AD-A037 530 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
ESTIMATION AND TESTS FOR UNKNOWN LINEAR RESTRICTION IN MULTIVAR--ETC(U)
DEC 76 J D HEALY AF-AFOSR-2350-72
UNCLASSIFIED MIMEOGRAPH SER-471 AFOSR-TR-77-0175 NL

1 OF 2
AD
A037530

AFOSR - TR - 77 - 0175

ADA037530

AD
DDC FILE COPY

Approved for public release;
distribution unlimited.

PURDUE UNIVERSITY

DDC
RECEIVED
MAR 31 1977
C

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

(4)

ESTIMATION AND TESTS FOR UNKNOWN LINEAR RESTRICTIONS IN MULTIVARIATE LINEAR MODELS*

by

John Douglas Healy
Purdue University

DDC
RECEIVED
MAR 31 1977
C

Department of Statistics
Division of Mathematical Sciences

Mimeograph Series #471

December 1976

*Research supported in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR-72-2350 at Purdue University. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright violation hereon.

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

INTRODUCTION

Almost every statistician has used simple linear regression many times; it probably is the most well-used statistical procedure. If there is more than one dependent variable present, we enter into the realm of multivariate regression. In both univariate and multivariate regression, we can estimate regression coefficients, find confidence intervals for the regression coefficients, and test whether the regression coefficients are equal to a known matrix. However another kind of problem exists in multivariate regression, but does not exist in univariate regression. In multivariate regression, the regression coefficient matrix may not be of full row rank, i.e., there may exist unknown linear restrictions on the regression coefficient matrix. We may want to estimate the regression coefficient matrix and the unknown linear restrictions under the hypothesis that the linear restrictions do exist. For instance, when we estimate one linear restriction, we usually are trying to find the linear combination of the elements of each column of the regression coefficient matrix which equal some unknown quantity.

We now define precisely the model and hypothesis to which we have been referring:

$$x_i = \Xi f_i + e_i, \qquad i = 1,2,\ldots,N,$$

$$B\Xi = \alpha a,$$

where x_i is a p-dimensional vector of observations, Ξ is the unknown pxk ($k \geq p$) regression coefficient matrix, f_i is a k-dimensional vector of dependent variables, e_i is a p-dimensional error vector, B is a rxp($r \leq p$) matrix of linear restrictions, α is an unknown rxs($s \leq r$) matrix which provides a basis for the space spanned by the columns of $B\Xi$, and a is a known sxk matrix. The matrix form of the above equations is

(0.0.1) $$X = \Xi F + E,$$

(0.0.2) $$B\Xi = \alpha a,$$

where

$$X = (x_1, x_2, \ldots, x_N),$$

$$F = (f_1, f_2, \ldots, f_N),$$

$$E = (e_1, e_2, \ldots, e_N).$$

T. W. Anderson [1951a] found the maximum likelihood estimators (MLE's) of the parameters B, Ξ, and Σ when a is the zero matrix. Later, Villegas [1961] found the MLE's of B, Ξ, Σ, and α in the above model when F is the design matrix associated with the MANOVA model and when B is a lxp matrix. Villegas's model can be called the single linear functional relationship model with replications (Moran [1971], Madansky [1959]). When F is the design matrix associated with the MANOVA model, each column of Ξ is the mean

vector for a group of observations. In many cases the number of groups increases when the sample size increases. This situation is itself a special case of the more general case where the number of parameters increases as the sample size increases. Villegas does discuss the consistency of his estimators when the number of groups increases with the sample size.

In Chapter 1, we estimate the parameters in the model and hypothesis specified by (0.0.1) and (0.0.2). We also give several special cases of our model, including several models which resemble a model discussed by Gleser and Watson [1973]. Our discussion of the consistency of the estimators is directed mainly to cases when the number of parameters does not stay fixed as the sample size increases.

One of the biggest advantages of getting maximum likelihood estimators is that we can usually use these estimators in deriving likelihood ratio tests. For many multivariate problems, the exact distribution of the likelihood ratio test statistic is exceedingly complicated. However the asymptotic distribution of $-2 \log \Lambda$, where Λ is the likelihood ratio test statistic is usually a chi-square distribution. In Chapter 2, we use the estimators we derived in Chapter 1 to get the likelihood ratio test statistic for testing

$$H_0: B\Xi = \alpha a \quad \text{versus} \quad H_1: B\Xi \neq \alpha a.$$

Since the exact distribution of this statistic is intractable, we find its asymptotic distribution. Our results show that the asymptotic distribution of the test statistic depends on how the

number of parameters increases with the sample size. It is noteworthy that in several cases, -2 log Λ,where Λ is the likelihood ratio test statistic,does not have an asymptotic chi-square distribution.

The basic model discussed in the first two chapters is commonly called the classical multivariate linear regression model. Another type of linear model, which has been discussed in the literature, is the "growth curves" model (Cochran and Bliss [1948], Shrikhande [1954], and Gleser and Olkin [1964, 1969]). In this model we observe N independent pxl column vectors x_i: i = 1,2,...,N, which satisfy

$$x_i = F\Xi + e_i,$$

where F is a known pxq matrix, Ξ is an unknown q-dimensional vector and e_i is a p-dimensional error vector. This model has been generalized by Gleser and Olkin [1966] in their discussion of k sample growth curves.

All these models, the classical multivariate linear model and the growth curves models, can be generalized to a model first discussed by Potthoff and Roy [1964] and later by Rao [1965] and Gleser and Olkin [1969]. We may write the model which we refer to as the Potthoff-Roy model in the following way:

(0.0.3) $X = F_1 \Xi F_2 + E$

where X is a cxN matrix of observations, F_1 and F_2 are known cxp ($p \leq c$) and mxN ($m \leq N$) matrices respectively, Ξ is an unknown pxm matrix, and E is a cxN error matrix. Each column of E is distributed independently with mean vector 0 and unknown covariance matrix Σ.

Potthoff and Roy [1964] gave ad hoc tests of the hypothesis

(0.0.4) $\quad F_3 \Xi F_4 = \xi_0$

where F_3, F_4 and ξ_0 are known rxp ($r \leq p$), mxk ($k \leq m$), and rxk matrices respectively. F_1 and F_4 are assumed to have full column rank, and F_2 and F_3 are assumed to have full row rank. Rao [1965] found the conditional likelihood ratio test of the hypothesis stated above, and Gleser and Olkin [1969] showed that Rao's conditional likelihood ratio test is actually the unconditional likelihood ratio test.

In Chapter 3, we work with the Potthoff-Roy model (0.0.3) and estimate parameters under a hypothesis similar to (0.0.4). The hypothesis we discuss is concerned with unknown linear restrictions on the regression coefficient matrix. This hypothesis can be written the following way:

(0.0.5) $\quad U_1 \Xi F_4 = \alpha b,$

where U_1 is an unknown rxp ($r \leq p$) matrix, F_4 is a known mxk matrix, α is an unknown rxs matrix, and b is a known sxk matrix. We assume that the unknown covariance matrix Σ has the form $\sigma^2 \cdot I_c$ where σ^2 is an unknown. In Chapter 3 we reduce the Potthoff-Roy model and the above hypothesis (0.0.5) to a canonical form. We also find the MLE's of the parameters in the general model (0.0.3), (0.0.5) and in the reduced model. As in Chapter 1, we discuss consistency of the estimators when the number of parameters is allowed to increase with the sample size.

Chapter 4 bears the same relationship to Chapter 3 that Chapter 2 bears to Chapter 1. In Chapter 4, we derive the likelihood ratio test statistic for testing

$$H_0: U_1 \equiv F_4 = \alpha b \quad \text{versus} \quad H_1: U_1 \equiv F_4 \neq \alpha b.$$

We find the asymptotic distributions of the likelihood ratio test statistic; these depend on how the number of parameters increases with the sample size. In several cases, the asymptotic distribution is not the usual chi-square distribution.

CHAPTER I

ESTIMATION OF UNKNOWN LINEAR RESTRICTIONS ON THE PARAMETERS OF THE CLASSICAL MULTIVARIATE LINEAR REGRESSION MODEL

1.0 Introduction

In this chapter, we discuss estimation of the parameters of the classical multivariate linear regression model (Anderson [1958; Chapter 8]) when an hypothesis concerned with unknown linear restrictions on the parameters is assumed to be true. Section 1.1 contains derivation of the maximum likelihood estimators (MLE's) of the parameters; while Section 1.2 derives consistency properties of the MLE's. We show that some of the estimators are not consistent when the number of parameters in the model increases with the sample size. Several special cases of our model are discussed in Section 1.3 including the multivariate linear functional model (Madansky [1959], Moran [1971], Sprent [1969], Villegas [1961]), and models proposed by Kristoff [1973] and Rao [1973]. In all of our special cases, the independent variables in the regression model are dummy variables.

1.1 Maximum Likelihood Estimation

Let our model be:

(1.1.1) $$x_i = \equiv f_i + e_i, \quad i = 1,2,\ldots,N,$$

where each x_i is a p-dimensional vector of dependent variables, each f_i is a k-dimensional vector of independent variables or covariates ($k \geq p$), Ξ is an unknown pxk parameter matrix of regression coefficients, and the e_i's are p-dimensional vectors of errors.

We assume that the e_i's are statistically independent of one another, and have the same normal distribution with mean vector 0 and unknown covariance matrix Σ. We will be finding the maximum likelihood estimators (MLE) of Σ, Ξ and two other matrices B and α which satisfy,

(1.1.2) $$B \Xi = \alpha a,$$

where a is a known sxk matrix ($s \leq k$) ($k-s \geq p$), B is an unknown rxp ($r \leq p$) matrix and α is an unknown rxs($s \leq r$) matrix. We are concerned with cases in which either a has full row rank or a is the zero matrix, i.e., we are testing $B \Xi = 0$. It should be noted that if a is not the zero matrix and is not full row rank, we can reparametrize so that our resulting matrix will be full row rank. We derive the MLE's of the parameters when a is full row rank. Since the proof is similar (actually easier) when a is the zero matrix, we will merely state the results in this case. In all of our special cases (see Section 1.3), a = (1,1,...,1) or a is the zero matrix.

Anderson [1951a] considered the above problem when a is the zero matrix. His derivation of the MLE's uses Lagrange multipliers and differentiation of the likelihood function. A derivation, similar to the one we give when a has full row rank, could be used as an alternative method of obtaining and verifying the MLE's in Anderson's problem. We believe that that derivation would be simpler and more

intuitive than Anderson's. Since we would not employ differentiation, we would not have to worry about saddle points, etc. In his paper, Anderson [1951a] also gives methods of generating confidence intervals and likelihood ratio tests of various hypotheses.

Our computations will be simplified greatly if we write (1.1.1) in the following way:

(1.1.3) $$X = \equiv F + E,$$

where

$$X = (x_1, x_2, \ldots, x_N),$$

$$F = (f_1, f_2, \ldots, f_N),$$

$$E = (e_1, e_2, \ldots, e_N).$$

We will call X the observation matrix, F the covariate matrix and E the error matrix. We will assume that F and a have full row rank.

Maximizing the likelihood with respect to many parameters can be done in several ways. One way is to: 1) fix one of the parameters (i.e. treat one of the parameters as fixed or given); 2) maximize the likelihood with respect to the other parameters (note: the derived MLE's of the other parameters will be functions of the fixed parameter); 3) substitute the derived MLE's of the other parameters back into the likelihood; and finally 4) maximize the likelihood with respect to the parameter that had been fixed. We will be following this method, with B treated as the fixed parameter.

Part I. B fixed or given

We now transform X into a form in which the proper estimators of the parameters are easy to see. Let C be a p-rxp matrix which satisfies $CC' = I_{p-r}$ and $CB' = 0$. Let

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} BX \\ CX \end{pmatrix} . \tag{1.1.4}$$

Each column of Z is distributed independently with a p-dimensional normal distribution having covariance matrix

$$\Psi = \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix} = \begin{pmatrix} B\Sigma B' & B\Sigma C' \\ C\Sigma B' & C\Sigma C' \end{pmatrix} .$$

The mean of Z is

$$E(Z) = \begin{pmatrix} B\Xi F \\ C\Xi F \end{pmatrix} = \begin{pmatrix} \alpha a F \\ C\Xi F \end{pmatrix} .$$

Let

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} (F'(FF')^{-\frac{1}{2}}, L),$$

where L is a NxN-k matrix which satisfies $L'L = I_{N-k}$ and $L'F = 0$. Note that

$$\begin{aligned} E(Y) &= E[\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} (F'(FF')^{-\frac{1}{2}}, L)], \\ &= \begin{pmatrix} \alpha a F \\ C\Xi F \end{pmatrix} (F'(FF')^{-\frac{1}{2}}, L), \\ &= \begin{pmatrix} \alpha a (FF')^{\frac{1}{2}} & 0 \\ C\Xi (FF')^{\frac{1}{2}} & 0 \end{pmatrix} . \end{aligned}$$

Since $(F'(FF')^{-\frac{1}{2}},L)$ is an orthogonal matrix, each column of Y is independently normally distributed with covariance matrix ψ.

We now have transformed the data X into a form in which it is easy to find the estimators. Let us write the joint distribution of Y in the following way:

$$(1.1.5) \quad f(Y) = f(Y_{21}|Y_{11})f(Y_{11})\cdot f(Y_{22}|Y_{12})f(Y_{12}),$$

where $f(Y_{21}|Y_{11})$ indicates the conditional density of Y_{21} given Y_{11}, $f(Y_{11})$ indicates the marginal density of Y_{11}, etc. Since the columns of Y are independent normally distributed random variables, all of the densities in (1.1.5) are normal densities.

The parameters in our transformed model are $\alpha, C\Xi$, and ψ. An equivalent parametrization is

$$\alpha, C\Xi, \psi_{11}, \psi_{21}\psi_{11}^{-1}, \text{ and } \psi_{22.1} = \psi_{22}-\psi_{21}\psi_{11}^{-1}\psi_{12}.$$

We note that in (1.1.5) only $f(Y_{21}|Y_{11})$ depends on $C\Xi$, and only $f(Y_{11})$ and $f(Y_{21}|Y_{11})$ depend on α in their parameterizations.

Thus, we begin by finding the MLE of $C\Xi$ assuming that $\alpha_1\psi_{21}(\psi_{11}^{-1})$, ψ_{11}, and $\psi_{22.1}$ are fixed. We know that

$$f(Y_{21}|Y_{11}) = \frac{1}{|\psi_{22.1}|^{k/2}(2\pi)^{(p-r)k/2}} e^{-\frac{1}{2}(\mathrm{tr}\,\psi_{22.1}^{-1}(Y_{21}-\mu_{21})(Y_{21}-\mu_{21})')},$$

$$(1.1.6) \quad \leq \frac{1}{|\psi_{22.1}|^{k/2}(2\pi)^{(p-r)k/2}},$$

where $\mu_{21} = E(Y_{21}|Y_{22}) = C\Xi(FF')^{\frac{1}{2}} + \psi_{21}\psi_{11}^{-1}(Y_{11}-\alpha a(FF')^{\frac{1}{2}})$.
If we pick $\hat{C\Xi}$ so that

$$(1.1.7) \quad Y_{21} = \hat{C\Xi}\,(FF')^{\frac{1}{2}} + \psi_{21}(\psi_{11}^{-1})(Y_{11}-\alpha a(FF')^{\frac{1}{2}}) = \hat{\mu}_{21},$$

then it is clear that $f(Y_{21}|Y_{22})$ attains its maximum (1.1.6). We may rewrite (1.1.5) to get

$$(1.1.8) \quad f(Y) \leq \frac{1}{|\psi_{22.1}|^{k/2}(2\pi)^{(p-r)k/2}} f(Y_{11})f(Y_{22}|Y_{21})\ f(Y_{22}),$$

with equality when (1.1.7) holds.

We next maximize the right-hand side of (1.1.8) with respect to α, treating $\psi_{21}\psi_{11}^{-1},\psi_{11},\psi_{22.1}$ as fixed. We know that

$$f(Y_{11}) = \frac{1}{|\psi_{11}|^{k/2}(2\pi)^{rk/2}} e^{-\frac{1}{2}\,tr\psi_{11}^{-1}[(Y_{11}-\alpha a(FF')^{\frac{1}{2}})(Y_{11}-\alpha a(FF')^{\frac{1}{2}})']}.$$

Using the theory of multivariate regression, we get

$$(1.1.9) \quad f(Y_{11}) \leq \frac{1}{|\psi_{11}|^{k/2}(2\pi)^{rk/2}} e^{-\frac{1}{2}\,tr\psi_{11}^{-1}(Y_{11}MY_{11}')},$$

where $M = I-(FF')^{\frac{1}{2}}\, a'(aFF'a')^{-1}\, a(FF')^{\frac{1}{2}}$. Equality in (1.1.9) occurs only when

$$(1.1.10) \qquad \hat{\alpha} = Y_{11}(FF')^{\frac{1}{2}}\, a'(aFF'a')^{-1}.$$

Substituting (1.1.9) into (1.1.8), we get

$$(1.1.11) \quad f(Y) \leq \frac{e^{-\frac{1}{2}\,tr\psi_{11}^{-1}(Y_{11}MY_{11}')}}{|\psi_{11}|^{k/2}(2\pi)^{pk/2}|\psi_{22.1}|^{k/2}} f(Y_{22}|Y_{12})\cdot f(Y_{12}).$$

We now maximize the right-hand side of (1.1.11) with respect to ψ_{11} keeping $\psi_{22.1}$ and $\psi_{21}(\psi_{11})^{-1}$ fixed. Since

$$f(Y_{12}) = \frac{e^{\frac{1}{2}\,tr\psi_{11}^{-1}\,Y_{12}Y_{12}'}}{|\psi_{11}|^{(n-k)/2}(2\pi)^{(N-k)r/2}},$$

(1.1.11) can be written

$$f(Y) \le \frac{e^{-\frac{1}{2}\,\mathrm{tr}\psi_{11}^{-1}(Y_{12}Y_{12}'+Y_{11}MY_{11}')}}{|\psi_{11}|^{N/2}(2\pi)^{(pk+(N-k)r)/2}|\psi_{22.1}|^{k/2}} f(Y_{22}|Y_{12}),$$

where $f(Y_{22}|Y_{12})$ does not depend on ψ_{11}. Using Lemma 3.2.2 of Anderson [1958], we have

$$(1.1.12) \qquad f(Y) \le \frac{e^{-\frac{1}{2}Nr}\cdot f(Y_{22}|Y_{12})}{|\hat{\psi}_{11}|^{N/2}(2\pi)^{(pk+(N-k)r)/2}|\psi_{22.1}|^{k/2}},$$

where $\hat{\psi}_{11} = (Y_{12}Y_{12}' + Y_{11}MY_{11}')/N$.

Finally, we maximize the right-hand side of (1.1.12) with respect to $\psi_{21}(\psi_{11})^{-1}$ and $\psi_{22.1}$. We know that

$$\frac{f(Y_{22}|Y_{12})}{|\psi_{22.1}|^{k/2}} = \frac{e^{-\frac{1}{2}\,\mathrm{tr}\psi_{22.1}^{-1}(Y_{22}-\psi_{21}\psi_{11}^{-1}Y_{12})(Y_{22}-\psi_{21}\psi_{11}^{-1}Y_{12})'}}{|\psi_{22.1}|^{N/2}(2\pi)^{(N-k)(p-r)/2}},$$

$$(1.1.13) \qquad \le \frac{e^{-\frac{1}{2}\,\mathrm{tr}\psi_{22.1}^{-1}(Y_{22}(I_{N-k}-Y_{12}'(Y_{12}Y_{12}')^{-1}Y_{12})Y_{22}')}}{|\psi_{22.1}|^{N/2}(2\pi)^{(N-k)(p-r)/2}},$$

with equality only when

$$\psi_{21}(\hat{\psi}_{11})^{-1} = Y_{22}Y_{12}'(Y_{12}Y_{12}')^{-1}.$$

Using Lemma 3.2.2 of Anderson [1958], we have

$$(1.1.14) \qquad \frac{f(Y_{22}|Y_{12})}{|\psi_{22.1}|^{k/2}} \le \frac{e^{-\frac{1}{2}N(p-r)}}{|\hat{\psi}_{22.1}|^{N/2}(2\pi)^{(N-k)(p-r)/2}},$$

where

$$\hat{\psi}_{22.1} = \frac{1}{N}\,(Y_{22}(I-Y_{12}'(Y_{12}Y_{12}')^{-1}Y_{12})Y_{22}').$$

Combining (1.1.12) through (1.1.14), we get

$$(1.1.15)\qquad f(Y) \leq \frac{1}{(2\pi)^{(Np)/2}(|\hat{\psi}_{11}||\hat{\psi}_{22.1}|)^{N/2}}\cdot e^{-\frac{1}{2}Np}.$$

There will be equality in (1.1.15) if

$$\hat{C\Xi} = [Y_{21}(FF')^{-\frac{1}{2}} - (\hat{\psi}_{21}\hat{\psi}_{11}^{-1})(Y_{11}-\hat{\alpha}a(FF')^{\frac{1}{2}})](FF')^{-\frac{1}{2}},$$

$$\hat{\alpha} = Y_{11}(FF')^{\frac{1}{2}}\,a'(aFF'a')^{-1},$$

$$\hat{\psi}_{21}(\hat{\psi}_{11})^{-1} = Y_{22}Y_{12}'(Y_{12}Y_{12}')^{-1},$$

$$\hat{\psi}_{11} = \frac{1}{N}\,(Y_{12}Y_{12}' + Y_{11}(I_k-(FF')^{\frac{1}{2}}\,a'(aFF'a')^{-1}\,a(FF')^{\frac{1}{2}})Y_{11}'),$$

$$\hat{\psi}_{22.1} = \frac{1}{N}\,(Y_{22}(I_{N-k}-Y_{12}'(Y_{12}Y_{12}')^{-1}Y_{12})Y_{22}').$$

Now we go backwards and express $\hat{\Xi}$ and $\hat{\alpha}$ in terms of X. After a little simplification, using the facts that

$$\binom{B}{C}^{-1} = (B'(BB')^{-1},C'),$$

$$C'C = I_p-B'(BB')^{-1}B,$$

$$LL' = I_N - F'(FF')^{-1}F,$$

we obtain

$$(1.1.16)\qquad \hat{\Xi} = XA - X(I_N-AF)X'B'(BX(I_N-AF)X'B')^{-1}BX(A-G),$$

$$(1.1.17)\qquad \hat{\alpha} = BX(F'a'(aFF'a')^{-1}),$$

where

(1.1.18) $A = F'(FF')^{-1}$,

$G = F'a'(aFF'a')^{-1}a$.

We can also go backwards and find $\hat{\Sigma}$. When we do this, we get

$$\hat{\Sigma} = N^{-1}(X-\hat{\Xi}F)(X-\hat{\Xi}F)'.$$

We may summarize our results so far in the following theorem.

Theorem 1.1.1. When B is fixed, the MLE of α, Ξ, and Σ in the model given by (1.1.2) and (1.1.3) are

$$\hat{\alpha} = BXF'a'(aFF'a')^{-1},$$

$$\hat{\Xi} = XA-X(I_N-AF)XB'(BX(I_N-AF)X'B')^{-1}BX(A-G),$$

$$\hat{\Sigma} = N^{-1}(X-\hat{\Xi}F)(X-\hat{\Xi}F),$$

where A and G are given by (1.1.18).

Part II. Substitution of parameters back into the likelihood and maximization with respect to B.

If we substitute the estimators of α, Ξ, Σ given in Theorem 1.1.1 (note: they are functions of B) into the likelihood for X, we find that

(1.1.19) $$\max_{C\Xi,\alpha,\psi} \log f(X) = -\frac{1}{2}pN \log 2\pi - \frac{1}{2}N \log |\hat{\Sigma}| - \frac{1}{2}pN.$$

Maximizing (1.1.19) with respect to B is equivalent to minimizing $|\hat{\Sigma}|$ with respect to B. After simplification we get

(1.1.20) $$|\hat{\Sigma}| = \frac{1}{N^p}|W| \cdot \frac{|BTB'|}{|BWB'|},$$

where

(1.1.21) $W = X(I_N - F'(FF')^{-1}F)X'$,

(1.1.22) $T = X(I_N - F'a'(aFF'a')^{-1}aF)X'$.

Note that in terms of MANOVA concepts, W can be thought of as the within covariance matrix and T as the total covariance matrix.

Let $U = N^{-\frac{1}{2}} BW^{\frac{1}{2}}$. Then (1.1.20) becomes

$$(1.1.23) \qquad |\hat{\Sigma}| = \frac{1}{N^p} |W| \frac{|UW^{-\frac{1}{2}} TW^{-\frac{1}{2}} U'|}{|UU'|} .$$

For purposes of minimizing (1.1.23), we might as well assume that $UU' = I_r$, for if UU' doesn't equal the identity matrix, there exists an invertible matrix H such that $U^* = HU$ also minimizes (1.1.23) and $U^*U^{*\prime} = I_r$.

If $UU' = I_r$, Theorem 10, page 129 of Bellman [1970] tells us that the minimum value of $\hat{\Sigma}$ is

$$(1.1.24) \qquad |\hat{\Sigma}| = \frac{1}{N^p} |W| \cdot (\lambda_p \cdot \lambda_{p-1} \cdot \ldots \cdot \lambda_{p-r+1}),$$

where λ_i is the ith largest eigenvalue of $W^{-\frac{1}{2}} T W^{-\frac{1}{2}}$. Let Γ' be a matrix whose columns are the eigenvectors associated with the r smallest eigenvalues of $W^{-\frac{1}{2}} TW^{-\frac{1}{2}}$. If we choose U to be Γ, then the right-hand side of (1.1.23) achieves the minimum value of $|\hat{\Sigma}|$ as seen in (1.1.24). Thus, if we let

$$(1.1.25) \qquad \hat{B} = N^{-\frac{1}{2}} \Gamma W^{-\frac{1}{2}},$$

then the likelihood function is maximized. It is easy to show that the columns of $\hat{B}'$ are themselves eignevectors of $W^{-1} T$ corresponding to the r smallest eigenvalues of $W^{-1} T$.

We summarize our results in the following theorem.

Theorem 1.1.2. The MLE of B, α, Ξ, and Σ in the model given by (1.1.2) and (1.1.3) assuming a and F are full row rank are:

$$\hat{\alpha} = \hat{B}XF'a'(aFF'a')^{-1},$$

$$\hat{\Xi} = X(F'(FF')^{-1}) - W\hat{B}(\hat{B}W\hat{B}')^{-1}(\hat{B}X(F'(FF')^{-1} - F'a'(aFF'a')^{-1}a)),$$

$$\hat{\Sigma} = N^{-1}(X - \hat{\Xi}F)(X - \hat{\Xi}F)',$$

where

$$W = X(I_N - F'(FF')^{-1}F)X',$$

$$T = X(I_N - F'a'(aFF'a')^{-1}aF)X',$$

and the columns of $\hat{B}'$ are the eigenvectors corresponding to the r smallest eigenvalues of $W^{-1}T$.

Remark I. If we multiply $\hat{B}$ on the right by any invertible matrix, the resulting matrix also maximizes the likelihood since if $B^* = H\hat{B}$, $|H| \neq 0$, then

$$\frac{|B^*TB^{*\prime}|}{|B^*WB^{*\prime}|} = \frac{|H\hat{B}T\hat{B}'H'|}{|H\hat{B}W\hat{B}'H'|} = \frac{|H|\cdot|\hat{B}T\hat{B}'|\cdot|H'|}{|H|\cdot|\hat{B}W\hat{B}'|\cdot|H'|} = \frac{|\hat{B}T\hat{B}'|}{|\hat{B}W\hat{B}'|}.$$

Remark II. All matrices which maximize the likelihood are of the form $H\hat{B}$ for some invertible H. We will not prove this, since a proof of the assertion is straightforward.

Remark III. We have been assuming that F has full row rank. We now demonstrate how to reparametrize so that the results in Theorem 1.1.2 can be applied when F is not of full row rank. Assume c(c<k) is the

rank of F, and c-s $\geq$ p. Let

$$F = (\Gamma_1\Gamma_2)\begin{pmatrix} D & O \\ O & O \end{pmatrix}U. \tag{1.1.26}$$

The right-hand side of (1.1.26) is the Eckart-Young decomposition where U and $(\Gamma_1\Gamma_2)$ are orthogonal matrices and D is a diagonal invertible cxc matrix. Now

$$\begin{aligned} \Xi F &= \Xi(\Gamma_1\Gamma_2)\begin{pmatrix} D & O \\ O & O \end{pmatrix}U, \\ &= (\Xi\Gamma_1)(D,O)U, \\ &= \Xi^*(D,O)U = \Xi^*F^*, \end{aligned}$$

where $\Xi^* = \Xi\Gamma_1$ and $F^* = (D,O)U$. Since F^* is full row rank, we may use Theorem 1.1.2 to get the MLE's of the parameters. If $\hat{\Xi}^*$ is the MLE of Ξ^*, we have

$$\begin{aligned} \hat{\Xi}^* &= \hat{\Xi}\Gamma_1, \\ \hat{\Xi} &= (\hat{\Xi}^*,P)\begin{pmatrix} \Gamma_1' \\ \Gamma_2 \end{pmatrix}, \end{aligned}$$

where P is any finite pxk-c matrix. Usually when F is not of full row rank there are restrictions on Ξ. We can pick P so that $\hat{\Xi}$ satisfies those restrictions.

We now state a theorem which gives us the MLE's for our model when a is the zero matrix:

Theorem 1.1.3. The MLE of B, Ξ, and Σ in the model given by (1.1.2) and (1.1.3) when a is the zero matrix,i.e.,(1.1.2) becomes $B\Xi = 0$ are:

$$\hat{\Xi} = X(F'(FF')^{-1}) - W\hat{B}(\hat{B}W\hat{B}')^{-1}(\hat{B}X(F'(FF')^{-1})),$$

$$\hat{\Sigma} = N^{-1}(X-\hat{\Xi}F)(X-\hat{\Xi}F)',$$

where

$$W = X(I_N-F'(FF')^{-1}F)X',$$

$$T = XX',$$

and the columns of $\hat{B}'$ are the eigenvectors corresponding to the r smallest eigenvalues of $W^{-1}T$.

Let us now consider the model of Theorem 1.1.2 with one change - namely, instead of assuming that each e_i is independently normally distributed with common covariance matrix Σ, we now allow the e_i's to be jointly normally distributed with mean vector 0 and

(1.1.27) $$\text{cov}(e_i,e_j) = k_{ij}\cdot\Sigma,$$

where $K = (k_{ij})$ is a known invertible matrix. The maximum likelihood estimators of α, B, Ξ, and Σ are easy to compute, using Theorem 1.1.2 and the following lemma:

Lemma 1. Let $Z = XK^{-\frac{1}{2}}$ (X comes from our new model), then $E(Z) = \Xi FK^{-\frac{1}{2}}$ and each column of Z is independent with a p-dimensional normal distribution having covariance matrix Σ.

Proof. Since Z is a linear combination of normally distributed random variables, it is itself normally distributed. Further,

$$E(Z) = E(XK^{-\frac{1}{2}}) = (E(X))K^{-\frac{1}{2}} = \Xi FK^{-\frac{1}{2}}.$$

Let $(m^{ij}) = K^{-\frac{1}{2}}$. Then

$$\begin{aligned}\operatorname{cov}(Z_\alpha Z_\beta') &= E(Z_\alpha - V_\alpha)(Z_\beta - V_\beta)' \\ &= E\Big(\sum_{ij}(x_i - \Xi f_i)m^{\alpha i}m^{\beta j}(x_j - \Xi f_j)'\Big) \\ &= \sum_{ij} m^{\alpha i}m^{\beta j}E(x_i - \Xi f_i)(x_j - \Xi f_j) \\ &= \Sigma m^{\alpha i}m^{\beta j}k_{ij} - \Sigma \\ &= \delta_{\alpha\beta} - \Sigma,\end{aligned}$$

where $\sigma_{\alpha\beta}$ is the Kronecker Delta function. Q.E.D.

If we transform X as prescribed in Lemma 1, the resulting model exactly corresponds to the model in Theorem 1.1.2. We therefore have the following result.

Theorem 1.1.4. The maximum likelihood estimators in the model given by (1.1.2) and (1.1.3) with the following change,

$$\operatorname{cov}(e_i \cdot e_j) = k_{ij} \cdot \Sigma,$$

where $(k_{ij}) = K$ is known, are:

$$\hat{\alpha} = \hat{B}X(K^{-1}F'a')(aFK^{-1}F'a')^{-1},$$

$$\hat{\Xi} = XK^{-1}F'(FK^{-1}F')^{-1} - W\hat{B}'(\hat{B}W\hat{B}')^{-1}(\hat{B}X \cdot [K^{-1}F'(FK^{-1}F')^{-1} - K^{-1}F'a'(aFK^{-1}F'a')^{-1}a]),$$

$$\hat{\Sigma} = N^{-1}(X - \hat{\Xi}F)(X - \hat{\Xi}F)',$$

where

$$W = X(I_N - K^{-1}F'(FK^{-1}F')^{-1}FK^{-1})X',$$

$$T = X(I_N - K^{-1}F'a'(aFK^{-1}F'a')^{-1}aFK^{-1})X',$$

the columns of $\hat{B}'$ are the r eigenvectors associated with the r smallest eigenvalues of $W^{-1}T$.

1.2 Consistency of the Estimators

As the number of observations gets large, it is important to know what our estimators converge to. In most statistical problems the number of parameters stays fixed as the sample size increases. However, in this section we will be finding out what our estimators from Theorem 1.1.2 converge to when the number k of columns of Ξ is allowed to increase with the sample size. The elements of our Ξ matrix are what Neyman and Scott [1948] have called "incidental parameters". When there are incidental parameters present, some estimators (as in our case) may turn out to be inconsistent. We will not discuss the consistency of the estimators in Theorem 1.1.3 or Theorem 1.1.4 since it is clear that we have analagous results. In our discussion, p (the dimension of the dependent variable), r (the row rank of B) and s (the column rank of α) are assumed to be fixed. It is evident that

$$t = \lim_{N\to\infty} \frac{N-k}{N}$$

is a measure of how fast the number of parameters increases with the sample size, N. We will assume that t is always greater than zero and less than or equal to one. If the number of parameters stays fixed, t will equal one. We will be concerned with the consistency of $\hat{B}$, $\hat{\alpha}$, and $\hat{\Sigma}$. We will first discuss the consistency of $\hat{B}$ and $\hat{\alpha}$.

In order to make a discussion of the consistency of $\hat{B},\hat{\alpha}$ meaningful, we will have to place restrictions on B and $\hat{B}$ which will make these matrices unique. It should be remembered that if $\hat{B},\hat{\alpha}$ maximize the likelihood, then so do $H\hat{B}$, $H\hat{\alpha}$ where H is an

invertible matrix. In fact all MLE of B,α will be of the form $\hat{HB}$, $\hat{H\alpha}$ for some invertible matrix H. Similarly (B,α) satisfy

(1.2.1) $\qquad B\Xi = \alpha a,$

if and only if $HB,H\alpha$ satisfy $HB\Xi = H\alpha$, where H is an invertible matrix.

Let B,α be a pair of matrices which satisfy (1.2.1). By requiring B to satisfy a number of restrictions, (B,α) will be the unique matrices which satisfy (1.2.1). We will show that if $\hat{B}$ and $\hat{\alpha}$ are MLE of B and α, and if $\hat{B}$ satisfies the same restrictions as B, then $\hat{B},\hat{\alpha}$ converge almost surely to B,α. We will be showing the above for only one particular set of restrictions. However, it is clear that if one set of MLE $(\hat{B}_1,\hat{\alpha}_1)$ converge almost surely to B_1,α_1, where $\hat{B}_1$ and B_1 satisfy one group of restrictions, then any other set of MLE $\hat{B}_2,\hat{\alpha}_2$ will converge almost surely to B_2,α_2, where $\hat{B}_2$ and B_2 satisfy another group of restrictions, provided the respective restrictions make B_1 and B_2 unique.

Let B,α be a set of matrices which satisfy (1.2.1) and let

$$B^* = (B_2^{-1}B_1, I_r) = B_2^{-1}(B_1,B_2) = B_2^{-1}(B),$$

$$\alpha^* = B_2^{-1}\alpha,$$

where $B = (B_1,B_2)$, B_1: rxp-r, and B_2: rxr. B^* is the only matrix with its last r columns being the identity which satisfies (1.2.1). Similarly, if $\hat{B}_1,\hat{\alpha}$ are maximum likelihood estimators, we can generate another set of maximum likelihood estimators $\hat{B}^*,\hat{\alpha}^*$ where $\hat{B}^*$ has the identity matrix as its last r columns:

$$\hat{B}^* = \hat{B}_2^{-1}(\hat{B}_1,\hat{B}_2) = \hat{B}_2^{-1}\hat{B},$$

$$\hat{\alpha}^* = \hat{B}_2^{-1}\hat{\alpha}.$$

Hence, $\hat{B}^*$ is the only MLE of B which has the identity for its last r columns. We will show that $\hat{B}^*,\hat{\alpha}^*$ converge almost surely to B^*,α^*.

Lemma 1. If $N-k \to \infty$ then $(N-k)^{-1}W$ goes almost surely to Σ.

Proof. Recall that

$$W = X(I_N - F'(FF')^{-1}F)X',$$

$$= (\Xi F+E)(I_N - F'(FF')^{-1}F)(\Xi F + E)',$$

$$= E(I_N - F'(FF')^{-1}F)E'.$$

Each column of E has an independent normal distribution with mean vector 0 and covariance matrix Σ. By Theorem 4.3.2 in Anderson [1958], W is distributed the same way as

$$\sum_{i=1}^{N-k} u_i\, u_i',$$

where u_i are independent $N(0,\Sigma)$ random variables. We can conclude that

$$(N-k)^{-1} \sum_{i=1}^{N-k} u_i\, u_i'$$

converges almost surely to Σ. Therefore $(N-k)^{-1}(W)$ goes almost surely to Σ. Q.E.D.

Lemma 2. Let $z_1 z_2 \ldots$ be independent identically distributed random variables with means 0 and common finite variances. Let $b_{n,m}$ be any

array of real numbers $m \leq n$ $n = 1,2,\ldots$ satisfying

$$\lim_{n\to\infty} \sum_{m=1}^{n} b_{nm}^2 = v, \qquad 0 < v < \infty,$$

then $n^{-\frac{1}{2}} \sum_{m=1}^{n} b_{nm} z_m$ goes to 0 almost surely.

Proof. The proof is in Chow [1966]. Q.E.D.

Lemma 3. Assume that

$$R = \lim_{N\to\infty} N^{-1} \Xi F(I_N - F'a'(aFF'a')^{-1}aF)F'\Xi'$$

exists and is finite, then

$$(1.2.2) \qquad N^{-1} E(I_N - F'a'(aFF'a')^{-1}aF)F'\Xi'$$

goes almost surely to zero.

Proof. Consider the i,jth element of (1.2.2). That element is the product of the ith row of $N^{-\frac{1}{2}}$ E and the jth column of

$$(1.2.3) \qquad N^{\frac{1}{2}}(I_N - F'a'(aFF'a')^{-1}aF)F'\Xi'.$$

Each element in the ith row of E is independent with mean 0 and common variance. The sum of the squares of the elements in the jth column of (1.2.3) is the j,jth element of

$$N^{-1} \Xi F(I_N - F'a'(aFF'a')^{-1}aF)F'\Xi'.$$

By our hypothesis, this element converges to something finite as N goes to infinity. By Lemma 2, the i,jth element of (1.2.2) goes almost surely to zero. Q.E.D.

Lemma 4. Assume that R (as defined in Lemma 3) exists and is finite, then $N^{-1}T$ goes almost surely to $R+\Sigma$.

Proof. Recall that

$$N^{-1}T = N^{-1}X(I_N - F'a'(aFF'a')^{-1}aF)X',$$

$$= N^{-1}(\Xi F+E)(I_N - F'a'(aFF'a')^{-1}aF)(\Xi F+E)',$$

$$(1.2.4) \quad = N^{-1}\Xi F(I_N - F'a'(aFF'a')^{-1}aF)F'\Xi' + N^{-1}E(I_N - F'a'(aFF'a')^{-1}aF)F'\Xi' +$$

$$N^{-1}\Xi F(I_N - F'a'(aFF'a')^{-1}aF)E' + N^{-1}E(I_N - F'a'(aFF'a')^{-1}aF)E'.$$

By our hypothesis the first term in (1.2.4) converges to R. By Lemma 3, the second and third terms go almost surely to zero. If we use Theorem 4.3.2 in Anderson [1958], we find that the fourth term in (1.2.4) has the same distribution as

$$\sum_{i=1}^{N-s} u_i u_i',$$

where u_i has a normal distribution with mean vector 0 and covariance matrix Σ. u_i and u_j are independent if $i \neq j$. We know that $(N-s)^{-1} \sum_{i=1}^{N-s} u_\alpha u_\alpha'$ goes almost surely to Σ. Since s is fixed as N goes to infinity we have that $N^{-1}E(I_N - F'a'(aFF'a')^{-1}aF)E' \xrightarrow{a.s.} \Sigma$. Using all of the above arguments, we have $N^{-1}T$ goes almost surely to $R+\Sigma$. Q.E.D.

Lemma 5. The columns of B*' are eigenvectors of $I_p - \Sigma^{-1}R$ corresponding to eigenvalue one.

Proof. We know that for every N

$$N^{-1}\Xi F(I_N - F'a'(aFF'a')^{-1}aF)F'\Xi'B^{*\prime} =$$

$$N^{-1}\Xi F(I_N - F'a'(aFF'a')^{-1}aF)F'a'\alpha^{*\prime} = 0.$$

Because of the above $RB^{*\prime} = 0$. We therefore have

$$(I_p - \Sigma^{-1}R)B^{*\prime} = B^{*\prime}. \qquad \text{Q.E.D.}$$

Theorem 1.2.1. Under the assumptions of Lemma 4 and assuming R is of rank p-r, $\hat{B}^*$ is a strongly consistent estimator of $\hat{B}^*$.

Proof. By Lemma 1, we have $(N-k)^{-1}W \xrightarrow{a.s.} \Sigma$. By Lemma 4, $N^{-1}T \xrightarrow{a.s.} R+\Sigma$. Combining these statements we get

$$((N-k)W)^{-1}(N^{-1}T) \xrightarrow{a.s.} \Sigma^{-1}(\Sigma+R) = I_p+\Sigma^{-1}R.$$

Since $\lim_{N\to\infty} \frac{N-k}{N} = t > 0$, we have

$$W^{-1}(T) \xrightarrow{a.s.} (1/t)(I_p+\Sigma^{-1}R).$$

Since the eigenvalues of a matrix are continuous functions of the elements of that matrix, the eigenvalues of $W^{-1}(T)$ coverge almost surely to the eigenvalues of $1/t(I_p+\Sigma^{-1}R)$. Since R is positive semidefinite of rank p-r and Σ is positive definite, the smallest eigenvalue of $1/t(I_p+\Sigma^{-1}R)$ is 1/t. It has multiplicity r. The r smallest eigenvalues of $W^{-1}(T)$ must go almost surely to 1/t. Let $\hat{B}_N^*$ be the estimator of B^* if we have N observations. Let $\hat{B}_N$ be the estimator given in Theorem 1.1.2 ($\hat{B}_N$ satisfies $N^{-1}\hat{B}_N W\hat{B}_N' = I_r$) used to generate $\hat{B}_N^*$, i.e.,

$$\hat{B}_N^* = (\hat{B}_N^{(2)^{-1}}\hat{B}_N^{(1)}, I_r) = \hat{B}_N^{(2)^{-1}}(\hat{B}_N^{(1)}, \hat{B}_N^{(2)}) = \hat{B}_N^{(2)^{-1}}(\hat{B}_N),$$

where $\hat{B}_N = (\hat{B}_N^{(1)}, \hat{B}_N^{(2)})$. Because $N^{-1}\hat{B}_N W \hat{B}_N' = I_r$ and $N^{-1}W$ converges almost surely to $t\cdot\Sigma$, $\hat{B}_N$ is bounded almost surely. Let us pick any subsequence of $\hat{B}_N$. Since $\hat{B}_N$ is almost surely bounded, there must exist a subsequence of this subsequence which converges. Let $\hat{B}_{\pi_N}$ denote the convergent subsequence. Let C be defined by

$$\lim_{N\to\infty} \hat{B}_{\pi_N} = C.$$

Every column of C' is the limit of a sequence of eigenvectors of $W^{-1}(T)$ associated with an eigenvalue which goes almost surely to $1/t$. Since $W^{-1}(T)$ converges almost surely to $1/t\ (I+\Sigma^{-1}R)$, each column of C must equal some eigenvector of $1/t(I+\Sigma^{-1}R)$ associated with eigenvalue $1/t$. Since

$$\lim_{N\to\infty} (\pi_N)^{-1}\hat{B}_{\pi_N}(W)\hat{B}'_{\pi_N} = tC\Sigma C' = I_r,$$

C is of full row rank. C must span the space of eigenvectors of $(1/t)(I+\Sigma^{-1}R)$ associated with $1/t$. By Lemma 5, B* also spans this space. Therefore there exists an invertible matrix V such that

$$B^* = (B^{(2)^{-1}}B^{(1)}, I) = VC.$$

If $C = (C^{(1)}, C^{(2)})$, V must equal $(C^{(2)})^{-1}$ and

$$B^* = (C^{(2)})^{-1}C.$$

Let $||A||$ denote the largest value of any element in A. We know that

$$||\hat{B}^*_{\pi_N} - B^*|| = ||(\hat{B}^{(2)}_{\pi_N})^{-1}\cdot\hat{B}_{\pi_N} - (C^{(2)})^{-1}C||,$$

$$||\hat{B}^*_{\pi_N} - B^*|| \leq ||(\hat{B}^{(2)}_{\pi_N})^{-1}\hat{B}_{\pi_N} - (C^{(2)})^{-1}\hat{B}_{\pi_N}|| + ||(C^{(2)})^{-1}\hat{B}_{\pi_N} - (C^{(2)})^{-1}C||,$$

(1.2.5)

$$||\hat{B}^*_{\pi_N} - B^*|| \leq ||(\hat{B}_{\pi_N})||\ ||(\hat{B}^{(2)}_{\pi_N})^{-1} - (C^{(2)})^{-1}|| + ||(C^{(2)})^{-1}||\ ||\hat{B}_{\pi_N} - C||.$$

The first term on the right-hand side of (1.2.5) is arbitrarily small since $||\hat{B}_{\pi_N}||$ is almost surely bounded and $\hat{B}_{\pi_N}$ differs from C by an arbitrarily small amount when N is large. The second term vanishes since $(C^{(2)})^{-1}$ is bounded and $\hat{B}_{\pi_N}$ goes almost surely to C. We therefore have that $\hat{B}^*_{\pi_N}$ goes almost surely to B*. We have shown that for any subsequence of $\hat{B}^*_N$, there exists a subsequence of that subsequence which converges to B* almost surely. $\hat{B}^*$ must converge almost surely to B*. Q.E.D.

Theorem 1.2.2. If $N(aFF'a')^{-1}$ converges to a matrix with all elements finite then $\hat{\alpha}^*$ is a strongly consistent estimate of α^*.

Proof. Note that

$$\hat{\alpha}^* = \hat{B}^*XF'a'(aFF'a')^{-1},$$

$$= \hat{B}^*(\Xi F+E)(F'a'(aFF'a')^{-1}),$$

(1.2.6)
$$= \hat{B}^*\Xi FF'a'(aFF'a')^{-1} + \hat{B}^*E(F'a'(aFF'a')^{-1}.$$

Since $\hat{B}^*$ goes almost surely to B*, the first term on the right of (1.2.6) goes almost surely to

$$B^* \equiv FF'a'(aFF'a')^{-1} = \alpha^*aFF'a'(aFF'a')^{-1} = \alpha^*.$$

By applying Lemma 2 in a way similar to what we did in Lemma 3, we know that $N^{-1}(aFF'a')^{-1}$ converging to a finite matrix implies that $E\ F'a'(aFF'a')^{-1}$ goes almost surely to zero. We can conclude that $\hat{B}^*EF'a'(aFF'a')^{-1}$ converges almost surely to zero. Q.E.D.

We now must discuss the consistency of $\hat{\Sigma}$. It should be noted that the MLE's of Ξ and Σ are unique; they do not depend on the choice of MLE of B and α. Because of this, we will use $\hat{B}^*$ as the MLE of B and $\hat{\alpha}^*$ as the MLE of α. We have seen that

$$\hat{\Sigma} = N^{-1}(X-\hat{\Xi}F)(X-\hat{\Xi}F)',$$

$$= N^{-1}(X-XF'(FF')^{-1}F+W\hat{B}^*(\hat{B}^*W\hat{B}^{*\prime})^{-1}(\hat{B}^*X(F'(FF')^{-1}F - F'a'(aFF'a')^{-1}aF)\cdot$$
$$(X-XF'(FF')^{-1}F+W\hat{B}^*(\hat{B}^*W\hat{B}^{*\prime})^{-1}(\hat{B}X(F'(FF')^{-1}F - F'a'(aFF'a')^{-1}aF)'.$$

After a little simplification which uses the definitions of W and T, we get

$$\hat{\Sigma} = N^{-1}W+N^{-1}W\hat{B}^*(\hat{B}^*W\hat{B}^{*\prime})^{-1}\hat{B}^*(T-W)\hat{B}^{*\prime}(\hat{B}^*W\hat{B}^{*\prime})^{-1}\hat{B}^{*\prime}W.$$

From our previous lemmas and theorems we know that

$$N^{-1}W \overset{a.s.}{\to} t\cdot\Sigma, \qquad \hat{B}^* \overset{a.s.}{\to} B^*,$$

$$N^{-1}T \overset{a.s.}{\to} \Sigma+R, \qquad RB^{*\prime} = 0.$$

Using the above we have

$$\hat{\Sigma} \overset{a.s.}{\to} t\cdot\Sigma+\Sigma B^{*\prime}(B^*\Sigma B^{*\prime})^{-1}B^*(\Sigma+R-t\Sigma)B^{*\prime}(B^*\Sigma B^{*\prime})^{-1}B^{*\prime}\Sigma$$
$$B^*\Sigma B^{*\prime}(B^*(\Sigma+R)B^{*\prime})^{-1}B^{*\prime}\Sigma,$$

$$= t\cdot\Sigma+(1-t)(\Sigma B^*(B^*\Sigma B^{*\prime})^{-1}B^{*\prime}\Sigma).$$

Since the above expression is valid regardless of which B in the class of B's which satisfy $B\Xi = \alpha a$ we take, we have the following theorem:

Theorem 1.2.3. If we assume the conditions given in Lemmas 3 and 4 and in Theorem 2.2.2, then $\hat{\Sigma}$ goes almost surely to

$$(1.2.7) \qquad t\cdot\Sigma + (1-t)\Sigma B'(B\Sigma B')^{-1}B'\Sigma.$$

The most startling thing about the above is not that $\hat{\Sigma}$ is not a consistent estimate; when the number of parameters gets large, the estimate of the covariance matrix is usually inconsistent. What makes the above unusual is the fact that the matrix $\hat{\Sigma}$ goes to is a function of B. The second term in (1.2.7) is very unusual.

We can not discuss the consistency of $\hat{\Xi}$, since it is not a fixed matrix of parameters. It is interesting to consider to what

$$N^{-1}\hat{\Xi}F(I-F'a'(aFF'a')^{-1}aF)F'\hat{\Xi}$$

converges almost surely. We might expect it to converge almost surely to R as

$$N^{-1}\Xi F(I-F'a'(aFF'a')^{-1}aF)F'\Xi$$

does. However, if we went through a proof, we would find it actually goes almost surely to

$$R+(1-t)\Sigma + -(1-t)\Sigma B'(B\Sigma B')^{-1}B'\Sigma.$$

1.3. Special Cases

Special cases of the models we discussed have come up many times in the literature. We will be discussing cases when the F matrix has the following form:

$$(1.3.1) \qquad F = \begin{pmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots & 1 \end{pmatrix}.$$

If the F matrix has the above form, our additional information consists of knowing some of the observations come from the same mean, i.e., we have replications at each mean. The model could be written this way:

$$(1.3.2) \quad x_{ij} = \xi_i + e_{ij};\ i = 1,2,\dots,k;\ j = 1,2,\dots,n_i;$$

$$\Xi = (\xi_1, \xi_2, \dots, \xi_k);$$

$$E = (e_{11}, e_{12}, \dots, e_{in_1} \ \dots \ e_{kn_k}).$$

Note: In all of our special cases,

$$N = \sum_{i=1}^{k} n_i,\ \bar{x} = \Big(\sum_{i=1}^{k} n_i\Big)^{-1}\Big(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\Big),\ \bar{x}_i = (n_i)^{-1}\sum_{j=1}^{n_i} x_{ij}.$$

We will need the MLE's in the following two cases. The first case specifies that the set of mean vectors is in a lower (p-r) dimensional space passing through the origin:

$$(1.3.3) \qquad B\xi_i = 0, \quad \forall_i.$$

The second case specifies that the set of mean vectors is in a

lower dimensional space which can pass through any point:

(1.3.4) $\quad B\xi_i = \alpha, \ \forall_i.$

For the first case, we will apply Theorem 1.1.3 with F as defined by (1.3.1). Our result is:

Application 1. When our model is

$$x_{ij} = \xi_i + e_{ij}; \ i = 1,2,\ldots,k; \ j = 1,2,\ldots,n_i;$$
$$B\xi_i = 0;$$

then the MLE of B, ξ_i and Σ are

$$\hat{\xi}_i = \bar{x}_i - W\hat{B}'(\hat{B}W\hat{B}')^{-1}\hat{B}\bar{x}_i,$$

$$\hat{\Sigma} = N^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij}-\hat{\xi}_i)(x_{ij}-\hat{\xi}_i)',$$

where

$$W = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)',$$

$$T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij})(x_{ij}'),$$

and the columns of $\hat{B}$ are the eigenvectors corresponding to the r smallest eigenvalues of $W^{-1}T$.

For the second case we can apply Theorem 1.1.2 with a = (1,1,...1) and F as defined by (1.3.1) to get:

Application 2. When our model is

$$x_{ij} = \xi_i + e_{ij}; \ i = 1,2,\ldots,k; \ j = 1,2,\ldots,n_i;$$
$$B\xi_i = \alpha;$$

then the MLE of B, α, ξ_i, and Σ are

$$\hat{\alpha} = \hat{B}\bar{x},$$

$$\hat{\xi}_i = \bar{x}_i - W\hat{B}(\hat{B}W\hat{B}')^{-1}B(\bar{x}_i - \bar{x}),$$

$$\hat{\Sigma} = N^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \hat{\xi}_i)(x_{ij} - \hat{\xi}_i)',$$

where

$$W = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)',$$

$$T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})(x_{ij} - \bar{x})',$$

and the columns of $\hat{B}'$ are the eigenvectors corresponding to the r smallest eigenvalues of $W^{-1}T$.

The model in Application 2 is the same model Rao [1973] considers when he talks about a test for dimensionality. His test of dimensionality is a test of the hypothesis that $B\xi_i = \alpha$ versus the hypothesis $B\xi_i \neq \alpha$. His test statistic turns out to be similar to the likelihood ratio test statistic although he neither mentions nor proves this. He does find the likelihood ratio test when Σ is known.

Villegas [1961] considers both Application 1 and 2 - the first of which he calls a homogeneous linear functional relationship. All of Villegas's results are only valid when we are talking about a single functional relationship, i.e., B is a row vector. Through geometrical arguments similar to the techniques used by Max Van Uven [1930] who derived estimates of $\hat{B}$ and $\hat{\Xi}$ when Σ is known, Villegas

derived maximum likelihood estimators which agree with Anderson's and ours. $\hat{B}$ turns out to be the eigenvector associated with the smallest eigenvalue of $W^{-1}T$. Villegas also talks about cases in which Theorem 1.1.4 applies, i.e., when $K \neq I_N$. He shows that the covariance matrix has the form needed in Theorem 1.1.4 when it arises from certain experimental designs (mainly incomplete block designs.) Since our results are valid when B is any rank ($\leq$ p-1), our results can be thought of as extensions of Villegas's results for a single functional relationship.

We can give another application which fits directly into a one-way analysis of variance. Let our model be

$$x_{ij} = \mu+\xi_i+e_{ij},$$

where μ is the unknown grand mean. We will make the common assumption that $\sum_{i=1}^{k} \xi_i n_i = 0$. We will be fitting parameters under the hypothesis that,

$$B\xi_i = 0.$$

It should be noted that $B\xi_i$ can not equal anything but zero when $\Sigma\xi_i n_i = 0$. The MLE of μ is

$$\hat{\mu} = \bar{x}.$$

If we substitute $\hat{\mu}$ into the likelihood we have exactly the same maximization problem that is solved in Theorem 1.1.3 except that we will use

$$X^* = X-\bar{x}(1,1,\ldots,1) = (x_{11}-\bar{x}, x_{12}-\bar{x},\ldots,x_{kn_k}-\bar{x}),$$

instead of X. If we use X* and F as defined by (1.3.1) in Theorem 1.1.3,we get the following application:

Application 3. When our model is

$$x_{ij} = \mu+\xi_i+e_{ij}; \quad i = 1,2,\ldots,k; \; j = 1,2,\ldots,n_i;$$
$$B\xi_i = 0;$$

then the MLE's of μ, B, ξ_i and Σ are

$$\hat{\mu} = \bar{x},$$
$$\hat{\xi}_i = \bar{x}_i-\bar{x} - W\hat{B}'(\hat{B}W\hat{B}')^{-1}\hat{B}(\bar{x}_i-\bar{x}),$$
$$\hat{\Sigma} = N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\hat{\xi}_i)(x_{ij}-\hat{\xi}_i)',$$

where

$$W = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)',$$

$$T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})(x_{ij}-\bar{x})',$$

and the columns of $\hat{B}'$ are the r eigenvectors corresponding to the r smallest eigenvalues of $W^{-1}T$.

The model considered in Application 3 is a generalization of the model given by Kristoff [1973]. Kristoff gives an ad hoc goodness of fit test for his model which is actually equivalent to likelihood ratio test statistic.

In all applications so far, the estimate of B which was given and which maximizes the likelihood was unique only up to multiplication

on the left by a nonsingular matrix. By picking a unique member from the class of maximum likelihood estimators as we did in our section on the consistency of the estimators, we will show another class of models can be handled with our method.

Consider the following model:

$$(1.3.5) \qquad Y_{ij} = v_i + m_{ij};\ i = 1,2,\ldots,k;\ j = 1,2,\ldots,n_i;$$
$$z_{ij} = Hv_i + g_{ij};$$

where y_{ij} and z_{ij} are p-r and r dimensional vectors of observations respectively, v_i is a p-r dimensional parameter vector, H is a unknown rxp-r parameter matrix, and $\binom{m_{ij}}{g_{ij}}$ is the error term. We will be trying to estimate v_i and H. The most reasonable assumption (according to Acton [1959]) about the distribution of the errors is that each $\binom{m_{ij}}{g_{ij}}$ have a joint normal distribution with mean 0 and unknown covariance matrix Σ. Errors arising from different observations are independent of each other. We will now show that our new model (1.3.5) is just another application of the model in Theorem 1.1.3.

If we let

$$x_{ij} = \binom{y_{ij}}{z_{ij}}, \quad e_{ij} = \binom{m_{ij}}{g_{ij}}, \quad \xi_i = \binom{v_i}{Hv_i},$$

our new model (1.3.5) can be rewritten as

$$x_{ij} = \xi_i + e_{ij}.$$

We also have a side condition that

$$(H, I)\xi_i = 0, \ \forall_i.$$

This formulation of (1.3.5) is very similar to Application 1, the only difference being that now we want the last columns of $\hat{B}$ to form the identity matrix as we did in the section on the consistency of the estimators. If we let

$$\hat{B}^* = (\hat{B}_2^{-1}B_1, I) = \hat{B}_2^{-1}(\hat{B}_1, \hat{B}_2) = \hat{B}_2(\hat{B}),$$

where $\hat{B} = (\hat{B}_1, \hat{B}_2)$ is the estimate of B from Application 1, then $\hat{B}^*$ is the only matrix with the correct form which maximizes the likelihood. $\hat{B}_2$ will be invertible with probability one; however if it is close to being singular (one of its eignevalues is very small), our results will be misleading. It would indicate that there is a strong internal relationship between the p-r variables composing y_{ij}. Since $\hat{B}^*$ is the only matrix of the correct form which is a maximum likelihood estimate, $-(\hat{B}^{(2)})^{-1}\hat{B}^{(1)}$ must be the maximum likelihood estimate of H. From Application 1 we can also get the MLE of ξ_i and Σ. Since v_i is the top p-r rows of ξ_i, we have the MLE of v_i. If we summarize the preceding statements, we have:

Application 4: If our model is given by

$$y_{ij} = v_i + m_{ij}; \ i = 1,2,\ldots,k; \ j = 1,2,\ldots,n_i$$
$$z_{ij} = Hv_i + g_{ij};$$

where y_{ij}, v_i, m_{ij}, z_{ij}, H and g_{ij} are defined in the paragraph following (1.3.5), then the MLE of H, v_i and Σ are given by

$$\hat{H} = (\hat{B}^{(2)})^{-1}\hat{B}^{(1)},$$

$$\hat{v}_i = \bar{y}_i - W\hat{B}'(\hat{B}W\hat{B}')^{-1}\hat{B}\begin{pmatrix}\bar{y}_i\\ \bar{z}_i\end{pmatrix},$$

$$\hat{\Sigma} = N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\begin{pmatrix}y_{ij}-\hat{v}_i\\ z_{ij}-\hat{H}\hat{v}_i\end{pmatrix}\begin{pmatrix}y_{ij}-\hat{v}_i\\ z_{ij}-\hat{H}\hat{v}_i\end{pmatrix},$$

where

$$W = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\begin{pmatrix}y_{ij}\\ z_{ij}\end{pmatrix} - \begin{pmatrix}\bar{y}_i\\ \bar{z}_i\end{pmatrix}\begin{pmatrix}y_{ij}\\ z_{ij}\end{pmatrix} - \begin{pmatrix}\bar{y}_i\\ \bar{z}_i\end{pmatrix})',$$

$$T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\begin{pmatrix}x_{ij}\\ z_{ij}\end{pmatrix}\begin{pmatrix}y_{ij}\\ z_{ij}\end{pmatrix})',$$

and the columns of $\hat{B}' = (\hat{B}^{(1)},\hat{B}^{(2)})'$ are the eigenvectors associated with the r smallest eigenvalues of $W^{-1}T$.

Remark: Application 4 is very similar to a model discussed by Gleser and Watson [1973]. In Chapter 3, we will be discussing models which are generalizations of Gleser and Watson's model.

We could make minor alterations on the model we just discussed. For instance, we could estimate parameters in the following model:

$$y_{ij} = v_i + m_{ij},$$
$$z_{ij} = Hv_i + \alpha + g_{ij},$$

where all the terms (except α) are defined in the previous application. The maximum likelihood estimates can be derived from Application 2 in a manner analagous to the way we derived the estimates for Application 4 from Application 1.

Similarly Application 3 could be extended to cover the following model:

$$y_{ij} = \mu_1 + v_i + m_{ij},$$
$$z_{ij} = \mu_2 + Hv_i + g_{ij},$$

where all the terms (except μ_1 and μ_2) are the same as in Application 1.

CHAPTER 2

TESTING THE EXISTENCE OF UNKNOWN LINEAR RESTRICTIONS IN THE CLASSICAL MULTIVARIATE LINEAR REGRESSION MODEL

2.0 Introduction

Let our model be the same model we considered in Chapter 1:

(2.0.1) $\quad X = \Xi F + E,$

where X is a pxN matrix of observations, Ξ is an unknown pxk($p \leq k \leq N$) parameter matrix, F is a known kxN matrix of covariates, and E is a pxN matrix of errors. We assume that each column of E is independent of any other column. We also assume that each column of E has a normal distribution with mean vector 0 and unknown covariance matrix Σ. In this chapter we are concerned with testing

(2.0.2) $\quad H_0: B\Xi = \alpha a$ against $H_1: B\Xi \neq \alpha a,$

where B is an unknown rxp matrix, α is an unknown rxs($s \leq r < p$) matrix, a is a known sxk matrix. We will derive results when a is of full row rank. For the case a is the zero matrix, i.e., when we test $B\Xi = 0$ versus $B\Xi \neq 0$, we will merely state our results since in this case all results can be derived in a way analagous to the case when a is of full row rank.

In Section 2.1 we will find the likelihood ratio test statistic of H_0 versus H_1, and mention similarities to test statistics of Rao [1965] and Kristoff [1973]. Section 2.2 will contain a discussion of the asymptotic distribution of the roots of which likelihood ratio statistic is a function. We will be concerned with cases when the number of parameters increases with the sample size. In Section 2.3, we will use the results of the preceding section to get the asymptotic distributions of the likelihood ratio test statistic and therefore asymptotic tests of H_0 vs. H_1. Section 2.4 will contain a proof that the tests described in Section 2.3. are consistent.

2.1. Likelihood Ratio Test Statistic

In this section we will be finding the likelihood ratio test of H_0: $B\Xi = \alpha a$ versus H_1: $B\Xi \neq \alpha a$ when our model is

$$X = \Xi F + E. \tag{2.1.1.}$$

All variables are defined in the introduction of this chapter.

In Chapter 1 we derived the maximum likelihood estimators of the parameters under H_0. If we substitute those estimators into the likelihood function (see (1.1.24) and (1.1.19)), we have

$$\max_{H_0} L(X,B,\Xi,\alpha,\Sigma) = (2\pi)^{-\frac{1}{2}pN} e^{-\frac{1}{2}pN} |W|^{-\frac{1}{2}N} (\lambda_p \cdot \lambda_{p-1} \cdots \cdot \lambda_{p-r+1})^{-\frac{1}{2}N}, \tag{2.1.2}$$

where

$$W = X(I-F'(FF')^{-1}F)X', \tag{2.1.3}$$

$$T = X(I-F'a'(aFF'a')^{-1}aF)X', \tag{2.1.4}$$

and λ_i is the ith largest eigenvalue of $W^{-1}T$. We may use standard multivariate regression procedures to get the maximum value of the likelihood function when H_1 is true:

$$(2.1.5) \qquad \max_{H_1} L(X,B,\Xi,\alpha,\Sigma) = (2\pi)^{-\frac{1}{2}pN} e^{-\frac{1}{2}pN} |W|^{-\frac{1}{2}N},$$

where W is defined above. If we combine (2.1.2) and (2.1.5) we will be able to get the likelihood ratio test statistic of H_0 versus H_1. Our result is summarized in the following theorem.

Theorem 2.1.1. If our model is given by (2.1.1) and we wish to test the hypothesis H_0: $B\Xi=\alpha a$ versus H_1: $B\Xi\neq\alpha a$ (a has full row rank), then the likelihood ratio test statistic is

$$\Lambda = \frac{\max_{H_0} L(X,B,\Xi,\alpha,\Sigma)}{\max_{H_1} L(X,B,\Xi,\alpha,\Sigma)} = (\lambda_p \cdot \lambda_{p-1} \cdot \ldots \cdot \lambda_{p-r+1})^{-\frac{1}{2}N},$$

where λ_i is the ith largest eigenvalue of $W^{-1}T$ and W and T are defined by (2.1.3) and (2.1.4) respectively.

Remark: When a is the zero matrix, the likelihood ratio test statistic is identical to that given in Theorem 2.1.1 except that T is equal to XX'.

We also have the following corollary:

Corollary 2.1.1. Let our model be

$$x_{ij} = \xi_i + e_{ij}; \; i = 1,2,\ldots,k; \; j = 1,2,\ldots,n_i;$$

where x_{ij} is a p-dimensional vector of observed values, ξ_i is the mean of the ith group of observations and e_{ij} is a p-dimensional

error vector. We assume that the errors are independently distributed with a normal distribution having mean vector 0 and unknown covariance matrix Σ. The likelihood ratio test statistic of the hypothesis H_0: $B\xi_i = \alpha$ versus H_1: $B\xi_i \neq \alpha$, where B is an unknown rxp matrix and α is an unknown rxs vector, is

$$\Lambda = (\lambda_p \cdot \lambda_{p-1} \cdot \ldots \cdot \lambda_{p-r+1})^{-\frac{1}{2}N},$$

where λ_i is the ith eigenvalue of $W^{-1}T$ and

$$W = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)(x'_{ij}-\bar{x}_i)',$$

$$T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij}-\bar{x})(x_{ij}-\bar{x})',$$

$$\bar{x}_i = (n_i)^{-1} \sum_{j=1}^{n_i} x_{ij}, \quad \bar{x} = N^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij},$$

$$N = \sum_{i=1}^{k} n_i.$$

Corollary 2.1.1 follows from Theorem 2.1.1 just as Application 2 followed from Theorem 1.1.2 in Chapter 1.

The reason we mentioned Corollary 2.1.1 is that the hypothesis we are testing in that corollary is exactly the hypothesis of dimensionality in Rao [1973]. Rao derives the likelihood ratio test statistic when Σ is known. He does not derive the likelihood ratio test when Σ is unknown however he does give an alternative test which is also based on the smallest roots of $W^{-1}T$. He gives an asymptotic test based on his test statistic which is valid only when k the number of groups is fixed. If we use the model in

Corollary 2.1.1 we may believe that the number of groups should increase when the sample size increases. The asymptotic test in this case would not be the same as when k is fixed (see Section 2.3).

Kristoff [1973] considered testing an unspecified linear relationship in several models. In the basic model (his case 1), we measure a person's scores on two tests. We assume there is an equivalent form of each test available. A person's scores are equal to that person's abilities (true scores) plus an error term. We summarize this model with the following equation:

$$x_{ij} = \mu+\xi_i+e_{ij};\ i = 1,2,\ldots,k;\ j = 1,2;\ N = 2k;$$

where x_{ij} is a 2-dimensional vector whose elements are the ith person's scores on the jth form of the two tests, μ is the average person's true scores on the two tests (it is the same for either form of the two tests), ξ_i is the difference between the ith person's true scores and the average person's true scores, and e_{ij} is the error term. The error terms are all pairwise independent. Each has a normal distribution with mean vector 0 and unknown covariance matrix Σ. We wish to test the hypothesis that a single unspecified linear relation exists against the hypothesis that none exists, i.e., we test

$$H_0:\ B\xi_i = 0,\ \forall_i \quad \text{versus} \quad H_1:\ B\xi_i \neq 0,\ \text{for some } i,$$

where B is an unknown 2-dimensional row vector. We found the maximum likelihood estimators of the parameters when H_0 is true in Application 3 of Chapter 1 under the assumption that $\sum_{i=1}^{k} \xi_i = 0$. We could also get

the MLE's when H_1 is true using the usual theory of multivariate linear regression. If we put these together we would get the likelihood ratio test statistic which turns out to be a function of the smallest eignevalue of $W^{-1}T$; W and T are given in Application 3. The smallest eigenvalue of $W^{-1}T$ is the same statistic Kristoff recommends. If we were to increase our sample size in this example, we would probably increase the number of people in our sample and not the number of equivalent forms of each test, that is, we would assume that k increases as N does and that

$$\lim_{N\to\infty} \frac{N-k}{N} = \lim_{k\to\infty} \frac{2k-k}{2k} = 1/2.$$

This example, therefore, provides us with a situation in which the number of parameters does not stay fixed as the sample size increases.

In case 2 of Kristoff, we have exactly the same model as the above with one minor change - the difference between the score on one form of the two tests and the score on the other form of the two tests need not have expectation 0. Our model is

$$X_{ij} = \xi_i + \sigma_j + e_{ij};\ i = 1,2,\ldots,k;\ j = 1,2;$$

where X_{ij}, ξ_i, and e_i; are defined as before and σ_j is the expected true score on the jth form of the two tests. Again we will test H_0: $B\xi_i = 0$, $\forall_i$ versus H_1: $B\xi_i \neq 0$, for some i, where B is an unknown 2-dimensional vector. If we estimate σ_j first, we can find the MLE's of the parameters when the hypothesis is true. (See Application 3 of Chapter 1 for the type of argument needed.)

When the hypothesis is false, it is also easy to get the MLE's. The likelihood ratio test statistic is a function of the smallest eigenvalue of $W^{-1}T$, where

$$W = \sum_{i=1}^{k} \sum_{j=1}^{2} (x_{ij}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x})(x_{ij}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x})',$$

$$T = \sum_{i=1}^{k} \sum_{j=1}^{2} (x_{ij}-\bar{x}_{\cdot j})(x_{ij}-\bar{x}_{\cdot j})',$$

$$\bar{x}_{i\cdot} = (1/2)(x_{i1}+x_{i2}),\ \bar{x}_{\cdot j} = k^{-1} \sum_{i=1}^{k} x_{ij},$$

$$\bar{x} = (2k)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{2} x_{ij}.$$

The smallest eigenvalue of $W^{-1}T$ is also the statistic Kristoff recommends.

2.2. Asymptotic Distributions of the Roots

In this section, we find the asymptotic distribution of the roots needed in the likelihood ratio tests under the null hypothesis, H_0: $B\Xi = \alpha a$. The roots in which we are interested are the smallest roots of

$$|T-\lambda_N W| = 0, \tag{2.2.1}$$

where

$$T = X(I_N-F'a'(aFF'a')^{-1}aF)X', \tag{2.2.2}$$

$$W = X(I_N-F'(FF')^{-1}F)X'. \tag{2.2.3}$$

Throughout this section, certain variables are subscripted with an "N" indicating that those variables are connected with a sample of

size N. We will let λ_{iN} be the ith root of (2.2.1) when our sample is of size N.

It is helpful to find the distribution of the smallest roots of

$$|T-W-\phi_N W| = 0.$$

Note that if ϕ_{iN} is the ith largest root of the above expression, then $\phi_{iN} + 1 = \lambda_{iN}$, where λ_{iN} is the ith largest root of (2.2.1). All our theorems are results in terms of ϕ_{iN}. We will assume that a is of full row rank.

In this section, we discuss cases when the number of parameters increases with the sample size. We already mentioned that the models of Kristoff [1973] provide us with examples where it is reasonable to assume that the number of parameters increases with the sample size. A measure of how fast the number of parameters increases will be

$$\lim_{N\to\infty} \frac{N-k}{N} = 1-\lim_{N\to\infty} \frac{K}{N} = 1-(1-t) = t.$$

There are three possible cases:

Case 1: k is fixed;

Case 2: $t \neq 1$;

Case 3: k goes to infinity as N does; $t = 1$.

We will always assume that r (the number of rows of B), p (the number of rows of X) and s (the row rank of a) are fixed.

When k is fixed, the asymptotic distribution of the r smallest roots (from Anderson [1951b] and from Hsu [1941]) is the following:

Theorem 2.2.1. Let $\rho_{iN} = N-\phi_{iN}$; $i = p-r+1,\ldots,p$; where ϕ_{iN} is the ith largest root of $|T-W-\phi_N W| = 0$. Then the limiting distribution of $(\rho_{p-r+1,N},\rho_{p-r+2,N},\ldots,\rho_{pN})$ when k is fixed is

$$\frac{2^{-\frac{1}{2}} r(k-s-p+r)_{(\pi)}\frac{1}{2}(r) \prod_{i=p-r+1}^{p} \rho_{iN}^{\frac{1}{2}(k-s-p-1)} e^{-\sum_{i=p-r+1}^{p} \rho_{iN}/2}}{\prod_{i=1}^{r} \Gamma(\frac{1}{2}(k-s-p+r-1-i))\Gamma(\frac{1}{2}(r+1-i))} \cdot$$

$$\prod_{i=p-r+1}^{p} \prod_{j=i+1}^{p} (\rho_{iN}-\rho_{jN}).$$

The above distribution is the same as the joint distribution of ρ_i where ρ_i is the ith largest root of

$$|J - \rho I| = 0,$$

and J is defined by

$$J = \sum_{i=1}^{k-s-p+r} u_i u_i',$$

where the u_i are independently distributed with a normal distribution with mean 0 and covariance matrix I_r.

Remark. Theorem 2.2.1 can also be used when a is the zero matrix by letting s = 0.

We now derive the asymptotic distribution of the roots in Case 2 and Case 3. The asymptotic distribution of the smallest roots in these cases is markedly different than the distribution of the roots given in Theorem 2.2.1. Before we state and prove several theorems which give the asymptotic distribution of the roots in Cases 2 and 3, we need to derive several lemmas.

Lemma 1. Let our model and hypothesis be given by

$$X = \Xi F + E,$$

$$B\Xi F = \alpha a,$$

where X, Ξ, F, E, B, α, and a are defined in the introduction to this chapter. The roots of

$$|T-W-\phi_N W| = 0,$$

where T,W are given by (2.2.2) and (2.2.3), have the same distribution as the roots of

(2.2.4) $$|(N-k)^{-1} U^*U^{*\prime} + N^{\frac{1}{2}}(N-k)^{-1}C-\phi_N(N-k)^{-\frac{1}{2}} Z+D_1| = 0,$$

where

$$C = \begin{pmatrix} C_{11} & \cdots & C_{p-r,1} & E_1' \\ \vdots & & \vdots & \vdots \\ C_{p-r,1} & \cdots & C_{p-r,p-r} & E_{p-r}' \\ E_1 & \cdots & E_{p-r} & 0 \end{pmatrix},$$

$$C_{hh'} = \sqrt{\gamma_{hN}}\, U^*_{h'h} + \sqrt{\gamma_{h'N}}\, U^*_{hh'},$$

$$E_h = \sqrt{\gamma_{hN}} \begin{pmatrix} U^*_{h,p-r+1} \\ U^*_{h,p-r+2} \\ \vdots \\ U^*_{hp} \end{pmatrix},$$

$$D_1 = \begin{pmatrix} \frac{N}{N-k}\gamma_{1N}-\phi_N & 0 & \cdots & 0 \\ 0 & \frac{N}{N-k}\gamma_{2N}-\phi_N & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & -\phi_N I_r \end{pmatrix},$$

$$Z = (N-k)^{\frac{1}{2}}(VV' - (N-k)I_p),$$

and γ_{iN} is the ith largest eigenvalue of

$$N^{-1}(\Xi F(I-F'a'(aFF'a')^{-1}aF)F'\Xi)\Sigma^{-1},$$

and finally U*, V are pxk-s and pxN-k respectively whose columns have independent normal distributions with mean vector 0 and covariance matrix I_p.

Proof. For any invertible pxp matrix θ we know that

$$|\theta(T-W)\theta'-\phi\theta W\theta'| = |T-W-\phi W| = 0;$$

the roots ϕ are the same whether we observe X or θX. Since we may pick θ so that $\theta\Sigma\theta' = I_p$, we may assume, without loss of generality, that the columns of X have p-variate normal distributions with mean vectors equal to the respective columns of $\theta\Xi F$ and with common covariance matrix I_p.

Next we will let V_1, V_2, V_3 be column orthogonal matrices such that

$$V_1 V_1' = F'a'(aFF'a')^{-1}aF,$$

$$V_2 V_2' = I-F'(FF')^{-1}F,$$

$$V_3 V_3' = F'(FF')^{-1}F-F'a'(aFF'a')^{-1}aF.$$

It is easy to see that such matrices exist. Let

$$Y = (Y_1,Y_2,Y_3) = (XV_1,XV_2,XV_3) = X(V_1,V_2,V_3),$$

where Y_1 is pxs, Y_2 is pxk-s, and Y_3 is pxN-k. Since (V_1,V_2,V_3) is an orthogonal matrix, each column of Y has an independent normal distribution with covariance matrix I_p.

The roots for which we are getting the asymptotic distribution are functions of Y_2 and Y_3. Since Y_1 is independent of (Y_2,Y_3), we may eliminate it from our considerations. The distribution of Y_2, Y_3 is constant $\cdot \exp -\frac{1}{2}[\mathrm{tr}(Y_2-\theta\equiv FV_2)(Y_2-\theta\equiv FV_2)' + (Y_3-\theta\equiv FV_3)(Y_3-\theta\equiv FV_3)'] =$ constant $\cdot \exp -\frac{1}{2}[\mathrm{tr}\ Y_2Y_2' - (Y_3-\theta\equiv FV_3)(Y_3-\theta\equiv FV_3)']$.

We want the distribution of the roots $(\phi_{1N},\phi_{2N},\ldots,\phi_{pN})$ of

$$|T-W-\phi_N W| = |Y_3Y_3'-\phi_N Y_2Y_2'|.$$

We know that Y_2Y_2' has a central Wishart distribution and that Y_3Y_3' has a noncentral Wishart distribution.

We now let

(2.2.5) $$G = \theta\equiv FV_3.$$

We can write the noncentrality parameter in the distribution of Y_3Y_3' in terms of GG':

(2.2.6) $$GG' = \theta\equiv FV_3V_3'\equiv'\theta'=\theta\equiv F(F'(FF')^{-1}F-F'a'(aFF'a')^{-1}aF)F'\equiv'\theta',$$

$$= \theta\equiv F(I-F'a'(aFF'a')^{-1}aF)F'\equiv'\theta'.$$

Let $(\gamma_{1N},\gamma_{2N},\ldots,\gamma_{pN})$ be the ordered eigenvalues of $N^{-1}GG'$. Under the hypothesis that $B\equiv = \alpha a$, the r smallest eigenvalues of $N^{-1}GG'$ will equal zero.

Next, we transform Y_2 and Y_3 in such a way that only a few elements of the resulting matrices are dependent on the non zero eigenvalues of $N^{-1}GG'$. Consider

$$U = \Gamma_1 Y_3 \Gamma_2, \text{ and } V = \Gamma_1 Y_2,$$

where Γ_1 and Γ_2 are orthogonal matrices which make

$$\Gamma_1 G \Gamma_2 = \sqrt{N} \begin{pmatrix} \sqrt{\gamma_{1N}} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\gamma_{2N}} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{\gamma_{p-r,N}} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} .$$

Since Γ_1 and Γ_2 are both orthogonal matrices, the distribution of U and V is

$$(2.2.7) \quad \text{constant} \cdot \exp - \frac{1}{2}\left(t - (UU' + VV') + \sum_{i=1}^{p-r} \sqrt{N \cdot \gamma_{iN}}\ u_{ii} + N \sum_{i=1}^{p-r} \gamma_{iN}\right),$$

where u_{ii} is the i,ith element of U.

We want the distribution of the roots of

$$(2.2.8) \quad |UU' - \phi_N VV'| = |\Gamma_1 Y_3 Y_3' \Gamma_1' - \phi_N \Gamma_1 Y_2 Y_2' \Gamma_1'| = |Y_3 Y_3' - \phi_N Y_2 Y_2'| = 0.$$

We should mention that U is pxk-s and that V is pxN-k.

Finally, we will make several substitutions which will give us our lemma. Let

$$U^* = U - \Gamma_1 G \Gamma_2 .$$

The joint distribution of U* and V is

$$(2.2.9) \quad \text{constant} \cdot \exp\left[-\frac{1}{2} \operatorname{tr}(U^* U^{*\prime} + VV')\right],$$

i.e., the columns of U*, which is pxk-s, and the columns of V, which is pxN-k, are independently distributed with covariance matrix I_p and mean vector 0. We therefore have

(2.2.10) $UU' = U^*U^{*\prime} + \sqrt{N}\, C + N(D_1+\phi_N \cdot I_p)$,

where C and D_1 are given in the statement of this lemma. Our final substitution is

(2.2.11) $Z = (N-k)^{-\frac{1}{2}}(VV'-(N-k)I_p)$,

$= (N-k)^{-\frac{1}{2}}(VV')-(N-k)^{\frac{1}{2}} I_p$.

The lemma now follows through a substitution of (2.2.10) and (2.2.11) into (2.2.8). Q.E.D.

Lemma 2: If P is a pxk matrix and each column of P has a normal distribution with mean 0 and covariance matrix I_p, then each element of the matrix $(k)^{-\frac{1}{2}}(PP'-kI_p)$ asymptotically has a normal distribution with mean 0, variance 1 for diagonal elements and variance 2 for off-diagonal elements. All elements are asymptotically independent. We will call this asymptotic distribution the p-dimensional matrix normal distribution.

Proof. Use Theorem 4.2.4 in Anderson [1958].

Remark. If we let $(PP')_{22}$ be the rxr matrix which comprises the lower right hand corner of PP', then

$$k^{-\frac{1}{2}}((PP')_{22}-kI_r)$$

has an r dimensional matrix normal distribution.

We now state several assumptions which we will make when we discuss the asymptotic distribution of the roots for Cases 2 and 3.

Assumption 1. The matrix $N^{-1}GG'$ defined by (2.2.6) converges to a finite matrix R which has rank p-r.

Assumption 2. $\gamma_{iN} = \gamma_i + o(\sqrt{N})$, where γ_i is the ith largest eigenvalue of R and γ_{iN} is the ith largest eigenvalue of $N^{-1}GG'$.

Assumption 3. The non zero roots of R have multiplicity one.

Assumption 3 is not necessary; the proof given below would have to be altered to apply when the non zero roots of R do not have multiplicity one. Since the alterations only complicate matters, and since they do not affect the distribution of the smallest roots, we will omit them.

Part 1. Case 2: $\lim_{N\to\infty} \frac{N-k}{N} = t \neq 1$.

In the following theorem, we give the asymptotic distribution of the roots when $\lim_{N\to\infty} \frac{N-k}{N} = t \neq 1$.

Theorem 2.2.2. Assume that $\lim_{N\to\infty} \frac{N-k}{N} = t < 1$, and that Assumptions 1, 2 and 3 hold. Let

$$\nu_{p-r+i,N} = (N-k)^{\frac{1}{2}}(\phi_{p-r+i,N} - (k/N-k)); \quad i = 1,2,\ldots,r;$$

where ϕ_{iN} is the ith largest root of $TW^{-1} - I_p$. The limiting distribution of $(\nu_{p-r+1,N}, \nu_{p-r+2,N}, \ldots, \nu_{pN})$ is the same as the distribution of the r roots from

$$|(1/t-1)^{\frac{1}{2}} Q_1 - (1/t-1)Q_2 - \nu I_r| = 0,$$

where Q_1 and Q_2 have the r-dimensional matrix normal distribution (see Lemma 2).

Proof. By Lemma 1, we only have to consider the distribution of the r smallest roots $(\phi_{iN};\ i = p-r+1, p-r+2, \ldots, p)$ of

$$(2.2.12)\quad |(N-k)^{-1}U^*U^{*\prime}-N^{\frac{1}{2}}(N-k)^{-1}C-\phi_N(N-k)^{\frac{1}{2}}Z+D_1| = 0,$$

where C, Z, and D_1 are defined in Lemma 1, and the columns of U*, V have independent normal distributions with mean vector 0 and covariance matrix I_p.

Consider the following matrix:

$$A = k^{-\frac{1}{2}}(U^*U^{*\prime}-kI_p).$$

If we substitute A into (2.2.12), we get the following equation:

$$(2.2.13)\quad |(k)^{\frac{1}{2}}(N-k)^{-1}A+N^{\frac{1}{2}}(N-k)^{-1}C-\phi_N(N-k)^{-\frac{1}{2}}Z+D_1+k(N-k)^{-1}I_p| = 0.$$

By Lemma 2, A and Z have p-dimensional matrix normal distributions. It is easy to see that C is asymptotically independent of A and Z. The elements of C are functions of the first p columns of U*. Since A is the same asymptotically if we delete the first p columns of U*, C and A can be thought of as functions of different variables asymptotically. The asymptotic distribution of C can be obtained by using the definition of C.

Consider the following variable:

$$(k/N-k)+\nu_{p-r+i}/\sqrt{N-k}.$$

We can substitute the above expression into (2.2.13) for ϕ_N and try to find out what ν_{p-r+i} must be distributed as when $N \to \infty$. Our result is

$$(2.2.14)\quad |k^{\frac{1}{2}}(N-k)^{-1}A+N^{\frac{1}{2}}(N-k)^{-1}C+(k(N-k)^{-1}+(N-k)^{-\frac{1}{2}}\nu_{p-r+i})Z+D_2|=0,$$

where

$$D_2 = \begin{pmatrix} \frac{N}{N-k}\gamma_{1N} - \frac{\nu_{p-r+i}}{\sqrt{N-k}} & 0 & \cdots & 0 \\ 0 & \frac{N}{N-k}\gamma_{2N} - \frac{\nu_{p-r+i}}{\sqrt{N-k}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{-\nu_{p-r+i}}{\sqrt{N-k}} \cdot I_r \end{pmatrix}.$$

Let us write

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad C = \begin{pmatrix} \mathcal{C}_{11} & \mathcal{C}_{12} \\ \mathcal{C}_{21} & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix},$$

where A_{11}, $\mathcal{C}_{11}$, Z_{11} are all p-rxp-r; A_{12}, $\mathcal{C}_{12}$ and Z_{12} are all p-rxr, etc. We now discuss the upper left hand block of p-rxp-r elements inside (2.2.14). When N is large, $k^{\frac{1}{2}}(N-k)^{-1}A_{11}$, $N^{\frac{1}{2}}(N-k)^{-1}\mathcal{C}_{11}$ and $(N-k)^{-\frac{1}{2}}Z_{11}$ all are arbitrarily small. The only matrix which remains is

$$\begin{pmatrix} \frac{N}{N-k}\gamma_{1N} & 0 & \cdots & 0 \\ 0 & \frac{N}{N-k}\gamma_{2N} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \frac{N}{N-k}\gamma_{p-r,N} \end{pmatrix} \to \begin{pmatrix} 1/t\,\gamma_1 & 0 & \cdots & 0 \\ 0 & 1/t\,\gamma_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/t\,\gamma_{p-r} \end{pmatrix}.$$

The elements in the last r rows and columns all go to 0 when N gets large. If we multiply the last r rows and columns by $(N-k)^{\frac{1}{4}}$, we will be able to find the terms that dominate. First of all,

$(N-k)^{\frac{1}{4}}[k^{\frac{1}{2}}(N-k)^{-1}A_{12}] \overset{a.s.}{\to} 0$, $(N-k)^{\frac{1}{4}}[(N)^{\frac{1}{2}}(N-k)^{-1}\mathcal{C}_{12}] \overset{a.s.}{\to} 0$ and

$(N-k)^{\frac{1}{4}}[(N-k)^{-\frac{1}{2}}-Z_{12}] \overset{a.s.}{\to} 0.$

Almost sure convergence is indicated by $\overset{a.s.}{\to}$.

When we multiply the r rows and columns by $(N-k)^{\frac{1}{4}}$, we multiply the lower right-hand corner by $(N-k)^{\frac{1}{2}}$. By Lemma 2 we know that

$$\frac{\sqrt{N-k}}{N-k}\sqrt{k}\,A_{22} \xrightarrow{L} (1/t-1)^{\frac{1}{2}} Q_1,$$

where $\xrightarrow{L}$ indicates convergence in distribution and Q_1 has a r-dimensional matrix normal distribution. Similarly

$$\frac{k}{N-k} Z_{22} \xrightarrow{L} (1/t-1)Q_2,$$

where Q_2 has a r-dimensional matrix normal distribution. All other terms go to zero.

If we combine the above statements, we get that when N is large (2.2.14) becomes

$$\left|\begin{pmatrix} 1/t\,\gamma_1 & 0 & \cdots & 0 \\ 0 & 1/t\,\gamma_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & Q_3 \end{pmatrix}\right| = 0,$$

where

$$Q_3 = (1/t-1)^{\frac{1}{2}} Q_1-(1/t-1)Q_2-\nu_{p-r+i}I_r.$$

Therefore ν_{p-r+i} is a root of $|Q_3| = 0$. The distribution of $(\nu_{p-r+1}, \nu_{p-r+2},\ldots, \nu_p)$ is the distribution of the roots of

$$|(1/t-1)^{\frac{1}{2}} Q_1-(1/t-1)Q_2-\nu I_r| = 0,$$

where Q_1 and Q_2 have r-dimensional matrix normal distributions.

Consider

$$\phi_{iN} = \frac{k}{N-k} + \nu_{iN}/\sqrt{N-k}\ ;\quad i = p-r+1,p-r+2,\ldots,p;$$

where ν_{iN} is picked so that we have equality in the above expression, i.e., ν_{iN} is defined by ϕ_{iN}:

$$\nu_{iN} = (\phi_{iN} - \frac{k}{N-k}) \sqrt{N-k} = f_{iN}(A,Z,C).$$

We now show that the distribution of ν_{iN} goes to the distribution of ν_i.

Our preceding discussion shows that when

$$\lim_{N\to\infty} A_{22} = Q_1, \quad \lim_{N\to\infty} Z_{22} = Q_2,$$

and A, Z and C all converge to finite matrices,

$$\lim_{N\to\infty} \nu_{iN} = \nu_i = f_i(Q_1,Q_2).$$

We have mentioned the limiting distributions of A, Z, C. The set of discontinuities of

$$(\nu_{p-r+1},\nu_{p-r+2},\ldots,\nu_p) = (f_{p-r+1}(Q_1,Q_2),\ldots,f_p(Q_1,Q_2)).$$

occur only when one or more of the roots $(\nu_{p-r+1},\nu_{p-r+2},\ldots,\nu_p)$ are equal; the set of discontinuities has measure 0 since the probability any of the roots are equal is zero. By applying Rubin's theorem (see Anderson [1951b]), we have that the asymptotic distribution of $(\nu_{p-r+1,N},\nu_{p-r+2,N}\ldots,\nu_{pN})$ is the same as the distribution of $(\nu_{p-r+1},\nu_{p-r+2},\ldots,\nu_p)$. Q.E.D.

Using Theorem 2 in Anderson [1951b] and the above theorem, Theorem 2.2.2, we conclude the following:

Theorem 2.2.3. Assume that $\lim_{N\to\infty} \frac{N-k}{N} = t < 1$ and that Assumptions 1, 2 and 3 hold. Let

$$\rho_{iN} = (\phi_{iN}-k/N-k)\ \frac{\sqrt{N}\ \sqrt{k}}{\sqrt{N-k}}\ ;\quad i = p-r+1,p-r+2,\ldots,p;$$

where ϕ_{iN} is the largest root of $TW^{-1}-I_p$. The limiting distribution of $(\rho_{p-r+1,N},\rho_{p-r+2,N},\ldots,\rho_{pN})$ is

$$2^{-r/2}\Big[\prod_{i=1}^{r}(\Gamma\tfrac{1}{2}(r+1-i))\Big]^{-1} e^{-\sum_{i=p-r+1}^{p}\rho_{iN}} \prod_{i=p-r+1}^{p}\prod_{j=i+1}^{r}(\rho_{iN}-\rho_{jN}).$$

Part 2: Case 3: $k \to \infty$ as $N \to \infty$ but $\lim_{N\to\infty} \frac{N-k}{N} = 1$.

The following theorem will contain the asymptotic distribution of the roots when k goes to infinity as N does and $\lim_{N\to\infty}(N-k)/N = 1$.

Theorem 2.2.4. Assume that $k \to \infty$ as $N \to \infty$, that $\lim_{N\to\infty} \frac{N-k}{N} = 1$ and that Assumptions 1, 2 and 3 hold. Let

$$\nu_{iN} = (\phi_{iN} - \frac{k}{N-k})Nk^{-\frac{1}{2}};\quad i = p-r+1,\ p-r+2,\ldots,p;$$

where ϕ_{iN} is the ith largest root of $TW^{-1}-I_p$. Then the limiting distribution of $(\nu_{p-r+1,N},\nu_{p-r+2,N},\ldots,\nu_{pN})$ is the same as the distribution of the r roots of

$$|Q-\nu\ I_r| = 0$$

where Q has an r-dimensional matrix normal distribution (see Lemma 2).

Proof. Because the proof of this theorem is very similar to the proof of Theorem 2.2.2, we will only give an outline of the proof.

By Lemma 1, we only have to consider the distribution of the r smallest roots $(\phi_{iN};\ i = p-r+1,\ p-r+2,\ldots,p)$ of

$$(2.2.15)\quad |(N-k)^{-1}U^*U^{*\prime}+N^{\frac{1}{2}}(N-k)^{-1}C-\phi_N(N-k)^{-\frac{1}{2}}Z+D_1| = 0,$$

where C, Z and D_1 are defined in Lemma 1 and the columns of U^*, V have independent normal distribution with mean vector 0 and covariance matrix I_p.

Consider the following matrix:

$$A = k^{-\frac{1}{2}}(U^*U^{*\prime}-kI_p).$$

Substituting A into (2.2.15) yields

$$(2.2.16)\quad \left|\sqrt{k}/N-k\ A + \sqrt{N}/N-k\ C-\phi_N \frac{Z}{\sqrt{N-k}} +D_1 + \frac{k}{N-k} I_p\right| = 0.$$

We now consider the following variable,

$$k/N-k + v_{p-r+i}/Nk^{-\frac{1}{2}}.$$

Substituting the above into (2.2.16), we obtain

$$(2.2.17)\quad \left|\frac{\sqrt{k}}{N-k} A + \frac{\sqrt{N}}{N-k} C - \left(\frac{k}{N-k} + \frac{v_{p-r+i}}{Nk^{-\frac{1}{2}}}\right)Z/\sqrt{N-k} + D_2\right| = 0,$$

where

$$D_2 = \begin{pmatrix} \frac{N}{N-k}\gamma_{1N} - \frac{v_{p-r+i}}{Nk^{-\frac{1}{2}}} & 0 & \cdots & 0 \\ 0 & \frac{N}{N-k}\gamma_{2N} - \frac{v_{p-r+i}}{Nk^{-\frac{1}{2}}} & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & \frac{-v_{p-r+i}}{Nk^{-\frac{1}{2}}} I_r \end{pmatrix}.$$

If we multiply the last r rows and columns of the matrix inside the determinant of (2.2.17) by $N^{\frac{1}{2}}k^{-\frac{1}{4}}$, and let N go to infinity, we get

$$\left|\begin{pmatrix} \gamma_1 & 0 & \dots & 0 \\ 0 & \gamma_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & Q \end{pmatrix}\right| = 0,$$

where

$$Q = Q_1 - \nu_{p-r+i} \cdot I_r = \lim_{N\to\infty} A_{22} - \nu_{p-r+i} \cdot I_r,$$

$$Q_1 = \lim_{N\to\infty} A_{22},$$

and A_{22} is the lower rxr right-hand corner of A. Therefore, ν_{p-r+i} is a root of $|Q| = 0$. By Lemma 2, the distribution of $(\nu_{p-r+1}, \nu_{p-r+2}, \dots, \nu_p)$ is the distribution of the roots of

$$|Q_1 - \nu I_r| = 0$$

where Q_1 has the r dimensional matrix normal distribution.

All that we have to show is that the $\nu_{p-r+i,N}$ which gives us equality in

$$\phi_{p-r+iN} = \frac{k}{N-k} + \frac{\nu_{p-r+i,N}}{Nk^{-\frac{1}{2}}}$$

goes in law to ν_{p-r+i}. The demonstration of this fact for Case 3 is the same as for Case 2. We therefore have our theorem. Q.E.D.

If we use Theorem 2 in Anderson [1951b] and Theorem 2.2.5, we can conclude the following:

Theorem 2.2.5. Assume that $\lim_{N\to\infty} \frac{N-k}{N} = 1$, that $k \to \infty$ as $N \to \infty$, and that Assumptions 1, 2, 3 hold. Let

$$\nu_{iN} = (\phi_{iN} - k/N-k)Nk^{-\frac{1}{2}}; \quad i = p-r+1, p-r+2, \dots, p;$$

where ϕ_{iN} is the ith largest root of $TW^{-1} - I_p$.

Then the limiting distribution of the set $(\nu_{iN};\ i = p-r+1,\ldots,p)$ is

$$2^{-r/2}\Big[\prod_{i=1}^{r}\Gamma(\tfrac{1}{2}(r+1-i))\Big]^{-1} e^{-\sum_{i=p-r+1}^{p}\nu_{iN}} \prod_{i=p-r+1}^{p}\prod_{j=i+1}^{p}(\nu_{iN}-\nu_{jN}).$$

Remark: Theorems 2.2.2 through 2.2.5 are valid when a is the zero matrix.

2.3. Asymptotic Tests of $B\Xi = \alpha a$ Versus $B\Xi \neq \alpha a$

In this section we use the asymptotic distribution given in the previous section to get asymptotic tests based on the likelihood ratio test statistic. It should be recalled from Theorem 2.1.1 that the likelihood ratio test statistic is given by

$$\Lambda = \prod_{i=p-r+1}^{p}(1/\lambda_{iN})^{\frac{1}{2}N} = \prod_{i=p-r+1}^{p}(1/1+\phi_{iN})^{\frac{1}{2}N},$$

where λ_{iN} is the ith largest eigenvalue of TW^{-1} and ϕ_{iN} is the ith largest eigenvalue of $TW^{-1}-I_p$.

First, let us consider the case when k is a fixed quantity:

Theorem 2.3.1. (Anderson [1951a]) If our model is given by (2.0.1) and we wish to test the hypothesis that H_0: $B\Xi = \alpha a$ versus $B\Xi \neq \alpha a$, then the asymptotic null distribution of

$$\Psi = -2 \log \Lambda$$

is a χ^2 distribution with $r(k-s-(p-r))$ degrees of freedom. The α-level asymptotic test of H_0: $B\Xi = \alpha a$ versus $H_1 = B\Xi \neq \alpha a$ would be to reject the hypothesis H_0 when

$$\Psi \geq \chi^2_{r(k-s-p+r)}(1-\alpha)$$

where $\chi^2_d(\beta)$ is the βth fractile of a χ^2 distribution with d degrees of freedom.

Remark: Theorem 2.3.1 holds when a is the zero matrix if we let $s = 0$.

Let us assume we are actually in Case 2, i.e., $\lim_{N\to\infty} (N-k)/N = t < 1$, and we (mistakenly) try to use the test given in Theorem 2.3.1. We now examine what happens to Ψ under H_0 when N is large. Note that

$$\Psi = -2 \log \Lambda = -2 \log \prod_{i=p-r+1}^{p} (1/1+\phi_{iN})^{\frac{1}{2}N}$$

$$= N \sum_{i=p-r+1}^{p} \log(1+\phi_{iN}).$$

Using Theorem 2.2.2 we can show that $1+\phi_{iN}$ goes almost surely to $1/t$ for $i = p-r+1, p-r+2,\ldots,p$. We therefore know that $\sum_{i=p-r+1}^{p} \log(1+\phi_{iN})$ goes almost surely to $r \log(1/t)$. Since $N \cdot r(\log 1/t)$ goes to positive infinity, we conclude that when H_0 is true, Ψ gets arbitrarily large in this case as N goes to infinity. If we were to apply the test given in Theorem 2.3.1 for Case 2, our probability of rejecting H_0, even when it is true, approaches 1 as N approaches infinity. For Case 3 we have a similar result. We summarize the preceding statements in the following theorem:

Theorem 2.3.2. If our model is given by (2.0.1) and we wish to test $B\Xi = \alpha a$ versus $B\Xi \neq \alpha a$ under the assumption that k goes to

infinity as N does, then $\Psi = -2 \log \Lambda$ where Λ is the likelihood ratio test statistic goes almost surely to positive infinity. The test given in Theorem 2.3.1 is meaningless in this case; when H_0 holds, we would reject H_0 almost surely in large samples.

Since Ψ does not have an asymptotic chi square distribution in either Case 2 or Case 3, we have to derive separate asymptotic tests for Case 2 and for Case 3.

Assume that we are in Case 2. For this case, we have the following theorem:

Theorem 2.3.3. If our model is given by (2.0.1) and we wish to test the hypothesis H_0: $B\Xi = \alpha a$ versus H_1: $B\Xi \neq \alpha a$ when $\lim_{N\to\infty} (N-k)/N = t < 1$ then the asymptotic null distribution of

$$((N)(N-k))^{\frac{1}{2}}(2rk)^{-\frac{1}{2}}((\frac{N}{N-k})^{r}\Lambda^{2/N} - 1),$$

where Λ is the likelihood ratio test statistic, is a normal distribution with mean 0 and variance 1. The asymptotic test of H_0: $B\Xi = \alpha a$ versus H_1: $B\Xi \neq \alpha a$ would be to reject H_0 when

$$(N(N-k))^{\frac{1}{2}}(2kr)^{-\frac{1}{2}}((\frac{N}{N-k})^{r}\Lambda^{2/N} - 1) > Z_{1-\alpha},$$

and do not reject otherwise, where Z_β is the β fractile of a standard normal distribution.

Proof: Consider the following sequence of statements:

$$\Lambda^{2/N} = \prod_{i=p-r+1}^{p} (1/\lambda_{iN}) = \prod_{i=p-r+1}^{p} (1/1+\phi_{iN}),$$

$$= \prod_{i=p-r+1}^{p} (1/(1+k/(N-k)+\nu_{iN}/(N-k)^{\frac{1}{2}})),$$

$$= \prod_{i=p-r+1}^{p} (\frac{N-k}{N})(1/(1+(N-k)^{\frac{1}{2}}N^{-1}\nu_{iN})),$$

$$= \prod_{i=p-r+1}^{p} (\frac{N-k}{N})(1-(N-k)^{\frac{1}{2}}N^{-1}\nu_{iN} + O(N)),$$

$$= (\frac{N-k}{N})^{r}(1-(N-k)^{\frac{1}{2}}N^{-1} \sum_{i=p-r+1}^{p} \nu_{iN} + O(N)).$$

The above equality can be written:

$$((\frac{N}{N-k})^{r}(\Lambda^{2/N}) -1)N/(N-k)^{\frac{1}{2}} = - \sum_{i=p-r+1}^{p} \nu_{iN} + O(N^{\frac{1}{2}}).$$

The asymptotic distribution of $\sum_{i=p-r+1}^{p} \nu_{iN}$ can be easily obtained from Theorem 2.2.2. The limiting distribution of $\sum_{i=p-r+1}^{r} \nu_{iN}$ is the distribution of

$$\sum_{i=p-r+1}^{r} \nu_i = -tr((\frac{1}{t} -1)\ Q_1-(\frac{1}{t} -1)Q_2).$$

The diagonal elements of Q_1 and Q_2 are all independent, each with a normal distribution with mean 0 and variance 2. Since the trace of a matrix is the sum of the diagonal elements, we know that $\sum_{i=p-r+1}^{r} \nu_i$ is normally distributed with mean 0 and variance $2r(\frac{1}{t})(\frac{1}{t} -1)$. We therefore conclude that

$$[(N-k)/(2rNk)^{\frac{1}{2}}] \sum_{i=p-r+1}^{p} \nu_{iN}$$

has an asymptotic normal distribution with mean 0 and variance 1.

Finally we may state that

$$(N(N-k))^{\frac{1}{2}}(2rk)^{-\frac{1}{2}}((\frac{N}{N-k})^{r}\Lambda^{2/N}-1)$$

has an asymptotic normal distribution with mean 0 and variance 1.

Q.E.D.

We now talk about the case where $k \to \infty$ as $N \to \infty$, but $\lim_{N\to\infty} (N-k)/N = 1$. In this case we have the following theorem:

Theorem 2.3.4. If our model is given by (2.0.1) and we wish to test the hypothesis H_0: $B\Xi = \alpha a$ versus H_1: $B\Xi \neq \alpha a$ when $\lim_{N\to\infty}(N-k)/N = t = 1$ and $k \to \infty$ as $N \to \infty$, then the asymptotic null distribution of

$$N(2rk)^{-\frac{1}{2}}((\frac{N}{N-k})^{r}\Lambda^{2/N}-1)$$

is a normal distribution with mean 0 and variance 1. The asymptotic test of H_0: $B\Xi = \alpha a$ versus H_1: $B\Xi \neq \alpha a$ would be to reject H_0 when

$$N(2rk)^{-\frac{1}{2}}((N/N-k)^{r}\Lambda^{2/N} -1) > Z_{1-\alpha},$$

and do not reject otherwise, where Z_β is the β fractile of a standard normal distribution.

Proof: Consider the following sequence of equations:

$$\Lambda^{2/N} = \prod_{i=p-r+1}^{p} (1/\lambda_{iN}) = \prod_{i=p-r+1}^{p} (1/1+\phi_{iN})$$

$$= \prod_{i=p-r+1}^{p} (1/(1+(k/N-k)+v_{iN}/Nk^{-\frac{1}{2}}))$$

$$= \prod_{i=p-r+1}^{p} (\frac{N-k}{N})(1/(1+(N-k)k^{\frac{1}{2}}N^{-2}v_{iN})$$

$$= (\frac{N-k}{N})^{r} \prod_{i=p-r+1}^{p} (1-k^{\frac{1}{2}}N^{-1}v_{iN} + O(N^{2}k^{-1}))$$

$$= (\frac{N-k}{N})^{r}(1-k^{\frac{1}{2}}N^{-1} \sum_{i=p-r+1}^{p} v_{iN}) + O(N^{2}k^{-1})$$

The above equality may be written

$$Nk^{-\frac{1}{2}}((\frac{N}{N-k})^{r}(\Lambda^{2/N}-1) = \sum_{i=p-r+1}^{p} v_{iN} + O(Nk^{-\frac{1}{2}}).$$

The asymptotic null distribution of $\sum_{i=p-r+1}^{p} v_{iN}$ can be obtained using Theorem 2.2.4. The limiting distribution of $\sum_{i=p-r+1}^{p} v_{iN}$ is the distribution of the trace of Q which has a normal distribution with mean 0 and variance 2r. The theorem now follows. Q.E.D.

Remark: When a is the zero matrix, Theorems 2.2.2 - 2.2.4 are all still valid.

2.4. Consistency of the Tests

In this section, we discuss the consistency of the tests from the preceding section, i.e., we show that the power of the tests goes to one as the sample size increases when a fixed alternative is assumed to be true. We will use the following theorem to show the consistency of the tests.

Theorem 2.4.1. Assume that an α-level asymptotic test rejects when a test statistic is greater than a constant and does not reject otherwise. If the test statistic goes to infinity almost surely as N does for a fixed alternative, the test is consistent.

Proof. This theorem follows from the definition of a consistent test. Q.E.D.

In the next theorem, we discuss the consistency of the test given in Theorem 2.3.1.

Theorem 2.4.2. If $N^{-1}F(I_N-F'a'(aFF'a')^{-1}aF)F'$ goes to a finite matrix of full rank, the test given in Theorem 2.3.1 is consistent

Proof. For our fixed alternative, let us consider $\Xi = \Xi_0$, where Ξ_0 is a matrix whose row rank is greater than p-r. Since we assumed that $N^{-1}F(I_N-F'a'(aFF'a')^{-1}aF)F'$ goes to a finite matrix of full rank, the matrix

$$N^{-1}\Xi_0F(I-F'a'(aFF'a')^{-1}aF)F'\Xi_0'$$

goes to a matrix R_0 of rank greater than p-r. We can show (see

the proof of Theorem 1.2.1) that $W^{-1}T$ converges almost surely to $1/t(I_p+\Sigma^{-1}R_0)$, and that the rth smallest eigenvalue of $W^{-1}T$ goes almost surely to the rth smallest eigenvalue of $1/t(I_p+\Sigma^{-1}R_0)$. In this case, $t = 1$. Since R_0 is of rank greater than p-r, the rth smallest eigenvalue of $W^{-1}T$ goes almost surely to a number greater than 0. We therefore have that

$$\lambda_p \cdot \lambda_{p-1} \cdot \ldots \cdot \lambda_{p-r+1}$$

goes almost surely to a number greater than one. We can now state that

$$-2 \log \Lambda = -2 \log(\lambda_p \cdot \lambda_{p-1} \cdot \ldots \cdot \lambda_{p-r+1})^{-\frac{1}{2}N}$$

goes almost surely to positive infinity. The theorem follows through an application of Theorem 2.4.1. Q.E.D.

For Case 2 and Case 3, we have to change what our fixed alternative is. In these cases, the number of parameters is assumed to increase with the sample size. It is fairly evident that the fixed alternative we picked when k is fixed makes no sense for Case 2 or Case 3.

We now describe what our fixed alternative will be. For each N, let us pick $\Xi = \Xi_{0N}$ so that the rth smallest eigenvalue of

$$(2.4.1) \quad N^{-1}\Xi_{0N}F(I-F'a'(aFF'a')^{-1}aF)F'\Xi_{0N}'$$

is fixed at $\gamma_0 > 0$. We are fixing the rth smallest eigenvalue of the noncentrality parameter of T. Let us also pick $\Xi = \Xi_{0N}$ so that the matrix given by (2.4.1) converges to a finite matrix.

We now consider the following theorem which is concerned with the asymptotic test for Case 2.

Theorem 2.4.3. The asymptotic test given in Theorem 2.3.3 (Case 2) is consistent.

Proof. We can show (see the proof of Theorem 1.2.1) that $W^{-1}T$ converges almost surely to $1/t(I_p+\Sigma^{-1}R)$ and that the rth smallest eigenvalue of $W^{-1}T$ goes almost surely to the rth smallest eigenvalue of $1/t(I_p+\Sigma^{-1}R)$. For our fixed alternative (see paragraph preceding this theorem), $R = R_0$ and the rth smallest eigenvalue of $1/t(I_p+\Sigma^{-1}R_0)$ is greater than $1/t$. We know that

$$\Lambda^{-2/N} = \lambda_p \cdot \lambda_{p-1} \cdot \ldots \cdot \lambda_{p-r+1}$$

goes almost surely to a quantity greater than $((N-k)/N)^r$. Therefore, since $(N(N-k)/k)^{\frac{1}{2}}$ goes to infinity as N does,

$$(N(N-k)/(2rk))^{\frac{1}{2}}((\frac{N}{N-k})^r \Lambda^{2/N}-1)$$

goes almost surely to positive infinity. The theorem follows after we apply Theorem 2.4.1. Q.E.D.

For Case 3, we have a similar result:

Theorem 2.4.4. The asymptotic test given in Theorem 2.3.5 (Case 3) is consistent.

Proof. We omit the proof since it is almost identical to the proof of Theorem 2.4.3.

CHAPTER 3

ESTIMATION OF UNKNOWN LINEAR RESTRICTIONS ON THE PARAMETERS OF A GENERAL LINEAR MODEL

3.0 Introduction

In this chapter, we discuss a very general linear model called the Potthoff-Roy model. This model can be formulated in the following matrix equation:

$$X = F_1 \Xi F_2 + E, \tag{3.0.1}$$

where X is a cxN matrix of observed values, F_1 is a known cxp ($c \geq p$) matrix, Ξ *is an unknown pxm matrix,* F_2 *is a known* mxN ($N \geq m$) matrix, and E is a cxN matrix of errors. The columns of E are independent with the same normal distribution having mean vector 0 and covariance matrix Σ. We require that F_1 and F_2 are of full column rank and full row rank respectively.

The classical multivariate linear regression model can be seen to be a special case of the Potthoff-Roy model by letting $F_1 = I_c$. If we let $F_2 = (1,1,\ldots,1)$ then the Potthoff-Roy model reduces to a simple "growth curves" model (Gleser and Olkin [1964]). Estimation of the parameters in the Potthoff-Roy model under various hypotheses has been discussed by Potthoff-Roy [1964], Rao [1965], and Gleser and Olkin [1969].

We want to find the MLE of Ξ, and of two other matrices U_1 and α which satisfy

(3.0.2) $\qquad U_1 \equiv F_3 = \alpha b,$

where U_1 is an unknown rxp ($r < p$) matrix, F_3 is a known mxk ($m \geq k$) matrix, α is a unknown rxs($s \leq r$) matrix, and b is a known sxk matrix. Throughout this chapter, we assume that

(3.0.3) $\qquad \Sigma = \sigma^2 \cdot I_c,$

where σ^2 is an unknown constant.

In Section 3.1, we reduce our model (3.0.1) and hypothesis (3.0.2) to a canonical form. Section 3.2 contains a derivation of the MLE's for the reduced model, and also gives the MLE's for the general model. Section 3.3 discusses several special cases of our reduced model. In Section 3.4, we consider consistency of the estimators in our models.

3.1. Reduction of the Model to a Canonical Form

Consider the following model and hypothesis:

(3.1.1) $\qquad X = F_1 \equiv F_2 + E,$

(3.1.2) $\qquad U_1 \equiv F_3 = \alpha b,$

where $X, F_1, \equiv, F_2, E, U_1, F_3, \alpha$, and b are defined in the introduction to this chapter. In this section, we reduce (3.1.1) and (3.1.2) to a simpler, or canonical form.

Let us discuss the following transformation:

$$(3.1.3)\quad Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{bmatrix} (F_1'F_1)^{-\frac{1}{2}} F_1' \\ V_1' \end{bmatrix} X,$$

where V_1 is a cxc-p column orthogonal matrix which satisfies $V_1'F_1 = 0$. Whenever we write the square root (or negative square root) of a matrix, we mean the unique, symmetric, positive definite square root. The columns of Y are independently distributed with a normal distribution having covariance matrix $\sigma^2 \cdot I_c$.

The mean of Y is

$$E(Y) = E\left[\begin{pmatrix} (F_1'F_1)^{-\frac{1}{2}} F_1' \\ V_1' \end{pmatrix} X\right] = \begin{pmatrix} (F_1'F_1)^{\frac{1}{2}} \equiv F_2 \\ 0 \end{pmatrix}.$$

Let

$$X^* = \begin{pmatrix} X_1^* & X_3^* \\ X_2^* & X_4^* \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} (F_2'(F_2F_2')^{-\frac{1}{2}}, V_2'),$$

where V_2 is a N-mxN column orthogonal matrix which satisfies $V_2F_2' = 0$. By Theorem 3.3.1 of Anderson [1958], the columns of X* will have independent normal distributions with covariance matrix $\sigma^2 \cdot I_c$. The mean of X* is

$$(3.1.4)\quad E(X^*) = E\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} (F_2'(F_2F_2')^{-\frac{1}{2}}, V_2) = \begin{pmatrix} (F_1'F_1)^{\frac{1}{2}} \equiv (F_2F_2')^{\frac{1}{2}} & 0 \\ 0 & 0 \end{pmatrix}.$$

If we let

$$(3.1.5)\qquad \equiv^* = (F_1'F_1)^{\frac{1}{2}} \equiv (F_2F_2')^{\frac{1}{2}},$$

our hypothesis (3.1.2) in terms of $\equiv^*$ becomes:

$$(3.1.6) \qquad U_1(F_1'F_1)^{-\frac{1}{2}}\,\Xi^*(F_2F_2')^{-\frac{1}{2}}\,F_3 = \alpha b.$$

We will now make several substitutions which will make our hypothesis (3.1.6) simpler.

Let

$$(3.1.7) \qquad U_2 = U_1(F_1'F_1)^{-\frac{1}{2}},$$

$$F_4 = (F_2F_2')^{-\frac{1}{2}}\,F_3(F_3'(F_2F_2')^{-1}F_3)^{-\frac{1}{2}},$$

$$d = b(F_3'(F_2F_2')^{-1}F_3)^{-\frac{1}{2}}.$$

With these substitutions our hypothesis becomes:

$$(3.1.8) \qquad U_2\,\Xi^*F_4 = \alpha d,$$

where F_4 is a known column orthogonal matrix.

We now write the joint distribution of X_1^*, X_2^*, X_3^* and X_4^*:

$$(3.1.9) \quad f(X_1^*,X_2^*,X_3^*,X_4^*)=(2\pi\sigma^2)^{-\frac{1}{2}Np}\exp(-2\sigma^2[\mathrm{tr}(X_1^*-\Xi^*)(X_1^*-\Xi^*)' + \mathrm{tr}\,X_2^*X_2^{*\prime}+\mathrm{tr}\,X_3^*X_3^{*\prime}+\mathrm{tr}\,X_4^*X_4^{*\prime}]).$$

From a quick examination of (3.1.9), we conclude that we could get the MLE ($\hat{\sigma}^2$) of σ^2 if we knew the MLE ($\hat{\Xi}^*$) of Ξ^*. Our result would be

$$(3.1.10) \qquad \hat{\sigma}^2 = \frac{1}{Nc}[\mathrm{tr}(X_1^*-\hat{\Xi}^*)(X_1^*-\hat{\Xi}^*)'+\mathrm{tr}\,X_2^*X_2^{*\prime}+\mathrm{tr}\,X_3^*X_3^{*\prime}+\mathrm{tr}\,X_4^*X_4^{*\prime}].$$

From (3.1.9), we also know that X_1^* is a sufficient statistic for Ξ^*, U_2 and α when σ^2 is treated as a fixed quantity. It is clear that finding the estimators of U_2, α, and Ξ^*, which satisfy (3.1.8) and which maximize (3.1.9) is equivalent to finding the estimators of U_2, α, Ξ^* which satisfy (3.1.8) and which minimize

$$\mathrm{tr}(X_1^* - \Xi^*)(X_1 - \Xi^*)'.$$

Therefore, we need only consider functions of X_1^* when we find MLE's of U_2, α, and Ξ^*.

We have reduced our estimation problem to the following problem. Let our model be

(3.1.11) $$X_1^* = \Xi^* + E^*,$$

where X_1^* is a pxm matrix of observed values, Ξ^* is an unknown pxm($p \leq m$) matrix, and E^* is a pxm error matrix. Each column of E^* is distributed as an independent p-dimensional normal distribution with mean vector 0 and covariance matrix $\sigma^2 \cdot I_p$, where σ^2 is unknown. We want the MLE's of Ξ^*, and of two other matrices U_2 and α which satisfy

(3.1.12) $$U_2 \Xi^* F_4 = \alpha d,$$

where U_2 is an unknown rxp matrix, F_4 is a known mxk column orthogonal matrix, α is a unknown rxs matrix and d is a known sxk matrix. We refer to (3.1.11) and (3.1.12) as either the reduced model or the model in canonical form. Note that $s \leq r < p \leq k$ and $k - s \geq p$.

In the next section, we will find the MLE's of the parameters in the reduced model. We will also use the MLE's of Ξ^*, α, and U_2 in the reduced model to get the MLE's of Ξ, α, and U_1 in the general model (3.1.1) and (3.1.2). It should be noted that the MLE of σ^2 for the reduced model is <u>not</u> the MLE of σ^2 for the general model. Equation (3.1.10) gives us the MLE of σ^2 for the general model.

3.2. Maximum Likelihood Estimators for the Model in Canonical Form

In this section, we will get the MLE for the parameters of the reduced model described at the end of Section 3.1. We will also give the MLE for the parameters of the general model (3.1.1), (3.1.2).

Let our model be the model described at the end of Section 3.1. As in Chapter 1, it is clear that if we find one set of MLE's $(\hat{U}_2,\hat{\alpha})$ of U_2,α then $A\hat{U}_2,A\hat{\alpha}$ is also a set of MLE's of U_2,α where A is any invertible matrix. Because of this, we will require that U_2 be row orthogonal.

The method of finding the MLE's of U_2, Ξ^*, and α will be similar to what we did in Chapter 1. We will 1) fix U_2,σ^2; 2) find the MLE's of Ξ^* and α as functions of the fixed values of U_2,σ^2; 3) substitute this estimate of Ξ^* back into the likelihood; and 4) find the maximum likelihood estimator of U_2,σ^2.

Part 1. U_2,σ^2 fixed or given

We will now transform X_1^* into a form in which the estimators of Ξ^* and α are easy to see. Let

$$(3.2.1) \qquad P = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{pmatrix} U_2 \\ V_4 \end{pmatrix} X_1^*,$$

where V_4 is a p-rxp row orthogonal matrix which satisfies $V_4U_2' = 0$. Each column of P has an independent p-dimensional normal distribution with covariance matrix $\sigma^2 \cdot I_p$. The mean of P is

$$(3.2.2) \quad E(P) = E\begin{pmatrix} U_2 \\ V_4 \end{pmatrix}(X_1^*) = \begin{pmatrix} U_2 \\ V_4 \end{pmatrix}E(X^*) = \begin{pmatrix} U_2\,\Xi^* \\ V_4\,\Xi^* \end{pmatrix}.$$

Let

$$(3.2.3)\quad R = \begin{pmatrix} R_1 & R_3 \\ R_2 & R_4 \end{pmatrix} = P(F_4 \;\; V_5),$$

$$= \begin{pmatrix} P_1 \cdot F_4 & P_1 \cdot V_5 \\ P_2 \cdot F_4 & P_2 \cdot V_5 \end{pmatrix},$$

where V_5 is a mxm-k column orthogonal matrix that makes (F_4, V_5) an orthogonal matrix. By Theorem 3.3.1 in Anderson [1958], the columns of R have independent p-dimensional normal distributions with covariance matrix $\sigma^2 \cdot I_p$. The mean of R is

$$(3.2.4)\quad E(R) = E\begin{pmatrix} R_1 & R_2 \\ R_3 & R_4 \end{pmatrix} = \begin{pmatrix} U_2 \,\Xi^*\, F_4 & U_2 \,\Xi^*\, V_5 \\ V_4 \,\Xi^*\, F_4 & V_4 \,\Xi^*\, V_5 \end{pmatrix} = \begin{pmatrix} \alpha\, d & U_2 \,\Xi^*\, V_5 \\ V_4 \Xi^*\, F_4 & V_4 \,\Xi^*\, V_5 \end{pmatrix}.$$

From the above expression it is easy to get the MLE's. Since all elements of R are distributed independently, we have that the MLE of $V_4 \,\Xi^*\, F_4$ is R_3, of $U_2 \,\Xi^*\, V_5$ is R_2, and of $V_4 \,\Xi^*\, V_5$ is R_4. We can apply a standard theorem in multivariate regression to get the MLE of α:

$$(3.2.5)\qquad \hat{\alpha} = R_1 d'(dd')^{-1}.$$

We now can get the MLE of Ξ^*:

$$(3.2.6)\quad \hat{\Xi}^* = \begin{pmatrix} U_2 \\ V_4 \end{pmatrix}^{-1} \begin{pmatrix} R_1 d'(dd')^{-1} d & R_2 \\ R_3 & R_4 \end{pmatrix} (F_4, V_5)^{-1},$$

$$= (U_2', V_4') \begin{pmatrix} R_1 d'(dd')^{-1} d & R_2 \\ R_3 & R_4 \end{pmatrix} \begin{pmatrix} F_4' \\ V_5' \end{pmatrix}.$$

If we go backwards, using first (3.2.3) and then (3.2.1), we get

$$\hat{\Xi}^* = U_2'U_2X^*F_4d'(dd')^{-1}d\,F_4'+V_4'V_4X_1^*F_4F_4' + U_2'U_2X_1^*V_5V_5' + V_4'V_4X_1^*V_5V_5'.$$

Using the facts that $I_p - V_4'V_4 = U_2'U_2$ and $I_k - V_5V_5' = F_4F_4'$, we get

$$\hat{\Xi}^* = X_1^*-U_2'U_2X_1^*F_4(I_k-d'(dd')^{-1}d)F_4',$$

$$\hat{\alpha} = U_2X_1^*F_4d'(dd')^{-1}.$$

It should be noted that neither $\hat{\Xi}^*$ nor $\hat{\alpha}$ is a function of σ^2.

We summarize our results so far.

Theorem 3.2.1. If our model is $X_1^* = \Xi^* + E^*$ where each column of E is distributed independently with a p-dimensional normal distribution having mean vector 0 and covariance matrix $\sigma^2 \cdot I_p$ (σ^2 is a fixed quantity), then the MLE's of Ξ^* and α which satisfy the hypothesis $U_2\,\Xi^*\,F_4 = \alpha d$, where U_2 is a fixed rxp row orthogonal matrix, F_4 is a known mxk column orthogonal matrix and d is a known sxk matrix are

$$\hat{\Xi}^* = X_1^*-U_2'U_2X_1^*F_4(I_k-d'(dd')^{-1}d)F_4', \tag{3.2.7}$$

$$\hat{\alpha} = U_2X_1^*F_4d'(dd')^{-1}. \tag{3.2.8}$$

Part 2. Substitution of our derived MLE's back into the likelihood and maximization with respect to U_2, σ^2.

In this part, we find the MLE's of U_2 and σ^2 using $\hat{\Xi}^*$, $\hat{\alpha}$ as defined by (3.2.7) and (3.2.8). We now write the distribution of X_1^* after substituting $\hat{\Xi}^*$ for Ξ^*.

$$(3.2.9)\quad f(X_1^*,\hat{\Xi}^*,\sigma^2,U_2) = (2\pi\sigma^2)^{-\frac{1}{2}mp}\exp(-2\sigma^2)\operatorname{tr}(X_1^*-\hat{\Xi}^*)(X_1^*-\hat{\Xi}^*)'$$

If we want to maximize (3.2.9), all we have to do is minimize

$$Q = \operatorname{tr}(X_1^*-\hat{\Xi}^*)(X_1^*-\hat{\Xi}^*)',$$

$$= \operatorname{tr}(U_2'U_2X_1^*F_4(I_k-d'(dd')^{-1}d)F_4')(U_2'U_2X_1^*F_4(I_k-d'(dd')^{-1}d)F_4')',$$

$$(3.2.10)\quad = \operatorname{tr}(U_2X_1^*F_4(I_k-d'(dd')^{-1}d)F_4'X_1^{*\prime}U_2').$$

Minimizing Q subject to the condition that U_3 is row orthogonal (i.e., $U_2U_2' = I_r$) is a straight forward application of the Courant-Fischer Min-Max Theorem (see Bellman [1970]). If we let the columns of $\hat{U}_2'$ be the eigenvectors associated with the r smallest eigenvalues of

$$(3.2.11)\quad M = X_1^*F_4(I_k-d'(dd')^{-1}d)F_4'X_1^{*\prime},$$

then $\hat{U}_2$ minimizes Q and therefore is the MLE of U_2. The minimum value of Q is $\sum_{i=p-r+1}^{p}\lambda_i$ where λ_i is the ith largest eigenvalue of M.

At this point we should talk about zero eigenvalues of M. Since the rank of $I_k-d'(dd')^{-1}d$ is k-s, M will have full rank with probability one if and only if $k-s \geq p$, i.e., M will have zero eigenvalues with probability one if and only if $k-s < p$. In all cases in the succeeding sections, we assume that $k-s \geq p$.

The MLE of σ^2 is easy to get since we know the minimum value of Q. The MLE ($\hat{\sigma}^2$) of σ^2 in our reduced model is

$$\hat{\sigma}^2 = \frac{1}{mp}\sum_{i=p-r+1}^{p}\lambda_i,$$

where λ_i is the ith largest eigenvalue of M.

Let us summarize our results in the following theorem:

Theorem 3.2.2. The MLE's of U_2, α, Ξ^*, and σ^2 in the reduced model (3.1.11), (3.1.12) are:

$$\hat{\alpha} = \hat{U}_2 X_1^* F_4 d'(dd')^{-1},$$

$$\hat{\Xi}^* = X_1^* - \hat{U}_2' \hat{U}_2 X_1^* F_4 (I_k - d'(dd')^{-1} d) F_4',$$

$$\hat{\sigma}^2 = \left(\frac{1}{mp}\right) \sum_{i=p-r+1}^{p} \lambda_i,$$

where λ_i is the ith largest eigenvalue of M, the rows of $\hat{U}_2$ are the eigenvectors associated with the r smallest eigenvalues of M, and

$$M = X_1^* F_4 (I_k - d'(dd')^{-1} d) F_4' X_1^{*\prime}.$$

Remark I. If we multiply $\hat{U}_2$ and $\hat{\alpha}$ on the left by any invertible matrix, the resulting matrices would also be MLE's.

Remark II. All matrices which are MLE's of U_2, α are of the form $H\hat{U}_2$, $H\hat{\alpha}$ for some invertible matrix H.

Theorem 3.2.2 gives us the MLE's of the parameters in our reduced model. If we use the MLE's of Ξ^*, U_2 and α given in Theorem 3.2.2 for our reduced model, and also use (3.1.3), (3.1.4), (3.1.5), and (3.1.7), we can get the MLE's of Ξ, U_1, and α in the general model.

Recall that the MLE for σ^2 in the general model is given by (3.1.10):

$$\hat{\sigma}^2 = \frac{1}{Nc}[tr(X_1^*-\hat{\Xi}^*)(X_1^*-\hat{\Xi}^*)'+tr\ X_2^*X_2^{*\prime}+tr\ X_3^*X_3^{*\prime}+tr\ X_4^*X_4^{*\prime}].$$

Following (3.2.11), we found that the minimum value of $tr(X_1^*-\hat{\Xi}^*)(X_1^*-\hat{\Xi}^*)'$ is $\sum_{i=p-r+1}^{p} \lambda_i$, where λ_i is the ith largest eigenvalue of M which is defined in Theorem 3.2.2. If we use the definitions of X_2^*, X_3^* and X_4^*, we get

$$tr\ X_2^*X_2^{*\prime}+tr\ X_3^*X_3^{*\prime}+tr\ X_4^*X_4^{*\prime}=tr(XX'-F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X').$$

Combining the preceding arguments, we finally have

$$\hat{\sigma}^2 = \frac{1}{Nc}(\sum_{i=p-r+1}^{p} \lambda_i+tr(XX'-F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')),$$

where λ_i is the ith largest eigenvalue of M.

We now give the MLE's of U_1, Ξ, α, and σ^2 for the general model in the following theorem.

Theorem 3.2.3. The MLE's of U_2, α, Ξ^*, and σ^2 in the general model (3.1.1) (3.1.2) are:

$$\hat{U}_1 = \hat{U}_2(F_1'F_1)^{\frac{1}{2}},$$

$$\hat{\alpha} = \hat{U}_2X_1^*F_4d'(dd')^{-1},$$

$$\hat{\Xi} = (F_1'F_1)^{-\frac{1}{2}}\ \hat{\Xi}^*(F_2F_2')^{-\frac{1}{2}},$$

$$\hat{\sigma}^2 = \frac{1}{Nc}(\sum_{i=p-r+1}^{p} \lambda_i+tr(XX'-F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X'))$$

where λ_i is the ith largest eigenvalue of M, the rows of $\hat{U}_2$ are the eigenvectors associated with the r smallest eigenvalues of M, and

$$M = X_1^* F_4 (I_k - d'(dd')^{-1} d) F_4' X_1^{*\prime},$$

$$X_1^* = (F_1'F_1)^{-\frac{1}{2}} F_1' X F_2' (F_2 F_2')^{-\frac{1}{2}},$$

$$F_4 = (F_2 F_2')^{-\frac{1}{2}} F_3 (F_3'(F_2 F_2')^{-1} F_3)^{-\frac{1}{2}},$$

$$d = b(F_3'(F_2 F_2')^{-1} F_3)^{-\frac{1}{2}},$$

$$\hat{\Xi}^* = X_1^* - \hat{U}_2' \hat{U}_2 X_1^* F_4 (I_k - d'(dd')^{-1} d) F_4'.$$

Remark I. If we multiply $\hat{U}_1$ and $\hat{\alpha}$ on the left by any invertible matrix, the resulting matrices are also MLE's.

Remark II. All MLE's of U_1 and α are of the form $A\hat{U}_1$, $A\hat{\alpha}$ where A is some invertible matrix.

Remark III. The rows of $\hat{U}_1$ are themselves eigenvectors corresponding to the r smallest eigenvalues of

$$(F_1'F_1)^{-\frac{1}{2}} M (F_1'F_1)^{\frac{1}{2}}.$$

3.3. Special Cases

The models we consider in this section are all special cases of our reduced model. It should be noted that our reduced model can be considered as a special case of our general model if we take $F_1 = I_p$, $F_2 = I_m$, and F_3 to be a column orthogonal matrix.

Consider the following situation:

$$(3.3.1) \qquad x_i = \xi_i + e_i; \quad i = 1,2,\ldots,m;$$

where x_i is a p-dimensional vector of observations, ξ_i is an

unknown p-dimensional mean vector, and e_i is a p-dimensional error vector. Each e_i is distributed independently with a normal distribution having mean vector 0 and covariance matrix $\sigma^2 \cdot I_p$ (σ^2 is unknown). We want to estimate ξ_i under the hypothesis that

(3.3.2) $\quad U_2\xi_i = \alpha; \quad i = 1,2,\ldots,m;$

where α, U_2 are unknown rx1 and rxp matrices respectively. If we let

$$X_1^* = (x_1,x_2,\ldots,x_m),\ \Xi^* = (\xi_1,\xi_2,\ldots,\xi_m),$$

$$E^* = (e_1,e_2,\ldots,e_m),$$

then (3.3.1) and (3.3.2) can be written

$$X_1^* = \Xi^* + E^*,$$

$$U_2\,\Xi^* F_4 = \alpha d,$$

where $F_4 = I_m$ and $d = (1,1,\ldots,1)$. In this form our model looks identical to the model in Theorem 3.2.2. Using Theorem 3.2.2, we get the following application.

Application 1. Assume our model is $x_i = \xi_i + e_i$ and we want to estimate U_2, ξ_i, and α subject to $U_2\xi_i = \alpha$, where x_i, ξ_i, e_i, U_2, and α are defined above. Then the MLE's of U_2, α, ξ_i, and σ^2 are:

$$\hat{\alpha} = \hat{U}_2\bar{x},$$

$$\hat{\xi}_i = x_i - \hat{U}_2'\hat{U}_2(x_i-\bar{x}),$$

$$\hat{\sigma}^2 = \frac{1}{mp}\sum_{i=p-r+1}^{p}\lambda_i,$$

where λ_i is the ith largest eigenvalue of

$$M = \sum_{i=1}^{m} (x_i-\bar{x})(x_i-\bar{x})',$$

the rows of $\hat{U}_2$ are eigenvectors corresponding to the r smallest eigenvalues of M, and

$$\bar{x} = \sum_{i=1}^{m} x_i/m.$$

Remark I. $A\hat{U}_2$ and $A\hat{\alpha}$ where A is an invertible matrix are also MLE's of U_2 and α.

In all the theorems and applications discussed so far, we have remarked that the estimator of the unknown linear restrictions ($\hat{U}_1$ or $\hat{U}_2$) is not unique. In fact any invertible matrix times $\hat{U}_1$ or $\hat{U}_2$ would also be a MLE. In the application we now discuss, we require that the last r columns of our maximum likelihood estimator of U_2 are the identity matrix (see the beginning of Section 1.2 and also the discussion preceding Application 4 in Section 1.3).

We will now consider the following model:

(3.3.3) $$y_i = v_i+f_i; \quad i = 1,2,\ldots,m;$$

$$z_i = Hv_i+\alpha+g_i; \quad i = 1,2,\ldots,m;$$

where y_i and z_i are p-r and r dimensional vectors of observed values, v_i is an unknown p-rxl vector, H is an unknown rxp-r parameter matrix, and f_i, g_i are p-r and r dimensional error vectors which are distributed independently of one another with a normal distribution having mean vector 0 and covariance matrix $\sigma^2 \cdot I_{p-r}$ and $\sigma^2 \cdot I_r$ respectively.

We will now rewrite the above model in such a way that it can be easily seen to be a special case of Theorem 3.2.2. Let

$$X_1^* = \begin{pmatrix} y_1, & y_2, \ldots, & y_m \\ z_1, & z_2, \ldots, & z_m \end{pmatrix}, \quad \Xi^* = \begin{pmatrix} v_1 & , v_2, \ldots, & v_m \\ Bv_1+\alpha, & Bv_2+\alpha, \ldots, & Bv_m+\alpha \end{pmatrix},$$

$$E^* = \begin{pmatrix} f_1, & f_2, \ldots, & f_m \\ g_1, & g_2, \ldots, & g_m \end{pmatrix},$$

then (3.3.3) can be formulated in the following way:

$$X_1^* = \Xi^* + E^*,$$

$$(-H, I)\,\Xi^* = \alpha(1, 1, \ldots, 1).$$

It is clear that the above model and hypothesis is exactly the same as in Application 1, with the exception that U_2 must have the identity as its last r columns. If $\hat{U}_2 = (\hat{U}_{12}, \hat{U}_{22})$ - $\hat{U}_{22}$ is rxr - is the estimate of U_2 in Application 1, then we can get the MLE of H from the following expression:

$$(-\hat{H}, I) = (\hat{U}_{22}^{-1}\hat{U}_{21}, I) = \hat{U}_{22}^{-1}(\hat{U}_{21}, \hat{U}_{22}) = \hat{U}_{22}^{-1}(\hat{U}_2).$$

Since $(-\hat{H}, I)$ is an invertible matrix times $\hat{U}_2$, it is also a MLE of U_2. It is clear that when we substitute $(-\hat{H}, I)$ into (3.2.10) for U_2, that Q is minimized. Since $(-\hat{H}, I)$ has the right form

$$\hat{H} = -\hat{U}_{22}^{-1}\,\hat{U}_{21}.$$

We summarize our results in the following application:

Application 2. Assume that our model is

$$y_i = v_i + f_i; \qquad i = 1,2,\ldots,m;$$

$$z_i = Hv_i + \alpha + g_i; \qquad i = 1,2,\ldots,m;$$

where y_i, z_i, v_i, α, H, f_i, and g_i are defined above. Then the MLE's of H, α, v_i, and σ^2 are:

$$\hat{H} = -\hat{U}_{22}^{-1}(\hat{U}_{21}),$$

$$\hat{\alpha} = (-\hat{H},I)\begin{pmatrix}\bar{y}\\ \bar{z}\end{pmatrix},$$

$$\begin{pmatrix}\hat{v}_i\\ \hat{Hv}_i+\hat{\alpha}\end{pmatrix} = \begin{pmatrix}y_i\\ z_i\end{pmatrix} - (-\hat{H},I)'(HH'+I_r)^{-1}(-H,I)\begin{pmatrix}y_i-\bar{y}\\ z_i-\bar{z}\end{pmatrix},$$

$$\hat{\sigma}^2 = \sum_{i=p-r+1}^{p} \lambda_i/mp,$$

where λ_i is the ith largest eigenvalue of M, the rows of $(\hat{U}_{21},\hat{U}_{22})$ are the eigenvectors associated with the r smallest eigenvalues of M, and

$$M = \sum_{i=1}^{m} \begin{pmatrix}y_i-\bar{y}\\ z_i-\bar{z}\end{pmatrix}\begin{pmatrix}y_i-\bar{y}\\ z_i-\bar{z}\end{pmatrix}'.$$

Remark I. $\hat{H},\hat{\alpha}$ are unique.

Application 2 is a generalization of the model considered first by Gleser and Watson [1973] and later by Bhargava [1975]. The proof utilized in these papers cannot be generalized to cover our case. Their model is a special case of Application 2, where $\alpha = 0$.

We will conclude this section with a model which can be considered a combination of the "error-in-variables" model and the usual linear regression model. In the model we discuss there are some variables which are measured with error and other variables which are measured perfectly. Consider the following model:

$$y_i = v_i + f_i; \qquad i = 1,2,\ldots,m;$$

$$z_i = Hv_i + \alpha d_i + g_i; \qquad i = 1,2,\ldots,m;$$

where y_i, v_i, z_i, H, v_i, f_i and g_i are the same as in Application 2, α is an unknown rxs matrix, and d_i is a known sxl matrix. v_i is the variable which is measured with error and d_i is the variable which is measured perfectly. We may apply Theorem 3.2.2 in a manner similar to what we did for Application 2 to get MLE's of H, v_i, and α. If we do this and use the fact that $d = (d_1, d_2, \ldots, d_m)$, we get

Application 3. Let our model be

$$y_i = v_i + f_i; \qquad i = 1,2,\ldots,m;$$

$$z_i = Hv_i + \alpha d_i + g_i; \qquad i = 1,2,\ldots,m;$$

where y_i, v_i, z_i, H, v_i, f_i, and g_i are the same as in Application 2, α is an unknown rxs matrix, and d_i is a known sxl matrix. The MLE of H, α, v_i, and σ^2 are:

$$\hat{H} = -\hat{U}_{22}^{-1}(\hat{U}_{21}),$$

$$\hat{\alpha} = (-\hat{H}, I)\left(\sum_{i=1}^{m} \begin{pmatrix} y_i \\ z_i \end{pmatrix} d_i'\right)\left(\sum_{i=1}^{m} d_i d_i'\right)^{-1},$$

$$\begin{pmatrix} \hat{v}_i \\ \hat{H}v_i + \hat{\alpha}d_i \end{pmatrix} = \begin{pmatrix} y_i \\ z_i \end{pmatrix} - (-\hat{H},I)'[(\hat{H},I)(\hat{H},I)']^{-1}[(-\hat{H},I)\begin{pmatrix} y_i \\ z_i \end{pmatrix} - \hat{\alpha}d_i],$$

$$\hat{\sigma}^2 = \sum_{i=p-r+1}^{p} \lambda_i/mp,$$

where λ_i is the ith largest eigenvalue of M, the rows of $(\hat{U}_{21},\hat{U}_{22})$ are eigenvectors associated with the r smallest eigenvalues of M, and

$$M = \sum_{i=1}^{m} \left(\begin{pmatrix} y_i \\ z_i \end{pmatrix} - \left(\sum_{j=1}^{m} \begin{pmatrix} y_j \\ z_j \end{pmatrix} d_j'\right)\left(\sum_{j=1}^{m} d_j d_j'\right)^{-1} d_i\right)\left(\begin{pmatrix} y_i \\ z_i \end{pmatrix} - \left(\sum_{j=1}^{m} \begin{pmatrix} y_j \\ z_j \end{pmatrix} d_j'\right)\left(\sum^{m} d_j d_j'\right)^{-1} d_i\right)'.$$

Remark I. $\hat{H}$, $\hat{\alpha}$ are unique MLE's.

3.4 Consistency of the Estimators

In this section we discuss the consistency of the estimators from Section 3.2. We first work with our reduced model. All the results for the reduced model are rigorously proved. For the general model, we merely state our results since they follow from the results for the reduced model.

Let us consider $\hat{U}_2,\hat{\alpha}$ the estimators of U_2,α in our model. In order to make a discussion of the consistency of $\hat{U}_2$ and $\hat{\alpha}$ meaningful, we have to place restrictions on U_2 and $\hat{U}_2$ which will make them unique. Our arguments here are the same as in Section 1.2 of Chapter 1. Let (U_2^*,α^*) be the unique members of the class of matrices (U_2,α) which satisfy $U_2^* \equiv^* F_4 = \alpha d$, where U_2^* has the identity matrix as its last r columns. Let $\hat{U}_2^*$ be the unique MLE of U_2 which has the identity matrix as its last r columns. In

Application 2 and Application 3 of the previous section, we satisfy this requirement. We will show that $\hat{U}_2^*$, $\hat{\alpha}^*$ are strongly consistent estimators of U_2^*, α^*. First, we will prove some useful lemmas.

Lemma 1. Assume that

$$m^{-1}\,\Xi^* F_4(I_k - d'(dd')^{-1}d)\Xi^{*\prime}$$

converges to a finite matrix R. Then

$$m^{-1}M = m^{-1}X^* F_4(I_k - d'(dd')^{-1}d)F_4'X^{*\prime}$$

goes almost surely to $R+(1-t_1)\sigma^2 \cdot I_p$ where

$$t_1 = \lim_{m\to\infty}(m-(k-s))m^{-1} = \lim_{m\to\infty}(m-k)m^{-1}.$$

Proof. Consider X_1^* which is a pxm matrix. Each column of X_1^* has an *independent* p-dimensional normal distribution with covariance matrix $\sigma^2 \cdot I_p$. The mean of X_1^* is Ξ^*. $X_1^* F_4$ is a pxk matrix. Since F_4 is a column orthogonal matrix, each column of $X_1^* F_4$ is distributed independently with a p-dimensional normal distribution with covariance matrix $\sigma^2 \cdot I_p$. The mean of $X^* F_4$ is $\Xi^* F_4$. We have

$$m^{-1}X^* F_4(I_k - d'(dd')^{-1}d)F_4'X^* = m^{-1}(\Xi^*+E^*)F_4(I_k - d'(dd')^{-1}d)F_4(\Xi^*+E^*)'$$

$$(3.4.1)\quad = m^{-1}\Xi^* F_4(I_k - d'(dd')^{-1}d)F_4'\Xi^{*\prime} + m^{-1}E^* F_4(I_k - d'(dd')^{-1}d)F_4'\Xi^{*\prime}$$

$$+ m^{-1}\Xi^* F_4(I_k - d'(dd')^{-1}d)F_4'E^{*\prime} + m^{-1}E^* F_4(I_k - d'(dd')^{-1}d)F_4'E^{*\prime}.$$

By our assumptions and Lemma 2 of Chapter 1, we have that the 2nd and 3rd terms on the right-hand side of (3.4.1) go almost surely to 0.

By our assumptions, the 1st term goes to R. If we use Theorem 4.3.2 in Anderson [1958] we find that the last term has the same distribution as

$$\sum_{i=1}^{k-s} u_i u_i',$$

where u_i has a normal distribution independent of u_j $(i \neq j)$ with mean vector 0 and covariance matrix $\sigma^2 \cdot I_p$. We know that

$$\frac{1}{m} \sum_{i=1}^{k-s} u_i u_i' = \frac{k-s}{m} \sum_{i=1}^{k-s} u_i u_i'/k-s$$

goes almost surely to $(1-t_1)\sigma^2 \cdot I_p$. If we combine all the above statements, we get

$$\frac{1}{m} X^* F_4 (I_k - d'(dd')^{-1}d) F_4' X^{*\prime} \xrightarrow{a.s.} R + (1-t_1)\sigma^2 I_p.$$

Q.E.D.

Lemma 2. If R is finite of rank p-r, then the only matrix which has the identity as its last r columns and when multiplied by R yields 0 is U_2^*.

Proof. $m^{-1} U_2^* (\Xi^* F_4 (I - d'(dd')^{-1}d) F_4' \Xi^{*\prime}) = m^{-1} \alpha d((I - d'(dd')^{-1}d) F_4' \Xi^{*\prime}) = 0$. Since this is true for every m, it is true in the limit, i.e., $U_2^* R = 0$. Since R has rank p-r, it has a unique r-dimensional space of eigenvectors associated with eigenvalue 0. Let us consider a matrix whose rows form a basis for this eigenspace. If this matrix is to have the identity matrix as its last r columns, it is clear that this matrix must be U_2^*.

Q.E.D.

Theorem 3.4.1. Under the assumptions of Lemma 1 and Lemma 2, $\hat{U}_2^*$ is a strongly consistent estimate of U_2^*.

Proof. By Lemma 1, we know that $m^{-1}M$ goes almost surely to $R+(1-t_1)\sigma^2 I_p$. Since the eigenvalues of a matrix are continuous functions of that matrix, we are able to conclude that the r smallest eigenvalues of $m^{-1}M$ converge almost surely to the smallest eigenvalue of $R+(1-t_1)\sigma^2 I_p$ which is $(1-t_1)\sigma^2$. By Lemma 2, U_2^* is the only matrix with the identity as its last r columns which satisfies $U_2^* R = 0$. We may conclude that U_2^* is the only matrix of the right form whose rows are eigenvectors associated with $(1-t_1)\sigma^2$ the smallest eigenvalue of $R+(1-t_1)\sigma^2 \cdot I_p$.

Let $\hat{U}_{2m}^*$ be the estimate of U_2^* if we have m observations. Let $\hat{U}_{2m}$ be the estimate given in Theorem 3.2.2 used to generate $\hat{U}_{2m}^*$, i.e.,

$$\hat{U}_{2m}^* = ((\hat{U}_{2m}^{(2)})^{-1}\hat{U}_{2m}^{(1)}, I_r) = (\hat{U}_{2m}^{(2)})^{-1}(\hat{U}_{2m}^{(1)}, \hat{U}_{2m}^{(1)}) = (\hat{U}_{2m}^{(2)})^{-1}(\hat{U}_{2m})$$

where $\hat{U}_{2m} = (\hat{U}_{2m}^{(1)}, \hat{U}_{2m}^{(2)})$. Since $(\hat{U}_{2m})\cdot(\hat{U}_{2m})' = I_r$, $\hat{U}_{2m}$ is bounded almost surely. Let us pick any subsequence of $\hat{U}_{2m}$. Since $\hat{U}_{2m}$ is bounded almost surely, there must exist a subsequence of this subsequence which converges. Let $\hat{U}_{2\pi_m}$ denote the convergent subsequence. Also let

$$C = \lim_{m\to\infty} \hat{U}_{2\pi_m}.$$

Every row of C is the limit of a sequence of eigenvectors of $m^{-1}M$ associated with one of the r smallest eigenvalues. Since

$m^{-1}M$ converges almost surely to $R+(1-t_1)\sigma^2 \cdot I_p$, each row of C must equal some eigenvector of $R+(1-t_1)\sigma^2 I_p$ associated with $(1-t_1)\sigma^2$. Since

$$\lim_{m\to\infty} \hat{U}_{2\pi_m} \cdot \hat{U}'_{2\pi_m} = CC' = I_r,$$

C is of full row rank and therefore its rows must span the space of eigenvectors of $R+(1-t_1)\sigma^2 I_p$ associated with $(1-t_1)\sigma^2$. We already showed that U_2^* spans that same space. We therefore have

$$U_2^* = (C^{(2)})^{-1}(C^{(1)}, C^{(2)}) = (C^{(2)})^{-1}C$$

where $C = (C^{(1)}, C^{(2)})$.

Let $||A||$ denote the largest element of A. We will now show that $||\hat{U}^*_{2\pi_m} - U_2^*||$ goes to 0 almost surely.

$$||\hat{U}^*_{2\pi_m} - U_2^*|| = ||(\hat{U}^{(2)}_{2\pi_m})^{-1} \cdot \hat{U}_{2\pi_m} - (C^{(2)})^{-1}C||$$

$$\leq ||(\hat{U}^{(2)}_{2\pi_m})^{-1}\hat{U}_{2\pi_m} - (C^{(2)})^{-1}\hat{U}_{2\pi_m}|| + ||(C^{(2)})^{-1}\hat{U}_{2\pi_m} - (C^{(2)})^{-1}C||$$

$$\leq ||\hat{U}_{2\pi_m}|| \; ||(\hat{U}^{(2)}_{2\pi_m})^{-1} - (C^{(2)})^{-1}|| + ||(C^{(2)})^{-1}|| \; ||\hat{U}_{2\pi_m} - C||. \tag{3.4.4}$$

The first term on the right-hand side of (3.4.4), $||\hat{U}_{2\pi_m}|| \cdot ||(\hat{U}^{(2)}_{2\pi_m})^{-1} - (C^{(2)})^{-1}||$, is arbitrarily small since $\hat{U}_{2\pi_m}$ is bounded almost surely and $\hat{U}^{(2)}_{2\pi_m}$ converges to $\hat{U}^{(2)}_{2\pi_m}$. Since $(C^{(2)})^{-1}$ is bounded and $\hat{U}_{2\pi_m}$ goes almost surely to C, $||C^{(2)}||^{-1}||\hat{U}_{2\pi_m} - C||$ goes almost surely to zero. Combining the above statements, we have that $\hat{U}^*_{2\pi_m}$ goes almost surely to U_2^*. We have shown that for any subsequence

of $\hat{U}^*_{2m}$ there exists a subsequence of that subsequence which converges to U^*_2 almost surely. $\hat{U}^*_2$ must converge almost surely to U^*_2. Q.E.D.

We now discuss the consistency of $\hat{\alpha}^* = \hat{U}^*_2 X^*_1 F_4 d'(dd')^{-1}$.

Theorem 3.4.2. If $(dd')^{-1}.m$ converges to a matrix with all elements finite, then $\hat{\alpha}^*$ is a strongly consistent estimate of α^* where α^* satisfies $U^*_2 \Xi^* F_4 = \alpha^* d$.

Proof. Note that

$$\hat{\alpha}^* = \hat{U}^*_2 X^*_1 F_4 d'(dd')^{-1},$$

$$= \hat{U}^*_2 (\Xi^* + E^*) F_4 d'(dd')^{-1},$$

$$= \hat{U}^*_2 \Xi^* F_4 d'(dd')^{-1} + \hat{U}^*_2 E^* F_4 d'(dd')^{-1}.$$

Since $\hat{U}^*_2$ goes almost surely to U^*_2, $\hat{U}^*_2 \Xi^* F_4 d'(dd')^{-1}$ goes almost surely to

$$U_2 \Xi^* F_4 d'(dd')^{-1} = \alpha^* dd'(dd')^{-1} = \alpha^*.$$

If we apply Lemma 2 of Chapter 1, we get that $E^* F_4 d'(dd')^{-1}$ goes almost surely to 0. We therefore have that $\hat{U}^*_2 E^* F_4 d'(dd')^{-1}$ goes almost surely to zero. Q.E.D.

We now show that the MLE of σ^2 in the reduced model is not consistent. We have already mentioned that the r smallest eigenvalues of $m^{-1}M$ go almost surely to $t_1\sigma^2$. Since λ_i/k for $i = p-r+1,\ldots,p$, are the r smallest eigenvalues of

AD-A037 530 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
ESTIMATION AND TESTS FOR UNKNOWN LINEAR RESTRICTION IN MULTIVAR--ETC(U)
DEC 76 J D HEALY AF-AFOSR-2350-72
UNCLASSIFIED MIMEOGRAPH SER-471 AFOSR-TR-77-0175 NL

2 OF 2
AD
A037530

END

DATE
FILMED
4-77

$X^*F_4(I_r-d'(dd')^{-1}d)F_4'X^{*\prime}/m$, we have

$$\hat{\sigma}^2 = \frac{1}{mp}\sum_{i=p-r+1}^{p}\lambda_i = \frac{1}{p}\sum_{i=p-r+1}^{p}(\lambda_i/m) \xrightarrow{a.s.} \frac{1}{p}(1-t_1)\sigma^2.$$

Straightforward substitution yields the following result:

$$\hat{\Xi}^*F_4(I_k-d'(dd')^{-1}d)F_4'\hat{\Xi}^{*\prime}/m \xrightarrow{a.s.} R+(1-t_1)\sigma^2(I_p-U_2'(U_2U_2')^{-1}U_2).$$

Let us now consider the parameters in the general model. For the definitions of all terms see Theorem 3.2.3 and the beginning of Section 3.1. Let $\hat{U}_1^*$ be the MLE of U_1 which has the identity as its last r columns. Let $\hat{\alpha}^*$ be the corresponding value of $\hat{\alpha}$. Let U_1^*, α^* be the parameter matrices in the population which satisfy $U_1^*\Xi F_3 = \alpha^* b$, and U_1^* has the identity matrices as its last r columns. We could prove the following theorem in an analagous way to what we did for the reduced model.

Theorem 3.4.3. If our model is the model of Theorem 3.2.3 and if

$$N^{-1}\Xi^*F_4(I_k-d'(dd')^{-1}d)F_4'\Xi^{*\prime} =$$
$$N^{-1}(F_1'F_1)^{\frac{1}{2}\prime}\Xi(F_2F_2')^{\frac{1}{2}\prime}F_4(I_k-d'(dd')^{-1}d)F_4'(F_2F_2')^{\frac{1}{2}\prime}\Xi(F_1'F_1)^{\frac{1}{2}\prime}$$

converges to a finite matrix R of rank p-r, and if $t_1=\lim_{N\to\infty}(N-(k-s))N^{-1}$, then

i) $N^{-1}(X^*(F_4(I_k-d'(dd')^{-1}d)F_4')X^{*\prime})$ goes almost surely to $R+(1-t_1)\sigma^2\cdot I_p$;

ii) the rows of $\hat{U}_1^*$ are eigenvectors of $(F_1'F_1)^{\frac{1}{2}\prime}R$ corresponding to eigenvalue 0;

iii) $\hat{U}_1^*$ is a strongly consistent estimate of U_1^*;

iv) if $N\cdot(dd')^{-1}$ converges to a finite matrix, then $\hat{\alpha}^*$ is a strongly consistent estimate of α^*;

v) $\frac{1}{Nc}\sum_{i=p-r+1}^{p}\lambda_i$ goes almost surely to $(1-t_1)\sigma^2 r/c$.

Since

$$(\sigma^2)^{-1}\mathrm{tr}(XX'-F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')$$

has a chi square distribution with cN-pm degrees of freedom, we have

$$(3.4.6)\quad \mathrm{tr}(XX'-F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')/\sigma^2(cN-pm) \xrightarrow{a.s.} 1,$$

provided that cN-pm goes to ∞ as N does. If $\lim_{N\to\infty}\frac{N-m}{N} = t_2$, we have

$$\mathrm{tr}(XX'-F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')/Nc \xrightarrow{a.s.} \sigma^2(1-s/c(1-t_2)).$$

It we combine the above statement and v) of Theorem 3.4.3, we get that

$$\hat{\sigma}^2 \xrightarrow{a.s.} \sigma^2(1+\frac{1}{c}((1-t_1)r-(1-t_2)p)).$$

Since $p > r$, and $1-t_2 = \lim_{N\to\infty}\frac{m}{N} \geq \lim_{N\to\infty}\frac{k-s}{N} = 1-t_1 (m \geq k)$, $\hat{\sigma}^2$ underestimates σ^2.

CHAPTER 4

TESTING THE EXISTENCE OF UNKNOWN LINEAR RESTRICTIONS IN A GENERAL LINEAR MODEL

4.0. Introduction

Let our model be the Potthoff-Roy model:

(4.0.1) $X = F_1 \Xi F_2 + E$

where X is a cxN matrix of observations, F_1 is a known cxp matrix, Ξ is an unknown pxm parameter matrix, F_2 is a known mxN ($N \geq m$) matrix and E is a cxN error matrix whose columns are distributed independently with a normal distribution having mean vector 0 and covariance matrix $\sigma^2 \cdot I_c$ (σ^2 is unknown). In this chapter we will be concerned with testing

(4.0.2) $H_0: U_1 \Xi F_3 = \alpha b$ versus $H_1: U_1 \Xi F_3 \neq \alpha b$,

where U_1 is an unknown rxp matrix, F_3 is a known mxk ($m \geq k$) matrix, α is an unknown rxs matrix and b is a known sxk matrix. In Section 4.1, we derive the likelihood ratio test statistic for H_0 versus H_1. In Section 4.2, we find the asymptotic distribution of the roots needed in the likelihood ratio criterion. In Section 4.3, we use the asymptotic distributions of the likelihood ratio test statistic to get asymptotic tests of H_0 versus H_1. In Section 4.4, we show the tests from the preceding section are consistent.

4.1. Likelihood Ratio Test Statistic

In this section we find the likelihood ratio test of H_0: $U_1 \equiv F_3 = \alpha b$ versus H_1: $U_1 \equiv F_3 \neq \alpha b$, where our model is given by

(4.1.1) $\quad X = F_1 \equiv F_2 + E.$

All variables are defined in the introduction to this chapter. Our result can be summarized in the following theorem.

Theorem 4.1.1. If our model is given by (4.1.1) and we wish to test the hypothesis H_0: $U_1 \equiv F_3 = \alpha b$ versus H_1: $U_1 \equiv F_3 \neq \alpha b$, then the likelihood ratio test statistic is

$$\Lambda_1 = \left[\frac{\operatorname{tr}(XX' - F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')}{\sum_{i=p-r+1}^{p} \lambda_i + \operatorname{tr}(XX' - F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')} \right]^{\frac{1}{2}cN},$$

where λ_i is ith largest eigenvalue of M, and

$$M = X_1^* F_4 (I_k - d'(dd')^{-1}d) F_4' X_1^{*\prime},$$

$$X_1^* = (F_1'F_1)^{-\frac{1}{2}} F_1' X F_2' (F_2F_2')^{-\frac{1}{2}},$$

$$F_4 = (F_2F_2')^{-\frac{1}{2}} F_3 (F_3'(F_2F_2')^{-1}F_3)^{-\frac{1}{2}},$$

$$d = b(F_3'(F_2F_2')^{-1}F_3)^{-\frac{1}{2}}.$$

Proof. We need the maximum value of the likelihood when H_0 is true and when H_1 is true. In Chapter 3, we derived the MLE's of U_1, $\equiv$, α and σ^2 when the H_0 is true (see Theorem 3.2.3). If we substitute these estimators into the likelihood we get:

$$\max_{H_0} L = (2\pi\hat{\sigma}^2)^{-\frac{1}{2}cN} e^{-\frac{1}{2}\hat{\sigma}^2[(X-F_1\hat{\Xi}F_2)(X-F_1\hat{\Xi}F_2)']},$$

$$= (2\pi)^{-\frac{1}{2}cN}(\hat{\sigma}^2)^{-\frac{1}{2}cN} e^{-\frac{1}{2}cN}$$

$$= (2\pi)^{-\frac{1}{2}cN} e^{-\frac{1}{2}cN}\left[\frac{1}{Nc}\left(\sum_{i=p-r+1}^{p}\lambda_i + tr(X_2^*X_2^{*\prime}) + tr(X_3^*X_3^{*\prime}) + tr(X_4^*X_4^{*\prime})\right)\right]^{-\frac{1}{2}cN},$$

$$(4.1.2) \quad = (2\pi e)^{-\frac{1}{2}cN}\left[\frac{1}{Nc}\left(\sum_{i=p-r+1}^{p}\lambda_i + tr(XX' - F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')\right)\right]^{-\frac{1}{2}cN},$$

where λ_i is the ith largest root of M, and M and the variables which define it are given in Theorem 4.1.1. For definitions of X_2^*, X_3^*, X_4^*, see Section 3.1.

We now get the maximum value of the likelihood when the alternative is true. When the alternative is true, our model is just $X = F_1\Xi F_2 + E$ *with no restrictions* on Ξ. The columns of E have the same distribution as under H_0. The likelihood function is

$$(4.1.3) \quad L(X,\Xi,\sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}cN} e^{-\frac{1}{2}\sigma^2(tr(X-F_1\Xi F_2)(X-F_1\Xi F_2)')}.$$

If we use standard multivariate regression procedures, we get that the MLE of Ξ is

$$(4.1.4) \quad \hat{\Xi} = (F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}.$$

The MLE of σ^2 is also easy to get:

$$(4.1.5) \quad \hat{\sigma}^2 = \frac{1}{Nc}tr(X-F_1\hat{\Xi}F_2)(X-F_1\hat{\Xi}F_2)',$$

$$= \frac{1}{Nc}[tr(XX' - F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')].$$

When we substitute (4.1.5) (4.1.4) into (4.1.3), we get that the maximum value of the likelihood when the alternative is true is

$$(4.1.6)\quad \max_{H_1} L = (2\pi e)^{-\frac{1}{2}cN}\left[\frac{1}{Nc}\,\mathrm{tr}(XX'-F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')\right]^{-\frac{1}{2}cN}.$$

If we combine (4.1.2) and (4.1.6) we will get the likelihood ratio test statistic of H_0 versus H_1:

$$\Lambda_1 = \frac{\max_{H_0} L}{\max_{H_1} L} = \frac{[\mathrm{tr}(XX'-F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')]^{\frac{1}{2}cN}}{[\sum_{i=p-r+1}^{p}\lambda_i+\mathrm{tr}(XX'-F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')]^{\frac{1}{2}cN}}.$$

Q.E.D.

Remark. It is clear that the likelihood ratio test statistic is a function of

$$\Lambda_2 = \sum_{i=p-r+1}^{p}\lambda_i/[\mathrm{tr}(XX'-F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X')] = \Lambda_1^{-2/cN}-1.$$

The numerator and denominator of the above expression are independent since $\sum_{i=p-r+1}^{p}\lambda_i$ is a function of X_1^*, the denominator is a function of X_2^*, X_3^*, and X_4^*, and X_1^* is independent X_2^*, X_3^*, and X_4^* (see Section 3.1 for definitions of X_2^*, X_3^*, and X_4^*).

Remark II. The likelihood function can be made arbitrarily large if $F_1 = I_p$ and $F_2 = I_N$ by taking $\hat{\Xi} = X$ and $\hat{\sigma}^2 = \epsilon$ where ϵ is an arbitrarily small positive number. Because of this, there does not exist a test of the hypothesis $U_2\Xi^*F_4 = \alpha d$ versus $U_2\Xi^*F \neq \alpha d$ in the reduced model. What causes the problem is that under the alternative hypothesis, there is nothing left to estimate σ^2 after we fit Ξ. We

will therefore assume that $\lim_{N\to\infty} Nc-mp = \infty$ when we test

$U_1 \equiv F_3 = \alpha b$ versus $U_1 \equiv F_3 \neq \alpha b$.

4.2. Asymptotic Distribution of the Roots

In this section, we find the asymptotic distribution of the roots needed in the likelihood ratio tests. We are interested in the r smallest roots of

$$|M-\lambda I_p| = 0,$$

where

$$(4.2.1) \quad M = X_1^* F_4 (I_k - d'(dd')^{-1}d) F_4' X_1^{*\prime}.$$

It is helpful to work with the r smallest roots of

$$|(N\sigma^2)^{-1}M-\phi^* I_p| = 0.$$

It should be noted that $\phi^* = (N\sigma^2)^{-1}\lambda$.

We now prove a useful lemma which is similar to Lemma 1 of Chapter 2.

Lemma 1. Let our model and hypothesis be given by

$$X = F_1 \equiv F_2 + E,$$
$$U_1 \equiv F_3 = \alpha b,$$

where X, F_1, $\equiv$, F_2, E, U_1, F_3, α, and b are defined in the introduction of this chapter. The roots of

$$(4.2.2) \quad |(N\sigma^2)^{-1}M - \phi^* I_p| = 0,$$

where M is given by (4.2.1) have the same distribution as the roots of

(4.2.3) $|N^{-1}U^*U^*+N^{-\frac{1}{2}}C+D_0-\phi^*I_p| = 0,$

where

$$C = \begin{pmatrix} C_{1,1} & \cdots & C_{1,p-r} & \cdots & E_1' \\ \vdots & & \vdots & & \vdots \\ C_{1,p-r} & \cdots & C_{p-r,p-r} & \cdots & E_{p-r}' \\ E_1 & \cdots & E_{p-r} & \cdots & 0 \end{pmatrix},$$

$$C_{hh'} = \sqrt{\gamma_{hN}}\, U^*_{h'h} + \sqrt{\gamma_{h'N}}\, U^*_{hh'},$$

$$E_h = \sqrt{\gamma_{hN}} \cdot \begin{pmatrix} U^*_{h,p-r+1} \\ U^*_{h,p-r+2} \\ \vdots \\ U^*_{h,p} \end{pmatrix},$$

$$D_0 = \begin{pmatrix} \gamma_{1N} & 0 & \cdots & 0 \\ 0 & \gamma_{2N} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

and γ_{iN} is the ith largest eigenvalue of

$$(N\sigma^2)^{-1}(\Xi^*F(I-d'(dd^*)^{-1}d)F_4'\Xi^{*\prime}),$$

and U* is a pxk-s matrix whose columns have independent normal distributions with mean vector 0 and covariance matrix I_p.

Proof. First, consider

$$X_1^* = (F_1'F_1)^{-\frac{1}{2}}F_1'XF_2'(F_2F_2')^{-\frac{1}{2}}.$$

When we take the square root (negative square root) of a matrix, it is always the unique symmetric square root (negative square root). The columns of X_1^* are independently normally distributed with covariance matrix $\sigma^2 \cdot I_p$. The mean of X_1^* is

$$\Xi^* = (F_1'F_1)^{\frac{1}{2}}\Xi(F_2F_2')^{\frac{1}{2}}.$$

Now consider $\sigma^{-1}X_1^*F_4$, where

$$F_4 = (F_2F_2')^{-\frac{1}{2}}F_3(F_3'(F_2F_2')^{-1}F_3)^{-\frac{1}{2}}.$$

Since F_4 is a column orthogonal matrix, each column of $\sigma^{-1}X_1^*F_4$ is distributed independently with a normal distribution having covariance matrix I_p. Next consider $\sigma^{-1}X_1^*F_4V_6$ where V_6 is a matrix such that

$$V_6'V_6 = I_{k-s}, \quad V_6V_6' = I_k - d'(dd')^{-1}d,$$

$$d = b(F_3'(F_2F_2')^{-1}F_3)^{-\frac{1}{2}}.$$

Since V_6 is a column orthogonal matrix, each column of $\sigma^{-1}X_1^*F_4V_6$ (which is pxk-s) is distributed with an independent normal distribution having covariance matrix I_p. The mean of $\sigma^{-1}X_1^*F_4V_6$ is

$$E(\sigma^{-1}X_1^*F_4V_6) = \sigma^{-1}\Xi^*F_4V_6.$$

Consider

$$U = \sigma^{-1}\Gamma_1X_1^*F_4V_6\Gamma_2,$$

where Γ_1 and Γ_2 are orthogonal matrices such that

$$(N\sigma^2)^{-1}\Gamma_1\Xi^*F_4V_6\Gamma_2 = \begin{pmatrix} \sqrt{\gamma_{1N}} & 0 & \cdots & 0 \\ 0 & \sqrt{\gamma_{2N}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 0 \end{pmatrix},$$

and γ_{iN} is the ith largest eigenvalue of

$$(\sigma^2N)^{-1}\Xi^*F_4V_6V_6'F_4'\Xi^{*\prime}=(\sigma^2N)^{-1}\Xi^*F_4(I_k-d'(dd')^{-1}d)F_4'\Xi^{*\prime}.$$

It should be noted that the r smallest eigenvalues of the above expression equal 0 by our hypothesis. We may write (4.2.2) the following way:

$$|(N\sigma^2)^{-1}M - \phi^*I_p| = |N^{-1}UU'-\phi^*I_p| = 0.$$

Finally, we make the following substitution. Let

$$U^* = U-\sigma^{-1}\Gamma_1\Xi^*F_4V_5\Gamma_2.$$

Then each column of U^* has a normal distribution with mean vector 0 and covariance matrix I_p. We also have

$$UU' = U^*U^{*\prime}+\sqrt{N}\ C+D_0,$$

where C and D_0 are defined in the statement of the lemma. The lemma now follows. Q.E.D.

At this point we separate into three cases

Case 1: k is fixed;

Case 2: $t_1 \neq 1$;

Case 3: k goes to infinity as N does, $t_1 = 1$;

where $t_1 = \lim_{N\to\infty} (N-k+s)N^{-1}$. We always assume that r (the number of rows of U_2), p (the number of rows in Ξ), and s (the row rank of b and d) are fixed quantities.

For each case, we now present important results about the asymptotic distribution of the roots. For Case 1, we have the following theorem:

Theorem 4.2.1. Assume that k is fixed. Let $\nu_i = \lambda_i(\sigma^2)^{-1} = \phi_i^* \cdot N$, where ϕ_i^* is the ith largest root of $|(N\sigma^2)^{-1}M - \phi^* I_p| = 0$. Then the limiting distribution of $(\nu_{p-r+1}, \nu_{p-r+2}, \ldots, \nu_p)$ is

$$\frac{2^{-\frac{1}{2}r(k-s-p+r)}\pi^{\frac{1}{2}r}\left(\prod_{i=p-r+1}^{p}\nu_i\right)^{\frac{1}{2}(k-s-p-1)}e^{-\frac{1}{2}\sum_{i=p-r+1}^{p}\nu_i}}{\prod_{i=1}^{r}\Gamma(\frac{1}{2}(k-s-p+r-1-i))\Gamma(\frac{1}{2}(r+1-i))} \cdot \prod_{i=p-r+1}^{p}\prod_{j=i+1}^{p}(\nu_i-\nu_n)$$

Proof. By Lemma 1, we only have to consider the distribution of the r smallest roots (ϕ_i^* = i=p-r+1,p-r+2,...,p) of

$$|N^{-1}U^*U^{*\prime} + N^{-\frac{1}{2}}C + D_0 - \phi^* I_s| = 0, \tag{4.2.4}$$

where C, U*, and D_0 are defined in Lemma 1.

For Case 1 we can utilize the proof given in Hsu [1940]. Equation (17) in Hsu is identical to our equation (4.2.4) with the following correspondences:

Hsu (17)	(4.2.3.)
A	$U^*U^{*'}$
C	C
U	0
D	$\phi^* I_s - D_0$
V	$N^{-\frac{1}{2}}$

All that we have to do is follow the steps in Hsu's proof. Q.E.D.

Remark. The distribution of the roots given in Theorem 4.2.1 is the same as the joint distribution of ρ_i where ρ_i is the ith largest root of

$$|B - \rho I| = 0$$

and B is defined by

$$B = \sum_{i=1}^{k-s-p+r} u_i u_i'$$

where the u_i are independently distributed with a normal distribution with mean vector 0 and covariance matrix I_r.

For Case 2 and Case 3, we make the following assumptions.

Assumption 1. The matrix

(4.2.5) $$(\sigma^2 N)^{-1} \Xi^* F_4 (I_k - d'(dd')^{-1} d) F_4' \Xi^{*'}$$

converges to a finite matrix $(\sigma^2)^{-1} R$ of rank p-r.

Assumption 2. $\gamma_{iN} = \gamma_i + O(\sqrt{N})$ where γ_{iN} is the ith largest eigenvalue of (4.2.5) and γ_i is the ith largest eigenvalue of $(\sigma^2)^{-1} R$.

We now state and prove several theorems for Case 2.

Theorem 4.2.2. Assume that $\lim_{N\to\infty} (N-k)N^{-1} = t_1 \neq 1$, and that Assumptions 1 and 2 hold. Let

$$\nu_i = (N\phi_i^*-k)(N-k)^{-\frac{1}{2}};\ i = p-r+1,p-r+2,\ldots,p;$$

where ϕ_i^* is the ith largest eigenvalue of $(N\sigma^2)^{-1}M$. The limiting distribution of $(\nu_{p-r+1},\nu_{p-r+2},\ldots,\nu_p)$ is the same as the distribution of the r roots from

$$|(1/t_1-1)^{\frac{1}{2}}Q_1 - \nu I_r| = 0$$

where Q_1 has the r-dimensional matrix normal distribution (see Lemma 2 of Chapter 2).

Proof. By Lemma 1, we only have to consider the distribution of the r smallest roots $(\phi_{iN}^*:\ i = p-r+1,\ p-r+2,\ldots,p)$ of

$$(4.2.5)\qquad |N^{-1}U^*U^{*\prime}+N^{-\frac{1}{2}}C+D_0-\phi^*I_p| = 0,$$

where C, U*, and D_0 are defined in Lemma 1.

If we multiple each matrix inside (4.2.5) by $N(N-k)^{-1}$ and let

$$D_1 = N(N-k)^{-1}D_0-N(N-k)^{-1}\phi^*I_p,$$

$$\phi = N(N-k)^{-1}\phi^*,$$

then (4.2.5) becomes

$$|(N-k)^{-1}U^*U^{*\prime}+N^{\frac{1}{2}}(N-k)^{-1}C-D_1| = 0.$$

The above equation is exactly the same as equation (2.2.12) with Z = 0.

We therefore may follow the proof of Theorem 2.2.2 with Z = 0. The theorem is therefore proved. Q.E.D.

If we use Theorem 2 in Anderson [1951b], we have

Theorem 4.2.3. Assume $\lim_{N\to\infty} (N-k)N^{-1} = t_1 \neq 1$ and Assumptions 1 and 2 hold. Let

$$\rho_i = (\frac{N}{\sqrt{k}}\phi_i^* - \sqrt{k}); \quad i = p-r+1, p-r+2, \ldots, p;$$

where ϕ_i^* is the ith largest eigenvalue if $(N\sigma^2)^{-1}M$. The limiting distribution of $(\rho_{p-r+1}, \rho_{p-r+2}, \ldots, \rho_p)$ is

$$2^{-r/2}[\prod_{i=1}^{r} \Gamma(\tfrac{1}{2}(r+1+i))]^{-1} e^{-\sum_{i=1}^{r}\rho_{p-r+i}} \prod_{i=p-r+1}^{p}\prod_{j=i+1}^{p}(\rho_i - \rho_j).$$

We conclude with several theorems for Case 3.

Theorem 4.2.4. Assume that $k \to \infty$ as $N \to \infty$, that $\lim_{N\to\infty} (N-k)N^{-1} = 1$, and that Assumptions 1 and 2 hold. Let

$$\nu_i = (\frac{N}{N-k}\phi_i^* - \frac{k}{N-k})Nk^{-\frac{1}{2}}; \quad i = p-r+1, p-r+2, \ldots, p;$$

where ϕ_i^* is the ith largest eigenvalue of $(N\sigma^2)^{-1}M$. The limiting distribution of $(\nu_{p-r+1}, \nu_{p-r+2}, \ldots, \nu_p)$ is the distribution of the r roots of

$$|Q - \nu I_r| = 0$$

where Q has the r-dimensional matrix normal distribution (see Lemma 2 of Chapter 2).

Proof. By Lemma 1, we only have to consider the distribution of the r smallest roots $(\phi_i^*: i = p-r+1, p-r+2, \ldots, p)$ of

$$(4.2.6) \quad |N^{-1}U^*U^{*\prime} + N^{-\frac{1}{2}}C + D_0 - \phi^* I_p| = 0,$$

where C, U*, and D_0 are defined in Lemma 1.

If we multiply each matrix inside (4.2.5) by $N(N-k)^{-1}$ and let

$$D_1 = N(N-k)^{-1}D_0 - N(N-k)^{-1}\phi^* I_p,$$

$$\phi = N(N-k)^{-1}\phi^*,$$

then (4.2.5) becomes

$$|(N-k)^{-1}U^*U^{*\prime} + N^{\frac{1}{2}}(N-k)^{-1}C + D_1| = 0.$$

The above equation is identical to equation (2.2.15) with Z = 0. We may follow the proof of Theorem 2.2.4 with Z = 0 to get the required result. Q.E.D.

Using Theorem 2 in Anderson [1951b], we have:

Theorem 4.2.5. Assume that $k \to \infty$ as $N \to \infty$, that $\lim_{N\to\infty} (N-k)N^{-1} = 1$, and that Assumptions 1 and 2 hold. Let

$$\nu_i = \left(\frac{N}{N-k}\phi_i^* - \frac{k}{N-k}\right)Nk^{-\frac{1}{2}}; \quad i = p-r+1, p-r+2, \ldots, p;$$

where ϕ_i^* is the ith largest eigenvalue of $(N\sigma^2)^{-1}M$. Then the limiting distribution of $(\nu_{p-r+1}, \nu_{p-r+2}, \ldots, \nu_p)$ is

$$2^{-r/2}\left[\prod_{i=1}^{r} \Gamma\left(\tfrac{1}{2}(r+1-i)\right)\right]^{-1} e^{-\sum_{i=p-r+1}^{p} \nu_i} \prod_{i=p-r+1}^{p} \prod_{j=i+1}^{p} (\nu_i - \nu_j).$$

4.3. Asymptotic Tests of $U_1 \Xi F_3 = \alpha b$ Versus $U_1 \Xi F_3 \neq \alpha b$

In this section, we use the asymptotic distributions of the smallest roots given in the proceding section to get the asymptotic tests based on the likelihood ratio statistic derived in Theorem 4.1.1. We also use the following lemma:

Lemma 1. Let

$$(4.3.1) \quad \theta = (\sigma^2)^{-1} \mathrm{tr}(XX' - F_1(F_1'F_1)^{-1}F_1'XF_2'(F_2F_2')^{-1}F_2X').$$

Then $(Nc-mp)^{-\frac{1}{2}}(\theta-(Nc-mp))$ converges in law to a normal random variable with mean 0 and variance 2. We also have that $(Nc-mp)^{-1}\theta$ goes almost surely to 1.

Proof. We have shown that θ has a chi-square distribution with Nc-mp degrees of freedom (see the end of Section 3.4). The lemma now follows from standard theorems. Q.E.D.

By Theorem 4.1.1, the likelihood ratio test statistic is

$$\Lambda_1 = (\Lambda_2+1)^{-\frac{1}{2}cN},$$

where

$$\Lambda_2 = \sum_{i=p-r+1}^{p} \lambda_i/(\sigma^2\theta),$$

λ_i is the ith largest eigenvalue of M, and θ is defined by (4.3.1). In terms of the eigenvalues $(\phi_1^*,\phi_2^*,\ldots,\phi_p^*)$ of $(N\sigma^2)^{-1}M$ we have

$$\Lambda_2 = N \sum_{i=p-r+1}^{p} \phi_i^*/\theta.$$

We now break up our discussion of the asymptotic tests into three parts which correspond to the three cases discussed in the preceding section.

Part 1. Case 1: k fixed.

When k is fixed, we have the following theorem:

Theorem 4.3.1. If our model is given by (4.0.1), and we wish to test the hypothesis H_0: $U_1 \equiv F_3 = \alpha b$ versus H_1: $U_1 \equiv F_3 \neq \alpha b$ when k is fixed, then the asymptotic null distribution of

$$(cN-mp)\Lambda_2 = (cN-mp)(\Lambda_1^{-2/cN}-1),$$

where Λ_1 is the likelihood ratio test statistic, is a chi-square distribution with r(k-s-p+r) degrees of freedom. The α level asymptotic test of H_0 versus H_1 would be to reject H_0 when

$$(cN-mp)(\Lambda_1^{-2/cN}-1) \geq \chi^2_{r(k-s-p+r)}(1-\alpha),$$

and do not reject otherwise, where $\chi^2_d(B)$ is the Bth fractile of a chi square distribution with d degrees of freedom.

Proof: When k is fixed, the asymptotic distribution of $N \sum_{i=p-r+1}^{p} \phi_i^*$ can be easily obtained using the remark following Theorem 4.2.1. The limiting distribution of $N \sum_{i=p-r+1}^{p} \phi_i^*$ is the same as the distribution of

$$tr(B) = tr \sum_{i=1}^{k-s-p+r} u_i u_i',$$

where u_i are independently distributed with a normal distribution

having mean vector 0 and covariance matrix I_r. Since each diagonal element of $\sum_{i=1}^{k-s-p+r} u_i u_i'$ has a chi-square distribution with k-s-p+r degrees of freedom, and since there are r independent diagonal elements, the distribution of the tr(B) is a chi-square distribution with r(k-s-p+r) degrees of freedom. We conclude that the limiting distribution of $N \sum_{i=p-r+1}^{p} \phi_i^*$ (for Case 1) is a chi-square distribution with r(k-s-p-r) degrees of freedom.

We know by Lemma 1 that $\theta/(Nc-mp)$ goes almost surely to one. Since θ and $\sum_{i=p-r+1}^{p} \phi_i^*$ are independent, we get that

$$(Nc-mp)(\Lambda_1^{-2/cN}-1) = N \sum_{i=p-r+1}^{p} \phi_i^*/(\theta/(Nc-mp))$$

has a limiting chi-square distribution with r(k-s-p+r) degrees of freedom. Q.E.D.

Part 2. Case 2: $t_1 = \lim_{N\to\infty} (N-k)N^{-1} \neq 1$.

When the number of parameters increases with the sample size in such a way that $t_1 = \lim_{N\to\infty} (N-k)N^{-1} \neq 1$, we use the following theorem which gives us the needed asymptotic test:

Theorem 4.3.2. If our model is given by (4.0.1), and we wish to test the hypothesis H_0: $U_1 \equiv F_3 = \alpha b$ versus H_1: $U_1 \equiv F_3 \neq \alpha b$ when $t_1 \neq 1$, then the asymptotic null distribution of

$$\Lambda_3 = \left(\frac{Nc-pm}{2r(Nc-pm)+2kr^2}\right)^{\frac{2}{2}} (k^{-\frac{2}{2}}(Nc-pm)\Lambda_2 - k^{\frac{2}{2}}r),$$

where $\Lambda_2 = \Lambda_1^{-2/cN} - 1$, and Λ_1 is the likelihood ratio test statistic, is a standard normal distribution. The α-level asymptotic test would be to reject H_0 when

$$\Lambda_3 > Z_{1-\alpha},$$

and do not reject otherwise, where Z_B is the B fractile of a standard normal distribution.

Proof. Consider

$$k^{-\frac{1}{2}}(Nc-pm)\Lambda_2 - k^{\frac{1}{2}}r = \frac{\left(\sum_{i=p-r+1}^{p} N\phi_i^*\right)k^{-\frac{1}{2}}}{\theta/(Nc-pm)} - k^{\frac{1}{2}}r,$$

$$= \frac{\sum_{i=p-r+1}^{p} (N\phi_i^* - k)k^{-\frac{1}{2}} - k^{\frac{1}{2}}r[(\theta/(Nc-pm)) - 1]}{\theta/(Nc-pm)}. \tag{4.3.2}$$

Since $\theta/(Nc-pm)$ goes almost surely to one, we have that the limiting distribution of

$$k^{-\frac{1}{2}}(Nc-pm)\Lambda_2 - k^{\frac{1}{2}}r$$

is the limiting distribution of

$$\sum_{i=p-r+1}^{p} (N\phi_i^* - k)k^{-\frac{1}{2}} - k^{\frac{1}{2}}r((\theta/(Nc-pm)) - 1). \tag{4.3.3}$$

When $t_1 \neq 1$, the asymptotic distribution of

$$\sum_{i=p-r+1}^{p} (N\phi_i^* - k)(k)^{-\frac{1}{2}} = ((N-k)/k)^{\frac{1}{2}} \sum_{i=1}^{r} \nu_{p-r+i}$$

can be easily obtained from Theorem 4.2.2. The limiting distribution of $\sum_{i=p-r+1}^{p} \nu_i$ is the same as the distribution of $tr(1/t_1 - 1)^{\frac{1}{2}}Q_1$, where

Q_1 has an r dimensional matrix normal distribution. Since the distribution of $tr(1/t_1-1)^{\frac{1}{2}}Q_1$ is a normal distribution with mean zero and variance $2r(1/t_1-1)$, it follows that the limiting distribution of

$$\sum_{i=p-r+1}^{p}(N\phi_i^*-k)k^{-\frac{1}{2}}$$

is a normal distribution with mean 0 and variance 2r.

We also know from Lemma 1 that

$$(Nc-pm)^{-\frac{1}{2}}((\theta/(Nc-pm))-1)$$

is asymptotically distributed as a normal random variable with mean 0 and variance 2. We therefore have that

$$k^{\frac{1}{2}}r((\theta/(Nc-pm))-1)$$

has a limiting normal distribution with mean 0 and variance

$$\lim_{N\to\infty} 2kr^2/(Nc-pm).$$

If we combine the above three paragraphs and recall that θ and $\sum_{i=p-r+1}^{p}\phi_i^*$ are independent, we have that

$$k^{-\frac{1}{2}}(Nc-pm)\Lambda_2-k^{\frac{1}{2}}r$$

has a limiting normal distribution with mean 0 and variance

$$2r+2r^2\lim_{N\to\infty} k/(Nc-pm)$$

since

$$[2r+(2r^2k/(Nc-pm)]^{-1}=(Nc-pm)/(2r(Nc-pm)+2kr^2).$$

We have finally that

$$\left(\frac{Nc-pm}{2r(Nc-pm)+2kr^2}\right)^{\frac{1}{2}}\left(k^{-\frac{1}{2}}(Nc-pm)\Lambda_2-k^{\frac{1}{2}}r\right)$$

has a limiting normal distribution with mean 0 and variance 1.

Q.E.D.

Part 3. Case 3: $k \to \infty$ as $N \to \infty$ but $t_1 = \lim_{N\to\infty} (N-k)N^{-1} = 1$.

We conclude this section with a theorem which gives the asymptotic test of H_0 versus H_1 for Case 3.

Theorem 4.3.3. If our model is given by (4.0.1), and we wish to test the hypothesis H_0: $U_1 \equiv F_3 = \alpha b$ versus H_1: $U_1 \equiv F_3 \neq \alpha b$ when $\lim_{N\to\infty} (N-k)/N = t_1 = 1$ and $k \to \infty$ as $N \to \infty$, then the asymptotic null distribution of

$$\Lambda_3 = (Nc-pm)(2kr)^{-\frac{1}{2}}\Lambda_2-(kr/2)^{\frac{1}{2}},$$

where $\Lambda_2 = \Lambda_1^{-2/cN}$ and Λ_1 is the likelihood ratio test statistic is a standard normal distribution. The α-level asymptotic test would be to reject H_0 when $\Lambda_3 > Z_{(1-\alpha)}$, and do not reject otherwise, where Z_B is the Bth fractile of a standard normal distribution.

Proof. Consider

$$\begin{aligned}\Lambda_3 &= (Nc-pm)(2kr)^{-\frac{1}{2}}\Lambda_2-(kr/2)^{\frac{1}{2}},\\ &= \frac{N(2kr)^{-\frac{1}{2}}\sum_{i=p-r+1}^{p}\phi_i^*}{\theta/(Nc-pm)} - (kr/2)^{\frac{1}{2}},\\ &= \frac{N(2kr)^{-\frac{1}{2}}\sum_{i=p-r+1}^{p}\phi_i^*-(kr/2)^{\frac{1}{2}}-(kr/2)^{\frac{1}{2}}[(\theta/(Nc-pm))-1]}{\theta/(Nc-pm)}.\end{aligned}$$

Since $N((\theta/(Nc-pm))-1)$ goes in law (by Lemma 1) to a normal random variable with mean 0 and finite variance, and since for Case 3, $\lim_{N\to\infty} k/N = 0$, we have that

$$(k)^{-\frac{1}{2}}((\theta/(Nc-pm))-1)$$

goes in law to a random variable which is constant at zero.

By Lemma 1, we may state that $\theta/(Nc-pm)$ goes almost surely to 1. Since θ and $\sum_{i=p-r+1}^{p} \phi_i^*$ are independent, we know that the asymptotic distribution of Λ_3 is the same as the asymptotic distribution of

$$(4.3.4) \qquad N(2kr)^{-\frac{1}{2}} \sum_{i=p-r+1}^{p} \phi_i^* - (kr/2)^{\frac{1}{2}}.$$

For Case 3, the asymptotic distribution of the above expression can be easily obtained from Theorem 4.2.4. Since the limiting distribution of

$$\sum_{i=p-r+1}^{p} \nu_i = (N/N-k)[Nk^{-\frac{1}{2}}\Sigma\phi_i^* - rk^{\frac{1}{2}}]$$

is the same as the distribution of tr Q where Q has the r dimensional matrix normal distribution, $\sum_{i=p-r+1}^{p} \nu_i$ has a limiting normal distribution with mean 0 and variance 2r. Since $\lim_{N\to\infty} \frac{N-k}{N} = 1$, we can conclude that the limiting distribution of (4.3.4) is a standard normal distribution. Q.E.D.

4.4. Consistency of the Tests

In this section, we discuss the consistency of the tests from the preceding section. A test is consistent if the power of a test goes to one as the sample size increases when a fixed alternative is assumed to be true.

We now give a description of what our fixed alternative will be. For each N, let us pick $\Xi = \Xi_{0N}$, so that the rth smallest eigenvalue of

$$(4.4.1) \qquad N^{-1}\Xi^*_{0N}F_4(I_k - d'(dd')^{-1}d)F_4'\Xi^{*\prime}_{0N},$$

is a fixed positive number. F_4 and d are defined in Theorem 4.1.1, and

$$\Xi^*_{0N} = (F_1'F_1)\,\Xi_{0N}(F_2F_2') \ .$$

This is a very reasonable definition of fixed alternative. We also assume that the matrix given by (4.4.1) converges to a finite matrix R.

For Case 1, we have the following theorem:

Theorem 4.4.1. The asymptotic test given in Theorem 4.3.1 (Case 1: k fixed) is consistent.

Proof. The test statistic given in Theorem 4.3.1 is

$$(Nc-mp)\Lambda_2 = (Nc-mp)(\Lambda_1^{-2/cN}-1)$$

$$= N\sum_{i=p-r+1}^{p}\phi_i^*/(\theta/(Nc-mp))$$

where θ is given by (4.3.1) and ϕ_i^* are the eigenvalues of M (see Theorem 4.1.1). By Lemma 1 of Section 3.2, $\theta/(Nc-mp)$ goes almost surely to one. By i) of Theorem 3.4.3, we have that

$$N^{-1}M = N^{-1}X^*F_4(I_k - d'(dd')^{-1}d)F_4'X^{*\prime}$$

goes almost surely to $R+(1-t_1)\sigma^2 I_p$, where $t_1 = \lim_{N\to\infty} (N-k)N^{-1}$. Since k is fixed, $t_1 = 1$. Therefore, in this case, M goes almost surely to R. Since the eigenvalues of a matrix are continuous functions of that matrix, the rth smallest eigenvalue of M goes almost surely to the rth smallest eigenvalue of R. For our fixed alternative (see the paragraph preceding this theorem), $R = R_0$ and the rth smallest eigenvalue of R_0 is some positive number. We can conclude that $N\phi^*_{p-r+1}$ goes almost surely to positive infinity. Therefore $(Nc-mp)\Lambda_2$ goes almost surely to positive infinity. If we apply Theorem 2.4.1, our theorem (Theorem 4.4.1) follows. Q.E.D.

For Case 2, $t_1 \neq 1$, we have a similar result:

Theorem 4.4.2. The asymptotic test given in Theorem 4.3.2 (Case 2) is consistent.

Proof. Let us consider

$$k^{-\frac{1}{2}}(Nc-pm)\Lambda_2 - k^{\frac{1}{2}}r = \frac{\sum_{i=p-r+1}^{p} (N\phi_i^*-k)k^{-\frac{1}{2}} - k^{\frac{1}{2}}r[(\theta/(Nc-pm))-1]}{\theta/(Nc-pm)}$$

which is equation (4.3.2). By Lemma 1, $\theta/(Nc-pm)$ goes almost surely to 1. By i) of Theorem 3.4.3 we have that

$$N^{-1}M = N^{-1}X^*F_4(I_k - d'(dd')^{-1}d)F_4'X^{*\prime}$$

goes almost surely to $R+(1-t_1)\sigma^2 I_p$, where $t_1 = \lim_{N\to\infty} (N-k)N^{-1}$. Since the eigenvalues of a matrix are continuous functions of that matrix, the rth smallest eigenvalues of M goes almost surely to the rth

smallest eigenvalue of R. For our fixed alternative (see the paragraph preceding Theorem 4.4.1), $R = R_0$ and the rth smallest eigenvalue of $N^{-1}M$ goes almost surely to a quantity greater than $(1-t_1)\sigma^2$. We therefore have that ϕ^*_{p-r+1} goes almost surely to a quantity greater than $1-t_1$, and that $\phi^*_{p-r+1} - \frac{k}{N}$ goes almost surely to a quantity greater than 0. We conclude that

$$\sum_{i=p-r+1}^{p} (N\phi_i^*-k)k^{-\frac{1}{2}}$$

goes almost surely to positive infinity. Λ_3 then goes to positive infinity. We now may apply Theorem 2.4.1 to complete the proof of this theorem. Q.E.D.

For Case 3, we have the following theorem.

Theorem 4.4.3. The asymptotic test described in Theorem 4.3.3 is consistent.

Proof. We omit a proof since the proof is similar to the proofs for Theorem 4.4.1 and Theorem 4.4.2.

BIBLIOGRAPHY

1. Acton, F. S. (1959). Analysis of Straight Line Data, John Wiley and Sons Inc., New York.

2. Anderson, T. W. (1951a). Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. Ann. Math. Statist. 22, 327-351.

3. Anderson, T. W. (1951b). The Asymptotic Distribution of Certain Characteristic Roots and Vectors. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of Berkeley Press, Berkeley and Los Angeles, 103-130.

4. Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis, John Wiley and Sons Inc., New York.

5. Bellman, R. (1970). Introduction to Matrix Analysis, McGraw-Hill, New York.

6. Chow, Y. S. (1966). Some Convergence Theorems for Independent Random Variables. Ann. Math. Statist. 37, 1482-1493.

7. Cochran, W. G. and Bliss, C. I. (1948). Discriminant Functions with Covariance. Ann. Math. Statist., 19, 151-176.

8. Gleser, L. J. and Olkin, I. (1964). Estimation for a Regression Model with Covariance. Technical Report No. 15 (NSF Grant GP-214), Stanford University.

9. Gleser, L. J. and Olkin, J. (1966). A k-Sample Regression Model with Covariance. Multivariate Analysis (P. R. Krishnaiah, Editor). Academic Press, New York.

10. Gleser, L. J. and Olkin, J. (1969). Linear Models in Multivariate Analysis. Essays in Probability and Statistics (R. C. Bose and others, Editors) University of North Carolina Press, Chapel Hill.

11. Gleser, L. J. and Watson, G. S. (1973). Estimation of a Linear Transformation. Biometrika, 60, 525-534.

12. Hsu, P. L. (1941). On the Limiting Distribution of Roots of a Determinantal Equation. J. London Math. Soc. 16, 183-194.

13. Kristoff, W. (1973). Testing a Linear Relation between True Scores of Two Measures. Psychometrika, 38, 101-111.

14. Madansky, A. (1959). The Fitting of Straight Lines When Both Variables Are Subject to Error. J. Amer. Statist. Assoc. 54, 173-205.

15. Moran, P. A. P. (1971). Estimating Structural and Functional Relationships. J. Multivariate Anal. 1, 232-255.

16. Neyman, J. and Scott, E. (1946). Consistent Estimates Based on Partially Consistent Observations. Econometrika 16, 1-32.

17. Olkin, I. and Shrikhande, S. S. (1954). On a Modified T^2 Problem. Ann. Math. Statist. 25, 808.

18. Potthoff, R. F. and Roy, S. N. (1964). A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems. Biometrika 51, 313-26.

19. Rao, C. Radhakrishna (1965). The Theory of Least Squares When the Parameters are Stochastic and Its Application to the Analysis of Growth Curves. Biometrika 50, 447-458.

20. Rao, C. Radhakrishna (1973). Linear Statistical Inference and Its Applications, John Wiley and Sons Inc., New York.

21. Sprent, P. (1966). A Generalized Least Squares Approach to Linear Functional Relationships. J. Roy. Statist. Soc. Ser. B 28, 278-297.

22. Sprent, P. (1969). Models in Regression and Related Topics. Methuen, London.

23. van Uven, M. J. (1930). Koninklijke Akademie van Wetenschappen te Amsterdam, Proceedings of the Selection of Sciences 33, 143-157, 307-326.

24. Villegas, C. (1961). Maximum Likelihood Estimation of a Linear Functional Relationship. Ann. Math. Statist. 32, 1048-1062.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(19) REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (18) AFOSR - TR - 77 - 0175	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) (6) ESTIMATION AND TESTS FOR UNKNOWN LINEAR RESTRICTIONS IN MULTIVARIATE LINEAR MODELS.	5. TYPE OF REPORT & PERIOD COVERED (9) Interim rept.
(14)	6. PERFORMING ORG. REPORT NUMBER Mimeograph Series #471
7. AUTHOR(s) John D. Healy (10) John Douglas Healy	8. CONTRACT OR GRANT NUMBER(s) AFOSR 72-2350
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332	12. REPORT DATE (11) Dec 76
	13. NUMBER OF PAGES 120
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) (12) 122p.	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited
(15) AF-AFOSR-2350-72

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
(16) 2304 (17) A5

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
linear models, linear restraints, estimation of linear restraints, tests for linear restraints, multivariate linear models, maximum likelihood, consistency of estimators, likelihood ratio tests, consistent tests, Potthoff-Roy model, errors-in-variables regression model

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
This report considers the multivariate linear regression model $X = F_1 \Xi F_2 + E$, where X is a $c \times N$ matrix of observations, F_1 is a known $c \times p$ matrix of covariates, F_2 is a known $m \times N$ design matrix (containing values of independent variables in the regression), and Ξ is an unknown $p \times m$ matrix of regression coefficients assumed under a null hypothesis H_0 to satisfy a system → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

291730

20 Abstract

of linear restraints of the form H_0: $U_1 \Xi = F_3 = \alpha b$, where F_3: $m \times k$ and b: $s \times h$ are known matrices, and U_1: $r \times p$ and α: $r \times s$ ($s \leq r \leq p$) are unknown matrices of restraint coefficients. The error matrix E: $c \times N$ is assumed to have columns which are statistically independent, multivariate normal random vectors with common mean vector 0 and common unknown matrix Σ.

Chapters 1 and 2 consider the no-covariate case; that is, $c=p$, $F_1 = I_p$, $U_1 = B$, $F_3 = I_k$, $m = k$. In Chapter 1, the maximum likelihood estimators (MLE's) of the unknown parameters B, α, Ξ, and Σ are found (under the assumption that H_0 is true). Consistency of these estimators is discussed and several special cases are presented. In Chapter 2, the likelihood ratio test statistic for testing H_0 against general alternatives is derived, and a sequence of asymptotically consistent tests is obtained. Here, "asymptotic" means that $N \to \infty$ and that k, the number of columns of Ξ, may or may not increase to infinity as $N \to \infty$.

Chapters 3 and 4 obtain corresponding MLE's and asymptotically consistent tests of H_0 in the general case, but under the additional assumption that $\Sigma = \sigma^2 I_c$, where $\sigma^2 > 0$ is unknown.