ \end{array} \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix}\begin{array}{l} 1 \\ n-1 \end{array} \sim N\left(\begin{pmatrix} n^{1/2}\mu_1', & n^{1/2}\mu_2' \\ 0 & 0 \end{pmatrix},\ I_n\otimes\Sigma\right)$$

and

p_2

$$(3.4)\qquad W_2 = Q\tilde{W}_2 = \begin{array}{c} \begin{pmatrix} W_{12} \\ W_{22} \end{pmatrix} \end{array}\begin{array}{l} 1 \\ m_2-1 \end{array} \sim N\left(\begin{pmatrix} m_2^{1/2}\mu_2' \\ 0 \end{pmatrix}, I_{m2}\otimes\Sigma_{22}\right).$$

An idea is to introduce a dummy random matrix W_1^* whose elements represent missing variables on the first p_1 coordinates or the counter part of W_2 as follows:

$$(3.5)\qquad (W_1^*, W_2) = \begin{pmatrix} W_{11}^* & W_{12} \\ W_{21}^* & W_{22} \end{pmatrix} \sim N\left(\begin{pmatrix} \eta' & m_2^{1/2}\mu_2' \\ \Delta & 0 \end{pmatrix}, I_n\otimes\Sigma\right)$$

where $\eta : p_1\times 1$ and $\Delta : (m_2-1)\times p_1$. Note that η and Δ are taken to be independent of μ. Then

$$(3.6)\quad Y = \begin{array}{c} \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \\ Y_{31} & Y_{32} \\ Y_{41} & Y_{42} \end{pmatrix} \\ \begin{array}{cc} p_1 & p_2 \end{array} \end{array} \begin{array}{l} 1 \\ 1 \\ m_2-1 \\ n-1 \end{array} = \begin{pmatrix} Z_{11} & Z_{12} \\ W_{11}^* & W_{12} \\ W_{21}^* & W_{22} \\ Z_{21} & Z_{22} \end{pmatrix} \sim N\left(\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \\ 0 & 0 \end{pmatrix}, I_{n+m_2}\otimes\Sigma\right)$$

where

$$B_{11}=n^{1/2}\mu_1',\quad B_{12}=n^{1/2}\mu_2',\quad B_{21}=\eta'$$
$$B_{22}=m_2^{1/2}\mu_2',\quad B_{31}=\Delta \text{ and } \quad B_{32}=0.$$

That is, using the dummy matrix W_1^*, the model (3.1) is formalized as

$$(3.7)\qquad Y = X_1B + E \quad\text{with}\quad E \sim N(0,\ I_{n+m_2}\otimes\Sigma)$$

where

$$X_1 = \begin{pmatrix} I_{m_2+1} \\ 0 \end{pmatrix} : (n+m_2)\times(m_2+1) \quad\text{and}\quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{pmatrix} : (m_2+1)\times p.$$

Further, the prior information : $B_{12}=cB_{22}$ and $B_{32}=0$ with $c=(m_2/n)^{1/2}$ is also formalized as

$$(3.8)\qquad X_3BX_4 = 0$$

where

$$(3.9)\qquad X_3 = \begin{pmatrix} -c & 1 & 0 \\ & 0\ 0 & I_{m_2-1} \end{pmatrix} : m_2\times(m_2+1) \quad\text{and}\quad X_4 = \begin{pmatrix} 0 \\ I_{p2} \end{pmatrix}$$

And the hypothesis $H: \mu_2=0$ in (3.2) is expressed as

$$(3.10)\qquad X_5BX_6=0 \quad\text{with}\quad X_5=(0,1,0) : 1\times(1+m_2) \text{ and } X_6=X_4,$$

Thus, it truns out that the introduction of the dummy matrix W_1^* enables us to put the problem in the framework of an extended GMANOVA problem. But the dummy variable is eventually by invariance deleted as will be seen below.

3.2. *Testing* (3.10). In the above, it is shown that the problem of testing $\mu_2=0$ is a special case of the extended GMANOVA problem treated in Section 2. However, the Assumption 2.1 that enables the extended GMANOVA problem to correspond to the GMANOVA problem does not hold because $M_3M_5\neq M_5M_3$, where in the present case $M_i=X_i'(X_iX_i')^{-1}X_i'$ $(i=3,5)$ because of the form of X_1. Hence the result in Section 2 is not applicable here. To directly analyze the problem (3.10) under the restriction (3.9), let

$$(3.11)\qquad P = \begin{pmatrix} a & -ca & 0 \\ ca & a & 0 \\ 0 & 0 & I_{m_2-1} \end{pmatrix} : (m_2+1)\times(m_2+1), \quad \left(a=\frac{1}{1+c^2}\right).$$

Then $P\in\mathcal{O}(m_2+1)$ and from (3.8) and (3.9)

$$(3.12)\qquad 0 = X_3BX_4 = X_3PP'BX_4 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & I \end{pmatrix}\Theta\begin{pmatrix} 0 \\ I \end{pmatrix} = \begin{pmatrix} \Theta_{22} \\ \Theta_{32} \end{pmatrix}$$

where

$$(3.13)\qquad \Theta = \begin{pmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \\ \Theta_{31} & \Theta_{32} \end{pmatrix} = \begin{pmatrix} a(B_{11}+cB_{21}) & a(B_{12}+cB_{22}) \\ a(-cB_{11}+B_{21}) & 0 \\ B_{31} & 0 \end{pmatrix} = P'B$$

Note that Θ_{12} does not depend on the dummy parameters η and Δ. Now letting $U=\begin{pmatrix} P' & 0 \\ 0 & I \end{pmatrix} Y$, the model (3.7) with (3.9) becomes

$$(3.14) \qquad U=(U_{ij}) \sim N\left(\begin{pmatrix} \Theta \\ 0 \end{pmatrix}, I \otimes \Sigma\right) \quad \text{with} \quad \Theta_{22}=0 \quad \text{and} \quad \Theta_{32}=0,$$

where $U_{ij}: n_i \times p_j$ $(i=1,\dots,4: j=1,2)$ with $n_1=n_2=1$, $n_3=m_2-1$ and $n_4=n-1$. On the other hand, the hypothesis (3.10) becomes

$$0=X_5BX_6=X_5P\Theta X_6=(ca,a,0)\Theta X_6=ca\Theta_{12}$$

because $\Theta_{22}=0$ from (3.12). Consequently the problem is now to test

$$H: \Theta_{12}=0$$

under the model (3.14) and it is slightly different from the GMANOVA problem in Chapter 3 and the extended GMANOVA problem in Section 2. But applying the invariance principle, the problem is reduced to the MANOVA problem. In fact, under the translation group

$$\mathcal{F}=\{F=(F_1, 0): (n+m_2)\times p \mid F_1'=(F_{11}', F_{12}', F_{13}', 0'),\ F_1: (n+m_2)\times p_1\}$$

acting on U by $U \to U+F$ for $F \in \mathcal{F}$, the model (3.14) is reduced to

$$U_2=(U_{12}', U_{22}', U_{32}', U_{42}')' \sim N((\Theta_{12}', 0', 0', 0')', I_{n+m_2} \otimes \Sigma_{22})$$

and there the problem of testing $\Theta_{12}=0$ is a special case of the MANOVA problem. Hence noting $n_1=1$, the following result is well known (see Section 7 of Chapter 3).

Theorem 3.1. The test with critical region

$$(3.15) \qquad T \equiv U_{12}(U_{22}'U_{22}+U_{32}'U_{32}+U_{42}'U_{42})^{-1}U_{12}' > c$$

is uniformly most powerful invariant (UMPI) for testing $\Theta_{12}=0$ under the model (3.14). The null distribution of $(n+m_2-1)T$ is

$F(1, n+m_2-1)$, i.e., F-distribution with degrees of freedom 1 and $n+m_2-1$.

It is remarked that the test (3.15) is the LRT under the model (3.14) and that the group here leaving the problem invariant is $\mathcal{G}=\mathcal{Gl}(p_1)\times\mathcal{Gl}(p_2)\times\mathcal{F}$ acting on U by $U \to UQ+F$ with $Q=\begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}$ where $(Q_1, Q_2, F)\in\mathcal{G}$.

In Eaton and Kariya (1983), the result is directly shown based on the original model (3.1). In the original term of (3.1), the test (3.15) is expressed as

$$T=\left(\frac{n}{n+m}\right)^2(\bar{z}_2+c\bar{w}_2)\left[\left(\frac{n}{n+m}\right)^2(\bar{w}_2-c\bar{z}_2)'(\bar{w}_2-c\bar{z}_2) + S_{22}+V_{22}\right]^{-1}(\bar{z}_2+c\bar{w}_2)',$$

where $c=(m_1/n)^{1/2}$, $e_n'\tilde{Z}/n=(\bar{z}_1, \bar{z}_2)$ with $\bar{z}_2: 1\times p_2$, $e_{m_2}'\tilde{W}_2/m_2=\bar{w}_2$, $S_{22}=\tilde{W}_2'[I-e_{m_2}e_{m_2}'/m_2]\tilde{W}_2$ and $V=Z'[I-e_ne_n'/n]Z=\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$ with $V_{22}: p_2 \times p_2$.

4. Testing on Regression Coefficients in a Two equations SUR Model

4.1. *Problem.* In Section 1 of Chapter 1, it is observed that an SUR (seemingly unrelated regression) model is regarded as an extended GMANOVA model. In this section, we treat a problem of testing on the regression coefficients in a two equations SUR model, which is an extended GMANOVA problem. Let

$$(4.1) \qquad y_i=\tilde{X}_i\beta_{ii}+\varepsilon_i, \quad E(\varepsilon_i)=0, \quad E(\varepsilon_i\varepsilon_j')=\sigma_{ij}I_n \ (i=1,2)$$

be an SUR model of two equations, where $\tilde{X}_i: n \times k_i$ is of rank k_i and normality is assumed for error terms ε_i's. Then as has been shown in Chapter 1, the model is put in a form of multivariate regression;

$$Y = XB + E, \quad E \sim N(0, I_n \otimes \Sigma) \tag{4.2}$$

where $n \geq k_1 + k_2$

$$Y = [y_1, y_2] : n\times 2, \quad X = [\tilde{X}_1, \tilde{X}_2] : n\times(k_1+k_2)$$

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} : (k_1+k_2)\times 2 \quad \text{with} \quad \beta_{12}=0 \text{ and } \beta_{21}=0 \tag{4.3}$$

$$E = [\varepsilon_1, \varepsilon_2] : n\times 2, \quad \Sigma = (\sigma_{ij}) : 2\times 2.$$

It is noted that the rank of X is not necessarily of full rank. The prior information $\beta_{12}=0$ and $\beta_{21}=0$ on the coefficient matrix B is expressed as

$$X_3 B X_4 = 0 \quad \text{and} \quad X_5 B X_6 = 0 \tag{4.4}$$

where

$$X_3 = [I_{k_1}, 0] : k_1\times(k_1+k_2), \quad X_4 = (0, 1)' : 2\times 1 \tag{4.5}$$

$$X_5 = [0, I_{k_2}] : k_2\times(k_1+k_2), \quad X_6 = (1, 0)' : 2\times 1 \tag{4.6}$$

Hence the model (4.2) with (4.4) is an extended GMANOVA model and under this model, we consider the problem of testing the hypothesis

$$H : \beta_{111}=0 \quad \text{versus} \quad K : \beta_{111} \neq 0 \tag{4.7}$$

where β_{111} is the $l\times 1$ vector of the first l coefficients of β_{11} or $\beta_{11} = [\beta'_{111}, \beta'_{112}]'$. The hypothesis H in (4.7) is clearly expressed as

$$H : X_7 B X_8 = 0 \quad \text{where} \tag{4.8}$$

$$X_7 = [I_l, 0, 0] : l\times(k_1+k_2) \text{ and } X_8 = (1, 0)' : 2\times 1.$$

In particular, $l=1$ corresponds to the hypothesis that a coefficient in the first equation is zero. Of course, this problem is an extended GMANOVA problem described in Chapter 1.

For the problem of testing (4.7) or (4.8), we obtain the F-test (or t-test when $l=1$) based on the first equation in (4.1) alone,

where the information on the nonzero correlation with the second equation is ignored. As is well known, if the correlated second equation does not exist, the F-test is a UMPI (uniformly most powerful invariant) test. However, an optimality of the F-test should be evaluated in the present model (4.2). The existence of the correlated second equation implies that a GLSE (generalized least square estimator) of β_{11} based on the combined model (4.2) is more efficient than the OLSE (ordinary LSE) based on the first equation alone, when $\sigma_{ij} \neq 0$ (see Chapter 1). In this section, it is questioned whether an LBI test exists and under what a condition the F test (or t-test when $l=1$) has an optimality.

4.2. *Canonical form and invariance.* To make the structure of the problem clear, we shall obtain a canonical form of the problem (not via the procedure in Section 2). Let

$$\begin{aligned} N_j &= I_n - \tilde{X}_j(\tilde{X}'_j\tilde{X}_j)^{-1}\tilde{X}'_j \quad (j=1,2) \\ N_0 &= I - X(X'X)^+X' \\ R_j &= X(X'X)^+X' - \tilde{X}_j(\tilde{X}'_j\tilde{X}_j)^{-1}\tilde{X}'_j \quad (j=1,2) \\ r_j &= \text{rank}(X) - k_j \quad (j=1,2) \end{aligned} \tag{4.9}$$

where $(X'X)^+$ denotes the Penrose generalized inverse. Further let L_0 be an $n\times q_0$ matrix satisfying

$$\begin{cases} N_0 = L_0 L'_0, \quad L'_0 L_0 = I_{q_0} \\ q_0 = n - \text{rank}(X) \end{cases} \tag{4.10}$$

and let H_j be an $n\times r_j$ matrix such that

$$H_j H'_j = R_j, \quad H'_j H_j = I_{r_j} \quad (j=1,2). \tag{4.11}$$

In addition, let

$$L_j = [L_0, H_j] : n\times q_j, \quad q_j = n - k_j \quad (j=1,2) \tag{4.12}$$

and observing $L_j'\tilde{X}_j = 0$, we set

$$P_j = [\tilde{X}_j(\tilde{X}_j'\tilde{X}_j)^{-1},\ L_j] : n\times n \tag{4.13}$$

$$P_j'y_j = \begin{pmatrix} (\tilde{X}_j'\tilde{X}_j)^{-1}\tilde{X}_j'y_j \\ L_0'y_j \\ H_j'y_j \end{pmatrix} = \begin{pmatrix} b_{jj} \\ w_j \\ u_j \end{pmatrix} \begin{matrix} k_j \\ q_0 \\ r_j \end{matrix} \tag{4.14}$$

where $j=1,2$. Then a canonical form of the model (2.1) is

$$W=(w_1, w_2) : q_0\times 2 \sim N(0, I\otimes\Sigma)$$

$$c = \begin{pmatrix} b_{11} \\ b_{22} \\ u_1 \\ u_2 \end{pmatrix} : (k_1+k_2+r_1+r_2)\times 1 \sim N\left(\begin{pmatrix} \beta_{11} \\ \beta_{22} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\right) \tag{4.15}$$

W and c are independent.

where

$$\Omega_{11} = \begin{pmatrix} \sigma_{11}(\tilde{X}_1'\tilde{X}_1)^{-1} & \sigma_{12}(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'\tilde{X}_2(\tilde{X}_2'\tilde{X}_2)^{-1} \\ \sigma_{21}(\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'\tilde{X}_1(\tilde{X}_1'\tilde{X}_1)^{-1} & \sigma_{22}(\tilde{X}_2'\tilde{X}_2)^{-1} \end{pmatrix} : (k_1+k_2)\times(k_1+k_2),$$

$$\Omega_{12} = \begin{pmatrix} 0 & \sigma_{12}(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'H_2 \\ \sigma_{21}(\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'H_1 & 0 \end{pmatrix} : (k_1+k_2)\times(r_1+r_2),$$

$$\Omega_{22} = \begin{pmatrix} \sigma_{11}I_{r_1} & \sigma_{12}H_1'H_2 \\ \sigma_{21}H_2'H_1 & \sigma_{22}I_{r_2} \end{pmatrix} : (r_1+r_2)\times(r_1+r_2),$$

and $\Omega_{21}=\Omega_{12}'$. Here letting the translation group $R^{k_1-l}\times R^{k_2}$ act on (b_{112}, b_{22}) where $b_{11}=(b_{111}', b_{112}')'$, the model (4.15) is reduced by invariance to

$$W \sim N(0, I_{q_0}\otimes\Sigma)$$

$$d = \begin{pmatrix} b_{111} \\ u_1 \\ u_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_{111} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}G_{11} & 0 & \sigma_{12}F_{12} \\ 0 & \sigma_{11}I & \sigma_{12}H_1'H_2 \\ \sigma_{21}F_{21} & \sigma_{21}H_2'H_1 & \sigma_{22}I \end{pmatrix}\right). \tag{4.16}$$

where V and d are independent

$$(\tilde{X}_1'\tilde{X}_1)^{-1} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \begin{matrix} l \\ k_1-l \end{matrix} \tag{4.17}$$

$$F_{21}' = F_{12} = (G_{11}\ G_{12})\tilde{X}_1'H_2 = (G_{11}\ G_{12})\tilde{X}_1'H_2H_2'H_2$$
$$= (G_{11}\ G_{12})[\tilde{X}_1' - \tilde{X}_1'\tilde{X}_2(\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2']H_2$$
$$H_1'H_2 = H_1'H_1H_1'H_2H_2'H_2$$
$$= H_1'[X(X'X)^{+}X' - \tilde{X}_1(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1' - \tilde{X}_2(\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'$$
$$+ \tilde{X}_1(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'\tilde{X}_2(\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2']H_2.$$

In this model, the problem of testing $\beta_{111}=0$ is left invariant under group $\mathscr{G}=(R_*)^2\times\mathscr{O}(q_0)$ acting on (5.16) by

$$b_{111}\to a_1b_{111},\quad u_i\to a_iu_i,\quad w_i\to a_iPw_i\ (i=1,2) \tag{4.18}$$

where $(a_1, a_2, P)\in\mathscr{G}$ and $R_*=\{x\in R\,|\,x\neq 0\}$. Arguing as in Section 5 of Chapter 3 and deriving the distribution of a maximal invariant under this group through Wijsman's (1967) theorem, we obtain

Theorem 4.1. No LBI test under $\mathscr{G}$ exists.

The proof is omitted. But a remark to be made is related to the following relation

$$\frac{dP^T_{(\beta_{111},\rho)}}{dP^T_{(0,0)}} = \frac{dP^T_{(\beta_{111},\rho)}}{dP^T_{(0,\rho)}}\,\frac{dP^T_{(0,\rho)}}{dP^T_{(0,0)}} \tag{4.19}$$

where $P^T_{(\beta_{111},\rho)}$ is the distribution of a maximal invariant $T=T(b_{111}, u_1, u_2, w_1, w_2)$ under (β_{111}, ρ) and $\rho=\sigma_{12}/(\sigma_{11}\sigma_{22})^{1/2}$. Note that under the transformation (4.18), the distribution depends on (β_{111}, Σ) only through (β_{111}, ρ). Now (4.19) denotes the density of T under (β_{111}, ρ) with respect to $P^T_{(0,0)}$ which is obtained by the ratio of the integrals over $\mathscr{G}$ as has been seen in the case of the GMANOVA problem in Chapter 3. Under the null hypothesis $\beta_{111}=0$, the