UNIVERSITY OF CAPE TOWN

DEPARTMENT OF MATHEMATICAL STATISTICS

****

DISTRIBUTIONS OF CERTAIN TEST STATISTICS IN MULTIVARIATE REGRESSION

BY

D. COUTSOURIDES

****

A thesis prepared under the supervision of Professor C.G. Troskie in fulfilment of the requirements for the degree of Doctor of Philosophy in Mathematical Statistics

****

Μαρια Γιαννη Λουρδες

PREFACE

This thesis is principally concerned with test criteria for testing different hypotheses for the multivariate regression.

In this preface a brief summary of each of the succeeding chapters is given.

In Chapter 1 the problem of testing the equality of two population multiple correlation coefficients in identical regression experiments has been studied. The author's results are extentions to those of Schuman and Bradley.

In Chapter 2 the results of Chapter 1 are extended to the multivariate case, in other words, the author has constructed tests in order to test the equality of two population generalized multiple correlation matrices.

In Chapter 3 the author shows that the Ridge Regression, Principal Components and Shrunken estimators yield the same central t and F statistics as the ordinary least square estimator.

In Chapter 4 using the results of Aitken, simultaneous tests for the C_p-criterion of Mallows are constructed. Some comments on extrapolation and prediction are made.

In Chapter 5 the Ridge and Principal components residuals are studied. Their use for detecting outliers, when multicollinearity is present, is examined.

I wish to take this opportunity of thanking Professor C.G. Troskie for the help and advice he has given me during the preparation of this thesis. I would also like to thank all the members of the staff of the Department of Mathematical Statistics, University of Cape Town, and especially Professor A.H. Money, for the encouragement that they have given to me. Thanks are also due to Mrs M.I. Cousins for her excellent and efficient typing, and for the time she spent typing the present thesis while she was on leave.

Finally I would like to thank Miss L. Teixeira for her help and encouragement.

I would like further to acknowledge the scholarship received from the University of Cape Town and the financial assistance from the C.S.I.R.

D. Coutsourides
January, 1980

CONTENTS

CHAPTER ONE

TESTING CERTAIN HYPOTHESIS WITH RESPECT TO THE MULTIPLE CORRELATION COEFFICIENT

1.1 INTRODUCTION

Let $X = \begin{pmatrix} X_1 \\ X^{(2)} \end{pmatrix}$ distributed according to a multivariate normal distribution $N(\mu,\Sigma)$ with $X : p\times 1$ and $X^{(2)} : (p-1)\times 1$. The population squared multiple correlation between X_1 and $X^{(2)}$ is given by

$$(1.1.1) \qquad P^2_{1.2,\ldots,p} = P^2 = \frac{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}{\Sigma_{11}}$$

where

$$(1.1.2) \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

and Σ_{11} is a scalar.

If $X_{(1)}, X_{(2)},\ldots,X_{(N)}$ is a random sample from X, then the maximum likelihood estimate (mℓe) of P^2 is given by the sample squared multiple correlation coefficient.

$$(1.1.3) \qquad R^2_{1.2,\ldots,p} = R^2 = \frac{A_{12}A_{22}^{-1}A_{21}}{A_{11}}$$

where

$$(1.1.4) \qquad A = \sum_{\alpha=1}^{N}(X_{(\alpha)}-\bar{X})(X_{(\alpha)}-\bar{X})' \ .$$

Consider now a second set of variables $X^* = \begin{pmatrix} X^*_1 \\ X^{*(2)} \end{pmatrix}$

which is distributed as $N(\mu^*,\Sigma^*)$ and let the population squared multiple correlation coefficient between X_1^* and $X^{*(2)}$ be given by

$$(1.1.5) \qquad P^{*2}_{1.2,\ldots,p} = P^{*2} = \frac{\Sigma^*_{12}\Sigma^{*-1}_{22}\Sigma^*_{21}}{\Sigma^*_{11}}$$

where Σ^* is defined in the same way as Σ.

Let $X^*_{(1)}, X^*_{(2)},\ldots,X^*_{(N)}$ be a random sample from X^*, then the mℓe of P^{*2} is given by

$$(1.1.6) \qquad R^{*2}_{1.2,\ldots,p} = R^{*2} = \frac{A^*_{12}A^{*-1}_{22}A^*_{21}}{A^*_{11}}$$

We make the assumption that X and X^* are independently distributed and so are the samples independent. Alternatively X^* could be the same variables as X, but the two samples could be from two different but independent time periods; or $X^*_{(1)},\ldots,X^*_{(N)}$ could be a repeated experiment (independent) of random vector X.

1.2 TESTING THE EQUALITY OF TWO MULTIPLE CORRELATION COEFFICIENTS

The problem of interest is to test

$$(1.2.1) \qquad H_0: P^2 = P^{*2} \quad \text{against} \quad H_1: P^2 \neq P^{*2}.$$

To test the above hypothesis we consider the following two test statistics based on the sample correlation coefficients

$$(1.2.2) \qquad V = \frac{R^2}{R^{*2}}$$

and

(1.2.3) $$W = \frac{R^2}{1-R^2} \Big/ \frac{R^{*2}}{1-R^{*2}} .$$

The following two situations are of importance. The first case is when we consider the variables $X^{(2)}$ and $X^{*(2)}$ to be held fixed. The conditional densities of V and W for fixed $X^{(2)}$ and $X^{*(2)}$ are then of interest. The second case is when $X^{(2)}$ and $X^{*(2)}$ are not fixed. Both situations are of practical importance.

If $R^2 = \frac{A_{12}A_{22}^{-1}A_{21}}{A_{11}}$, then $\frac{R^2}{1-R^2} \cdot \frac{N-p}{p-1}$ conditionally on A_{22} has a non-central F-distribution with $p-1$ and $N-p$ degrees of freedom and noncentrality parameter (Anderson (1958) p.93)).

(1.2.4) $$\lambda = \frac{1}{2\sigma^2} \beta A_{22} \beta'$$

where

(1.2.5) $$\sigma^2 = \sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

and

(1.2.6) $$\beta = \Sigma_{12}\Sigma_{22}^{-1}$$

Similarly $\frac{R^{*2}}{1-R^{*2}} \cdot \frac{N-p}{p-1}$ conditionally on A^*_{22}, has a noncentral F-distribution with $p-1$ and $N-p$ degrees of freedom and noncentrality parameter

(1.2.7) $$\lambda^* = \frac{1}{2\sigma^{*2}} \beta^* A^*_{22} \beta^{*\prime}$$

where

(1.2.8) $\beta^* = \Sigma^*_{12}\Sigma^{*-1}_{22}$

and

(1.2.9) $\sigma^{*2} = \sigma^*_{11.2} = \Sigma^*_{11}-\Sigma^*_{12}\Sigma^{*-1}_{22}\Sigma^*_{21}$

The statistic W given by (1.2.3) was first considered by Schumann and Bradley (1957) under the assumptions that

(i) $\sigma^2 = \sigma^{*2}$

and

(ii) $A_{22} = A^*_{22}$

Since $\sigma^2 = \Sigma_{11}(1-P^2)$ and $\sigma^{*2} = \Sigma^*_{11}(1-P^{*2})$ the assumption (i) is not a trivial one because a change in the multiple correlation coefficient will probably affect the conditional variance as well. The second assumption implies that the same set of independent variables is chosen for the second set. When the variables are fixed this assumption may be reasonable, but when the variables are random then the assumption is not feasible at all, since in this case A_{22} and A^*_{22} have independent Wishart distributions, which will have an effect on the distribution of W. The conditional density function of W for given $X^{(2)}$ and $X^{*(2)}$ is given by Schumann and Bradley (1957) as

(1.2.10) $g(w,a,b,\lambda\ ,\ \lambda^*) = e^{-\lambda-\lambda^*}\sum_{r=0}^{\infty}\sum_{s=0}^{\infty} \frac{\lambda^r}{r!} \cdot \frac{\lambda^{*s}}{s!}$

$$[B(a+r,b)B(a+s,b)]^{-1}w^{a+r-1} \cdot H(w;a,b,r,s)$$

where

$$(1.2.11) \quad \text{(i)} \quad 0 \le w \le \infty$$

$$\text{(ii)} \quad H(w) := H(w;a,b,r,s) = (w-1)^{-(2a+2b+r+s-1)} \int_1^w (x-1)^{2a+r+s-1}(w-x)^{2b-1}x^{-(a+b+r)}dx$$

with $\quad 2a = p-1 \quad \text{and} \quad 2b = N-p$

Schumann and Bradley also tabulated the function

$$(1.2.12) \quad G(w_0;\ a,b,\lambda = \lambda^*) = \int_0^{w_0} g(w)\ dw\ .$$

The question which arises is, when are we able to use the tables of Schumann and Bradley to test the hypothesis (1.2.1), i.e. $H_0:\ P^2 = P^{*2}$

For large N, we may suppose that $\frac{1}{N-1} A_{22} \simeq \Sigma_{22}$ and $\frac{1}{N-1} A^*_{22} \simeq \Sigma^*_{22}$

then

$$\lambda = \frac{N-1}{2} \cdot \frac{P^2}{(1-P^2)}\ , \quad \lambda^* = \frac{N-1}{2} \cdot \frac{P^{*2}}{(1-P^{*2})}$$

and hence, if H_0 is true we can use the above tables to test if H_0: $P^2 = P^{*2}$, or equivalently if H_0: $\lambda = \lambda^*$

Since the assumptions of Schumann and Bradley are rather restrictive in a practical situation, the unconditional distribution of W will be derived. Since 2λ and $2\lambda^*$ are independently distributed as $\frac{P^2}{(1-P^2)}\ \chi^2_{N-1}$ and $\frac{P^*}{(1-P^{*2})}\ \chi^2_{N-1}$ variables respectively, (refer to (1.2.4) and (1.2.5)). We have, after taking expected values over

$$(1.2.13)\quad E\left\{e^{-\frac{1}{2}\phi\chi^2_{N-1}}\left(\frac{\phi}{2}\chi^2_{N-1}\right)^r\right\}E\left\{e^{-\frac{1}{2}\phi^*\chi^2_{N-1}}\left(\frac{\phi^*}{2}\chi^2_{N-1}\right)^s\right\}$$

with $\phi = \dfrac{P^2}{(1-P^2)}$ and $\phi^* = \dfrac{P^{*2}}{(1-P^{*2})}$,

from (1.2.10) the marginal density of W (for the unconditional case)

$$(1.2.14)\quad g(w,a,b,P^2,P^{*2}) = (1-P^2)^{\frac{1}{2}(N-1)}(1-P^{*2})^{\frac{1}{2}(N-1)}$$

$$\cdot\sum_{r=o}^{\infty}\sum_{s=o}^{\infty}(P^2)^r(P^{*2})^s[B(a+r,b)B(a+s,b)]^{-1}/r!s!$$

$$\cdot w^{a+r-1}\ H(w;a,b,r,s)$$

$$\cdot\ \frac{\Gamma(\frac{1}{2}(N-1)+r)\cdot\Gamma(\frac{1}{2}(N-1)+s)}{\Gamma(\frac{1}{2}(N-1))\ \cdot\ \Gamma(\frac{1}{2}(N-1))}$$

with a,b and $H(w)$ defined as in (1.2.11)

To test the hypothesis H_0: $P^2 = P^{*2}$ it is necessary to tabulate

$$G(w_0,a,b,P^2 = P^{*2}) = \int_0^{w_0} g(w)dw\ .$$

It is extremely difficult to tabulate the above function because of the complicated expressions involved. In what follows the moments of W are given which may be useful for approximating the exact percentage points. An asymptotic result is also given in Section 1.4.

We compute now the h-moment of $g(w){:} = g(w;a,b,P^2 = P^{*2})$. Under H_0 we have

$$E(w^h) = \int_0^{\infty} g(w)w^h\ dw$$

or

$$(1.2.15)\qquad E(w^h) = (1-P^2)^{N-1}\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}(P^2)^{r+s}\{B(a+r,b)B(a+s,b)\}^{-1}$$

$$\cdot\ \frac{\Gamma(\frac{1}{2}(N-1)+r)\ \cdot\ \Gamma(\frac{1}{2}(N-1)+s)}{\Gamma(\frac{1}{2}(N-1))\ \cdot\ \Gamma(\frac{1}{2}(N-1))}\ \cdot\ I_w/r!s!$$

where $\quad I_w = \int_0^\infty w^{a+r+h-1}\, H(w;a,b,r,s)\, dw$.

So we need to compute the integral

$$(1.2.16)\qquad I_w = \int_0^\infty w^{a+r+h-1}\, H(w;a,b,r,s)\, dw$$

where $H(w)$ as in (1.2.11).

The integral $H(w)$ appearing in (1.2.16) may be written in terms of hypergeometric functions following Erdelyi (1953, p.115). Then with transformation $y = \frac{(x-1)}{(w-x)}$ when $0 \le w \le 1$ and $y = \frac{w(x-1)}{(w-x)}$ when $1 \le w \le \infty$.

$$(1.2.17)\qquad H(w) = \int_0^\infty y^{2a+r+s-1}\,(1+wy)^{-(a+b+r)}(1+y)^{-(a+b+s)}\,dy$$

$$= B(2a+r+s,2b)\,{}_2F_1[a+b+r,2a+r+s;\ 2a+2b+r+s;\ (1-w)],\qquad 0 \le w \le 1$$

and

$$(1.2.18)\qquad H(w) = w^{-(2a+r+s)}\int_0^\infty y^{2a+r+s-1}\left(1+\frac{y}{w}\right)^{-(a+b+s)}\cdot(1+y)^{-(a+b+r)}\,dy$$

$$= w^{-(2a+r+s)}\,B(2a+r+s,2b)\,{}_2F_1[a+b+s,2a+r+s,\ 2a+2b+r+s,\frac{(w-1)}{w}],\qquad 1 \le w \le \infty$$

Case (i) $0 \le w \le 1$

From (1.2.16) and (1.2.17) we have

$$(1.2.19) \qquad I_w = B(2a+r+s,2b) \int_0^1 w^{a+r+h-1} \, {}_2F_1[a+b+r,2a+r+s; 2a+2b+r+s; (1-w)]\, dw$$

or

$$(1.2.20) \qquad I_w = B(2a+r+s,2b)\sum_{\rho=0}^{\infty} \frac{(a+b+r)_\rho (2a+r+s)_\rho}{(2a+2b+r+s)_\rho \rho!} \int_0^1 w^{a+h+r-1} \cdot (1-w)^\rho \, dw$$

If we set

$$(1.2.21) \qquad I_a = \int_0^1 w^{a+h+r-1} (1-w)^\rho \, dw = B(a+h+r,\rho-1), \quad a+h+r,\rho \ge 1.$$

Combining (1.2.15), (1.2.20) and (1.2.21) we have

$$(1.2.22) \qquad E(w^h) = (1-P^2)^{N-1}\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}(P^2)^{r+s}[B(a+r,b)B(a+s,b)]^{-1} \frac{\Gamma(\tfrac{1}{2}(N-1)+r)\Gamma(\tfrac{1}{2}(N-1)+s)}{\Gamma(\tfrac{1}{2}(N-1))\Gamma(\tfrac{1}{2}(N-1))}/r!s! \cdot B(2a+r+s,2b)\sum_{\rho=0}^{\infty}\frac{(a+b+r)_\rho (2a+r+s)_\rho}{(2a+2b+r+s)_\rho \rho!} \cdot B(a+h+r,\rho-1)$$

Case (ii) $1 \le w \le \infty$

From (1.2.16) and (1.2.18) we have

$$(1.2.23) \qquad I_w^* = B(2a+r+s,2b)\sum_{\sigma=0}^{\infty}\frac{(a+b+s)_\sigma (2a+r+s)_\sigma}{(2a+2b+r+s)_\sigma \sigma!} \int_1^\infty w^{h-a-s-\sigma-1}(w-1)^\sigma \, dw$$

But

$$(1.2.24)\qquad I^*_{w_1} = \int_1^\infty w^{h-a-s-\sigma-1} (w-1)^\sigma \, dw$$

$$= \sum_{k=0}^{\sigma} \binom{\sigma}{k} (-1)^{\sigma-k} \int_1^\infty w^{h-a-s-\sigma-1+k} \, dw$$

$$= \sum_{k=0}^{\sigma} \binom{\sigma}{k} (-1)^{\sigma-k-1} \frac{1}{h-a-s-\sigma+k}, \quad h-a-s-\sigma-1+k<0$$

Combining (1.2.15), (1.2.23) and (1.2.24) we have

$$E(w^h) = (1-P^2)^{N-1} \sum_{r=0}^\infty \sum_{s=0}^\infty (P^2)^{r+s} [B(a+r,b)B(a+s,b)]^{-1}/r!s!$$

$$\frac{\Gamma(\frac{1}{2}(N-1)+r)\Gamma(\frac{1}{2}(N-1)+s)}{\Gamma(\frac{1}{2}(N-1))\Gamma(\frac{1}{2}(N-1))} B(2a+r+s,2b)$$

$$\sum_{\sigma=0}^\infty \frac{(a+b+s)_\sigma (2a+r+s)_\sigma}{(2a+2b+r+s)_\sigma \sigma!} \sum_{k=0}^{\sigma} (-1)^{\sigma-k-1} \binom{\sigma}{k}$$

$$\frac{1}{h-a-s-\sigma+k}$$

We assume $h-a-s-\sigma-1+k$ is an integer number with $h-a-s-\sigma-1+k<0$.

It is important to remember that both multiple correlation coefficients are assumed to be nonzero. If one of the coefficients, say, $P^2 = 0$, then it is only necessary to test if H_0: $P^{*2} = 0$ versus H_1: $P^{*2} \neq 0$. To test this hypothesis one uses

$$F = \frac{R^{*2}}{1-R^{*2}} \cdot \frac{N-p}{p-1}$$

and rejection of H_0: $P^{*2} = 0$ will also imply that $P^2 \neq P^{*2}$.

1.3 TESTING THE EQUALITY OF TWO MULTIPLE CORRELATIONS WITH CERTAIN ASSUMPTIONS ABOUT THE COVARIANCE MATRICES

The same assumptions are made as in Section 1.1 with the exception of the covariance matrices which have the following forms:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{and} \quad \Sigma^* = \begin{pmatrix} \Sigma_{11} & \Sigma^*_{12} \\ \Sigma^*_{21} & \Sigma_{22} \end{pmatrix}$$

We are interested in testing the hypothesis H_0: $P^2 = P^{*2}$ which is equivalent to testing H_0: $\Sigma_{12} = \Sigma^*_{12}$.

In order to test the above hypothesis we use the

$$(1.3.1) \qquad \frac{P^2}{P^{*2}} = \frac{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}{\Sigma^*_{12}\Sigma_{22}^{-1}\Sigma^*_{21}}$$

The maximum likelihood estimates of $\frac{P^2}{P^{*2}}$ in (1.3.1) is given by

$$(1.3.2) \qquad V = \frac{A_{12}A_{22}^{-1}A_{21}}{A^*_{12}A^{*-1}_{22}A^*_{21}} \; .$$

If $X^{(2)}$ and $X^{*(2)}$ are held fixed then the conditional distributions of $\frac{A_{12}A_{22}^{-1}A_{21}}{\sigma^2}$ and $\frac{A^*_{12}A^{*-1}_{22}A^*_{21}}{\sigma^{*2}}$ are noncentral χ^2-distributions, with $p-1$ degrees of freedom and non-centrality parameters $\lambda = \frac{1}{2\sigma^2}\beta A_{22}\beta'$ and $\lambda^* = \frac{1}{2\sigma^{*2}}\beta^* A^*_{22}\beta^{*'}$ where σ^2, β as in (1.2.5) and (1.2.6), while $\sigma^{*2} = \Sigma_{11}-\Sigma^*_{12}\Sigma_{22}^{-1}\Sigma^*_{21}$ and $\beta^* = \Sigma^*_{12}\Sigma_{22}^{-1}$.

Notice that under H_0: $P^2 = P^{*2}$ we have $\sigma^2 = \sigma^{*2}$ and $\beta = \beta^*$. But unfortunatley only when N is large we will have $\frac{1}{N-1} A_{22} \simeq \Sigma_{22}, \frac{1}{N-1} A^*_{22} \simeq \Sigma_{22}$ and consequently $\lambda \simeq \lambda^*$ under H_0. Thus the conditional as well as the unconditional density of V is of interest. Since the conditional density of V is the density of the ratio of two noncentral χ^2 variables, we have

$$(1.2.3) \quad f\left(v \mid X^{(2)} = x^{(2)}, X^{*(2)} = x^{*(2)}\right) = \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} e^{-\lambda} \frac{\lambda^r}{r!} e^{-\lambda^*} \frac{\lambda^{*s}}{s!} B[\tfrac{1}{2}(p-1)+r, \tfrac{1}{2}(p-1)+s] v^{\frac{1}{2}(p-1)+r-1}(1+v)^{-\frac{1}{2}(p-1+p-1)-r-s}$$

Hence $P(v \leq v_0 \mid \lambda = \lambda^*) = \int_0^{v_0} f(v \mid X^{(2)} = x^{(2)}, X^{*(2)} = x^{*(2)})\, dv$ can be found, (Johnson and Kotz (1970, p.197)). Provided that N is large and under the null hypothesis $P^2 = P^{*2}$ and $\lambda = \lambda^*$. To find the unconditional density of V we again proceed as in Section 1.2 (see (1.2.13)) by computing the necessary expectations.

This gives the unconditional density of V under H_0 as

$$(1.2.4) \quad f(v) = (1-P^2)^{N-1} v^{\frac{1}{2}(p-1)-1}(1+v)^{-(p-1)} \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} \frac{\Gamma(p-1+r+s)\Gamma(\frac{N-1}{2}+r)\Gamma(\frac{N-1}{2}+s)}{(\Gamma(\frac{N-1}{2}))^2\,\Gamma(\frac{(p-1)}{2}+r)\Gamma(\frac{(p-1)}{2})+s)r!s!} \cdot (P^2)^{r+s} v^r (1+v)^{-(r+s)}$$

The above density can be written in the following form

$$(1.2.5)\quad f(v) = (1-P^2)^{N-1} v^{\frac{1}{2}(p-1)-1}(1+v)^{-(p-1)}\{\Gamma(\tfrac{1}{2}(p-1))\cdot\Gamma(\tfrac{1}{2}(p-1))\}^{-1}\cdot\Gamma(p-1)$$

$$\cdot\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(p-1)_{r+s}(\frac{N-1}{2})_s(\frac{N-1}{2})_r}{(\frac{1}{2}(p-1))_r(\frac{1}{2}(p-1))_s r!s!}\cdot\left(\frac{vP^2}{1+v}\right)^r\left(\frac{P^2}{1+v}\right)^s$$

or finally

$$(1.2.6)\quad f(v) = \frac{(1-P^2)^{N-1}\, v^{\frac{1}{2}(p-1)-1}\Gamma(p-1)}{\Gamma(\frac{1}{2}(p-1)\Gamma(\frac{1}{2}(p-1))(1+v)^{(p-1)}}$$

$$_3F_2(p-1,\tfrac{N-1}{2},\tfrac{N-1}{2};\ \tfrac{1}{2}(p-1),\tfrac{1}{2}(p-1);\ \tfrac{vP^2}{1+v},\tfrac{P^2}{1+v})\ .$$

In order to compute the moments of $f(v)$, i.e.

$$(1.2.7)\quad E(v^h) = \int_0^\infty v^h f(v)\,dv$$

we need to compute the following integral

$$(1.2.8)\quad I = \int_0^\infty v^{\frac{1}{2}(p-1)+h+r-1}(1+v)^{-\{(p-1)+r+s\}}dv$$

$$= B(\tfrac{1}{2}(p-1)+h+r,\tfrac{1}{2}(p-1)+s-h),\quad \tfrac{p-1}{2}+h+r<p-1+r+s$$

For computation of (1.2.8) (see Erdelyi (1954, p.310)).
Combining (1.2.6), (1.2.7) and (1.2.8) we have

$$E(v^h) = \frac{\Gamma(p-1)(1-P^2)^{N-1}}{\Gamma(\frac{1}{2}(p-1)\Gamma(\frac{1}{2}(p-1))}\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(p-1)_{r+s}(\frac{N-1}{2})_s(\frac{N-1}{2})_r}{(\frac{1}{2}(p-1))_r(\frac{1}{2}(p-1))_s r!s!}\cdot(P^2)^{r+s}$$

$$\cdot B(\tfrac{1}{2}(p-1)+h+r,\tfrac{p-1}{2}+s+h)$$

$$\text{for}\quad \tfrac{p-1}{2}+h+r<p-1+r+s$$

1.4 ASYMPTOTIC RESULT

In this section we will give an asymptotic result for the ratio $V = \frac{R^2}{R^{*2}}$ which can be used for testing the hypothesis H_0: $P^2 = P^{*2}$.

Fisher (1928) has shown that for N-large NR^2 is asymptotically distributed as $\chi_{p-1}^{\prime(2)}(\lambda)$ with noncentrality parameter $\lambda = \frac{N}{2} P^2$.

We have

$$(1.4.1) \qquad V = \frac{NR^2}{NR^{*2}} = \frac{\chi_{p-1}^{\prime(2)}(\lambda)}{\chi_{p-1}^{\prime(2)}(\lambda^*)}$$

$$= \frac{\dfrac{\chi_{p-1}^{\prime(2)}(\lambda)}{p-1}}{\dfrac{\chi_{p-1}^{\prime(2)}(\lambda^*)}{p-1}}$$

which is distributed as doubly noncentral F with p-1, p-1 degrees of freedom and noncentrality parameters

$$\lambda = \frac{N}{2} P^2 \quad \text{and} \quad \lambda^* = \frac{N}{2} P^{*2}.$$

If both numerator and demoninator are approximated (Johnson and Kotz (1970, p.197)), the approximating distribution is that of

$$(1.4.2) \; V^* = \frac{1 + \lambda(p-1)^{-1}}{1 + \lambda^*(p-1)^{-1}} \; F_{\nu,\nu'}$$

where

(1.4.3) $\nu = (p-1+\lambda)^2(p-1+2\lambda)^{-1}$ and

$$\nu' = (p-1+\lambda^*)^2(p-1+2\lambda^*)^{-1}$$

We can express (1.4.2) and (1.4.3) in terms of the multiple population correlation coefficient taking into account that under H_0: $P^2 = P^{*2}$, so we have

$$(1.4.4)\quad V^* = \frac{1 + \frac{N}{2(p-1)} P^2}{1 + \frac{N}{2(p-1)} P^2} \cdot F_{\nu,\nu'}$$

with

$$\nu = \nu' = (p-1 + \frac{N}{2} P^2)^2(p-1+NP^2)^{-1}$$

1.5 AN EXAMPLE (Bradley and Schumann)

We consider regression equations of the form $y = a + b_1x_1 + b_2x_2$ for two solvents, cyclohexanol and methy ethyl ketone. We are interested in comparing the dependencies y, on the physical properties of the phases of the two different solvents. Suppose that two samples were taken and the following regression equations were found

1. Cyclohexanol

$$y = 1,0286 + 1,0931X_1 + 0,1736X_2$$

2. Methyl ethyl ketone

$$y = -2,8715 + 0,8712X_1 + 0,523X_2$$

The corresponding squared multiple correlation coefficients are $R^2 = 0,571$ and $R^{*2} = 0,486$ with 2 and 34 and 2 and 31 degrees of freedom respectively.

We would like to test

H_0: $P^2 = P^{*2}$ versus H_1: $P^2 \neq P^{*2}$ at 0,05% level.

$$V = \frac{NR}{N^*R^*} = \frac{0{,}571}{0{,}486} \cdot \frac{37}{34} = 1{,}28$$

We calculate now

$$V^* = \frac{1+\lambda(p-1)^{-1}}{1+\lambda^*(p-1)^{-1}} \quad F_{\nu,\nu^*}$$

where $\lambda = \frac{N}{2} P^2$, $\lambda^* = \frac{N^*}{2} P^{*2}$

$$\nu = (p-1+\lambda)^2(p-1+2\lambda)^{-1}, \quad \nu' = (p-1+\lambda^*)^2(p-1+2\lambda^*)^{-1}.$$

We have $\lambda = \frac{37}{2}\, 0{,}571 = 10{,}56$, $\lambda^* = \frac{34}{2}\, 0{,}486 = 8{,}26$

$$\frac{1+\lambda(p-1)^{-1}}{1+\lambda^*(p-1)^{-1}} = \frac{1+10{,}56\ 2^{-1}}{1+8{,}26\ 2^{-1}} = \frac{6{,}28}{5{,}13} = 1{,}22$$

$$\nu = (p-1+\lambda)^2(p-1+2\lambda)^{-1} = (2+10{,}26)^2(2+2\cdot 10{,}26)^{-1} = 6{,}67 \simeq 7$$

$$\nu' = (p-1+\lambda^*)^2(p-1+2\lambda^*)^{-1} = (2+8{,}26)^2(2+2\cdot 8{,}26)^{-1} = 5{,}6 \simeq 6$$

From the tally $F^{0,05}_{7,6} = 4{,}21$, so $V^* = 1{,}22 \cdot 4{,}21 = 5{,}14$

Since $V^* = 5{,}14 > 1{,}28 = V$ we do not reject H_0.

CHAPTER TWO

TESTING THE EQUALITY OF TWO MULTIPLE CORRELATION MATRICES

2.1 INTRODUCTION

Consider two p-dimensional vectors $X = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_p \end{bmatrix}$ and

$X^* = \begin{bmatrix} X_1^* \\ X_2^* \\ \cdot \\ \cdot \\ \cdot \\ X_p^* \end{bmatrix}$ distributed as $N(\mu,\Sigma)$ and $N(\mu^*,\Sigma^*)$ respectively.

Let $X_{(1)}, X_{(2)},\ldots,X_{(N)}$ and $X^*_{(1)}, X^*_{(2)},\ldots,X^*_{(N)}$, $N > p$ be two random samples of N-observations on X and X^* furthermore let $A = \sum_{\alpha=1}^{N}(X_{(\alpha)}-\overline{X})(X_{(\alpha)}-\overline{X})'$ and $A^* = \sum_{\beta=1}^{N}(X^*_{(\beta)}-\overline{X}^*)(X^*_{(\beta)}-\overline{X}^*)'$ be the Wishart matrices. Then $\overline{X}$, $\overline{X}^*$ and $\frac{A}{N}$, $\frac{A^*}{N}$ are the maximum likelihood estimates of μ, μ^* and Σ, Σ^* respectively.

Let X, X^* be partitioned into two sets, i.e.

$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$ and $X^* = \begin{pmatrix} X^{*(1)} \\ X^{*(2)} \end{pmatrix}$, where $X^{(1)}$ and $X^{*(1)}$ have q- components, while $X^{(2)}$ and $X^{*(2)}$ have r-components. Note that $q+r = p$ and $q \leq r$. Partition Σ, Σ^* and A, A^* accordingly, that is

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad , \quad \Sigma^* = \begin{pmatrix} \Sigma^*_{11} & \Sigma^*_{12} \\ \Sigma^*_{21} & \Sigma^*_{22} \end{pmatrix}$$

and

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad , \quad A^* = \begin{pmatrix} A^*_{11} & A^*_{12} \\ A^*_{21} & A^*_{22} \end{pmatrix}$$

where Σ_{11}, Σ^*_{11}, A_{11} and A^*_{11} are $q \times q$ matrices, Σ_{12}, Σ^*_{12}, A_{12} and A^*_{12} are $q \times r$ matrices, and Σ_{22}, Σ^*_{22}, A_{22} and A^*_{22} are $r \times r$ matrices.

Since A_{11} is positive definite (A is positive definite) let $A^{\frac{1}{2}}_{11}$ be the positive definite square root of A_{11}. Any type of matrix square root may be considered. For example $A^{\frac{1}{2}}_{11}$ may be defined as $A^{\frac{1}{2}}_{11} = \Delta D^{\frac{1}{2}} \Delta'$ where Δ is an orthogonal matrix, $D = \text{diag}(a_1, \ldots, a_q)$ and $a_1, \ldots, a_q$ are the characteristic roots of A_{11} or A_{11} may be defined as $A^{\frac{1}{2}}_{11} = T$ where $A_{11} = TT'$, T is a nonsingular matrix and can be triangular. With the first definition $A^{\frac{1}{2}}_{11}$ is symmetric but this need not be the case with the second definition. In an expression such as $A^{\frac{1}{2}}_{11} \; G \; A^{\frac{1}{2}}_{11}$ we adopt the convention that the postmultiplier is $(A^{\frac{1}{2}}_{11})'$.

The generalised multiple correlation matrix

(2.1.1) $R = A_{11}^{-\frac{1}{2}} A_{12} A_{22}^{-1} A_{21} A_{11}^{-\frac{1}{2}}$

was defined by Khatri (1964) as a measure of the correlation between the two sets of variables $X^{(1)}$ and $X^{(2)}$. The population generalised multiple correlation matrix can similarly be defined as the matrix

(2.1.2) $P = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$

and it is obvious that R is the maximum likelihood estimate of P.

The matrix R has many interesting properties. For $q = 1$, $r = 1$ and $p = 2$, (2.1.1) reduces to $R = a_{11}^{-1} a_{12} a_{22}^{-1} a_{21} = r_{12}^2$ the square of the correlation coefficient between X_1 and X_2. For $q = 1$ and $r = p-1$, $R = r_{1.2...p}^2$ the square of the multiple correlation coefficient between X_1 and $(X_2,...,X_p)$ (see Anderson (1958)). It is because of these relationships between R and the correlation and multiple correlation coefficients that R was defined as the generalised multiple correlation matrix (Khatri (1964)).

Note that for $X^* \sim N(\mu^*, \Sigma^*)$ the sample generalised multiple correlation matrix and the population generalised multiple correlation matrix are given by

(2.1.3) $R^* = A_{11}^{*-\frac{1}{2}} A_{12}^{*} A_{22}^{*-1} A_{21}^{*} A_{11}^{*-\frac{1}{2}}$

and

(2.1.4) $P^* = \Sigma_{11}^{*-\frac{1}{2}} \Sigma_{12}^{*} \Sigma_{22}^{*-1} \Sigma_{21}^{*} \Sigma_{11}^{*-\frac{1}{2}}$ respectively.

In what follows we assume:

(i) X and X^* are independently distributed although they could be the same variables but at a different time period.

(ii) The sample sizes are identical.

We are interested in testing the hypothesis

(2.1.5) $H_0: P = P^*$ versus $H_1: P \neq P^*$.

In order to test (2.1.5) we will consider scalar functions of the ratios $P \cdot P^{*-1}$ or $P(P+P^*)^{-1}$ given by

(2.1.6) (i) $|PP^{*-1}|$

(ii) $tr(PP^{*-1})$

(iii) $\frac{|P|}{|P+P^*|}$

(iv) $tr(P(P+P^*)^{-1})$

(v) $\frac{tr(P)}{tr(P^*)}$

(iv) $tr(P) - tr(P^*)$

The m.ℓ.e of the scalar functions in (2.1.6) are given by (2.1.7) accordingly

(2.1.7) (i) $|R \cdot R^{*-1}|$

(ii) $tr(RR^{*-1})$

(iii) $\frac{|R|}{|R+R^*|}$

(iv) $tr(R(R+R^*)^{-1})$

(v) $\frac{tr(R)}{tr(R^*)}$

(vi) $tr(R) - tr(R^*)$

2.2 ZONAL POLYNOMIALS

In this section we briefly state results on zonal polynomials, which will be used in sequel.

Zonal polynomials are by now too well known to require any special introduction. The fundamental theory can be found in James (1960 ,1961), Constantine (1963) and more recently in Subrahmanian (1976).

<u>Definition 2.2.1</u>

A partition $\kappa = (k_1, k_2, \ldots, k_p)$ of the integer k into p parts is a set of integers $k_1 \geq k_2 \geq \ldots \geq k_p > 0$ such that $\sum_{i=1}^{p} k_i = k$. (Here the Greek characters denote a partition of the corresponding Roman character.)

The zonal polynomial $C_\kappa(S)$ is defined as the component of $(\mathrm{tr}\, S)^k$ in the subspace V_k being the vector space of homogeneous polynomials $\phi(S)$ of degree k in the $n = \frac{1}{2}\, p(p+1)$ different elements of the $p \times p$ symmetric matrix S. Being invariant under the orthogonal group, i.e.

$$C_\kappa(HSH') = C_\kappa(S) \qquad H \in O(p)$$

it is a symmetric homogeneous polynomial of degree k in the latent roots of S. If S is symmetric and R is positive definite we define $C_\kappa(SR) = C_\kappa(R^{\frac{1}{2}} S R^{\frac{1}{2}})$ where $R^{\frac{1}{2}}$ is the positive definite square root. This definition is true since SR and $R^{\frac{1}{2}} S R^{\frac{1}{2}}$ have the same latent roots.

Note $C_\kappa(bS) = b^k C_\kappa(S)$ for b scalar.

We introduce now the notation

$$(2.2.1) \qquad (t)_\kappa = \Pi_{j=1}^{p}[t - \frac{(j-1)}{2}]_{k_j}$$

$$= \Pi_{j=1}^{p}\{\Gamma[t+k_j - \frac{(j-1)}{2}]/\Gamma[t-(j-1)/2]\}$$

We can also write (2.2.1) as

$$(2.2.2) \qquad (t)_\kappa = \frac{\Gamma_p(t,\kappa)}{\Gamma_p(t)}$$

where
$$\Gamma_p(t,\kappa) = \pi^{p(p-1)/4}\ \Pi_{j=1}^{p}\Gamma(t+k_j-(j-1)/2)$$
$$= \Gamma_p(t)(t)_\kappa$$

and
$$\Gamma_p(t) = \pi^{p(p-1)/4}\ \Pi_{j=1}^{p}\Gamma(t-(j-1)/2) \ .$$

<u>Lemma 2.2.1</u> (James (1960) and Constantine (1963))

Let S, T be $p\times p$ symmetric matrices.
Then for $H \in O_{(p)}$

$$\int_{O_{(p)}} C_\kappa(SHTH')dH = \frac{C_\kappa(S)C_\kappa(T)}{C_\kappa(I)}$$

where dH is the invariant Haar measure on $O_{(p)}$, normalized to make the volume of the group manifold unity.

<u>Lemma 2.2.2</u> (Constantine 1963)

Let $R : p\times p$ be a positive definite symmetric matrix and let $T : p\times p$ be an arbitrary symmetric matrix. Then

$$\int_{S>0} \operatorname{etr}(-RS)|S|^{t-\frac{1}{2}(p+1)} C_\kappa(TS)\,dS$$

$$= \Gamma_p(t,\kappa)|R|^{-t} C_\kappa(TR^{-1})$$

where the integration is over the space of all positive definite matrices, and valid for all $t > \frac{1}{2}(p-1)$.

Lemma 2.2.3 (Constantine 1963)

If R is any positive definite matrix $p\times p$ then

$$\int_0^I |S|^{t-\frac{1}{2}(p+1)} |I-S|^{u-\frac{1}{2}(p+1)} C_\kappa(RS)\,dS$$

$$= \frac{\Gamma_p(t,\kappa)\Gamma_p(u)}{\Gamma_p(t+u,\kappa)} C_\kappa(R)$$

$$= \frac{\Gamma_p(t)\Gamma_p(u)(t)_\kappa}{\Gamma_p(t+u)(t+u)_\kappa} C_\kappa(R)$$

Lemma 2.2.4 (Constantine 1966)

If $S : p\times p$ is a symmetric matrix, then

$$C_\kappa(S)C_\tau(S) = \Sigma_\delta g^\delta_{\kappa,\tau} C_\delta(S)$$

where κ,τ and δ are partitions of k,t and $d = k+t$, respectively, into not more than p parts. The coefficients $g^\delta_{\kappa,\tau}$ have been tabulated by Khatri and Pillai (1968) for all partition of $k+t$ up to order 7.

<u>Lemma 2.2.5</u> (James 1964)

If S is a symmetric $p \times p$ matrix, then

$$C_\kappa(I-S) = \Sigma_{\kappa^* \leq \kappa} \alpha_{\kappa^*} C_{\kappa^*}(S)$$

where κ^* is a partition of k into not more than p-parts. The order $\leq$ is lexicographic and is described in James (1964).

<u>Lemma 2.2.5*</u> (Khatri and Pillai 1968)

Let $G = \text{diag}(g_1, \ldots, g_r)$ and further let $G_1 = \text{diag}(1, G)$. Then

$$C_\kappa(G_1) = \sum_{t=0}^{k} \Sigma_\tau b_{\kappa,\tau} C_\tau(G).$$

<u>Definition 2.2.6</u> (Herz (1955), Constantine (1963), James (1964))

If $S : p \times p$ and $T : p \times p$ are symmetric matrices, we define the hypergeometric functions of single and double argument

(2.2.3) $\quad {}_mF_n\ (a_1, \ldots, a_m;\ b_1, \ldots, b_n;\ S)$

$$= \sum_{k=0}^{\infty} \Sigma_\kappa \frac{(a_1)_\kappa \ldots (a_m)_\kappa}{(b_1)_\kappa \ldots (b_n)_\kappa} \cdot \frac{C_\kappa(S)}{k!}$$

and

(2.2.4) $\quad {}_mF_n\ (a_1, \ldots, a_m;\ b_1, \ldots, b_n;\ S, T)$

$$= \sum_{k=0}^{\infty} \Sigma_\kappa \frac{(a_1)_\kappa \ldots (a_m)_\kappa}{(b_1)_\kappa \ldots (b_n)_\kappa} \cdot \frac{C_\kappa(S) C_\kappa(T)}{k! C_\kappa(I)}$$

Note the following conditions for convergence of the series.

(i) $m \leq n+1$; otherwise the series may only converge for $S = 0$.

(ii) If $m = n+1$ the series converge for $|\lambda_1| < 1$ where λ_1 is largest latent root of S in (2.2.3) or of $S \cdot T$ in the case (2.2.4).

(iii) If $m \leq n$ the series always converge.

(iv) The a_i and b_j are arbitrary numbers but none of the b_j may be integers or half integers less than $\frac{1}{2}(p-1)$, otherwise some of the denominators vanish.

(v) If one of the a_i is a negative integer, $a-q$ say, then for $k > p\,q+1$ all the coefficients vanish so that the function reduces to a finite polynomial of degree $p\,q$.

<u>Lemma 2.2.7</u> (James (1961))

For S : $p \times p$ symmetric

$$ {}_0F_0(S) = \mathrm{etr}(S) $$

$$ = \sum_{k=0}^{\infty} \Sigma_\kappa \frac{C_\kappa(S)}{k!} $$

<u>Lemma 2.2.8</u> (Herz (1955))

$$ {}_1F_0(\alpha, S) = |I-S|^{-\alpha} $$

$$ = \sum_{k=0}^{\infty} \Sigma_\kappa \frac{(\alpha)_\kappa C_\kappa(S)}{k!} $$

Lemma 2.2.9 (Anderson (1958) p. 318)

If the symmetric matrix S has a density function of the form $g(\lambda_1,\ldots,\lambda_p)$, where $\lambda_1 > \ldots > \lambda_p$ are the latent roots of S, i.e. the density function depends only on the roots, then the joint distribution of the ordered roots is

$$(2.2.5) \qquad h(\lambda_1,\ldots,\lambda_p) = \{\Gamma_p(\tfrac{1}{2}p)\}^{-1}\pi^{\frac{1}{2}p^2} g(\lambda_1,\ldots,\lambda_p)\alpha_p(\Delta_\lambda)$$

where $\Delta_\lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_p)$

and

$$\alpha_p(\Delta_\lambda) = \Pi_{i<j}\ (\lambda_i-\lambda_j)$$

Lemma 2.2.10 (Sugiyama (1967))

If $1 > \lambda_2 > \ldots > \lambda_p > 0$ and

$$\Delta_\lambda = \mathrm{diag}(\lambda_2,\ldots,\lambda_p)$$

$${}^1\Delta_\lambda = \mathrm{diag}(1,\lambda_2,\ldots,\lambda_p)$$

then

$$(2.2.6) \qquad \int\int_{1>\lambda_2>\ldots>\lambda_p>0}\cdots\int\int |\Delta_\lambda|^{t-\frac{1}{2}(p+1)} C_\kappa({}^1\Delta_\lambda)|I_{p-1}-\Delta_\lambda| \cdot\alpha_{p-1}(\Delta_\lambda)d\lambda_2 \ldots d\lambda_p$$

$$= (pt+k)\ \frac{\Gamma_p(\frac{1}{2}p)\Gamma_p(t,\kappa)\Gamma_p(\frac{1}{2}(p+1))}{\pi^{\frac{1}{2}p^2}\Gamma_p(t+\frac{1}{2}(p+1),\kappa)}\ C_\kappa(I_p)$$

Lemma 2.2.11 (Khatri and Pillai (1968))

If A_κ denotes the integral

(2.2.7) $$\int_D |S|^{\frac{1}{2}(t-p-1)} C_\kappa(S)\alpha_p(S)ds_2 \ldots ds_p$$

where $S = \text{diag}(s_1,\ldots,s_p)$,

$s_1 = 1 - \sum_{i=2}^{p} s_i$ and D is given by

$D = \{s_i | 0 < s_p < \ldots < s_1\}$, then

(2.2.8) $$A_\kappa = \{\Gamma_p(\tfrac{1}{2}t)\Gamma_p(\tfrac{1}{2}p)(\tfrac{1}{2}t)_\kappa C_\kappa(I_p)\} / \{\pi^{\frac{1}{2}p^2}(\tfrac{1}{2}tp)_\kappa \Gamma_p(\tfrac{1}{2}tp)\}$$

Definition (2.2.12) (Greenacre 1973)

The symmetrised density of the positive definite symmetric matrix $A : p\times p$ is defined as

$$f_s(A) = \int_{O(p)} f(HAH')dH$$

where A has the density function $f(A)$.

The symmetrised density satisfies the conditions of the density function and has the following property:

Theorem (2.2.13) (Greenacre 1973)

Let ϕ be a function of the random matrix $A : p\times p$ such that

$$\phi(A) = \phi(H'AH) \qquad H \in O(p) .$$

Then the distribution of $\phi(A)$ is invariant with respect to symmetrisation of the density of A.

Corrollary 2.2.14 (Greenacre 1973)

The moments of $|A|$, $|I-A|$, $tr(A)$ are identical for $f(A)$ and $f_s(A)$.

Corrolary 2.2.15 (Greenacre 1973)

The distribution of the latent roots of A is identical for $f_s(A)$ and $f(A)$.

The following two theorems are extensions of those given by Anderson (1958) for the central distribution.

Theorem 2.2.16 (Gupta 1971)

In the noncentral linear case the likelihood ratio statistic $U_{p,n,m}$ is distributed as $\Pi_{i=1}^{p} X_i$, where X_1 is independently distributed as

$$\beta(x_1;\lambda^2,m,n,1) = (B(\tfrac{1}{2}m,\tfrac{1}{2}n))^{-1} x_1^{\frac{1}{2}m-1}(1-x_1)^{\frac{1}{2}n-1}$$
$$e^{-\frac{1}{2}\lambda^2}\,{}_1F_1(\tfrac{1}{2}(m+n);\tfrac{1}{2}n;\tfrac{1}{2}\lambda^2(1-x_1))$$

and X_i $(i = 2,\ldots,p)$ are independently distributed as

$$\beta(x_i;\ m+1-i,n) = (B(\tfrac{1}{2}(m+1-i),\tfrac{1}{2}n))^{-1}$$
$$x_i^{\frac{1}{2}(m+1-i)}(1-x_i)^{\frac{1}{2}n-1}\ , \qquad 0 \leq x_i \leq 1, \quad m > i\ .$$

Theorem 2.2.17 (Gupta 1971)

In the noncentral linear case L.R statistic $U_{2r,n,m}$ is distributed like $X_1\Pi_{i=1}^{r-1} Y_i^2 X_{2r}$ where X_1 is independently distributed as $\beta(x_1;\lambda^2,m,n,1)$, Y_i $(i = 1,2,\ldots,r-1)$ are independently distributed as $\beta(y_i;\ 2(m-2i),2n)$ and X_{2r} is independently distributed as $\beta(x_{2r};\ m+1-p;n)$. $U_{2s+1,n,m}$ is distributed as $X_1\Pi_{i=1}^{s} Y_i^2$ where X_1 is distributed as $\beta(x_1;\lambda^2,m,n,1)$ and Y_i $(i = 1,2,\ldots,s)$ are independently distributed as $\beta(y_i;2(m-2i),2n)$.

It is important to note that since $U_{p,n,m}$ can be written as a product of Beta variables it follows

$$-\log U_{p,n,m} = \sum_{i=1}^{p}(-\log X_i) = \sum_{i=1}^{p} Y_i \ .$$

Clearly from the above relation it follows that $-\log U_{p,n,m}$ is the sum of independently distributed random variables and therefore the distribution can be obtained by taking successive convolutions, provided, as Gupta (1971) points out, the procedure yields expression which can be integrated easily at each stage.

2.3 TESTING THE EQUALITY OF TWO MULTIPLE CORRELATION MATRICES WHEN $\Sigma_{11} = \Sigma_{11}^*$, $\Sigma_{22} = \Sigma_{22}^*$ and $\Sigma_{12} = \Sigma_{21} = 0$

In this section the covariance matrices are given as

$$\Sigma = \begin{pmatrix}\Sigma_{11} & 0 \\ 0 & \Sigma_{22}\end{pmatrix} \quad \text{and} \quad \Sigma^* = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}\end{pmatrix}$$

The hypothesis given by (2.1.5) becomes

(2.3.1) H_0: $\Sigma_{12}^* = 0$ versus H_1: $\Sigma_{12}^* \neq 0$

In what follows we give some tests for testing (2.3.1).

(i) The likelihood ratio test, (Wilks (1932) and Anderson (1958)) is

$$W = \frac{|A^*|}{|A^*_{11}||A^*_{22}|} \qquad (2.3.2)$$

If the null hypothesis is true, then W is distributed as $U_{q,r,N-1}$ the criterion for testing a hypothesis about regression coefficients. For further details see (Anderson (1958) Chapter 9).

(ii) In terms of Roy's union-intersection technique (Roy (1957)) the hypothesis (2.3.1) can also be stated as

$H_0: \rho_1^2 = 0$ versus $H_1: \rho_1^2 \neq 0$

where ρ_1^2 is the largest canonical correlation coefficient between the two sets $X^{*(1)}$ and $X^{*(2)}$. See also (Morrison (1976) pp. 253-259).

(iii) Pillai (1958) proposes the following criteria

$V^{(q)} = \mathrm{tr}R^*$ and $V^{(q)} = \mathrm{tr}(R^*(I-R^*)^{-1})$.

2.4 TESTING THE EQUALITY OF TWO MULTIPLE CORRELATION MATRICES WHEN THE DIAGONAL SUBMATRICES ARE THE SAME, WHILE OFF DIAGONAL SUBMATRICES ARE DIFFERENT

For Σ, Σ^* we have

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{and} \quad \Sigma^* = \begin{pmatrix} \Sigma_{11} & \Sigma^*_{12} \\ \Sigma^*_{21} & \Sigma_{22} \end{pmatrix}$$

In order to test the hypothesis given by (2.1.5) we must test if

(2.4.1) $H_0: \Sigma_{12} = \Sigma^*_{12}$ versus $H_1: \Sigma_{12} \neq \Sigma^*_{12}$

We set $B_1 = A_{12}A_{22}^{-1}A_{21}$ and $B_2 = A^*_{12}A^{*-1}_{22}A^*_{21}$ then the mℓe

of $\theta = |PP^{*-1}| = \dfrac{|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|}{|\Sigma^*_{12}\Sigma_{22}^{-1}\Sigma^*_{21}|}$ is given by

(2.4.2) $\hat{\theta} = |B_1B_2^{-1}| = |B_2^{-\frac{1}{2}}B_1B_2^{-\frac{1}{2}}| = |V|.$

Conditionally on A_{22} and A^*_{22} we have that B_1 and B_2 are distributed as non-central Wishart distributions $W(\psi_1,\phi_1,r)$ and $W(\psi_2,\phi_2,r)$ respectively (Theorem 4.3.2 Anderson (1958), Constantine (1963), James (1964)).

Note that ψ_1,ϕ_1 and ψ_2,ϕ_2 are given by (2.4.3) and (2.4.4) respectively.

(2.4.3)
$$\psi_1 = \Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$
$$\phi_1 = \psi_1^{-1}\beta_1A_{22}\beta_1'$$
$$\beta_1 = \Sigma_{12}\Sigma_{22}^{-1}$$

and

(2.4.4)
$$\psi_2 = \Sigma^*_{11\cdot 2} = \Sigma_{11} - \Sigma^*_{12}\Sigma_{22}^{-1}\Sigma^*_{21}$$
$$\phi_2 = \psi_2^{-1}\beta_2A_{22}\beta_2'$$
$$\beta_2 = \Sigma^*_{12}\Sigma_{22}^{-1}$$

More specifically we have

(2.4.5)
$$f(B_1|A_{22}) = C_1|B_1|^{\frac{1}{2}(r-q-1)}\mathrm{etr}(-\tfrac{1}{2}\psi_1^{-1}B_1)$$
$$\mathrm{etr}(-\tfrac{1}{2}\phi_1)\,{}_0F_1(\tfrac{1}{2}r;\ \tfrac{1}{4}\phi_1\psi_1^{-1}B_1)$$

where $C_1 = \dfrac{1}{\Gamma_q(\frac{r}{2})|2\psi_1|^{n/2}}$ and $B_1 > 0$.

Similarly

(2.4.6) $f(B_2|A^*_{22}) = C_2|B_2|^{\frac{1}{2}(r-q-1)}\, etr(-\frac{1}{2}\psi_2^{-1}B_2)$

$$etr\,(-\tfrac{1}{2}\phi_2)\,{}_0F_1(\tfrac{1}{2}r;\tfrac{1}{4}\phi_2\psi_2^{-1}B_2)$$

where $C_2 = \dfrac{1}{\Gamma_q(\frac{r}{2})|2\psi_2|^{n/2}}$ and $B_2 > 0$.

In order to get the unconditional densities of B_1 and B_2, we will multiply by the densities of A_{22} and A^*_{22} accordingly and we will integrate with respect to A_{22} and A^*_{22}. But $A_{22} \sim W(\Sigma_{22},n)$ and $A^*_{22} \sim W(\Sigma_{22},n)$, i.e. Wishart distributions. So for the unconditional density of B_1 and B_2 we will have

(2.4.7) $f(B_1) = \displaystyle\int_{A_{22}>0} f(B_1,A_{22})\, dA_{22}$

$$= \int_{A_{22}>0} f(B_1|A_{22})\, f(A_{22})\, dA_{22}$$

and

(2.4.8) $f(B_2) = \displaystyle\int_{A^*_{22}>0} f(B_2,A^*_{22})\, dA^*_{22}$

$$= \int_{A^*_{22}>0} f(B_2|A^*_{22})\, f(A^*_{22})\, dA^*_{22}$$

Now we compute the density of B_1. We have that the joint density of B_1, A_{22} is

$$(2.4.9)\quad f(B_1,A_{22}) = C_1^*|B_1|^{\frac{1}{2}(r-q-1)}\ \mathrm{etr}(-\tfrac{1}{2}\psi_1^{-1}B_1)$$

$$\mathrm{etr}(-\tfrac{1}{2}\phi_1)\,{}_0F_1(\tfrac{1}{2}r,\tfrac{1}{4}\phi_1\psi_1^{-1}B_1)$$

$$|A_{22}|^{\frac{1}{2}(n-r-1)}\mathrm{etr}(-\tfrac{1}{2}\Sigma_{22}^{-1}A_{22})$$

where $$C_1^* = \frac{1}{\Gamma_q(\frac{r}{2})|2\psi_1|^{r/2}} \cdot \frac{1}{\Gamma_r(\frac{1}{2}n)|2\Sigma_{22}|^{\frac{1}{2}n}}$$

If we integrate now with respect to A_{22} we have

$$(2.4.10)\quad f(B_1) = C_1^*|B_1|^{\frac{1}{2}(r-q-1)}\mathrm{etr}(-\tfrac{1}{2}\psi_1^{-1}B_1)$$

$$\int_{A_{22}>0} |A_{22}|^{\frac{1}{2}(n-r-1)}\mathrm{etr}(-\tfrac{1}{2}\phi_1)\mathrm{etr}(-\tfrac{1}{2}\Sigma_{22}^{-1}A_{22})$$

$${}_0F_1(\tfrac{1}{2}r;\tfrac{1}{4}\phi_1\psi_1^{-1}B_1)\ dA_{22}$$

where C_1^* as in (2.4.9)

<u>Lemma 2.4.1</u>

Let $$A_k = \int_{A_{22}>0} |A_{22}|^{\frac{1}{2}(n-r-1)}\mathrm{etr}(-\tfrac{1}{2}\phi_1)\mathrm{etr}(-\tfrac{1}{2}\Sigma_{22}^{-1}A_{22})$$

$${}_0F_1(\tfrac{1}{2}r;\tfrac{1}{4}\phi_1\psi_1^{-1}B_1)\ dA_{22}$$

with ϕ_1, ψ_1 as in (2.4.3). Then

$$A_k = |2\Sigma_{22}|^{n/2}|I-P|^{n/2}\Gamma_r(\tfrac{1}{2}n)\,{}_1F_1(\tfrac{1}{2}n;\tfrac{1}{2}r;\tfrac{1}{2}\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}B_1)$$

where Σ^{12}, Σ^{22} and Σ^{21} are the submatrices of

$\Sigma^{-1} = \begin{pmatrix}\Sigma^{11} & \Sigma^{12}\\ \Sigma^{21} & \Sigma^{22}\end{pmatrix}$ and Σ^{-1} is the inverse of Σ.

<u>Proof</u>

We can write A_k as

(2.4.11) $$A_k = \int_{A_{22}>0} |A_{22}|^{\frac{1}{2}(n-r-1)} \mathrm{etr}(-\tfrac{1}{2}(\Sigma_{11\cdot 2}^{-1}\beta_1 A_{22}\beta_1' + \Sigma_{22}^{-1} A_{22})$$

$$ {}_0F_1(\tfrac{1}{2}r; \tfrac{1}{4}\Sigma_{11\cdot 2}^{-\frac{1}{2}}\beta_1 A_{22}\beta_1'\Sigma_{11\cdot 2}^{-\frac{1}{2}} B_1) dA_{22}$$

$$= \int_{A_{22}>0} |A_{22}|^{\frac{1}{2}(n-r-1)} \mathrm{etr}(-\tfrac{1}{2}(\beta_1'\Sigma_{11\cdot 2}^{-1}\beta_1 + \Sigma_{22}^{-1})A_{22})$$

$$ {}_0F_1(\tfrac{1}{2}r; \tfrac{1}{4}\ \Sigma_{11\cdot 2}^{-\frac{1}{2}}\beta_1 A_{22}\beta_1'\Sigma_{11\cdot 2}^{-\frac{1}{2}} B_1) dA_{22}$$

To evaluate (2.4.11) we make the transformation $A_{22} = \Sigma_{22}^{\frac{1}{2}}\ S\ \Sigma_{22}^{\frac{1}{2}}$ with $J(A_{22}\to S) = |\Sigma_{22}|^{\frac{1}{2}(r+1)}$ and then we use Lemma 2.2.2. So finally we have

(2.4.12) $$A_k = |2\Sigma_{22}|^{n/2}|I+\Lambda|^{-n/2}\ \Gamma_r(\tfrac{1}{2}n)$$

$$ {}_1F_1(\tfrac{1}{2}n; \tfrac{1}{2}r; \tfrac{1}{2}(I+\Lambda)^{-1}\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11\cdot 2}^{-1}B_1\Sigma_{11\cdot 2}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}})$$

where $\Lambda = \Sigma_{22}^{-\frac{1}{2}}\ \Sigma_{21}\Sigma_{11\cdot 2}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$.

The expression (2.4.12) can be simplified if $\Sigma\Sigma^{-1} = I$, then $\Sigma_{11\cdot 2} = (\Sigma^{11})^{-1}$ and $\beta_1 = \Sigma_{12}\Sigma_{22}^{-1} = -(\Sigma^{11})^{-1}\Sigma^{12}$.

It is easy to show that $|(I_r+\Lambda)|^{-\frac{1}{2}n} = |I-P|^{\frac{1}{2}n}$ consequently

(2.4.13) $${}_1F_1(\tfrac{1}{2}n; \tfrac{1}{2}r; \tfrac{1}{2}(I+\Lambda)^{-1}\Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11\cdot 2}^{-1}B_1\Sigma_{11\cdot 2}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}})$$

$$= {}_1F_1(\tfrac{1}{2}n; \tfrac{1}{2}r; \tfrac{1}{2}\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}B_1).$$

From (2.4.12) and (2.4.13) we have

$$A_k = |2\Sigma_{22}|^{n/2}|I - P|^{n/2}\Gamma_r(\tfrac{1}{2}n)\,{}_1F_1(\tfrac{1}{2}n;\tfrac{1}{2}r;\tfrac{1}{2}\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}B_1).$$

From (2.4.10) and Lemma 2.4.1 we have

$$(2.4.14)\quad f(B_1) = C_2^*|B_1|^{\frac{1}{2}(r-q-1)}\operatorname{etr}(-\tfrac{1}{2}\psi_1^{-1}B_1)\;{}_1F_1(\tfrac{1}{2}n;\tfrac{1}{2}r;\tfrac{1}{2}\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}B_1)$$

where $C_2^* = \dfrac{1}{\Gamma_q(\frac{r}{2})|2\psi_1|^{r/2}}\,|I-P|^{n/2}$.

Similarly for B_2 we have

$$(2.4.15)\quad f(B_2) = C_3^*|B_2|^{\frac{1}{2}(r-q-1)}\operatorname{etr}(-\tfrac{1}{2}\psi_2^{-1}B_2)\;{}_1F_1(\tfrac{1}{2}n;\tfrac{1}{2}r;\tfrac{1}{2}\Sigma^{*12}(\Sigma^{*22})^{-1}\Sigma^{*21}B_2)$$

where $C_3^* = \dfrac{1}{\Gamma_q(\frac{r}{2})|2\psi_2|^{r/2}}\,|I-P^*|^{n/2}$.

But B_1, B_2 are independently distributed and so we have

$$\begin{aligned}(2.4.16)\quad f(B_1,B_2) &= f(B_1)\cdot f(B_2)\\ &= C_0|B_1|^{\frac{1}{2}(r-q-1)}|B_2|^{\frac{1}{2}(r-q-1)}\\ &\quad \operatorname{etr}(-\tfrac{1}{2}\psi_1^{-1}B_1)\ \operatorname{etr}(-\tfrac{1}{2}\psi_2^{-1}B_2)\\ &\quad {}_1F_1(\tfrac{1}{2}n;\tfrac{1}{2}r;\tfrac{1}{2}\Omega_1B_1)\,{}_1F_1(\tfrac{1}{2}n;\tfrac{1}{2}r;\tfrac{1}{2}\Omega_2B_2)\end{aligned}$$

where $C_0 = \dfrac{|I-P|^{n/2}|I-P^*|^{n/2}}{\Gamma_q(\frac{r}{2})\Gamma_q(\frac{r}{2})|2\psi_1|^{r/2}|2\psi_2|^{r/2}}$,

$\Omega_1 = \Sigma^{12}(\Sigma^{22})^{-1}\ \Sigma^{21}$ and $\Omega_2 = \Sigma^{*12}(\Sigma^{*22})^{-1}\Sigma^{*21}$,

$\Sigma^{*-1}\Sigma^* = I$ and $\Sigma^{*-1} = \begin{pmatrix}\Sigma^{*11} & \Sigma^{*12}\\ \Sigma^{*21} & \Sigma^{*22}\end{pmatrix}$

We now make the transformation $B_1 = B_2^{\frac{1}{2}} V B_2^{\frac{1}{2}}$ with $J(B_1 \to V) = |B_2|^{\frac{1}{2}(q+1)}$, then we will get the joint density of B_2 and V. If we integrate over B_2 we finally get the density of V. So for the density of V we have

$$(2.4.17) \quad f(V) = C_0 |V|^{\frac{1}{2}(r-q-1)} \int_{B_2>0} |B_2|^{r-\frac{1}{2}(q+1)}$$

$$\text{etr}(-\tfrac{1}{2}\psi_1^{-1} B_2^{\frac{1}{2}} V B_2^{\frac{1}{2}}) \text{etr}(-\tfrac{1}{2}\psi_2^{-1} B_2)\, {}_1F_1(\tfrac{1}{2}n; \tfrac{1}{2}r; \tfrac{1}{2}\Omega_1 B_2^{\frac{1}{2}} V B_2^{\frac{1}{2}})$$

$${}_1F_1(\tfrac{1}{2}n; \tfrac{1}{2}r; \tfrac{1}{2}\Omega_2 B_2)\, dB_2$$

where C_0, Ω_1 and Ω_2 as in (2.4.16).

If we expand the hypergeometric functions, we will have

$$(2.4.18) \quad f(V) = C_0 |V|^{\frac{1}{2}(r-q-1)} \int_{B_2>0} |B_2|^{r-\frac{1}{2}(q+1)}$$

$$\text{etr}(-\tfrac{1}{2}\psi_1^{-1} B_2^{\frac{1}{2}} V B_2^{\frac{1}{2}}) \text{etr}(-\tfrac{1}{2}\psi_2^{-1} B_2)$$

$$\sum_{k=0}^{\infty} \sum_{\kappa} \frac{(\frac{1}{2}n)_\kappa}{(\frac{1}{2}r)_\kappa k!} C_\kappa(\tfrac{1}{2}\Omega_1 B_2^{\frac{1}{2}} V B_2^{\frac{1}{2}})$$

$$\sum_{s=0}^{\infty} \sum_{\sigma} \frac{(\frac{1}{2}n)_\sigma}{(\frac{1}{2}r)_\sigma s!} C_\sigma(\tfrac{1}{2}\Omega_2 B_2)\, dB_2$$

where C_0, Ω_1 and Ω_2 as in (2.4.16).

To find an explicit expression for the density of V, does not seem possible. However, we can compute the hth moment of $|V|$ as follows.

We have

$$(2.4.19) \quad E(|V|^h) = \int_{V>0} f(V)\ |V|^h\ dV$$

$$= C_0 \int_{V>0} |V|^{\frac{1}{2}(2h+r-q-1)} \int_{B_2>0} |B_2|^{r-\frac{1}{2}(q+1)}$$

$$\mathrm{etr}(-\tfrac{1}{2}\psi_1^{-1}B_2^{\frac{1}{2}}VB_2^{\frac{1}{2}})\ \mathrm{etr}(-\tfrac{1}{2}\psi_1^{-1}B_2)$$

$$\sum_{s=0}^{\infty}\sum_{\sigma}\frac{(\frac{1}{2}n)_\sigma(\frac{1}{2})^s}{(\frac{1}{2}r)_\sigma s!}\ C_\sigma(\Omega_2B_2)$$

$$\sum_{k=0}^{\infty}\sum_{\kappa}\frac{(\frac{1}{2}n)_\kappa(\frac{1}{2})^k}{(\frac{1}{2}n)_\kappa k!}\ C_\kappa(\Omega_1B_2^{\frac{1}{2}}VB_2^{\frac{1}{2}})\ dVdB_2$$

If we integrate first over V and then over B_2, we will obtain the hth moment of $|V|$. In order to integrate over V we need to compute the following integral

$$(2.4.20) \quad I_V = \int_{V>0} |V|^{\frac{1}{2}(2h+r)-\frac{1}{2}(q+1)}\ \mathrm{etr}(-\tfrac{1}{2}\psi_1^{-1}B_2^{\frac{1}{2}}VB_2^{\frac{1}{2}})C_\kappa(\Omega_1B_2^{\frac{1}{2}}VB_2^{\frac{1}{2}})dV$$

Using Lemma 2.2.2, we have

$$(2.4.21) \quad I_V = 2^k\Gamma_q(h+r/2\ ,\kappa)|2\psi_1|^{\frac{1}{2}(2h+r)}|B_2|^{-\frac{1}{2}(2h+r)}C_\kappa(\Omega_1\psi_1)$$

In order to integrate over B_2, we must compute the following integral

$$(2.4.22)\quad I_{B_2} = \int_{B_2>0} |B_2|^{\frac{1}{2}(r-2h)-\frac{1}{2}(q+1)} \operatorname{etr}(-\tfrac{1}{2}\psi_2^{-1}B_2)C_\sigma(\Omega_2 B_2)dB_2$$

$$= 2^s\ \Gamma_q(\tfrac{1}{2}(r-2h),\sigma)|2\psi_2|^{\frac{1}{2}(r-2h)}C_\sigma(\Omega_2\psi_2)$$

here we assume $\frac{1}{2}(r-2h) > \frac{1}{2}(q-1)$.

Combining (2.4.19), (2.4.21) and (2.4.22) we have finally

$$(2.4.23)\quad E(|V|^h) = C_0|2\psi_1|^{\frac{1}{2}(2h+r)}|2\psi_2|^{\frac{1}{2}(r-2h)}$$

$$\sum_{k=o}^{\infty}\sum_\kappa \frac{(\frac{1}{2}n)_\kappa}{(\frac{1}{2}r)_\kappa k!}\ \Gamma_q(\tfrac{1}{2}(2h+r),\kappa)\ C_\kappa(\Omega_1\psi_1)$$

$$\sum_{s=o}^{\infty}\sum_\sigma \frac{(\frac{1}{2}n)_\sigma}{(\frac{1}{2}r)_\sigma s!}\ \Gamma_q(\tfrac{1}{2}(r-2h),\sigma)\ C_\sigma(\Omega_2\psi_2)$$

Now using the well known relations, that if $\Sigma\Sigma^{-1} = I$ then

$\Sigma^{11} = \Sigma_{11\cdot2}^{-1}$ $\Sigma^{12} = -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot1}^{-1}$

$\Sigma^{22} = \Sigma_{22\cdot1}^{-1}$ $\Sigma^{21} = -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11\cdot2}^{-1}$

we have that $C_\kappa(\psi_1\Omega_1) = C_\kappa(\Sigma_{11\cdot2}\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}) =$ $C_\kappa(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) = C_\kappa(P)$. Similarly we have that $C_\sigma(\Omega_2\psi_2) = C_\sigma(P^*)$.

We can rewrite (2.4.23) in view of the above as

$$(2.4.24)\quad E(|V|^h) = C_0|2\psi_1|^{\frac{1}{2}(2h+r)}|2\psi_2|^{\frac{1}{2}(r-2h)}\Gamma_q(\tfrac{1}{2}(2h+r))$$

$$\Gamma_q(\tfrac{1}{2}(r-2h))\,{}_2F_1(\tfrac{1}{2}n,\tfrac{1}{2}(r+2h);\ \tfrac{1}{2}r;\ P)$$

$${}_2F_1(\tfrac{1}{2}n,\tfrac{1}{2}(r-2h);\ \tfrac{1}{2}r;\ P^*)$$

Given the moment sequence $E(|V|^h)$ of the random variable $|V|$ it can be shown that the density function which has these moments exists in the form of the inverse Mellin transform.

$$(2.4.25)\quad f(|V|) = \frac{1}{2\pi i}\oint_L E(|V|^h)|V|^{-h-1}\,dh.$$

$f(|V|)$ is the density function if we can show that the moments determine the density function uniquely. For example if $0 < |V| < 1$ we can show that the inverse Mellin transform is the density of $|V|$ by virtue of Carleman's theorem which we quote here.

<u>Theorem 2.4.2</u> If $\{\mu_h\}$ is a moment sequence such that $\sum_{h=1}^{\infty}\left(\frac{1}{\mu_{2h}}\right)^{\frac{1}{2h}}$ is divergent, then at most one distribution has the moment sequence $\{\mu_h\}$.

Without loss of generality we may assume $0 < |B_1| < |B_2|$, so that $0 < |V| < 1$, and hence from (2.4.24) and (2.4.25) we have

$$(2.4.26)\quad f(|V|) = C_\alpha \sum_{k=0}^{\infty}\Sigma_\kappa \frac{(\frac{1}{2}n)_\kappa}{(\frac{1}{2}r)_\kappa k!} C_\kappa(P) \sum_{s=0}^{\infty}\Sigma_\sigma \frac{(\frac{1}{2}n)_\sigma}{(\frac{1}{2}r)_\sigma s!} C_\sigma(P^*)$$
$$|V|^{-1}\frac{1}{2\pi i}\oint_L \left(\frac{|V||\psi_2|}{|\psi_1|}\right)^{-h}\Gamma_q(\tfrac{r}{2}+h,\kappa)\Gamma_q(\tfrac{r}{2}-h,\sigma)\,dh$$

where $$C_\alpha = \frac{|I-P|^{n/2}|I-P^*|^{n/2}}{\Gamma_q(\frac{r}{2})\Gamma_q(\frac{r}{2})}$$

$$(2.4.27)\quad I_L = \frac{1}{2\pi i}\oint_L \left(\frac{|V||\psi_2|}{|\psi_1|}\right)^{-h} \Gamma_q(h+\tfrac{r}{2},\kappa)\Gamma_q(\tfrac{r}{2}-h,\sigma)\, dh$$

$$= \pi^{q(q-1)/2}\ \frac{1}{2\pi i}\oint_L\left(\frac{|V||\psi_2|}{|\psi_1|}\right)^{-h} \Pi_{j=1}^{q}\Gamma(\tfrac{1}{2}r+h+k_j-\tfrac{1}{2}(j-1))$$

$$\Pi_{j=1}^{q}\Gamma(\tfrac{1}{2}r-h+s_j-\tfrac{1}{2}(j-1))\,dh$$

$$= \pi^{\frac{1}{2}q(q-1)}\frac{1}{2\pi i}\oint_L\left(\frac{|V||\psi_2|}{|\psi_1|}\right)^{-h}\frac{\Pi_{j=1}^{q}\Gamma(h+b_j)\Pi_{j=1}^{q}\Gamma(1-a_j-h)}{\Pi_{i=q+1}^{q}\Gamma(1-b_j-h)\Pi_{j=q+1}^{q}\Gamma(a_j+h)}dh$$

$$= \pi^{\frac{1}{2}q(q-1)} G^{qq}_{qq}\left(\frac{|V||\psi_2|}{|\psi_1|}\ \middle|\ \begin{matrix}a_1,\ldots,a_q\\ b_1,\ldots,b_q\end{matrix}\right)$$

where $a_j = -(\frac{r}{2} - 1 + s_j - \frac{1}{2}(j-1))$

and

$b_j = \frac{r}{2} + k_j - \frac{1}{2}(j-1)$

for $j = 1,2,\ldots,q$

Finally from (2.4.26) and (2.4.27) we have

$$(2.4.28)\quad f(|V|) = C_\alpha \pi^{\frac{1}{2}q(q-1)}\sum_{k=0}^{\infty}\Sigma_\kappa\frac{(\frac{1}{2}n)_\kappa}{(\frac{1}{2}r)_\kappa\, k!}C_\kappa(P)$$

$$\sum_{s=0}^{\infty}\Sigma_\sigma\frac{(\frac{1}{2}n)_\sigma}{(\frac{1}{2}r)_\sigma\, s!}C_\sigma(P^*)$$

$$|V|^{-1}\ G^{qq}_{qq}\left(\frac{|V||\psi_2|}{|\psi_1|}\ \middle|\ \begin{matrix}a_1,\ldots,a_q\\ b_1,\ldots,b_q\end{matrix}\right)$$

with C_α as in (2.4.26) and a_j, b_j for $j = 1,2,\ldots,q$ as in (2.4.27).

As is indicated above the densities of B_i, $(i = 1,2)$ are extremely complicated being unlike the densities of a noncentral Wishart distribution. From the fact that it was difficult to derive the density of $V = B_2^{-\frac{1}{2}} B_1 B_2^{-\frac{1}{2}}$ it is also difficult to derive the density of say $tr(V)$ or $tr(B_1(B_2+B_1)^{-1}) = tr(V(I+V)^{-1})$. But to test the hypothesis (2.4.1) one may also consider scalar functions of the following type $tr(B_1)/tr(B_2)$ or μ_1/μ_1^* where μ_1, μ_1^* are the largest roots of B_1 and B_2 respectively. That is, we are interested in scalar functions that only depend on the roots of B_1 and B_2. But since the distribution of the roots are invariant with respect to symmetrization it is possible to symmetrize the densities of the B_i first. The resulting expression will then be much simplified. It will be necessary to assume that $0 < tr(B_i) < 1$ for $i = 1,2$.

The author will also derive the density of $tr(B_1)/tr(B_2)$ and μ_1/μ_1^*, when $\Sigma_{11.2} = \psi_1$ and $\Sigma_{11.2}^* = \psi_2$ are known. In the latter case it will not be necessary to assume that $0 < tr(B_i) < 1$ for $i = 1,2$.

Lemma 2.4.4 The symmetrised density of B_1 is given as

$$f_s(B_1) = C_2^* |B_1|^{\frac{1}{2}(r-q-1)} \sum_{\ell=0}^{\infty} \Sigma_\lambda \frac{(-\frac{1}{2})^\ell}{\ell!} \sum_{k=0}^{\infty} \Sigma_\kappa \frac{(\frac{1}{2}n)_\kappa (\frac{1}{2})^k}{(\frac{1}{2}r)_\kappa k!}$$

$$\Sigma_\nu C_\nu(B_1) P_{\kappa,\lambda}^\nu(\psi_1^{-1}, \Omega_1)$$

with C_2^* as in (2.4.14), ν a partition of $n = k+\ell$ into not more than q parts and $P^{\nu}_{\kappa,\lambda}(\psi_1^{-1},\Omega_1)$ depends on ψ_1^{-1},Ω_1 only.

<u>Proof</u> Using Definition 2.2.12 from (2.4.14) we have

$$(2.4.29) \quad f_s(B_1) = C_2^*|B_1|^{\frac{1}{2}(r-q-1)} \int_{0(q)} \operatorname{etr}(-\tfrac{1}{2}\psi_1^{-1}HB_1H')\ {}_1F_1(\tfrac{1}{2}n;\ \tfrac{1}{2}r;\ \tfrac{1}{2}\Omega_1HB_1H')\ dH$$

$$= C_2^*|B_1|^{\frac{1}{2}(r-q-1)} \int_{0(q)} \sum_{\ell=0}^{\infty}\Sigma_\lambda \frac{(-\frac{1}{2})^\ell}{\ell!} C_\lambda(\psi_1^{-1}HB_1H')\ \sum_{k=0}^{\infty}\Sigma_\kappa \frac{(\frac{1}{2}n)_\kappa(\frac{1}{2})^k}{(\frac{1}{2}r)_\kappa k!}\ C_\kappa(\Omega_1HB_1H')\ dH$$

Using the following result by Khatri (1971)

$$(2.4.30) \quad \int_{0(q)} C_\lambda(H'\psi_1^{-1}HB_1)C_\kappa(H'\Omega_1HB_1)\ dH = \Sigma_\nu C_\nu(B_1)P^{\nu}_{\kappa,\lambda}(\psi_1^{-1},\Omega_1)$$

we finally have the symmetrised density of B_1.

<u>Theorem 2.4.5</u> The joint distribution of the latent roots $\mu_1 > \mu_2 > \dots > \mu_q > 0$ of B_1 such that $|B_1| \neq 0$ is

$$f(\mu_1,\mu_2,\dots,\mu_q) = C_3\ \Pi_{i<j}(\mu_i-\mu_j)\Pi_{i=1}^{q}\ \mu_i^{\frac{1}{2}(r-q-1)}$$

$$\sum_{\ell=0}^{\infty}\Sigma_\lambda \frac{(-\frac{1}{2})^\ell}{\ell!}\ \sum_{k=0}^{\infty}\Sigma_\kappa \frac{(\frac{1}{2}n)_\kappa(\frac{1}{2})^k}{(\frac{1}{2}r)_\kappa k!}\ \Sigma_\nu C_\nu(D\mu)P^{\nu}_{\kappa,\lambda}(\psi_1^{-1},\Omega_1)$$

with $C_3 = \dfrac{\pi^{\frac{1}{2}q^2}|I-P|^{n/2}}{\Gamma_q(\frac{1}{2}q)\Gamma_q(r/2)|2\psi_1|^{r/2}}$ and $D\mu = \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_q)$

Proof It follows from Corollary 2.2.15 and Lemma 2.2.9.

Theorem 2.4.6 The density of $tr(B_1) = W$, is

$$g(w) = \frac{|I-P|^{n/2}}{\Gamma_q(rq/2)|2\psi_1|^{r/2}} w^{\frac{1}{2}(rq-2)} \sum_{\ell=0}^{\infty} \Sigma_{\kappa} \frac{(-\frac{1}{2})^{\ell}}{\ell!} \sum_{k=0}^{\infty} \Sigma_{\kappa} \frac{(\frac{1}{2}n)_{\kappa}(\frac{1}{2})^{k}}{(\frac{1}{2}r)_{\kappa} k!}$$

$$w^{k+\ell} \Sigma_{\nu} \frac{(\frac{1}{2}r)_{\nu}}{(\frac{1}{2}rq)_{\nu}} C_{\nu}(I) P^{\nu}_{\kappa,\lambda}(\psi_1^{-1}, \Omega_1)$$

which is convergent for $0 < w < 1$.

Proof Consider the joint distribution of the latent roots of B_1 given by Theorem 2.4.5. We make the following transformation.

(2.4.31) $u_i = \mu_i/w, \quad i = 2,3,\ldots,q$

$u_1 = w(1-\sum_{i=2}^{q} u_i)$ and $\sum_{i=1}^{q} \mu_i = w$

The Jacobian of the above transformation is

$$J(\mu_1, \mu_2, \ldots, \mu_q \to w, u_2, \ldots, u_q) = w^{q-1}$$

Moreover we have

(2.4.32) $a_q(D\mu) = \Pi_{i<j}(\mu_i - \mu_j) = w^{\frac{1}{2}q(q-1)} \Pi_{i<j}(u_i - u_j)$

$$\Pi_{i=1}^{q} \mu_i^{\frac{1}{2}(r-q-1)} = w^{\frac{1}{2}(r-q-1)q}\, \Pi_{i=1}^{q} u_i^{\frac{1}{2}(r-q-1)}$$

The joint distribution of $u_2,\ldots,u_q$ and w is given by

(2.4.33) $$g(u_2,\ldots,u_q,w) = C_3 w^{q-1}\, w^{\frac{1}{2}q(q-1)}\, \Pi_{i<j}(u_i-u_j)$$

$$w^{\frac{1}{2}(r-q-1)q}\, \Pi_{i=1}^{q}\, u_i^{\frac{1}{2}(r-q-1)}$$

$$\sum_{\ell=0}^{\infty} \Sigma_\lambda \frac{(-\frac{1}{2})^\ell}{\ell!} \sum_{k=0}^{\infty} \Sigma_\kappa \frac{(\frac{1}{2}n)_\kappa(\frac{1}{2})^k}{(\frac{1}{2}r)_\kappa k!}\, w^{k+\ell}$$

$$\Sigma_\nu P^\nu_{\kappa,\lambda}(\psi_1^{-1},\Omega_1)\; C_\nu(Du)$$

with $Du = \mathrm{diag}(u_1,\ldots,u_q)$.

If we integrate now with respect to $u_2,\ldots,u_q$ using Lemma 2.2.11 we obtain the density of w.

<u>Theorem 2.4.7</u> The density function of the maximum latent root of B_1, i.e. μ_1 is

$$g(\mu_1) = \frac{\Gamma_q(\frac{1}{2}(q+1))\,|I-P|^{n/2}}{\Gamma_q(\frac{1}{2}(r+(q+1))\,|2\psi_1|^{r/2}} \sum_{\ell=0}^{\infty} \Sigma_\lambda \frac{(-\frac{1}{2})^\ell}{\ell!} \sum_{k=0}^{\infty} \Sigma_\kappa \frac{(\frac{1}{2}n)_\kappa(\frac{1}{2})^k}{(\frac{1}{2}r)_\kappa k!}$$

$$\mu_1^{\frac{1}{2}(qr)+k+\ell-1}\; (\tfrac{1}{2}rq+k+\ell)$$

$$\Sigma_\nu \frac{(r/2)_\nu}{(\frac{1}{2}(r+q+1))_\nu}\, C_\nu(I)\; P^\nu_{\kappa,\lambda}(\psi_1^{-1},\Omega_1)$$

<u>Proof</u> Consider the joint distribution of the latent roots given by Theorem 2.4.5. We make the transformation

(2.4.34) $m_i = \mu_i/\mu_1\,, \qquad i = 2,3,\ldots,q$

If we set $\Delta m = \mathrm{diag}(m_2,\ldots,m_q)$, ${}^1\Delta m = \mathrm{diag}(1,m_2,\ldots,m_q)$ then it is easy to show that

(2.4.35) $\quad a_q(D\mu) = \Pi_{i<j}(\mu_i-\mu_j) = \mu_1^{\frac{1}{2}q(q-1)}|I_{q-1}-\Delta m|\; a_{q-1}(\Delta m)$

$$J(\mu_1,\mu_2,\ldots,\mu_q \to \mu_1,\, m_2,\ldots,m_q) = \mu_1^{q-1}$$

$$|D\mu|^{\frac{1}{2}(r-q-1)} = \mu_1^{\frac{1}{2}q(r-q-1)}|\Delta m|^{\frac{1}{2}(r-q-1)}$$

The joint distribution of $m_2,\ldots,m_q$ and μ_1 is

(2.4.36) $\quad g(\mu_1,m_2,\ldots,m_q) = C_3\mu_1^{q-1}\mu_1^{\frac{1}{2}q(q-1)}|I_{q-1}-\Delta m|\; a_{q-1}(\Delta m)$

$$\mu_1^{\frac{1}{2}q(r-q-1)}\;|\Delta m|^{\frac{1}{2}(r-q-1)}$$

$$\sum_{\ell=0}^{\infty}\Sigma_\lambda\frac{(-\frac{1}{2})^\ell}{\ell!}\sum_{k=0}^{\infty}\Sigma_\kappa\frac{(\frac{1}{2}n)_\kappa(\frac{1}{2})^k}{(\frac{1}{2}r)_\kappa k!}\;\mu_1^{k+\ell}$$

$$\Sigma_\nu\; C_\nu({}^1\Delta m)\; P^\nu_{\kappa,\lambda}(\psi_1^{-1},\Omega_1)$$

Integrating with respect to $m_2,\ldots,m_q$ using Lemma 2.2.10, we finally have the density of μ_1.

<u>Theorem 2.4.8</u> The density function of $tr(B_1)/tr(B_2) = z_1$ is

$$f(z_1) = \frac{|I-P|^{n/2}|I-P*|^{n/2}}{\Gamma_q(\frac{rq}{2})\Gamma_q(\frac{rq}{2})|2\psi_1|^{r/2}|2\psi_2|^{r/2}}\sum_{\ell=0}^{\infty}\Sigma_\lambda\frac{(-\frac{1}{2})^\ell}{\ell!}$$

$$\sum_{k=0}^{\infty}\Sigma_\kappa\frac{(\frac{1}{2}n)_\kappa(\frac{1}{2})^k}{(\frac{1}{2}r)_\kappa k!}\;\Sigma_\nu\frac{(\frac{1}{2}r)_\nu}{(\frac{1}{2}rq)_\nu}C_\nu(I)\;P^\nu_{\kappa,\lambda}(\psi_1^{-1},\Omega_1)$$

$$\sum_{d=0}^{\infty}\Sigma_\delta\frac{(-\frac{1}{2})^d}{d!}\sum_{e=0}^{\infty}\Sigma_\varepsilon\frac{(\frac{1}{2}n)_\varepsilon(\frac{1}{2})^e}{(\frac{1}{2}r)_\varepsilon e!}\;(rq+k+\ell+d+e)^{-1}$$

$$z_1^{\frac{1}{2}(rq)+k+\ell-1} \Sigma_\xi \frac{(\frac{1}{2}r)_\xi}{(\frac{1}{2}rq)_\xi} C_\xi(I) \, P^\xi_{\delta,\varepsilon}(\psi_2^{-1},\Omega_2)$$

Hint: Without loss of generality we may assume $0 < tr(B_1) \leq tr(B_2) < 1$. Since $tr(B_1)$ and $tr(B_2)$ are independently distributed using Theorem 2.4.6 we can obtain the joint density of $tr(B_1)$ and $tr(B_2)$. If we make the transformation $\frac{tr(B_1)}{tr(B_2)} = z_1$ whose Jacobian is $tr(B_2)$ and integrate over $tr(B_2)$ we obtain the density of z_1.

Note: The density function of μ_1/μ_1^*, with μ_1, μ_1^* the largest latent roots of B_1 and B_2 respectively, can be obtained similarly.

We will now derive the joint density of the roots of B_1 when $\psi_1 = \Sigma_{11.2}$ is known. We have that B_1 conditionally on A_{22} is distributed as $W(\psi_1,\Omega_1,r)$, (refer to (2.4.5)). Consequently it is easy to show the joint distribution of the roots of B_1 conditionally on A_{22} is

(2.4.41) $$g(\theta_1,\theta_2,\ldots,\theta_q|A_{22}) = \frac{\pi^{\frac{1}{2}q^2}}{\Gamma_q(\frac{1}{2}q)\Gamma_q(\frac{1}{2}r)2^{\frac{1}{2}rq}} etr(-\tfrac{1}{2}\phi_1)$$

$$_0F_1(\tfrac{1}{2}r; \tfrac{1}{4}\phi_1, D\theta)$$

$$\Pi_{i=1}^q \theta_i^{\frac{1}{2}(r-q-1)} \, etr(-\tfrac{1}{2}D\theta) \, \Pi_{i<j}(\theta_i-\theta_j)$$

where $\theta_1 > \theta_2 > \ldots > \theta_q > 0$ are the roots of $|B_1-\theta\psi_1| = 0$ and $D\theta = diag(\theta_1,\theta_2,\ldots,\theta_q)$.

In order to obtain the unconditional density of the roots, we will multiply $g(\theta_1,\theta_2,\ldots,\theta_q|A_{22})$ in (2.4.41) by the density of $A_{22} \sim W(\Sigma_{22},n)$ and then we will integrate over A_{22}.

More analytically we have

$$(2.4.42)\quad g(\theta_1,\theta_2,\ldots,\theta_q) = C_\theta \Pi_{i=1}^{q}\ \theta_i^{\frac{1}{2}(r-q-1)} \mathrm{etr}(-\tfrac{1}{2}D\theta)\Pi_{i<j}(\theta_i-\theta_j)$$

$$\int_{A_{22}>0} |A_{22}|^{\frac{1}{2}(n-r-1)} \mathrm{etr}(-\tfrac{1}{2}\phi_1)\,{}_0F_1(\tfrac{1}{2}r;\ \tfrac{1}{4}\phi_1, D\theta)$$

$$\mathrm{etr}(-\tfrac{1}{2}\Sigma_{22}^{-1}A_{22})\ dA_{22}$$

with $\quad C_\theta = \pi^{\frac{1}{2}q^2}/\{\Gamma_q(\tfrac{1}{2}q)\Gamma_q)\tfrac{1}{2}r)2^{\frac{1}{2}rq}\Gamma_r(\tfrac{1}{2}n)|2\Sigma_{22}|^{\frac{1}{2}n}\}.$

We now compute the integral.

$$(2.4.43)\quad I_{A_{22}} = \int_{A_{22}>0} |A_{22}|^{\frac{1}{2}(n-r-1)} \mathrm{etr}\,(-\tfrac{1}{2}(\beta_1'\Sigma_{11\cdot2}^{-1}\beta_1+\Sigma_{22}^{-1})A_{22})$$

$${}_0F_1(\tfrac{1}{2}r;\ \tfrac{1}{4}\phi_1, D\theta)\ dA_{22}$$

If we make the transformation

$$A_{22} = \Sigma_{22}^{\frac{1}{2}} S \Sigma_{22}^{\frac{1}{2}} \quad \text{with} \quad J(A_{22} \to S) = |\Sigma_{22}|^{\frac{1}{2}(r+1)}$$

then (2.4.43) becomes

$$(2.4.44)\quad I_{A_{22}} = |\Sigma_{22}|^{\frac{1}{2}n} \int_{S>0} |S|^{\frac{1}{2}(n-r-1)}\ \mathrm{etr}(-\tfrac{1}{2}(I+\Lambda)S)$$

$$\sum_{k=0}^{\infty}\Sigma_\kappa \frac{((\tfrac{1}{2}r)_\kappa^{-1}(\tfrac{1}{4})^k(\tfrac{1}{2}r)_\kappa}{k!C_\kappa(I)} C_\kappa(D\theta)C_\kappa(\Lambda S)dS$$

with $\quad \Lambda = \Sigma_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11\cdot2}^{-1}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}}$

Using Lemma 2.2.2 we finally have

$$(2.4.45) \quad I_{A_{22}} = |2\Sigma_{22}|^{\frac{1}{2}n}\Gamma_r(\tfrac{1}{2}n)(I-P)^{n/2}{}_0F_2(\tfrac{1}{2}r;\ \tfrac{1}{2}n;\ \tfrac{1}{2}(I+\Lambda^{-1})^{-1},D\theta)$$

Combining (2.4.42) and (2.4.45) we have

$$(2.4.46) \quad g(\theta_1,\theta_2,\ldots,\theta_q) = C_\theta^*\Pi_{i=1}^{q}\theta_i^{\frac{1}{2}(r-q-1)}\mathrm{etr}(-\tfrac{1}{2}D\theta)\Pi_{i<j}(\theta_i-\theta_j)$$

$$ {}_0F_2(\tfrac{1}{2}r,\tfrac{1}{2}n;\ \tfrac{1}{2}(I+\Lambda^{-1})^{-1},\ D\theta)$$

with $$C_\theta^* = \frac{\pi^{\frac{1}{2}q^2}|I-P|^{n/2}}{\Gamma_q(\frac{1}{2}q)\Gamma_q(\frac{1}{2}r)2^{rq/2}}$$

Remarks

(i) The density functions of $tr(D\theta)$ and θ_1 can be easily obtained. Refer also to the proofs of Theorem 2.4.6 and Theorem 2.4.7.

(ii) The density functions of $tr(D\theta)/tr(D\theta^*)$ and θ_1/θ_1^* are also obtainable. Note that $tr(D\theta^*)$ is the trace of B_2 and θ_1^* is the maximum latent root of B_2.

2.5 TESTING THE EQUALITY OF TWO MULTIPLE CORRELATION MATRICES WHEN THE COVARIANCE MATRICES ARE DIFFERENT

In order to test the hypothesis given by (2.1.5), we consider the m.ℓ.e of $\theta = |P \cdot P^{*-1}|$, i.e.

$$\hat{\theta} = \frac{|A_{11}^{-\frac{1}{2}}A_{12}A_{22}^{-1}A_{21}A_{11}^{-\frac{1}{2}}|}{|A_{11}^{*-\frac{1}{2}}A_{12}^{*}A_{22}^{*-1}A_{21}^{*}A_{11}^{*-\frac{1}{2}}|} = |R||R^{*}|^{-1}$$

The m.ℓ.e of P, i.e. the sample correlation matrix $R = A_{11}^{-\frac{1}{2}}A_{12}A_{22}^{-1}A_{21}A_{11}^{-\frac{1}{2}}$ can be written as

(2.5.1) $$R = G^{-\frac{1}{2}}BG^{-\frac{1}{2}}$$

with $G = B+E$, $B = A_{12}A_{22}^{-1}A_{21}$ and $E = A_{11}-A_{12}A_{22}^{-1}A_{21}$.

The joint distribution of R, G when $P \neq 0$, i.e. $\Sigma_{12} \neq 0$ has been derived by Troskie (1969) and is given by

(2.5.2) $$f(R,G) = C_{\gamma}|I-P|^{n/2}|R|^{\frac{1}{2}(r-q-1)}|I-R|^{\frac{1}{2}(n-rq-1)}$$
$$G^{\frac{1}{2}(n-q-1)}\ \mathrm{etr}(-\tfrac{1}{2}\Sigma_{11.2}^{-1}G)$$
$${}_1F_1(\tfrac{1}{2}n;\ \tfrac{1}{2}r;\ \tfrac{1}{2}\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}G^{\frac{1}{2}}RG^{\frac{1}{2}})$$

with $R > 0$, $I-R > 0$, $\Sigma^{-1} = \begin{pmatrix}\Sigma^{11} & \Sigma^{21}\\ \Sigma^{21} & \Sigma^{22}\end{pmatrix}$ and

$$C_{\gamma} = \Gamma_q(\tfrac{1}{2}n)\{\Gamma_q(\tfrac{1}{2}r)\Gamma_q(\tfrac{1}{2}(n-r))\Gamma_q(\tfrac{1}{2}n)|2\Sigma_{11.2}|^{\frac{1}{2}n}\}^{-1}$$

The h-moment of $|R|$ can be found also in Troskie (1969) and is given by

(2.5.3) $$E(|R|^{h}) = \frac{\Gamma_q(\frac{1}{2}n)\Gamma_q(\frac{1}{2}r+h)}{\Gamma_q(\frac{1}{2}n+h)\Gamma_q(\frac{1}{2}r)}|I-P|^{\frac{1}{2}n}{}_3F_2(\tfrac{1}{2}n,\tfrac{1}{2}n,\tfrac{1}{2}r+h;\tfrac{1}{2}n+h,\tfrac{1}{2}r;P)$$

Similarly we have

$$(2.5.4) \qquad E(|R^*|^{-h}) = \frac{\Gamma_q(\frac{1}{2}n)\Gamma_q(\frac{1}{2}r-h)}{\Gamma_q(\frac{1}{2}n-h)\Gamma_q(\frac{1}{2}r)}|I-P^*|^{\frac{1}{2}n}$$

$$\qquad {}_3F_2(\tfrac{1}{2}n,\tfrac{1}{2}n,\tfrac{1}{2}r-h;\ \tfrac{1}{2}n-h,\tfrac{1}{2}r;\ P^*)$$

Without loss of generality we assume $0 < |R| < |R^*| < 1$, since $|R|$, $|R^*|$ are independently distributed we have

$$(2.5.5) \qquad E(|V|^h) = E(|R|/|R^*|)^h = \frac{\Gamma_q(\frac{1}{2}n)}{\Gamma_q(\frac{1}{2}r)}^2 |I-P|^{n/2}|I-P^*|^{n/2}$$

$$\frac{\Gamma_q(\frac{1}{2}r+h)\Gamma_q(\frac{1}{2}r-h)}{\Gamma_q(\frac{1}{2}n+h)\Gamma_q(\frac{1}{2}n-h)}\ {}_3F_2(\tfrac{1}{2}n,\tfrac{1}{2}n,\tfrac{1}{2}r+h;\ \tfrac{1}{2}n+h,\tfrac{1}{2}r;\ P)$$

$$\qquad {}_3F_2(\tfrac{1}{2}n,\tfrac{1}{2}n,\tfrac{1}{2}r-h;\ \tfrac{1}{2}n-h,\tfrac{1}{2}r;\ P^*)$$

The density of $|V|$ can be obtained using the inverse Mellin transform as follows, (see (2.4.25)). We have

$$(2.5.6) \qquad f(|V|) = \frac{1}{2\pi i}\oint_L E(|V|^h)|V|^{-h-1}\,dh$$

$$= C_\varepsilon |V|^{-1}\ \frac{1}{2\pi i}\oint_L \frac{\Gamma_q(\frac{1}{2}r+h)\Gamma_q(\frac{1}{2}r-h)}{\Gamma_q(\frac{1}{2}n+h)\Gamma_q(\frac{1}{2}n-h)}\ |V|^{-h}$$

$$\qquad {}_3F_2(\tfrac{1}{2}n,\tfrac{1}{2}n,\tfrac{1}{2}r+h;\ \tfrac{1}{2}n+h,\tfrac{1}{2}r;\ P)$$

$$\qquad {}_3F_2(\tfrac{1}{2}n,\tfrac{1}{2}n,\tfrac{1}{2}r-h;\ \tfrac{1}{2}n-h,\tfrac{1}{2}r;\ P^*)\ dh$$

with $\qquad C_\varepsilon = \left(\frac{\Gamma_q(\frac{1}{2}n)}{\Gamma_q(\frac{1}{2}r)}\right)^2 |I-P|^{n/2}|I-P^*|^{n/2}$

If we expand the hypergeometric functions of (2.5.6) and make use of (2.2.2), then in order to evaluate the integral of (2.5.6) it will be necessary to compute the integral I_L given by (2.5.7).

(2.5.7)
$$I_L = \frac{1}{2\pi i}\oint_L \frac{\Gamma_q(\tfrac{1}{2}r+h,\kappa)\Gamma_q(\tfrac{1}{2}r-h,\sigma)}{\Gamma_q(\tfrac{1}{2}n+h,\kappa)\Gamma_q(\tfrac{1}{2}n-h,\sigma)}|V|^{-h}\,dh$$

$$= \frac{1}{2\pi i}\oint_L |V|^{-h}\frac{\Pi_{i=1}^q\Gamma(\frac{r}{2}+h+k_j-\frac{1}{2}(j-1))\Pi_{i=1}^q\Gamma(\frac{r}{2}-h+s_j-\frac{1}{2}(j-1))}{\Pi_{j=1}^q\Gamma(\frac{n}{2}+h+k_j-\frac{1}{2}(j-1))\Pi_{j=1}^q\Gamma(\frac{n}{2}-h+s_j-\frac{1}{2})j-1))}dh$$

$$= \frac{1}{2\pi i}\oint_L |V|^{-h}\frac{\Pi_{j=1}^q\ (h+b_j)\Pi_{j=1}^q\ (1-h-a_j)}{\Pi_{j=q+1}^{2q}\Gamma(1-h-b_j)\Pi_{j=q+1}^{2q}\Gamma(a_j+h)}\,dh$$

$$= G_{2q\ 2q}^{q\ \ q}\left(|V|\ \middle|\ \begin{matrix}a_1,\ldots,a_{2q}\\ b_1,\ldots,b_{2q}\end{matrix}\right)$$

with
$$b_j = \begin{cases}\frac{r}{2}+k_j-\frac{1}{2}(j-1) & \text{for } j=1,2,\ldots,q\\ -(\frac{n}{2}-1+s_{j-q}-\frac{1}{2}(j-q-1)) & \text{for } j=q+1,\ldots,2q\end{cases}$$

and
$$a_j = \begin{cases}-(\frac{r}{2}-1+s_j-\frac{1}{2}(j-1)) & \text{for } j=1,2,\ldots,q\\ \frac{n}{2}+k_{j-q}-\frac{1}{2}(j-q-1) & \text{for } j=q+1,\ldots,2q\end{cases}$$

From (2.5.6) and (2.5.7) we finally have

(2.5.9)
$$f(|V|) = C_\varepsilon|V|^{-1}\sum_{k=0}^{\infty}\Sigma_\kappa\frac{(\frac{1}{2}n)_\kappa(\frac{1}{2}n)_\kappa}{(\frac{1}{2}r)_\kappa k!}C_\kappa(P)$$
$$\sum_{s=0}^{\infty}\Sigma_\sigma\frac{(\frac{1}{2}n)_\sigma(\frac{1}{2}n)_\sigma}{(\frac{1}{2}r)_\sigma s!}C_\sigma(P^*)$$
$$G_{2q\ 2q}^{q\ \ q}\left(|V|\ \middle|\ \begin{matrix}a_1,\ldots,a_{2q}\\ b_1,\ldots,b_{2q}\end{matrix}\right)$$

with C_ε and a_j, b_j as in (2.5.6) and (2.5.7) respectively.

Theorem 2.5.1 (de Waal 1969)

The density function of the largest sample canonical correlation coefficient r_1^2 of R is

$$h(r_1^2) = C_2(q,r,n)|I-P|^{n/2} \Sigma_{k=o}^{\infty}\Sigma_{\kappa}\Sigma_{t=o}^{\infty}\Sigma_{\tau}\Sigma_{\delta}(r_1^2)^{\frac{1}{2}rq+k+j-1}$$

$$(\frac{rq}{2} + k + j)A_{\kappa,\tau}^{\delta}(P)$$

with $C_2(q,r,n) = \dfrac{\Gamma_q(\frac{n}{2})\Gamma_q(\frac{q+1}{2})}{\Gamma_q(\frac{n-r}{2})\Gamma_q(\frac{r+q+1}{2})}$

and

$$A_{\kappa,\tau}^{\delta}(P) = g_{\kappa,\tau}^{\delta}(\tfrac{n}{2})_{\kappa}(\tfrac{n}{2})_{\kappa}(\tfrac{r}{2})_{\delta}(\tfrac{1}{2}(r-n+q+1))_{\tau}\frac{C_{\kappa}(P)C_{\sigma}(I)}{(\frac{r}{2})_{\kappa}(\frac{r+q+1}{2})_{\delta}C_{\kappa}(I)k!t!}$$

Theorem 2.5.2 (Troskie 1971)

The density function of $tr(R) = z$ such that $0 < z < 1$ is

$$h(z) = \frac{\Gamma_q(\frac{1}{2}n)}{\Gamma_q(\frac{1}{2}(n-r))\Gamma(\frac{1}{2}rq)}|I-P|^{n/2}z^{\frac{1}{2}rq-1}\Sigma_{k=o}^{\infty}\Sigma_{\kappa}\Sigma_{t=o}^{\infty}\Sigma_{\tau}\Sigma_{\delta}(\tfrac{1}{2}n)_{\kappa}(\tfrac{1}{2}n)_{\kappa}$$

$$(\tfrac{1}{2}(q-1+r-n))_\tau \; g^{\delta}_{\kappa,\tau} \; z^{k+t}$$

$$\frac{(\frac{1}{2}r)_\delta C_\kappa(P)C_\sigma(I)}{(\frac{1}{2}r)_\kappa(\frac{1}{2}rq)_\delta t!k!C_\kappa(I)}$$

Using the above mentioned theorems it will be easy to obtain the density functions of $tr(R)/tr(R^*)$ and r_1^2/r_1^{*2}, with r_1^{*2} the largest sample canonical correlation coefficient of R^*.

2.6 TESTING THE EQUALITY OF TWO MULTIPLE CORRELATION MATRICES IN THE LINEAR CASE

Let $\rho_1^2 \neq 0$ and $\rho_1^{*2} \neq 0$ be the population latent roots, i.e. the population canonical correlations of P and P^* respectively in the linear case, that is when Σ_{12} and Σ_{12}^* are of rank one. We can rewirte (2.1.5) as

(2.6.1) $\quad H_0 : \rho_1^2 = \rho_1^{*2}$ versus $H_1 : \rho_1^2 \neq \rho_1^{*2}$

As a test statistic for testing (2.6.1) we use

(2.6.2) $W_1 = \frac{|R|}{|R^*|}$ or $W_2 = \frac{|I-R|}{|I-R^*|}$

Since $|R|$ and $|R^*|$ are independently distributed we have

(2.6.3) $E(W_1^h) = E(|R|^h)\ E(|R^*|^{-h})$

similarly

(2.6.4) $E(W_2^h) = E(|I-R|^h)\ E(|I-R|^{-h})$

But the h^{th} moment of $|R|$, $|I-R|$ (see Troskie (1969)) is given by

(2.6.5) $$E(|R|^h) = \frac{\Gamma_q(\frac{1}{2}n)\Gamma_q(\frac{1}{2}r+h)}{\Gamma_q(\frac{1}{2}n+h)\Gamma_q(\frac{1}{2}r)}(1-\rho_1^2)^{\frac{1}{2}n}$$

$$_3F_2(\tfrac{1}{2}n,\tfrac{1}{2}n;\tfrac{1}{2}r+h;\tfrac{1}{2}n+h,\tfrac{1}{2}r;\rho_1^2)$$

and

(2.6.6.) $$E(|I-R|^h) = \frac{\Gamma_q(\frac{1}{2}n)\Gamma_q(\frac{1}{2}n-\frac{1}{2}r+h)}{\Gamma_q(\frac{1}{2}n+h)\Gamma_q(\frac{1}{2}n-\frac{1}{2}r)}(1-\rho_1^2)^{\frac{1}{2}n}$$

$$_2F_1(\tfrac{1}{2}n,\tfrac{1}{2}n;\tfrac{1}{2}n+h;\rho_1^2)$$

Similarly we have

(2.6.7) $$E(|R^*|^{-h}) = \frac{\Gamma_q(\frac{1}{2}n)\Gamma_q(\frac{1}{2}r-h)}{\Gamma_q(\frac{1}{2}n-h)\Gamma_q(\frac{1}{2}r)}(1-\rho_1^{*2})^{\frac{1}{2}n}$$

$$_3F_2(\tfrac{1}{2}n,\tfrac{1}{2}n;\tfrac{1}{2}r-h;\tfrac{1}{2}n-h,\tfrac{1}{2}r,\rho_1^2)$$

and

(2.6.8) $$E(|I-R^*|^{-h}) = \frac{\Gamma_q(\frac{1}{2}n)\Gamma_q(\frac{1}{2}n-\frac{1}{2}r-h)}{\Gamma_q(\frac{1}{2}n-h)\Gamma_q(\frac{1}{2}n-\frac{1}{2}r)}(1-\rho_1^{*2})$$

$$_2F_1(\tfrac{1}{2}n,\tfrac{1}{2}n;\tfrac{1}{2}n-h;\rho_1^{*2})$$

It is easy to show that

$$E(|R|^h) = \Pi_{j=1}^{q} E(X_j)^h$$

$$= E(\Pi_{j=1}^{q} X_j)^h$$

where $X_1, \ldots, X_q$ are independent Beta variables with density functions given by

$$(2.6.9) \quad \frac{\Gamma(\frac{1}{2}(n-j+1))}{\Gamma(\frac{1}{2}n-\frac{1}{2}r)\Gamma(\frac{1}{2}(r-j+1))} x_j^{\frac{1}{2}(r-j+1)-1}(1-x_j)^{\frac{1}{2}(n-r)-1}$$

for $j = 2,\ldots,q$

and

$$(2.6.10) \quad \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}r)\Gamma(\frac{1}{2}n-\frac{1}{2}r)} x_1^{\frac{1}{2}r-1}(1-x_1)^{\frac{1}{2}(n-r)-1}(1-\rho_1^2)^{\frac{1}{2}n}$$

$${}_2F_1(\tfrac{1}{2}n,\tfrac{1}{2}n;\tfrac{1}{2}r,\rho_1^2 x_1)$$

Similarly $E(|R^*|^{-h}) = \Pi_{j=1}^{q} E(X_j^*)^{-h}$

$$= E(\Pi_{j=1}^{q} X_j^*)^{-h}$$

where X_j^* are independent Beta variables with density functions given by

$$(2.6.11) \quad \frac{\Gamma(\frac{1}{2}(n-j+1))}{\Gamma(\frac{1}{2}n-\frac{1}{2}r)\Gamma(\frac{1}{2}(r-j+1))} x_j^{*\frac{1}{2}(r-j+1)-1}(1-x_j^*)^{\frac{1}{2}(n-r)-1}$$

for $j = 2,\ldots,q$

and

$$(2.6.12) \quad \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}r)\Gamma(\frac{1}{2}n-\frac{1}{2}r)} x_1^{*\frac{1}{2}r-1} (1-x_1^*)^{\frac{1}{2}(n-r)-1}(1-\rho_1^{*2})^{\frac{1}{2}n}$$

$${}_2F_1(\tfrac{1}{2}n,\tfrac{1}{2}n;\tfrac{1}{2}r;\rho_1^{*2} x_1^*)$$

Hence W_1 is distributed as a product of $2q$ independent real Beta variables with density functions given by (2.6.9), (2.6.10), (2.6.11) and (2.6.12). Note under $H_0 : \rho_1^2 = \rho_1^{*2}$.

It can be shown in the same way that W_2 is distributed as a product of $2q$ real independent variables $Z_1, Z_2, \ldots, Z_q$ and $Z_1^*, Z_2^*, \ldots, Z_q^*$ with density function given by (2.6.13), (2.6.14), (2.6.15) and (2.6.16)

$$(2.6.13) \quad \frac{\Gamma(\frac{1}{2}(n-j+1))}{\Gamma(\frac{1}{2}(n-r-j+1))\Gamma(\frac{1}{2}r)} z_j^{\frac{1}{2}(n-r-j+1)-1}(1-z_j)^{\frac{1}{2}r-1}$$

$$\text{for } j = 2,\ldots,q$$

$$(2.6.14) \quad \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n-\frac{1}{2}r)\Gamma(\frac{1}{2}r)} z_1^{\frac{1}{2}(n-r)-1}(1-z_1)^{\frac{1}{2}r-1}(1-\rho_1^2)^{\frac{1}{2}n}$$

$${}_2F_1(\tfrac{1}{2}n,\tfrac{1}{2}n;\tfrac{1}{2}r;\rho_1^2(1-z_1))$$

$$(2.6.15) \quad \frac{\Gamma(\frac{1}{2}(n-j+1))}{\Gamma(\frac{1}{2}(n-r-j+1))\Gamma(\frac{1}{2}r)} z_j^{*\frac{1}{2}(n-r-j+1)}(1-z_j^*)^{\frac{1}{2}r-1}$$

$$\text{for } j = 2,\ldots,q$$

$$(2.6.16) \quad \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n-\frac{1}{2}r)\Gamma(\frac{1}{2}r)} z_1^{*\frac{1}{2}(n-r)-1}(1-z_1^*)^{\frac{1}{2}r-1}(1-\rho_1^{*2})^{\frac{1}{2}n}$$

$${}_2F_1(\tfrac{1}{2}n,\tfrac{1}{2}n;\tfrac{1}{2}r;\rho_1^{*2}\ (1-z_1^*))$$

It is important to note that our results are valid, provided that the Gamma and Beta functions are defined

Denoting $|R|$ by $U_{q,n-r,r}$ for notational convenience we can state Theorem 2.2.17 as follows. In the non-central linear case the statistic $U_{2t,n-r,r}$ is distributed as $X_1 \ \Pi_{i=1}^{t-1} Y_i^2 X_{2t}$, where X_1 is independently distributed as in (2.6.12), Y_2 is independently distributed as $\beta(y_i; 2(r-2i); 2(n-r))$ and X_{2t} is independently distributed as $\beta(x_{2t}; r+1-q; n-r)$.

$U_{2s+1,n-r,r}$ is distributed as $X_1 \ \Pi_{i=1}^{s} Y_i^2$ where X_1 is distributed as in (2.6.12) and Y_i $i = 1,2,\ldots,s$ are independently distributed as $\beta(y_i; 2(r-2i); 2(n-r))$.

The above theorem can be found in Money (1972).

In order to derive the density of W_1 under H_0 we first derive the distribution of $|R|$ and $|R^*|$ under H_0. Since $|R|$, $|R^*|$ are independently distributed from their joint density, which is the product of their densities, we will be able to derive the density of W_1 by using the transformation $|R|/|R^*| = W_1$.

We give the following steps for the derivation of the density of $|R|$.

(i) We substitute for ${}_2F_1$ in (2.6.10) and make use of the binomial theorem in (2.6.9) and (2.6.10)

(ii) We make the transformation $y_i = -\log x_i \quad i = 1,2,\ldots,q.$ Consequently (2.6.9) and (2.6.10) become

$$(2.6.17) \quad f_{Y_1}(y_1) = (1-\rho_1^2)^{\frac{1}{2}n}[B(\tfrac{1}{2}r,\tfrac{1}{2}(n-r))]^{-1} \sum_{j=0}^{\infty} a_j \sum_{k=0}^{b=\frac{1}{2}(n-r-2)} (-1)^k \binom{b}{k} \exp(-\tfrac{1}{2}y_1(r+2j+2k))$$

with $y_1 \geq 0$ and $a_j = \dfrac{(\frac{1}{2}n)_j(\frac{1}{2}n)_j\rho_1^{2j}}{(\frac{1}{2}r)_j\, j!}$

and

$$(2.6.18) \quad f_{Y_i}(y_i) = K_i \sum_{\ell=0}^{b} (-1)^{\ell}\binom{b}{\ell}\exp(-\tfrac{1}{2}y_i(r-i+1+2\ell))$$

with $y_i \geq 0$ and $K_i = [B(\frac{1}{2}(r-i+1),\frac{1}{2}(n-r))]^{-1}$

(iii) Letting $Z_i = X_{2i}X_{2i+1}$ and making the transformation $y_i' = -\log z_i$, we get the density of Z_i and from the theorem on p.2.42, the density of Y_i' as

$$(2.6.19) \quad f_{Y_i'}(y_i') = (2B(r-2i,n-r))^{-1} \sum_{\ell=0}^{n-r-1} (-1)^{\ell} \binom{n-r-1}{\ell} \exp(-\tfrac{1}{2}y_i'(r+\ell-2i))$$

with $y_i' \geq 0.$

(iv) Use the convolution theorem successively as in Gupta (1971). It should be noted that when $n-r$ is an even integer then $b = \frac{1}{2}(n-r-2)$ is an integer. Here we assume that b is an integer. The calculations are very tedious even for $q = 2$

and therefore are omitted. It is also possible to obtain the densities of W_1 and W_2 by using the inverse Mellin transform.

It is important to note that all the results in the present chapter can be extended for the complex case. These results are not included here but they will be published in the near future.

CHAPTER THREE

F AND t TESTS FOR A GENERAL CLASS OF ESTIMATORS

3.1 INTRODUCTION

Consider the linear model

(3.1.1) $Y = \beta_0 1 + X\beta + \varepsilon$

where Y is an $n\times 1$ vector of response variables; 1 is an $n\times 1$ vector of ones; $X = (X_1, X_2,\ldots,X_p)$ is a full-column rank matrix, i.e. $r(X) = p$, of nonstochastic regressor variables, which are standardized so that $X_j'1 = 0$ and $X_j'X_j = 1$ for $j = 1,2,\ldots,p$; β_0 is an unknown scalar constant, while β is a $p\times 1$ vector of unknown regression coefficients; ε is an $n\times 1$ vector of unobservable randon error variables with $\varepsilon \sim N(0,\sigma^2 I)$.

Denote the latent roots of $X'X$ by $0 < \ell_1 \leq \ell_2 \leq \ldots \leq \ell_p$ and the corresponding orthonormal latent vectors by $V_1, V_2,\ldots,V_p$. Small latent roots and the corresponding latent vectors identify multicollinearities. This is because the $\ell_j = V_j'X'XV_j$ for $j = 1,2,\ldots,p$. It therefore follows that if r out of the p-latent roots are suitably small the above relation becomes

$$0 \simeq (XV_j)'(XV_j) \text{ and so } XV_j \simeq 0 \text{ for } j = 1,2,\ldots,r.$$

The problem of multicollinearity can be overcome by using biased estimation procedures such as Ridge, Principal Components, etc.

Consider the general form of the OLS, PC, SH and RR estimators which is given by

$$(3.1.2) \qquad \tilde{\beta} = \sum_{j=1}^{p} a_j C_j V_j$$

where $C_j = V_j'X'Y$ is the same for all estimators and a_j depends on the particular estimator. More specifically if

(3.1.3) (i) $a_j = \ell_j^{-1}$ then $\tilde{\beta}$ is the OLS estimator $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)$

(ii) $a_j = (\ell_j + k)^{-1}$ with $k > 0$, then $\tilde{\beta}$ is the RR estimator

(iii) $a_j = d^*\ell_j^{-1}$ where $0 \leq d^* < 1$. It is pointed out that although the choice of d^* depends on the data, in this paper it is treated as a constant; then $\tilde{\beta}$ is the deterministic SH estimator.

(iv) $a_j = 0$ for $j = 1,2,\ldots,r$ and $a_j = \ell_j^{-1}$ for $j = r+1,\ldots,p$ then $\tilde{\beta}$ is the PC estimator.

It is shown below that the t and F statistics for testing the general linear hypothesis using the estimator in (3.1.2) are the same for any choice of a_j ((3.1.3)(i) - (3.1.3)(iv)). Moreover we state some results concerning the sign of the estimator $\tilde{\beta}$ in (3.1.2).

It is important to stress that the theory which is presented here is classical rather than Bayesian. In what follows we make use of the Constrained Least Squares estimator (CLS).

where $(X'X)^{-1}_{ii}$ is that in (3.1.6) and S^2 is the least squares estimate for σ^2.

Note that the numerator of t_i is the i^{th} element of $\hat{\beta}$ minus zero, which is the expected value of $\hat{\beta}_i$ under H_0. The denominator is the square root of an unbiased estimator of the variance of the numerator. If we assume normality, under H_0, t_i of (3.2.2) has the central t-distribution with n-p-1 degrees of freedom.

A t-statistic for testing the hypothesis in (3.2.1) using $\tilde{\beta}_i$, which is the i^{th} component of $\tilde{\beta}$ in (3.1.2), for the cases (3.1.3)(ii) to (3.1.3)(iv), can be formed in direct analogy with (3.2.2), but it must be remembered that $\tilde{\beta}_i$ is biased.

For the i^{th} element of $\tilde{\beta}$ we have

$$\begin{aligned} (3.2.3) \qquad \tilde{\beta}_i &= \textstyle\sum_{j=1}^{p} a_j V_{ij} C_j \\ &= \textstyle\sum_{j=1}^{p} a_j V_{ij} (V_j' X'Y) \\ &= \textstyle\sum_{j=1}^{p} a_j V_{ij} (V_j' X'X\hat{\beta}) \end{aligned}$$

where V_{ij} is the i^{th} element of the j^{th} latent vector. If $\hat{e}_i$ denotes the linear unbiased, minimum variance estimator of $E(\tilde{\beta}_i)$, when $\beta_i = 0$ holds, then $\hat{e}_i$ is given by

$$(3.2.4)\qquad \hat{e}_i = \sum_{j=1}^{p} a_j V_{ij}(V_j' X' X\hat{\beta}_{ch})$$

$$= \sum_{j=1}^{p} a_j V_{ij}(V_j' XX[\hat{\beta}-\Delta_i^{-1}\ \hat{\beta}_i/(X'X)_{ii}^{-1}])$$

with $\hat{\beta}_{ch}$ as in (3.1.6).

The t-statistic for testing the hypothesis (3.2.1) is analogous to that of (3.2.2). Using $\tilde{\beta}_i$ of (3.2.3) this statistic is given by

$$(3.2.5)\qquad t_i^* = \frac{\tilde{\beta}_i-\hat{e}_i}{\{Var(\tilde{\beta}_i-\hat{e}_i)\}^{\frac{1}{2}}}$$

For the numerator of (3.2.5), using (3.2.3) and (3.2.4), we have

$$(3.2.6)\qquad \tilde{\beta}_i - \hat{e}_i = \left\{\sum_{j=1}^{p} a_j V_{ij}(V_j' X' X\hat{\beta})\right\} - \left\{\sum_{j=1}^{p} a_j V_{ij}(V_j' X' X [\hat{\beta}-\Delta_i^{-1}\hat{\beta}_i/(X'X)_{ii}^{-1}])\right\}$$

$$= \sum_{j=1}^{p} a_j V_{ij}\{(V_j'(X'X)\hat{\beta})-(V_j'(X'X)[\hat{\beta}-\Delta_i^{-1}\hat{\beta}_i/(X'X)_{ii}^{-1}])\}$$

$$= \sum_{j=1}^{p} a_j V_{ij} V_j'(X'X)\{\Delta_i^{-1}\hat{\beta}_i/(X'X)_{ii}^{-1}\}$$

$$= \sum_{j=1}^{p} a_j V_{ij}((X'X)V_j)'\{\Delta_i^{-1}\hat{\beta}_i/(X'X)_{ii}^{-1}\}$$

$$= \sum_{j=1}^{p} a_j V_{ij}(\ell_j V_j)'\{\Delta_i^{-1}\hat{\beta}_i/(X'X)_{ii}^{-1}\}$$

$$= \sum_{j=1}^{p} a_j V_{ij}\ell_j((\Delta_i^{-1})'V_j)'\hat{\beta}_i/(X'X)_{ii}^{-1}$$

$$= \sum_{j=1}^{p} a_j V_{ij}\ell_j(\ell_j^{-1}V_{ij})\hat{\beta}_i/(X'X)_{ii}^{-1}$$

$$= \sum_{j=1}^{p} a_j V_{ij}^2 \hat{\beta}_i/(X'X)_{ii}^{-1}$$

$$= d_i^2 \hat{\beta}_i/(X'X)_{ii}^{-1}$$

where $$d_i^2 = \sum_{j=1}^{p} a_j V_{ij}^2 .$$

In addition we have

(3.2.7) $$\mathrm{Var}(\tilde{\beta}_i - \hat{e}_i) = \mathrm{Var}(d_i^2 \hat{\beta}_i/(X'X)_{ii}^{-1})$$

$$= (d_i^2/(X'X)_{ii}^{-1})^2 \ \sigma^2 (X'X)_{ii}^{-1}$$

$$= \sigma^2 d_i^4/(X'X)_{ii}^{-1}$$

If the σ^2 in (3.2.7) is estimated by its unbiased estimator S^2 and we substitute together with (3.2.6) into (3.2.5) we obtain

(3.2.8) $$t_i^* = \frac{(d_i^2 \hat{\beta}_i/(X'X)_{ii}^{-1})}{(S^2 d_i^4/(X'X)_{ii}^{-1})^{\frac{1}{2}}}$$

$$= \hat{\beta}_i/(S^2 (X'X)_{ii}^{-1})^{\frac{1}{2}} = t_i$$

The d_i^2 is always positive except when $a_j = 0$ for $j = 1,2,\ldots,p$. It therefore follows that cancellation is permitted only when one or more of the a_j are positive.

The results of Leamer (1975) generalise to the following.

1. The estimate of β_j using the general estimator (3.1.2) and the CLS estimate of β_j, as given by (3.1.6), with $\beta_i = 0$ for $i \neq j$, have the same sign if the t-statistic of β_i is less in absolute value than the t-statistic of β_j.

2. The estimate of β_j using the $\tilde{\beta}$ in (3.1.2) and the CLS estimate $\hat{\beta}_c$ in (3.1.5), $\tilde{\beta}_j$ and $\hat{\beta}_{c,j}$ respectively, with $H\beta = h$, satisfy the inequality

$$|\tilde{\beta}_j - \hat{\beta}_{c,j}| \leq (S^2(X'X)^{-1}_{jj})^{\frac{1}{2}} \, r^{\frac{1}{2}} \, u_1^{\frac{1}{2}}$$

where r is the rank of H and u_1 is the statistic for testing the constraint $H\beta = h$ and is defined by (3.3.4) in Section 3 of the paper.

3.3 THE F-TEST FOR THE GENERAL ESTIMATOR

Suppose we want to test

(3.3.1) $H_0 : H\beta = h$ against $H_1 : H\beta \neq h$

with H, h as in (3.1.5).
Clearly we have $E(H\tilde{\beta}) \neq h$ when $\tilde{\beta}$ is not the OLS estimator of $\hat{\beta}$.

Let $\hat{e} = H \sum_{j=1}^{p} a_j V_j V_j'(X'X)\hat{\beta}_c$ be the linear unbiased minimum variance estimator of $E(H\tilde{\beta})$, when H_0 is true. Note that if $E(H\tilde{\beta})$ is known then $\mathrm{Var}(\hat{e}) = 0$. Consider now

$$\begin{aligned}
(3.3.2) \quad u^* &= H\tilde{\beta} - \hat{e} \\
&= H \textstyle\sum_{j=1}^{p} a_j V_j V_j'(X'X)\hat{\beta} - H \sum_{j=1}^{p} a_j V_j V_j'(X'X)\hat{\beta}_c \\
&= H \textstyle\sum_{j=1}^{p} a_j V_j V_j'(X'X)[\hat{\beta} - \hat{\beta}_c] \\
&= H \textstyle\sum_{j=1}^{p} a_j V_j V_j'(X'X)[\hat{\beta} - \{\hat{\beta} - (X'X)^{-1}H'(H(X'X)^{-1}H')^{-1} \\
&\qquad (H\hat{\beta} - h)\}]
\end{aligned}$$

$$= H \sum_{j=1}^{p} a_j V_j V_j' H'(H(X'X)^{-1}H')^{-1}(H\hat{\beta}-h)$$

$$= T\, u$$

where $\quad T = H \sum_{j=1}^{p} a_j V_j V_j' H'(H(X'X)^{-1}H')^{-1}$ and $u = H\hat{\beta}-h$.

Under H_0 we have

(3.3.3) (i) $E(u^*) = 0$

$$\text{(ii)}\quad Var(u^*) = T\,Var(u)T'$$
$$= T\,\sigma^2(H(X'X)^{-1}H')T'$$
$$= TQT' = Q^*$$

where $\quad Q = \sigma^2 H(X'X)^{-1}H'$.

If we substitute σ^2 by its estimate S^2 in (3.3.3)(ii) then the F-statistic for testing (3.3.1) will be

(3.3.4)
$$F^* = \frac{u^{*\prime}Q^{*-1}u^*}{r}$$
$$= \frac{u'T'(T')^{-1}Q^{-1}(T)^{-1}Tu}{r}$$
$$= \frac{u'Q^{-1}u}{r}$$
$$= \frac{(H\hat{\beta}-h)'(H(X'X)^{-1}H')^{-1}(H\hat{\beta}-h)}{rS^2}$$
$$= \frac{(\hat{\beta}-\hat{\beta}_c)(X'X)(\hat{\beta}-\hat{\beta}_c)}{rS^2} = F$$

In view of the above the following comments are made:

(i) The last relation represents the F-statistic for testing (3.3.1) when we make use of the least squares

CHAPTER FOUR

MULTIPLE INFERENCE AND SUBSET SELECTION BASED ON THE C_p-CRITERION OF MALLOWS

4.1 INTRODUCTION

Consider the general linear model

$$(4.1.1) \quad Y = X\beta + e$$

where Y is $n\times 1$ vector of observable random variables; X is $n\times(k+1)$ matrix of full rank with first column of unities; β is $(k+1)\times 1$ vector of regression coefficients; and e is $n\times 1$ vector of normal variables with $e \sim N(0,\sigma^2 I)$. We assume that X is fixed unless otherwise stated.

Let the model (4.1.1) be written as

$$(4.1.2) \quad Y = X_p\beta_p + X_r\beta_r + e$$

where X has been partitioned into X_p of dimension $n\times p$ and X_r of dimension $n\times r$. The β vector is partitioned conformably. Let $\hat{\beta} = \begin{pmatrix} \hat{\beta}_p \\ \hat{\beta}_r \end{pmatrix}$ denote the least squares estimate of β.

and let $\tilde{\beta}_p$ denote the subset least squares estimate of β_p, when the variables in X_r are deleted from the model. That is

$$(4.1.3) \quad \hat{\beta} = (X'X)^{-1}X'Y$$

and

(4.1.4) $\quad \tilde{\beta}_p = (X_p'X_p)^{-1}X_p'Y$

Furthermore let $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ represent the OLS estimates of variances for the two situations.

The C_p-criterion and the interpretation of C_p-plots were described by Mallows (1964, 1966) and subsequently by Corman and Toman (1966), Daniel and Wood (1971), Mallows (1973) and more recently by Spjøtvolle(1977).

In the C_p-criterion it is assumed that the k+1 variable model is the correct one, and using the mean square criterion of prediction we try to find out a p-variable model which explains as adequately the data as the k+1 variable model and having minimum C_p. Mallows (1966) suggested that the "standardized total squared error" be used as a criterion, given by

(4.1.5) $$\Gamma_p = \frac{SSB}{\sigma^2} + p$$
$$= \frac{E(SSE_p)}{\sigma^2} - n+2p$$

where SSB is the bias term, i.e. $SSB = (E(\tilde{Y}_p)-X\beta)'(E(\tilde{Y}_p)-X\beta)$, SSE_p is the residual sums of squares of the p-variable model and $\tilde{Y}_p = X_p\tilde{\beta}_p$.

As an estimate of Γ_p Mallows recommended the following statistic

$$(4.1.6) \quad C_p = \frac{SSE_p}{\hat{\sigma}^2} + 2p-n$$

Now if the p-variable model describes the data adequately then the bias will be near zero, in this case we have

$SSE_p \approx (n-p)\hat{\sigma}^2$ and so $C_p \approx \frac{(n-p)\hat{\sigma}^2}{\hat{\sigma}^2} + 2p-n = p.$

Now let $\bar{R}_p^2$ and $\bar{R}_k^2$ be the sample squared multiple correlation coefficients for the p and $k+1$ variables models; we can write (4.1.6) as (see Hocking (1976))

$$(4.1.7) \quad C_p = (n-k-1)\frac{(1-\bar{R}_p^2)}{(1-\bar{R}_k^2)} + 2p-n$$

or

$$(4.1.8) \quad C_p - p = r(F-1)$$

where F is the statistics for testing if $\beta_r = 0$.

4.2 SOME COMMENTS ON THE USE OF C_p

It has been stated by Hocking (1976) that when a regression model with parameters estimated from a given sample is used for prediction beyond the range of the data, the accuracy of such prediction can be very poor since the relationship between predictor and predictant in the sample does not necessarily remain the same outside the range of the sample. This is a very important point, which we should not overlook, since in the derivation of Γ_p it has been assumed that all future observations will arise in the same region as the regression sample.

In what follows we will discuss some relations between

C_p and p. These results follow from Hocking (1976). The potential uses of the regression equation are mainly parameter estimation and prediction. Now from Hocking (1976) we have that if

(4.2.1) $\quad Var(\hat{\beta}_r) - \beta_r\beta_r'$ is positive semi-definite (p.s.d.)

then

(4.2.2) $\quad Var(\hat{\beta}_p) - MSE(\tilde{\beta}_p)$ is p.s.d.

and

(4.2.3) $\quad VarP(\hat{y}) \geq MSEP(\tilde{y}_p)$.

Here we note that $MSE(\tilde{\beta}_p) = E(\tilde{\beta}_p-\beta_p)(\tilde{\beta}_p-\beta_p)'$, $Var(\hat{\beta}_r) = \sigma^2 B_{rr}$ with B_{rr} the appropriate submatrix of $(X'X)^{-1}$, $VarP(x'\hat{\beta}) = \sigma^2(1+x'(X'X)^{-1}x)$ with x' a $1\times(k+1)$ vector called predictor set, and finally $MSEP(\tilde{y}_p) = E(y-\tilde{y}_p)^2 =$ $\sigma^2(1+x_p'(X_p'X_p)^{-1}x_p) + (x_p'(X_p'X_p)^{-1}X_p'X_r\beta_r-x_r'\beta_r)^2$ with x_p' and x_r', $1\times p$ and $1\times r$ vectors respectively partitions of x'. In other words if (4.2.1) is satisfied then it is possible to estimate parameters and predict responses with smaller MSE using the subset equation. It is easy to establish that (4.2.1) is satisfied if the following relation holds.

$$(4.2.4) \qquad \frac{\beta_r' B_{rr}^{-1} \beta_r}{\sigma^2} \leq 1 \ .$$

Since β_r and σ^2 are unknown, estimating them from the full k+1 variable model we have,

$$(4.2.5) \qquad \frac{\hat{\beta}_r' B_{rr}^{-1} \hat{\beta}_r}{\hat{\sigma}^2} \leq 1$$

Notice that the test statistic used to test the hypothesis

H_0: $\beta_r = 0$ against H_1: $\beta_r \neq 0$ is given by

$$(4.2.6) \quad F = \frac{\hat{\beta}_r B_{rr}^{-1} \hat{\beta}_r}{r\hat{\sigma}^2}$$

and therefore (4.2.5) implies

$$(4.2.7) \quad F \leq \frac{1}{r}$$

Thus assuming the k+1-variable model is the correct one, then based on using the current data for fitting equations it seems reasonable to delete the variables in X_r if $F \leq \frac{1}{r}$. The claim is that with respect to MSE the subset equation yields better estimates of the parameter β_p and yields a better prediction equation. We must remember that results in (4.2.2) and (4.2.3) are valid for any predictor set x' (provided that (4.2.1) is satisfied) and therefore extrapolation beyond the range of the current data is permissible. Such extrapolation should, however, be done with caution since the above results based on the assumptions that

(i) The k+1-variable model is the correct one for all x'.

(ii) The p-term subset was selected without reference to the data.

These conditions are rarely met in practice.

A commonly used criterion for deleting variables, Efroymson (1966), is that the t-statistics associated with the parameter estimates for the full model are less than one in absolute value. Now a necessary condition for $\hat{\sigma}^2 B_{rr} - \hat{\beta}_r \hat{\beta}_r'$

to be p.s.d is that the t-statistic associated with the r-parameters in β_r are less than one in magnitude; thus it is clear that condition (4.2.7) is more restrictive. Pursuing the distinction between the predicting in the neighbourhood of the data and extrapolating outside the region, it may be argued that (4.2.7) is appropriate for extrapolation, but too restrictive for prediction, therefore relaxing the requirement (4.2.2) and considering only (4.2.3) we could use the subset for prediction if $\text{VarP}(\hat{y}) \geq \text{MSE}(\tilde{y}_p)$ provided that the prediction is done in the neighbourhood of the data. Using the matrix X as future input data we would require that:

$$(4.2.8) \qquad \text{VarP}(\hat{y}_i) - \text{MSEP}(\tilde{y}_{pi}) \geq 0$$

where the predictor set x' here is the ith row of X, i.e. X_i' and is partitioned to X_{ip}' and X_{ir}'.

From (4.2.8) we have

$$(4.2.9) \qquad \frac{1}{n}\sum_{i=1}^{n}[\text{VarP}(\hat{y}_i) - \text{MSEP}(\tilde{y}_{pi})] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}[\sigma^2(1+X_i'(X'X)^{-1}X_i)] - \frac{1}{n}\sum_{i=1}^{n}[\sigma^2(X_{ip}'(X_p'X_p)^{-1}X_{ip}+1) + (E(\tilde{y}_{pi}) - X_i'\beta)^2]$$

$$= \frac{1}{n}[n\sigma^2+\sigma^2\sum_{i=1}^{n}X_i'(X'X)^{-1}X_i] - \frac{1}{n}[n\sigma^2+\sigma^2\sum_{i=1}^{n}X_{ip}'(X_p'X_p)^{-1}X_{ip} + \sum_{i=1}^{n}(E(\tilde{y}_{pi})-X_i'\beta)^2]$$

$$= \frac{1}{n}[n\sigma^2+\sigma^2 tr(X'X)^{-1} \sum_{i=1}^n X_i X_i'] - \frac{1}{n}[n\sigma^2+\sigma^2 tr(X_p'X_p)^{-1}$$

$$\sum_{i=1}^n X_{ip}X_{ip}'] - \frac{1}{n}\sum_{i=1}^n (E(\tilde{y}_{p_i})-X_i'\beta)^2$$

$$= \frac{\sigma^2}{n}\{n+k+1-n-p-\frac{SSB}{\sigma^2}\}$$

$$= \frac{\sigma^2}{n}(r - \frac{SSB}{\sigma^2})$$

$$= \frac{r\sigma^2}{n}(1 - \frac{SSB}{r\sigma^2})$$

Since

(4.2.10) $$SSB = (E(\tilde{Y}_p)-X\beta)'(E(\tilde{Y}_p)-X\beta)$$

$$= (X\beta-E(\tilde{Y}_p))'(X\beta-E(\tilde{Y}_p))$$

$$= (X\beta-X_p(X_p'X_p)^{-1}X_p'X\beta)'(X\beta-X_p(X_p'X_p)^{-1}X_p'X\beta)$$

$$= \beta'X'(I-X_p(X_p'X_p)^{-1}X_p')X\beta$$

$$= \beta'\begin{pmatrix}X_p'\\X_r'\end{pmatrix}(I-X_p(X_p'X_p)^{-1}X_p')(X_p,X_r)\beta$$

$$= \beta_r'(X_r'X_r-X_r'X_p(X_p'X_p)^{-1}X_p'X_r)\beta_r$$

Using the partitioned inverse rule we have finally

(4.2.11) $$SSB = \beta_r'B_{rr}^{-1}\beta_r$$

Combining (4.2.8), (4.2.9) and (4.2.11) we have

(4.2.12) $$\frac{r\sigma^2}{n}\left(1 - \frac{\beta_r'B_{rr}^{-1}\beta_r}{r\sigma^2}\right) \geq 0$$

or

$$(4.2.13) \quad \frac{\beta_r' B_{rr}^{-1} \beta_r}{r\sigma^2} \leq 1$$

Replacing the parameters in (4.2.13) by their estimates we have

$$(4.2.14) \quad F = \frac{\hat{\beta}_r' B_{rr}^{-1} \hat{\beta}_r}{r\hat{\sigma}^2} \leq 1$$

Thus one would consider using $F \leq \frac{1}{r}$ if both extrapolation and precision in estimating β_p is the objective and the less restrictive objective condition $F \leq 1$, allowing for the deletion of more variables, if only prediction is the objective.

In view of the above discussion the relation of C_p with p will now be of interest.

From (4.1.8), since $0 < F < \infty$ we have $C_p - p \geq -r$ or $p - r \leq C_p$. If we are interested in extrapolation and precision, i.e. $F \leq \frac{1}{r}$ then from (4.1.8) we have $C_p \leq p+1-r = 2p-k$, i.e.

(4.2.15) $2p - k - 1 \leq C_p \leq 2p - k$ (extrapolation precision)

if prediction is the objective, i.e. $F \leq 1$ then from (4.1.8) we have

(4.2.16) $C_p \leq p$ (prediction).

If now $F \leq 2$, then we introduce bias; we are not allowed to extrapolate beyond the range of the data, but it is possible to find a subset with minimum mean square error of prediction. It is possible for the bias to offset the precision in estimating parameters. In terms of the C_p-statistic we have

(4.2.17) $C_p \leq k + 1$

The disadvantages, when we choose subsets based on (4.2.15), (4.2.16) and (4.2.17) can be summarised as follows:

(1) We do not have any statistical test, we rather have restrictive requirements.

(2) We have to specify the partition in advance; very impractical.

For these reasons we will consider tests, which we present in the next section.

4.3 SIMULTANEOUS TESTS

Aitken (1974) has given simultaneous procedures for the choice of variable subsets. His techniques are employed here.

From (4.1.5) we have $\Gamma_p = p + \frac{SSB}{\sigma^2}$ and from (4.2.11) we have $SSB = \beta_r' B_{rr}^{-1} \beta_r$, consequently the bias term will be zero if $\beta_r = 0$ and therefore $\Gamma_p = p$. Thus testing the hypothesis

(4.3.1) H_o: $\Gamma_p = p$ against H_1: $\Gamma_p > p$

is equivalent to testing

(4.3.2) H_0: $\beta_r = 0$ against H_1: $\beta_r \neq 0$.

The appropriate test statistic is given by

$$(4.3.3) \qquad F = \frac{\{Y'(X(X'X)^{-1}X' - X_p(X_p'X_p)^{-1}X_p')Y\}/r}{Y'(I-X(X'X)^{-1}X')Y/(n-k-1)}$$

$$= \frac{(\bar{R}_k^2 - \bar{R}_p^2)/r}{(1-\bar{R}_k^2)/(n-k-1)}$$

$$= \frac{(\hat{\beta}_r' B_{rr}^{-1} \hat{\beta}_r)/r}{\hat{\sigma}^2}$$

which has the F distribution with r and $n-k-1$ degrees of freedom.

The above test will have size α only if the partition of X into (X_p, X_r) is specified in advance. We now construct a simultaneous test for all partitions (X_p, X_r). Consider the statistic

$$(4.3.4) \qquad rF = u(X_p) = \frac{\bar{R}_k^2 - \bar{R}_p^2}{(1-\bar{R}_k^2)/(n-k-1)}$$

for testing (4.3.2). A simultaneous level α test for all hypotheses $\beta_r = 0$, for arbitrary partitions of β into (β_p, β_r) (including permutations of the elements of β) will be obtained by not rejecting when $u(X_p) < C_{n,k}^{\alpha}$ where $C_{n,k}^{\alpha}$ is the upper 100α% point of the null distribution of $u = \max_{X_p} u(X_p)$, the maximum being taken over all partitions of X into (X_p, X_r). The maximum of $u(X_p)$ occurs when X_p consists of the first column of X, i.e. column of 1's, so that

$$(4.3.5) \qquad u = \frac{\bar{R}_k^2}{(1-\bar{R}_k^2)/(n-k-1)}$$

But u/k has an $F'_{k,n-k-1}(\lambda)$ (non-central F) with $\lambda = \beta'X'X\beta/\sigma^2$, and when all hypotheses $\beta_r = 0$ are simultaneously true, then $\beta_r = 0$ and $\lambda = 0$ and thus $C^{\alpha}_{n,k} = k\,F^{(\alpha)}_{k,n-k-1}$ and the simultaneous test does not reject H_0: $\beta_r = 0$ for any partition if

$$(4.3.6)\qquad \frac{\bar{R}^2_k - \bar{R}^2_p}{(1-\bar{R}^2_k)/(n-k-1)} < k\,F^{(\alpha)}_{k,n-k-1}$$

or equivalently if

$$(4.3.7)\qquad \bar{R}^2_p > \bar{R}^2_0 = 1 - (1-\bar{R}^2_k)(1+k\,F^{(\alpha)}_{k,n-k-1} \times \frac{1}{(n-k-1)})$$

The subset of predictor variables corresponding to X_p will be called an $\bar{R}^2$- adequate (α) set if (4.3.7) is satisfied. Here adequate means "providing an $\bar{R}^2$ which is not significantly less than the $\bar{R}^2$ from the complete set of predictor variables." In many cases it will be sufficient to the $\bar{R}^2$ values for minimal adequate set, i.e. those for which $\bar{R}^2_p > \bar{R}^2_0$ but no subset of variables in X_p is adequate.

As with any simultaneous test procedure, the test for subhypotheses becomes increasingly conservative as the number of variables in the retained set decreases.

It is worth noticing that although the minimum C_p and the maximum $\bar{R}^2_p$ occur for the same set of variables, the value of p finally chosen may, of course, differ. The factor $(n-k-1)$ in (4.1.7) may cause sharp decrease in minimum C_p

values as p increases although $\bar{R}_p^2$ is only slowly increasing. Thus the C_p procedure may result in the selection of a larger set of variables than that from consideration of the $\bar{R}_p^2$ curve. Studies by Feiveson (1973) and by Radhakrishnan (1974), have indicated that essential variables may be deleted in the $\bar{R}_p^2$ procedure.

It is also possible to test whether the p-variable model is as good as the $k+1$ variable model, or even better. In order to do so, we test

(4.3.8) $\quad H_0: \Gamma_p \leq \Gamma_k$ against $H_1: \Gamma_p > \Gamma_k$

and since the $k+1$ variable model is assumed to be the correct one $\Gamma_k = k+1$. The above hypothesis using (4.1.5) can be written as

(4.3.9) $$H_0: \frac{E(SSE_p)}{\sigma^2} \leq k+1+n-2p \quad \text{against} \quad H_1: \frac{E(SSE_p)}{\sigma^2} > k+1+n-2p$$

or

(4.3.10) $$H_0: \frac{E(SSE_p)-(n-k-1)\sigma^2}{\sigma^2} \leq 2(k+1-p) = 2r$$

against

(4.3.11) $$H_1: \frac{E(SSE_p)-(n-k-1)\sigma^2}{\sigma^2} > 2r$$

As a test statistic we use the

(4.3.12) $$F_c = \frac{SSE_p-(n-k-1)\hat{\sigma}^2}{\hat{\sigma}^2} = \frac{\bar{R}_k^2 - \bar{R}_p^2}{(1-\bar{R}_k^2)/(n-k-1)}$$

Note that $F_c \sim r\, F'_{r,n-k-1}(\lambda^*)$ with $\lambda^* = \beta_r'(X_r'X_r - X_r'X_p(X_p'X_p)^{-1}X_p'X_r)\beta_r/\sigma^2$

A size α test is obtained by rejecting the hypothesis if

(4.3.13) $F_c > r\, F'^{(\alpha)}_{r,n-k-1}(2r)$

Since the hypothesis $\Gamma_p = p$ is more restricted than the hypothesis $\Gamma_p \leq k+1$, it is likely that more variables are to be selected using the above test. If now the partition is not specified in advance, by a similar argument as before, a simultaneous test of size α for all hypotheses over all partitions of X is obtained by rejecting any H_0 when

(4.3.14) $F_c > k\, F'^{(\alpha)}_{k,n-k-1}(2k)$

In terms of $\bar{R}_k^2$ and $\bar{R}_p^2$ (4.3.14) becomes

(4.3.15) $\bar{R}_p^2 > \bar{R}_0^{*2} = 1 - (1-\bar{R}_k^2)(1+k\, F'^{(\alpha)}_{k,n-k-1}(2k)\times(\frac{1}{(n-k-1)}))$

Since percentage points of the non-central F are not readily available, the central F approximation to the non-central F may be used. From Kendal and Stuart (1961, p.213) we have

(4.3.16) $F'_{v_1,v_2}(\lambda) \simeq (1 + \frac{\lambda}{v_1})\, F_{v^*,v_2}$

where $v^* = \frac{(v_1+\lambda)^2}{v_1+2\lambda}$

Now $v_1 = k$, $\lambda = 2k$ and $v_2 = n-k-1$, consequently where $e \sim N(0, \sigma^2 I)$.

For these data the total number of observations are $n = 13$ and the number of independent variables are five, i.e. $k+1 = 5$. Note that β_0 is included in all euqations in the table on the following page.

4.5 DISCUSSION

The procedure described here defines subsets, each one of them not being statistically poorer than the complete equation. The problem of the best subset still remains. It is up to the researcher to decide which subset is appropriate for his purpose. As it has been previously noted, if he is interested in prediction he would like to test if $H_0: \Gamma_p \leq k+1$, while for extrapolation and precision in estimating β's he would like to test if $H_0: \Gamma_p \leq p+1$. Once again simultaneous tests for the last hypothesis can be constructed as in Section 4.3. In the absence of any other considerations one of the minimal adequate equations might be considered.

With respect to choice of α, it is a matter for the experimenter to decide on the basis of power versus type one error considered, a level of $\alpha = 0{,}5$ has been occasionally recommended.

We also note that the procedure described here can be used to incorporate information such as whether some predictor variables should be included in the equation.

For example, suppose that prior information suggests the X_1 variable should be included in the equation. Then only subsets including variable X_1 should be included in the equation. Then only subsets including variable X_1 are to be tested, in this case the critical value will change, i.e. $k\,F'^{(\alpha)}_{k,n-k-1}(2k)$

$$(k-1)\ F'^{(\alpha)}_{k-1,n-k-1}(2(k-1))$$

REMARKS

1. It is possible for non adequate subsets to be classified as adequate. There are two reasons for this, the first being connected with the power of the test : if the hypothesis is not rejected it does not follow that it is true. Secondly, we are fitting a large number of equations, and by chance some of the incorrect ones may turn out to fit well.

2. If we construct simultaneous tests for the choice of subsets of variables based on the maximum $\bar{R}^2_p$ (Aitken (1974)) and on the minimum C_p then an $\bar{R}^2$- adequate (α) subset is also $C^0_p(\alpha)$ adequate but not vice versa since

$$F'^{(\alpha)}_{k,n-k-1}(2k) > F^{(\alpha)}_{k,n-k-1}\ .$$

3. It has been shown that the simultaneous tests for testing the hypothesis, i.e. there is no bias in the p-variable model (refer to (4.3.1)), are the same as those proposed by Aitken, based on the maximum $\bar{R}^2_p$.

CHAPTER 5

OUTLYING OBSERVATIONS USING GENERALISED INVERSE AND RIDGE RESIDUALS

5.1 INTRODUCTION

Consider the general linear model

(5.1.1) $Y = X\beta+e$

where Y is a $n\times 1$ vector of observed responses, X is the design matrix of dimension $n\times p$, assumed to have full rank p, i.e. $r(X) = p$ and to be standardised, β is a vector of unknown regression coefficients, and $e \sim N(0,\sigma^2 I)$ is an unobservable random error vector. We further make the assumption that $\beta'\beta < \infty$.

The presence of multicollinearity or near multicollinearity in the columns of X has deserved considerable attention in the recent literature, (see for example Hoerl and Kennard (1970, 1976), Mason, Gunst and Webster (1975), and Marquardt (1970)).

The general procedure is to compute the latent roots of $X'X$ and then focus attention on the small roots. If these roots are considered to be "too small" then a (biased) correction procedure is recommended. Alternatively if the "condi-

The "generalised inverse" of $X'X$ is given by

(5.2.8) $$A_r^+ = (X'X)^+ = V_r\Lambda_r^{-1}V_r' = \sum_{j=1}^{r} V_jV_j'/\lambda_j$$

and the generalized inverse or principal component estimator of β is given by

(5.2.9) $$\hat{\beta}_G = A_r^+X'Y$$

The generalized inverse (G.I.) residual is given by

(5.2.10) $$\hat{e}_G = Y - X\hat{\beta}_G = (I-XV_r\Lambda_r^{-1}V_r'X')Y = X(I-V_r\Lambda_r^{-1}V_r'X'X)\beta + (I-XV_r\Lambda_r^{-1}V_r'X')e$$

The variance of $\hat{e}_G$ is

(5.2.11) $$\mathrm{Var}(\hat{e}_G) = \sigma^2(I-XV_r\Lambda_r^{-1}V_r'X') = \sigma^2(I-XA_r^+X')$$

Lemma 5.2.1 If A is an $n\times n$ positive definite (p.d.) matrix and P is an $n\times m$ matrix with $r(P) = m$ then $P'AP$ is p.d. If A is non-negative definite (n.n.d) and P is any matrix then $P'AP$ is n.n.d.

Proof See for example Goldberger "Econometric Theory" pp 35-37.

<u>Theorem 5.2.2</u> Let $\hat{e}$ and $\hat{e}_R$ be the O.L.S. and R.R. residual vector respectively. Then $Var(\hat{e}_R) - Var(\hat{e})$ is a p.d. matrix.

<u>Proof</u> From (5.2.2) and (5.2.5) we have

$$Var(\hat{e}_R) - Var(\hat{e}) = \sigma^2\{(I-XWX'-kXWW'X') - (I-X(X'X)^{-1}X')\}$$
$$= \sigma^2 X\{(X'X)^{-1}-W-kWW'\}X' \ .$$

Since $r(X) = p$ the result will follow from Lemma 5.2.1 if B given by (5.2.12) is p.d.

(5.2.12) $B = (X'X)^{-1} - W - kWW'$

If we multiply (5.2.12) on the left by V' and on the right by V, we have

(5.2.13) $V'BV = V'(X'X)^{-1}V - V'WV - kV'WW'V$

Now the ith diagonal element of (5.2.13) is given by

(5.2.14) $\frac{1}{\lambda_i} - \frac{1}{\lambda_i+k} - \frac{k}{(\lambda_i+k)^2} = \frac{k}{\lambda_i(\lambda_i+k)^2} \geq 0$

Therefore if $k > 0$ then the roots of B are positive and so B is p.d.

<u>Theorem 5.2.3</u> If $\hat{e}$ and $\hat{e}_G$ are the O.L.S. and G.I. residual vectors then $Var(\hat{e}_G) - Var(\hat{e})$ is n.n.d.

<u>Proof</u> We have from (5.2.2) and (5.2.11)

(5.2.15) $Var(\hat{e}_G) - Var(\hat{e}) = \sigma^2 X\{(X'X)^{-1} - V_r\Lambda_r^{-1}V_r'\}X'$

Let $\Gamma = (X'X)^{-1} - V_r\Lambda_r^{-1}V_r'$ and multiply the left side by V' and the right side by V, then

$$V'\Gamma V = \begin{pmatrix} \Lambda_r^{-1} & 0 \\ 0 & \Lambda_{r-p}^{-1} \end{pmatrix} - \begin{pmatrix} \Lambda_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

Thus the roots of Γ are either zero or $\frac{1}{\lambda_i}$ for $i = r+1,\ldots,p$, where λ_i's are near zero but not exactly zero. This proves the result.

<u>Theorem 5.2.4</u> If $\hat{e}$, $\hat{e}_R$ and $\hat{e}_G$ are defined as before then

(i) Total variance of $\hat{e}$ is given by

$$tr(Var(\hat{e})) = \sigma^2(n-p)$$

(ii) Total variance of $\hat{e}_R$ is given by

$$tr(Var(\hat{e}_R)) = \sigma^2\{n-p + \textstyle\sum_{i=1}^{p}(\frac{k}{\lambda_i+k})^2\}$$

(iii) Total variance of $\hat{e}_G$ is given by

$$tr(Var(\hat{e}_G)) = \sigma^2(n-r).$$

<u>Proof</u> Follows directly using the trace operator.

From relations given by (5.2.1), (5.2.4), (5.2.10) we notice $\hat{e}$ is an unbiased estimator of e while $\hat{e}_R$ and $\hat{e}_G$ are biased estimators. From Theorems 5.2.2 and 5.2.3 the $\hat{e}_R$ and $\hat{e}_G$ estimators are more variable than the O.L.S. estimator $\hat{e}$. From Theorem 5.2.4 follows that the total variance of $\hat{e}$

is smaller than the total variance of the other two. Thus taking variance or total variance and the bias into account it seems hardly necessary to analyze R.R. or G.I. residuals. But what about the mean square error? Are the $\hat{e}_R$ and $\hat{e}_G$ closer to e than $\hat{e}$ is to e. As in the case with $\hat{\beta}_R$ and $\hat{\beta}$.

In what follows we examine $\hat{e}_R$, $\hat{e}_G$ and $\hat{e}$ with respect to the generalised mean square error as it has been defined by Theobald (1974).

Definition 5.2.5 (Theobald (1974)). Let $\tilde{\theta}_1$ and $\tilde{\theta}_2$ be two estimators of a vector parameter θ, and let $M_j = E(\tilde{\theta}_j-\theta)(\tilde{\theta}_j-\theta)'$. $j = 1,2$ be the second order moment matrices. Moreover let $m_j = E(\tilde{\theta}_j-\theta)'B(\tilde{\theta}_j-\theta)$, $j = 1,2$ where B is a n.n.d. matrix. The m_j is called generalized mean square error (g.m.s.e.). We say $\tilde{\theta}_2$ is a better estimator than $\tilde{\theta}_1$ if the g.m.s.e. of $\tilde{\theta}_2$ is less than the g.m.s.e. of $\tilde{\theta}_1$.

Lemma 5.2.6 (Theobald (1974)). The following two conditions are equivalent

(a) $M_1 - M_2$ is n.n.d.

(b) $m_1 - m_2 \geq 0$

for all n.n.d. B and m_j and M_j as in Definition 5.2.5.

Lemma 5.2.7 (Theobald (1974)). If

$$M_{LS} = E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$$
$$= \sigma^2 V\Lambda^{-1}V'$$

and
$$M_{RR} = E(\hat{\beta}_R-\beta)(\hat{\beta}_R-\beta)'$$
$$= \sigma^2 V\Lambda(\Lambda+kI)^{-2}V'+k^2V(\Lambda+kI)^{-1}V'\beta\beta'V(\Lambda+kI)^{-1}V'$$

then there exist a $K > 0$ such that

$$M_{LS} - M_{RR} \text{ is p.d.}$$

whenever $0 < k \leq K$.

Lemma 5.2.8 (Gunst and Mason (1976)). If M_{LS} as in Lemma 2.7 and

$$M_{GI} = E(\hat{\beta}_G-\beta)(\hat{\beta}_G-\beta)'$$
$$= \sigma^2 V_r\Lambda_r^{-1}V_r' + V_{p-r}V_{p-r}'\beta\beta V_{p-r}V_{p-r}'$$

then $M_{LS} - M_{GI}$ is p.d. if

$$\sum_{j=r+1}^{p} \frac{\lambda_j(V_j'\beta)^2}{\sigma^2} < 1 .$$

Theorem 5.2.9 There exist $K > 0$ such that the g.m.s.e. of $\hat{e}_R$ is less than the g.m.s.e. of $\hat{e}$, whenever $0 < k < K$.

Proof We have

$$\hat{e} - e = Y - X\hat{\beta} - (Y-X\beta)$$
$$= -X(\hat{\beta}-\beta)$$

and thus the second order moment matrix of $\hat{e}$ is

$$M_1 = E(\hat{e}-e)(\hat{e}-e)'$$
$$= E\{X(\hat{\beta}-\beta)(\hat{\beta}-\beta)'X'\}$$

$= X\ M_{LS}\ X'$

with M_{LS} as in Lemma 5.2.7.

Similarly

$$\hat{e}_R - e = -X(\hat{\beta}_R-\beta)$$

and so the second order matrix of $\hat{e}_R$ is

$$M_2 = E(\hat{e}_R-e)(\hat{e}_R-e)'$$
$$= X\ M_{RR}\ X'$$

with M_{RR} as in Lemma 5.2.7.

The result now follows from Definition 5.2.5 and Lemmas 5.2.6 and 5.2.7.

Theorem 5.2.10 The g.m.s.e. of $\hat{e}_G$ is less than the g.m.s.e. of e provided that

$$\sum_{j=r+1}^{p} \frac{\lambda_j (V_j'\beta)^2}{\sigma^2} < 1$$

Proof Let $M_3 = E(\hat{e}_G-e)(\hat{e}_G-e)'$

$$= X\{E(\hat{\beta}_G-\beta)(\hat{\beta}_G-\beta)\}X' = X\ M_{GI}\ X'$$

and M_{GI} be defined as in Theorem 5.2.8. The result follows using Definition 5.2.5 and Lemmas 5.2.6 and 5.2.8.

Theorems 5.2.9 and 5.2.10 show that although $\hat{e}_R$ and $\hat{e}_G$ are biased estimators of e, with respect to the g.m.s.e., they are under certain conditions better estimators of e than the O.L.S. estimator $\hat{e}$. This in fact supports Marquardt's recommendation that analysing R.R. and G.I. residuals may lead to profitable results and if both multicollinearity and erratic data points are present the problem must be tackled simultaneously. This will be attempted in the next section.

5.3 DETECTION OF OUTLIERS IN THE PRESENCE OF MULTICOLLINEARITY

Least Squares Residuals

We assume model (5.1.1) with e distributed as $N(0,\sigma^2 I)$. Let

(5.3.1) $\hat{e} = Y - X\hat{\beta} = (I - X(X'X)^{-1}X')e = Me$

with

$$\text{Var}(\hat{e}) = \sigma^2 M$$

The most common statistic for testing whether a single vector observation $(Y_i, X_{i1}, \ldots, X_{ip})'$ is an outlier for some i, is the ratio of the residual to the sample standard deviation thereof; i.e.

(5.3.2) $t_i = \hat{e}_i/(S^2 m_{ii})^{\frac{1}{2}}$

where $S^2 = \hat{\sigma}^2 = \Sigma\hat{e}_i/(n-p)$ and m_{ii} is the ith diagonal element of M. The statistic t_i is called the ith studentized residual. Let $T = \max|t_i|$ then a large value of T would indicate that the observation associated therewith is a possible outlier.

It is, however, not easy to find critical points for the statistic T. First attempt came from Tietjen, Moore and

Beckman (1973), who employed a gigantic simulation to estimate critical points for T in the univariate linear regression model $Y_i = \beta_0 + \beta_1 X_i + e_i$. This procedure is inadequate, as critical points are needed for the general model $Y = X\beta + e$, and it is difficult to justify a simulation of magnitude needed to produce reasonable tables.

However, two very interesting points emerge from the simulation study. Firstly, the authors were greatly concerned about the effects of the arrangement of the independent variables on the critical values of T. Hence they tested a number of rather extreme configurations of X. Their conclusion was that the arrangement of the independent variables does indeed influence the critical value, but the effect is for practical purposes negligible.

Secondly, the critical values for T estimated by Tietjen, Moore and Beckman are almost identical to the critical values calculated by Grubbs (1969) for the case of testing an outlier in a univariate sample.

This has led Prescott (1975) to develop a simple method for approximating critical values of T.

Assume that the variances of the residuals are constant. Let $T^* = \max|\hat{e}_i/\bar{S}|$, where $\bar{S}^2$ is the estimated average variance of the residuals.

Now

$$(5.3.3) \quad \overline{Var(\hat{e})} = \Sigma\, Var(\hat{e}_i)/n$$

$$= \Sigma[\sigma^2 - Var(\hat{y}_i)]/n$$

$$= n\sigma^2/n - tr[X(X'X)^{-1}X'\sigma^2]/n$$

$$= n\sigma^2/n - p\sigma^2/n$$

$$= (n-p)\sigma^2/n$$

Hence $\overline{s^2} = (n-p)s^2/n$

Therefore $T^* = \max|\hat{e}_i|/[(n-p)s^2/n]^{\frac{1}{2}}$

$$= \max|\hat{e}_i|/[\Sigma\ \hat{e}_i^2/n]^{\frac{1}{2}}$$

$$= n^{\frac{1}{2}} \max|\hat{e}_i|/[\Sigma\ \hat{e}_i^2]^{\frac{1}{2}}$$

$$= n^{\frac{1}{2}} \max|z_i| \quad \text{where} \quad z_i = \hat{e}_i/(\Sigma\ \hat{e}_i^2)^{\frac{1}{2}}$$

Stefansky (1971 and 1972) has proposed the maximum normed residual $|z| = \max|z_i|$ as a statistic for testing outliers, assuming that the residuals have constant variance. A method for calculating critical points of $|z|$ has been developed. It is based upon finding upper and lower bounds for the critical point, and then calculating successive improvements to the initial bounds, until the true critical point is reached.

A simple expression exists for the first upper bound of the critical point, U_1. In many cases the first upper bound will be equal to, or sufficiently close to, the critical point. If not, the first lower bound for the critical point L_1 must be calculated, then U_2, L_2, U_2,... etc., until the critical value is determined with sufficiency accuracy. However, L , U , L , ... must be determined by numerical methods.

The first upper bound for the $100(1-\alpha)$ percentile of $|z|$ is given by $U_1 = \{(n-p)F/[n(n-p-1+F)]\}^{\frac{1}{2}}$, where F is the $100(1-\alpha/n)$ percentage point of the F distribution with 1 and $n-p-1$ degrees of freedom, (Prescott (1975)).

Hence T^* is bounded above by $\{(n-p)F/(n-p-1+F)\}^{\frac{1}{2}}$

In the univariate linear regression model $Y_i = \beta_0 + \beta_1 X_i + e_i$, we have that $p = 2$ and hence T^* is bounded above by $\sqrt{(n-2)F/(n-3+F)}$. Prescott compares the critical values of T, computed by Tietjen, Moore and Beckman by simulation, with the upper bounds for T^* given by the above formula, and shows that they are almost identical. Hence he concludes that, apart from very extreme cases, the upper bound for T^* can be used as a reasonable approximation to the critical value for T.

The necessary points of the F distribution may be difficult to obtain, and so Lund (1975) has compiled tables of the upper bounds for T^* for α = 0,01, 0,05 and 0,10, sample sizes up to 100 and numbers of independent variables up to 25.

Generalized Inverse Residuals

Assuming that the first r roots of $X'X$ are significantly larger than zero, then the principal component estimator is $\hat{\beta}_G = A_r^+X'Y$ (refer to (5.2.8) and (5.2.9)). Notice that

$$(5.3.4) \qquad E(\hat{\beta}_G) = A_r^+X'X\beta = \Omega\beta$$

so that $\hat{\beta}_G$ is a biased estimator. If $A = X'X$ is of rank r, i.e. $r(A) = r$ then A_r^+ is the Moore-Penrose inverse and $\hat{\beta}_G$ is conditionally unbiased relative to the constraints implied by the columns of V_{p-r}. (See Chipman (1965)).

Now the generalized inverse residual $\hat{e}_G$ given by (5.2.10) can be written as

$$(5.3.5) \qquad \begin{aligned} \hat{e}_G &= (I-XA_r^+X')Y \\ &= (I-XA_r^+X')X\beta + (I-XA_r^+X')e \\ &\doteq NX\beta + Ne \\ &= \mu + Ne \end{aligned}$$

where

$$(5.3.6) \qquad N = I - XA_r^+X' \quad \text{and} \quad \mu = NX\beta$$

Notice that $E(\hat{e}_G) = NX\beta = \mu$ and $Var(\hat{e}_G) = \sigma^2NN' = \sigma^2N$, since N is symmetric idempotent.

Theorem 5.3.1

If $x \sim N(\mu, I)$ a set of necessary and sufficient conditions for $x'Qx + m'x + d$, to have a noncentral χ^2-distribution is

(a) $Q^2 \doteq Q$

(b) $m' = m'Q$

(c) $d = \frac{1}{4} m'm$

the degrees of freedom and the noncentrality parameter being given by $f = tr(Q)$ and $\lambda = d$.

Theorem 5.3.2 Let $\hat{e}_G = Y - X\hat{\beta}_G$, then $\hat{e}_G'\hat{e}_G/\sigma^2$ is distributed as noncentral $\chi_f^{2\prime}(\lambda)$, with $f = n-r$ and $\lambda = \beta'X'NX\beta/\sigma^2$.

Proof Since $\hat{e}_G = \mu + Ne$, with μ and N as in (5.3.6), we have

$$(5.3.7) \quad \hat{e}_G'\hat{e}_G = e'Ne + 2\mu'Ne + \mu'\mu$$

$$= e'Ne + m'e + d$$

We show now that the conditions of Theorem 5.3.1 are satisfied.

(a) $N^2 = N$

(b) $m'N = 2\mu'NN = 2\mu'N = m'$

(c) $\frac{1}{4} m'm = \frac{1}{4}(2\mu'N2N'\mu)$

$$= \mu'N\mu$$

$$= \beta'X'N'NX\beta$$

$$= \beta'X'NX\beta$$

$$= \mu'\mu = d$$

$$= \hat{e}_G'(I-N^*)\hat{e}_G$$

where $N^* = \begin{pmatrix} N_q^{-1} & 0 \\ 0 & 0 \end{pmatrix}$

<u>Theorem 5.3.3</u> Let S_q^2 be as in (5.3.10) then S_q^2/σ^2 is distributed as noncentral $\chi_f^{2\prime}(\lambda)$, where $f = n-r-q$ and $\lambda = \beta'X'(N-NN^*N)X\beta/\sigma^2$.

<u>Proof</u> We have from (5.3.5) and (5.3.10)

(5.3.11) $S_q^2 = \hat{e}_G'(I-N^*)\hat{e}_G$

$$= (\mu+Ne)'(I-N^*)(\mu+Ne)$$

$$= e'(N-NN^*N)e + 2\mu'(I-N^*)Ne + \mu'(I-N^*)\mu$$

$$= e'Qe + m'e + d$$

We show now that the conditions of Theorem 5.3.1 are satisfied.

(a) $Q^2 = Q = N-NN^*N$

(b) $m'Q = 2\mu'(I-N^*)N(N-NN^*N)$

$$= m' + D$$

with $D = -2\mu'(I-N^*)N(NN^*N)$

$$= -2\mu'(I-N^*)NN^*N$$

$$= -2\beta'X'N(I-N^*)NN^*N$$

$$= -2\beta'X\ NN^*N + 2\beta'X'NN^*NN^*N = 0$$

since $N^2 = N$ and $N^*NN^* = N^*$

$$(c)\ \tfrac{1}{4}\, m'm = \tfrac{1}{4}\{2\mu'(I-N^*)NN(I-N^*)2\mu\}$$

$$= \beta'X'(N-NN^*N)(N-NN^*N)X\beta$$

$$= \beta'X'QX\beta$$

since from (a) $N-NN^*N$ is idempotent. The degrees of freedom are $f = tr(N-NN^*N)$

$$= (n-r-q)$$

<u>Theorem 3.4</u> Let $\hat{e}_q$ and S_q^2 as in (5.3.8) and (5.3.10) respectively, then $\hat{e}_q$ and S_q^2 are independently distributed.

<u>Proof</u> We have

$$(5.3.12)\quad \hat{e}_q = [I_q, 0]\hat{e}_G$$

$$= [I_q, 0](\mu+Ne)$$

$$= \mu_q + [N_q, N_a]e$$

The result follows immediately since

$$[N_q, N_a](N-NN^*N) = 0.$$

Let the ith studentized G.I. residual be given by

$$(5.3.13)\quad t_{G,i} = \frac{\hat{e}_{G,i}}{(s_G^2 n_{ii})^{\frac{1}{2}}}$$

where $s_G^2 = \hat{e}_G'\hat{e}_G/(n-r)$

$$\doteq S_G^2/(n-r)$$

and $S_G^2 = \hat{e}_G'\hat{e}_G$. Note that n_{ii} is the diagonal element of N.

Let

$$(5.3.14)\quad \xi_i = \hat{e}_{G,i}/(S_G^2 n_{ii})^{\frac{1}{2}}, \qquad i : 1,\dots,n$$

<u>Theorem 5.3.5</u> The joint distribution of $\xi_1, \xi_2, \dots, \xi_q$ is given by

$$(5.3.15)\quad f(\xi_1,\dots,\xi_q) = (2\pi)^{-\frac{1}{2}q}|\sigma^2 N_q|^{-\frac{1}{2}}\ \Pi_{i=1}^{q}\ n_{ii}^{\frac{1}{2}}\ e^{-\frac{1}{2}(\lambda+\frac{1}{\sigma^2}\mu_q' N_q^{-1}\mu_q)}$$

$$\cdot\textstyle\sum_{\beta=0}^{\infty}(\frac{1}{2}\lambda)^{\beta}\ \{1-\Sigma_i\Sigma_j(n_{ii}n_{jj})^{\frac{1}{2}}n^{ij}\xi_i\xi_j\}^{\frac{1}{2}f+\beta-1}$$

$$\cdot\textstyle\sum_{\gamma=0}^{\infty}\{\Sigma_i\Sigma_j n^{ij}n_{ii}^{\frac{1}{2}}\mu_j\xi_i/\sigma^2\}^{\gamma}\ \Gamma(\frac{q+f+\gamma}{2}+\beta)$$

$$/\Gamma\ (\tfrac{1}{2}f+\beta)(2\sigma^2)^{-\frac{1}{2}(\gamma+q)}\beta!\gamma!$$

where $f = n-k-q$ and n^{ij} are elements of N_q^{-1}.

The probability density is defined over the range

$$(5.3.16)\quad \Sigma_i\Sigma_j(n_{ii}n_{jj})^{\frac{1}{2}}\ n^{ij}\xi_i\xi_j \le 1.$$

<u>Proof</u> Since $\hat{e}_q \sim N(\mu_q,\sigma^2 N_q)$ and $S_q^2/\sigma^2 \sim \chi_f'^2(\lambda)$ are independently distributed, their joint density is given by

$$(5.3.17)\quad f(\hat{e}_q,S_q^2) = (2\pi)^{-\frac{1}{2}q}|\sigma^2 N_q|^{-\frac{1}{2}}e^{-\frac{1}{2\sigma^2}(\hat{e}_q-\mu_q)'N_q^{-1}(\hat{e}_q-\mu_q)}$$

$$e^{-\frac{1}{2}\lambda}e^{-\frac{1}{2\sigma^2}S_q^2}\textstyle\sum_{\beta=0}^{\infty}\dfrac{(\frac{1}{2}\lambda)^{\beta}(S_q^2)^{\frac{1}{2}f+\beta-1}}{(2\sigma^2)^{\frac{1}{2}f+\beta}\Gamma(\frac{1}{2}f+\beta)\beta!}$$

Make the transformation

(5.3.18) $\xi_i = \frac{\hat{e}_{G,i}}{(S_G^2 n_{ii})^{\frac{1}{2}}}$ $i = 1,2,\ldots,q$

$S_G = \sqrt{S_G^2} = \sqrt{S_q^2 + \hat{e}_q' N_q^{-1} \hat{e}_q}$ (refer to (5.3.10))

the Jacobian being $2S_G^{q+1} \Pi_{i=1}^q n_{ii}^{\frac{1}{2}}$.

The joint density of $\xi_1,\ldots,\xi_q$ and S_G is

(5.3.19) $f(\xi_1,\xi_2,\ldots,\xi_q,S_G) = (2\pi)^{-\frac{1}{2}q}|\sigma^2 N_q|^{-\frac{1}{2}} \Pi_{i=1}^q n_{ii}^{\frac{1}{2}}$

$$\cdot e^{\frac{1}{2\sigma^2}\mu_q' N_q^{-1} \mu_q - \frac{1}{2}\lambda} \; e^{-\frac{1}{2\sigma^2} S_G^2} \; 2S_G^{q+1}$$

$$\cdot e^{\Sigma_i \Sigma_j \mu_j n^{ij} S_G n_{ii}^{\frac{1}{2}} \xi_i / \sigma^2}$$

$$\cdot \sum_{\beta=0}^{\infty} \frac{(2\sigma^2)^{-(\frac{1}{2}f+\beta)}(\frac{1}{2}\lambda)^\beta}{\Gamma(\frac{1}{2}f+\beta)\beta!} [S_G^2\{1-\Sigma_i\Sigma_j n^{ij}\xi_i\xi_j n_{ii}^{\frac{1}{2}} n_{jj}^{\frac{1}{2}}\}]^{\frac{1}{2}f+\beta-1}$$

with n^{ij} the (i,j) element of N_q^{-1}.

If we integrate now with respect to S_G we get the joint density of $\xi_1, \xi_2,\ldots,\xi_q$.

Let

(5.3.20) $I = \int_0^\infty 2S_G^{q+1} \, e^{-\frac{1}{2\sigma^2}S_G^2 + S_G \Sigma_i \Sigma_j \mu_j n^{ij} n_{ii}^{\frac{1}{2}} \xi_i / \sigma^2} (S_G^2)^{\frac{1}{2}f+\beta-1} dS_G$

Let $d^* = \Sigma_i \Sigma_j \mu_j \, n^{ij} n_{ii}^{\frac{1}{2}} \xi_i$

then (5.3.20) becomes

(5.3.21) $I = 2\int_0^\infty S_G^{(q+1)} \, e^{-\frac{1}{2\sigma^2} S_G^2 + \frac{d^*}{\sigma^2} S_G} (S_G^2)^{(\frac{1}{2}f+\beta)-1} \, dS_G$

$$= 2\int_0^\infty S_G^{2\left(\frac{q+f}{2}+\beta-\frac{1}{2}\right)} e^{-\frac{1}{2\sigma^2} S_G^2} \sum_{\gamma=0}^\infty \frac{(\frac{d^*}{\sigma^2})^\gamma S_G^\gamma}{\gamma!} \, dS_G$$

Let $W = S_G^2$ then $dW = 2S_G \, dS_G$ and integrating over W yields the result.

<u>Corollary 5.3.6</u> If $\lambda = \beta'X'(N-NN^*N)X\beta/\sigma^2 = 0$ then the density reduces to

$$(5.3.22) \quad f(\xi_1,\xi_2,\ldots,\xi_q) = (2\pi)^{-\frac{1}{2}q}|\sigma^2 N_q|^{-\frac{1}{2}}\Pi_{i=1}^q n_{ii}^{\frac{1}{2}} e^{-\frac{1}{2\sigma^2}\mu_q' N_q^{-1}\mu_q}$$

$$(1-\Sigma_i\Sigma_j\xi_i\xi_j n^{ij} n_{ii}^{\frac{1}{2}} n_{jj}^{\frac{1}{2}})^{\frac{1}{2}f-1} \sum_{\gamma=0}^\infty \{\Sigma_i\Sigma_j n^{ij} n_{ii}^{\frac{1}{2}}\mu_j\xi_i/\sigma^2\}^\gamma$$

$$\Gamma(\tfrac{1}{2}(f+q+\gamma))/(2\sigma^2)^{-\frac{1}{2}(\gamma+q)}\gamma! \; \Gamma(\tfrac{1}{2}f)$$

<u>Corrolary 5.3.7</u> If moreover $\mu = (I-XA_r^+X')X\beta = 0$, i.e. $\mu_q = 0$ then the joint distribution of the q-standardised errors is

$$(5.3.23) \quad f(\xi_1,\xi_2,\ldots,\xi_q) = \frac{\Gamma(\frac{f+q}{2})|N_q^{-1}|^{\frac{1}{2}}\Pi_{i=1}^q n_{ii}^{\frac{1}{2}}}{\pi^{q/2}\;\Gamma(\frac{f}{2})}$$

$$(1-\Sigma_i\Sigma_j(n_{ii}n_{jj})^{\frac{1}{2}} n^{ij}\xi_i\xi_j)^{\frac{1}{2}f-1}$$

In this last form the density was given by Ellenberg (1973). This is of course the case when $r(X) = r = p$, i.e. $\hat{\beta}_G = \hat{\beta}$, the O.L.S. estimator of β. Ellenberg (1973) further suggested that a test for deleting outliers in the O.L.S. model is $\xi^* = \max|\xi_i|$. The null hypothesis of no outliers would then be rejected if under the null-theory, ξ^* were too large. He then computed upper and lower bounds for the percentage points of ξ^* using a second order Bonferroni inequality (David (1956)). For any critical constant C, the inequality is given as

(5.3.24) $$\sum_{i=1}^{n} \Pr(|\xi_i| > C) - \sum_{i>j} \Pr(|\xi_i| > C, |\xi_j| > C)$$
$$\leq \Pr(\xi \;\; > C) \leq \sum_{i=1}^{n} \Pr(|\xi_i| > C)$$

Since $\xi_i = \dfrac{t_i}{(n-p)^{\frac{1}{2}}}$ (with $r = p$), we have

(5.3.25) $$\Pr\{\max|t_i| > Y_0\} = \Pr\{\max|\xi_i| \geq Y_0^*\}$$

with $$Y_0^* = \frac{Y_0}{(n-p)^{\frac{1}{2}}}$$

consequently Ellenberg's test is equivalent to that of the maximum studentized residual T.

Because of the non-centralities involved in (5.3.15), it is evidently clear that it will not be easy to compute an upper bound for $\Pr\{|\xi^*| \geq Y_0^*\}$. Let us examine under what conditions will the non-centrality parameters be equal to zero. Obviously if $\beta = 0$, both μ and λ will be zero, but this case is of no practical interest. If the $r(X)$ is indeed r, then A_r^+ is a Moore-Penrose inverse of $X'X$. Hence

(5.3.26) $$\hat{e}_G = (I - XA_r^+X')e$$

since $(X-XA_r^+X'X)\beta = 0$, as $XA_r^+X'X = X$. A_r^+ is the Moore-Penrose inverse of $X'X$. The joint density of $\xi_1, \xi_2, \ldots, \xi_q$ will be given by (5.3.23), so that Ellenberg's test could be used. But the performance of this test procedure for any given problem will depend not only on the size of the outlier but also on the correlation structure among the residuals that is of the n_{ij}. Therefore following Prescott's suggestion, let us instead of $T_G = \max_i |t_{G,i}| = \max_i |(n-r)^{\frac{1}{2}}\xi_i|$ use the statistic

$$(5.3.27) \quad T_G^* = \max \left| \frac{\hat{e}_{G,i}}{\bar{s}_G} \right|$$

where $\bar{s}_G$ is the estimated average variance of the G.I. residuals, that is

$$(5.3.28) \quad \overline{Var(\hat{e}_G)} = \textstyle\sum_{i=1}^n Var(\hat{e}_{G,i})/n$$

$$= \frac{\sigma^2}{n} \, tr(N)$$

$$= \sigma^2 \frac{(n-r)}{n}$$

and so

$$(5.3.29) \quad \bar{s}_G^2 = \left(\frac{n-r}{n}\right) s_G^2 = (Y-X\hat{\beta}_G)'(Y-X\hat{\beta}_G)/n$$

$$= \hat{e}_G'\hat{e}_G/n \, .$$

Hence

$$(5.3.30) \quad T_G^* = \max|\hat{e}_{G,i}/(\hat{e}_G'\hat{e}_G/n)^{\frac{1}{2}}|$$

$$= n^{\frac{1}{2}} \max|Z_{G,i}|$$

where $Z_{G,i}$ is the studentized G.I. residual.

Now if,

$$(5.3.31)\quad u_i^2 = \frac{(n-r-1)\hat{e}^2_{G,i}/n_{ii}}{\hat{e}'_G\hat{e}_G-\hat{e}^2_{G,i}/n_{ii}}$$

with $\quad n_{ii} = \mathrm{Var}(\hat{e}^2_{G,i})/\sigma^2$

then we have

$$(5.3.32)\quad u_i^2 = (n-r-1)\,\frac{Z^2_{G,i}/n_{ii}}{1-Z^2_{G,i}/n_{ii}}$$

But from Theorems 5.3.3 and 5.3.4, it follows that u_i^2 has a doubly noncentral F-distribution with non-centrality parameters given by

$$(5.3.33)\quad \lambda_1 = \frac{\mu_i^2}{n_{ii}\sigma^2}\ , \quad \lambda_2 = \frac{1}{\sigma^2}\,\beta'X\ (N-NN^*N)X\beta$$
$$= \frac{1}{\sigma^2}\,\mu'(I-N^*)\mu$$

Now if the variances are assumed to be equal, i.e. $\mathrm{Var}(\hat{e}_{G,i}) = \sigma^2 n_{ii} = \sigma^2 C_\alpha$ then since $\Sigma\ \mathrm{Var}(\hat{e}_{G,i}) = \sigma^2(n-r)$ it follows that $C_\alpha = \frac{n-r}{n}$.

In this case (5.3.33) becomes

$$(5.3.34)\quad u_i^2 = \frac{(n-r-1)Z^2_{G,i}/((n-r)/n)}{1-Z^2_{G,i}/(\frac{n-r}{n})}$$

$$= \frac{n(n-r-1)Z^2_{G,i}}{n-r-nZ^2_{G,i}}$$

and so

$$(5.3.35)\quad |u_i| = \left(\frac{n(n-r-1)Z_{G,i}^2}{n-r-nZ_{G,i}^2}\right)^{\frac{1}{2}} = g(|Z_{G,i}|)$$

Since g is a strictly increasing function, we have

$$(5.3.36)\quad |Z_{G,i}| = g^{-1}(|u_i|) = \left(\frac{(n-r)u_i^2}{n(n-r-1+u_i^2)}\right)^{\frac{1}{2}}$$

As we have seen u_i has a double noncentral F-distribution and an upper bound for the maximum normed G.I. residual is given by

$$(5.3.37)\quad \left(\frac{(n-r)F''_{\alpha/n}}{n(n-r-1+F''_{\alpha/n})}\right)^{\frac{1}{2}}$$

where $F''_{\alpha/n}$ is the significant point of the double non-central F-distribution with 1 and n-r-1 degrees of freedom and non-centrality parameters λ_1 and λ_2 given by (5.3.33).

Now if X is indeed of rank r and $\lambda_1 = \lambda_2 = 0$, then the upper bound for the critical point is given by

$$(5.3.38)\quad \left\{\frac{(n-r)F_{\alpha/n}}{n(n-r-1+F_{\alpha/n})}\right\}^{\frac{1}{2}}$$

which is exactly the result found by Stenfansky (1971, 1972), under the assumption that the variances of the $\hat{e}_{G,i}$'s are equal and the $r(X) = r$.

It is quite clear from the above result that to analyse G.I. residuals, when $r(X) = p$, but only r of the latent roots of $X'X$ are used in A_r^+, lead to complicated distributional results, for the test statistic for outliers.

Ridge Residuals

Suppose that a Ridge procedure is used to overcome the effect of multicollinearity. Let β be estimated by .

$$(5.3.39)\quad \hat{\beta}_R = (X'X+kI)^{-1}X'Y$$

$$= WX'Y \qquad 0 \leq k \leq 1$$

In general k is not known and is estimated from the data, if so, k will be a random variable. The resulting distributional results however become so complicated that for all practical purposes we will assume k known and fixed.

From (5.2.4) we have

$$(5.3.10)\quad \hat{e}_R = Y - X\hat{\beta}_R$$

$$= \mu_R + Re$$

with $\mu_R = X(I-Z)\beta$ and $R = I-XWX'$

Therefore we have

$$(5.3.41)\quad E(\hat{e}_R) = \mu_R$$

and

$$(5.3.42)\quad Var(\hat{e}_R) = \sigma^2 RR'$$

$$= \sigma^2(I-XWX'-kXWW'X')$$

$$= \sigma^2 Q \ .$$

Notice that μ_R will be zero only if $k = 0$. Further the matrix $R = I-XWX'$ is not idempotent. For fixed R since $e \sim N(0,\sigma^2 I)$, we have that $\hat{e}_R \sim N(\mu_R,\sigma^2 Q)$.

Let us now examine the use of Ridge residuals for detecting outliers.

Suppose that we use

$$(5.3.43)\quad T_R^* = \max\left|\frac{\hat{e}_{R,i}}{\bar{S}_R}\right|$$

with $\bar{S}_R$ the estimated average variance. More analytically we have

$$(5.3.44)\quad \overline{\mathrm{Var}(\hat{e}_R)} = \Sigma_i \mathrm{Var}(\hat{e}_{R,i})/n$$

$$= \mathrm{tr}(\sigma^2 Q/n)$$

and hence

$$(5.3.45)\quad \bar{S}_R^2 = \frac{\mathrm{tr}(Q)}{n} S_R^2$$

with $\quad S_R^2 = (Y-X\hat{\beta}_R)'(Y-X\hat{\beta}_R)/(n-p).$

Now

$$(5.3.46)\quad \{\frac{\mathrm{tr}(Q)}{n-p}\}^{\frac{1}{2}} T_R^* = \left(\frac{\mathrm{tr}(Q)}{n-p}\right)^{\frac{1}{2}} \max|\hat{e}_{R,i}/\left[\left(\frac{\mathrm{tr}(Q)}{n}\right)S_R^2\right]^{\frac{1}{2}}|$$

$$= \max|\hat{e}_{R,i}/\{\frac{n-p}{n} S_R^2\}^{\frac{1}{2}}|$$

$$= n^{\frac{1}{2}} \max|\hat{e}_{R,i}/\hat{e}_R'\hat{e}_R|$$

$$= n^{\frac{1}{2}} \max|z_{R,i}|$$

where $z_{R,i}$ is the normed ridge residual.

Let us form the following statistic u_i^* given by (5.3.47) analogous to that given by (5.3.31).

$$(5.3.47) \quad u_i^{*2} = \frac{(n-p-1)\hat{e}_{R,i}^2/q_{ii}}{\hat{e}_R'\hat{e}_R - \hat{e}_{R,i}^2/q_{ii}}$$

with $q_{ii} = \text{Var}(\hat{e}_{R,i})/\sigma^2$

We have that $\hat{e}_{R,i} \sim N(\mu_{R,i}\,\sigma^2 q_{ii})$, consequently u_i^{*2} will have an F-distribution with 1 and n-p-1 degrees of freedom iff $\mu_{R,i} = 0$, i.e. iff $k = 0$, but this is the least squares case. If $k \neq 0$, but fixed, the numerator will have a non-central χ^2 distribution, but it is not at all clear what the distribution of the denominator will be. Therefore the complete lack of distributional results makes it very difficult to use Ridge Residuals to determine outliers.

5.4 GENERALIZED INVERSE OR RIDGE RESIDUALS WITH LEAST SQUARES ESTIMATE FOR σ^2

Prescott suggested the use of

$$(5.4.1) \quad T^* = \max\left|\frac{\hat{e}_i}{\bar{S}}\right|$$

as a test for outliers where $\bar{S}^2$ is an estimate of the average variance assuming that the variances are all equal. If G.I. of R.R. residuals are used then the statistics proposed in the previous sections were

$$(5.4.2) \quad T_G^* = \max\left|\frac{\hat{e}_{G,i}}{\bar{S}_G}\right|$$

and

(5.4.3) $$T_R^* = \max\left|\frac{\hat{e}_{R,i}}{\tilde{S}_R}\right|$$

If in (5.3.28) we estimate σ^2 by $S^2 = (Y-X\hat{\beta})'(Y-X\hat{\beta})/(n-p)$ then we will have

(5.4.4) $$\tilde{S}_G^2 = \frac{n-r}{n} S^2 = \frac{n-r}{n}\,\frac{\hat{e}'\hat{e}}{n-p}$$

and so (5.4.2) becomes

(5.4.5) $$T_G^* = \max|\hat{e}_{G,i}/(\frac{n-r}{n(n-p)}\,\hat{e}'\hat{e})^{\frac{1}{2}}|$$

In view of the above results (5.3.31) becomes

(5.4.6) $$u_i^2 = (n-r-1)\,\frac{\hat{e}_{G,i}^2/n_{ii}}{\Sigma\hat{e}_i^2-\hat{e}_{G,i}^2/n_{ii}}$$

with $$n_{ii} = \frac{\mathrm{Var}(\hat{e}_{G,i})}{\sigma^2}$$

We must find the density of u_i^2. Now the numerator $\hat{e}_{G,i}^2/n_{ii}$ has a noncentral χ^2-distribution with 1 degree of freedom and noncentrality parameter $\lambda_1 = \mu_i^2/n_{ii}$.

To determine the distribution of the denominator let $A_q^2 = \hat{e}'\hat{e} - \hat{e}_q N_q^{-1}\hat{e}_q$, with $\hat{e}_q = [I_q,0]\hat{e}_G = [I_q,0](\mu+Ne)$. We have

(5.4.7) $$\hat{e}_q' N_q^{-1}\hat{e}_q = \{[I_q,0](\mu+Ne)\}'N_q^{-1}\{[I_q,0](\mu+Ne)\}$$

$$= e'NN^*Ne+2\mu_q'N_q^{-1}[I_q,0]Ne+\mu_q'N_q^{-1}\mu_q$$

and since

$\hat{e}'\hat{e} = e'Me$, with $M = I - X(X'X)^{-1}X'$, we have

(5.4.8) $$A_q^2 = e'(M-NN\ N)e - 2\mu_q' N_q^{-1}[I_q,0]\,Ne - \mu_q' N_q^{-1}\mu_q \;.$$

But $M - NN^*N$ is not idempotent so that the results of Theorem 5.3.1 are no longer applicable. The same difficulty arises when using as an estimate for σ^2 in (5.3.44) the O.L.S. one, instead of the $S_R^2 = (Y-X\hat{\beta}_R)'(Y-X\hat{\beta}_R)/(n-p)$.

Thus it is clear that there is no advantage by using a least squares estimate for σ^2 when analysing C.I. or R.R. residuals.

5.5 MULTICOLLINEARITY, HYPOTHESIS TESTING AND VARIABLE SELECTION PROCEDURES

In this section we shall try to cast more light on the effects of multicollinearity on hypothesis testing and on techniques for selecting variables which are commonly employed by the practitioner.

<u>Lemma 5.5.1</u> In the presence of multicollinearity $X\beta$, σ^2, and R^2 can be estimated relatively precisely although the same is not true for β.

<u>Proof</u> The estimate of σ^2, i.e. S^2 is

$$(5.5.1) \qquad S^2 = \frac{e'Me}{n-p}$$

with $M = I - X(X'X)^{-1}X'$. Since M is idempotent and $e \sim N(0,\sigma^2 I)$, it follows that $eMe' \sim \sigma^2\chi^2_{r_1}$, where $r(M) = r_1$. For the variance of S^2 we have

$$(5.5.2) \qquad Var(S^2) = \left(\frac{\sigma^2}{n-p}\right)^2 Var(\chi^2_{r_1})^2$$

$$= \left(\frac{\sigma^2}{n-p}\right)^2 2r_1 < \infty$$

From (5.5.2) it is obvious that the variance of S^2 does not depend on the latent roots of $X'X$, so even if $r(X'X)$ is less than p, the variance of S^2 will be bounded. Note that here $r_1 = n-p$.

The covariance matrix of $X\hat{\beta}$ is $\sigma^2 X(X'X)^{-1}X'$, but M is idempotent consequently $\sigma^2\gamma'M\gamma \geq 0$ for all $\gamma \neq 0$ and finally

$$(5.5.3) \qquad \sigma^2 I\gamma'\gamma \geq \sigma^2\gamma'X(X'X)^{-1}X'\gamma \quad \text{for all} \quad \gamma \neq 0.$$

Since $0 \leq \overline{R}^2 = \frac{\hat{\beta}'(X'X)\hat{\beta}}{\gamma'\gamma} \leq 1$ we have, $E(\overline{R}^2)^h \leq 1$ and so $Var(\overline{R}^2) < 1$, i.e. R^2 can be estimated well.

We show now that β can not be estimated precisely, i.e. the last $p-r$ component of β. We have

$$(5.5.4)' \quad Var(\hat\beta) = \sigma^2(X'X)^{-1}$$

$$= \sigma^2\sum_{j=1}^{r} \frac{1}{\lambda_j} V_jV_j' + \sigma^2\sum_{j=r-1}^{p} \frac{1}{\lambda_j} V_jV_j'$$

Since λ_j^{-1} is very large for $j = r+1,...,p$ this means large diagonal and off diagonal elements in last p-r rows and columns of $(X'X)^{-1}$, i.e. large variances and covariances for the last p-r components of $\hat\beta$.

Since the variance of $\hat\beta$ is affected by multicollinearity, it is obvious that any statistic, which is a function of the $Var(\hat\beta)$ is also affected. In what follows we examine some test statistics.

(i) When we test the hypothesis

$$(5.5.5) \quad H_0: \beta_j = 0 \text{ against } H_1: \beta_j \neq 0$$

we use a t-statistic given by

$$(5.5.6) \quad t_j = (1-\bar{R}_j^2)^{\frac{1}{2}} \frac{\hat\beta_j}{S} \qquad j = 1,2,...,p$$

with $\bar{R}_j$ the multiple correlation coefficient when the X_j variable is regressed on the other predictors. The last p-r variables involved in multicollinearities tend to have small t-statistics, since $(1-\bar{R}_j^2)^{\frac{1}{2}}$ is small, irrespective of what the population values of β_j might be for $j = r+1,...,p$. The noncentrality parameter of the t-statistic is also a function of $(1-\bar{R}_j^2)^{\frac{1}{2}}\beta_j$, so not only the numerical value of the test statistic but also its associated power is reduced for fixed β_j by strong multicollinearities.

(ii) Suppose that we are interested in testing the hypothesis

(5.5.7) $H_0: \beta = 0$ against $H_1: \beta \neq 0$

A 100(1-α)% confidence region for the parameter β, is given by

(5.5.8) $$\frac{(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})}{pS^2} \leq F_{p,n-p}^{(1-\alpha)}$$

where $F_{p,n-p}^{(1-\alpha)}$ is values of Fisher's F-distribution with p,n-p degrees of freedom. The region constitutes the surface and interior of a p-dimensional hyperellipsoid centered at $\hat{\beta}$, The volume of this hyperellipsoid is

(5.5.9) $$V_A = \frac{\pi^{p/2}}{\Gamma(\frac{p}{2}+1)} P_p^{p/2} |X'X|^{-\frac{1}{2}}$$

with $P_p = pS^2 F_{p,n-p}^{(1-\alpha)}$.

The semi-axes of the hyperellipsoid are

(5.5.10) $c_i = \lambda_i^{-\frac{1}{2}} P_p^{\frac{1}{2}}$ $i = 1,2,\ldots,p$.

From (5.5.9) and (5.5.10) it is clear that the confidence region centered at $\hat{\beta}$ is very wide because of the multicollinearity, consequently we tend to accept the null hypothesis irrespective of what the true value of β might be.

(iii) From Theil (1971) p.174, we have

(5.5.11) $$r_{y.i|p-1}^2 = \frac{t_i^2}{t_i^2+n-p} \quad \text{for } i = 1,2,\ldots,p$$

with $r_{y.i|p-1}$ being the partial correlation between Y and X_i variable keeping the remaining p-1 variables fixed. Note also that t_i of (5.5.11) is given by (5.5.6). Thus if t_i^2 is forced to zero due to multicollinearity then $r^2_{y.i|p-1}$ will be forced to zero.

From the above discussion it is clear that the usual stepwise procedures like the backward and forward procedures do not perform well when multicollinearity is present between the X's.

Several other variable selection procedures have, however, also been proposed in the literature. For excellent discussions the reader is referred to the papers by Hocking (1976) and Thompson (1978). Two of the most commonly used procedures are the S_p-criterion (originally proposed by Stein (1960)) and the C_p-criterion of Mallows (1966).

The S_p-criterion is used when all variables $(Y, X_1, \ldots, X_p)$ and considered to be random and to be distributed normally, while C_p is used when the variables $X_1, \ldots, X_p$ are known or fixed.

The S_p-criterion is given by

(5.5.12) $$S_{p'} = \frac{1}{(n-p')(n-p'-2)} SSE_{p'}, \quad p' \leq p$$

where $SSE_{p'}$ is the residual sum of square for the p'-variable model.

From Lemma (5.5.1) it is clear the statistic given by (5.5.12) is unaffected by multicollinearity. The same is true for the C_p-criterion as it can be seen from relation (4.1.6) of Chapter 4.

5.6 CONCLUSIONS

Marquardt's suggestion that the problem of erractic data points and multicollinearity should be tackled at the same time, appears to be difficult from a statistical point of view, since the distributional results are very complicated. In order to overcome the above mentioned difficulties, some heuristic methods are under investigation in the Department of Mathematical Statistics, University of Cape Town, for example, plots of the ridge residuals or the Studentized ridge residuals versus k. The author hopes that these results will soon be able to be published.

It is however important to stress that a form of Ridge trace should be used to detect possible outliers before proceeding with any further analysis. It may well be that due to the multicollinearity problem, the ordinary least squares estimate are so unreliable that outliers may escape detection or faultly identified simply because the estimate $\hat{e}$ is far removed from its true value e.

The author thinks it is worthwhile mentioning that further research is needed to give some guidelines to the practitioner especially when he is confronted with the detection of outliers in the linear model while multicollinearity and autocorrelation are present.

BIBLIOGRAPHY

AITKIN, M. (1974). Simultaneous inference and choice of variable subsets in multiple regression. Technometrics, 16, 221-227.

ANDERSON, T.W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley.

BRADLEY AND SCHUMAN, D. (1957). The Comparison of the sensitivities of similar experiments. Applications. Biometrica, 13, 496-510.

CHIPMAN, J.A. (1964). On least squares with insufficient observations. Journal of the American Statistical Association, 59, 1078-111.

CONSTANTINE, A.G. (1963). Some non-central distribution problems in multivariate analysis. Ann. Math. Statist., 34, 1270-1285.

CONSTANTINE, A.G. (1966). The distribution of Hotelling's generalized T_0^2. Ann. Math. Statist. 37, 215-226.

CORMAN, T. and TOMAN, R. (1966). Selection of variables for fitting equation to data. Technometrics, 8, 27-61.

DANIEL, C. and WOOD, F. (1971). Fitting equation to data. New York, Wiley Interscience.

DAVID, H.A. (1956). On the application to statistics of an elementary theorem in probability. Biometrica, 43, 85-90.

de WAAL, D.J. (1969). On the noncentral distribution of the largest canonical correlation coefficient. S. Afr. Statist. Jour., 3, 91-94.

EFROYMSON, N. (1966). Stepwise regression - a backward and forward look. Presented at the Eastern Regional meeting of the Inst. of Math. Statist., Florham Park, New Jersey.

ELLENBERG, J.H. (1973). The joint distribution of the standardized least squares residuals from a general linear regression. Journal of the American Statistical Association, 68, 941-943.

ERDÉLYI, A. (1953). Higher trancendental functions, Vol. 1, New York.

ERDÉLY, A. (1954). Tables of integral transforms, Vol. 1, New York.

FEIVESON, A. (1973). Selecting variables in regression and classification of thresholdings. Ph.D. dissertation, Texas University College Station, Texas.

GOLDBERGER, A. (1964). Econometric Theory. John Wiley and Sons, 255-260.

GREENACRE, M. (1973). Symmetrised multivariate distributions. S. Afr. Statist. Jour., 7, 95-101.

GRUBBS, F. (1969). Procedures for detecting outlying observations in samples. Technometrics, 17, 309-314.

GUNST, R.F. and MASON, R.L. (1976). Generalized mean square error properties of regression estimators. Communication in Statistics, A5(15), 1501-1508.

GUNST, R.F. and MASON, R.L. (1977). Biased estimation in Regression : An evaluation using Mean Squared Error. Journal of the American Statistical Association, 72, 616-628.

GUPTA, A.K. (1971). Noncentral distribution of Wilks' statistic in MANOVA. Ann. Math. Statist., 42, 1254-1261.

HALD, A. (1952). Statistical theory and engineering applications. Wiley.

HERZ, C.S. (1955). Bessel functions of matrix argument. Ann. Math., 61, 474-523.

HOCKING, R. (1976). The analysis and selection of variables in linear regression. Biometrics, 32, 1-49.

HOERL, A.E. and KENNARD, R.W. (1970). Ridge regression : Biased estimation for nonorthogonal problem. Technometrics, 12, 55-67.

HOERL, A.G. and KENNARD, R.W. (1970). Ridge regression : Application to nonorthogonal problems. Technometrics, 12, 61-82.

HOERL, A.G. and KENNARD, R.W. (1976). Ridge regression. Iterative estimation of biasing parameter. Communication in Statistics, 75(1), 77-88'

JAMES, A.T. (1960). The distribution of the latent roots of the covariance matrix. Ann. Math. Statist., 31, 151-158.

JAMES, A.T. (1961). Zonal polynomials of the real positive definite symmetric matrices. Ann. Math., 74, 456-469.

JAMES, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples. Ann. Math. Statist., 35, 475-501.

JOHNSON, N. and KOTZ, S. (1970). Continuous univariate distribution. Vol. 2. Wiley.

KENDAL, M. and STUART, A. (1961). The advanced theory of Statistics, Vol. 2.

KHATRI, C.G. (1964). Distribution of the Generalised Multiple Correlation matrix in the Dual Case. Ann. Math. Statist., 35, 1801-1806.

KHATRI, C.G. and PILLAI, K.C.S. (1968). On the noncentral distributions of two test criteria in multivariate analysis of variance. Ann. Math. Statist., 39, 215-226.

LEAMER, E.E. (1975). A Result on the sign of Restricted Least Squares Estimates. Journal of Econometrics, 3, 387-390.

LUND, R.E. (1975). Tables for an approximate test for outliers in linear models. Technometrics, 17, 473-476.

MALLOWS, C. (1964). Choosing variables in a linear regression - a graphical aid. Presented at the central regional meeting of the Inst. of Math. Statist., Manhattan, Kansas.

MALLOWS, C. (1966). Choosing a subset regression. Presented at a joint statistical meeting, Los Angeles, California.

MALLOWS, C. (1973). Some comments on C_p. Technometrics, 15, 661-675.

MARQUARDT, D.W. (1970). Generalized inverse, ridge regression, biased linear estimation and nonlinear estimation. Technometrics, 12, 591-612.

MASON, R.L., GUNST, R.F. and WEBSTER, J.T. (1975). Regression analysis and problems of multicollinearity. Communications in Statistics. A4(13), 277-292.

MONEY, A.H. (1972). Noncentral multivariate beta distributions and their applications. Unpublished Ph.D. thesis, University of Cape Town.

MORRISON, D.F. (1976). Multivariate Statistical Methods. McGraw-Hill.

OBENCHAIN, D.L. (1977). Classical F-tests and Confidence Regions for Ridge Regression. Technometrics, 4, 429-439.

PILLAI,K.C.S.(1955). Some new test criteria in multivariate analysis. Ann. Math. Statist., 26, 117-121.

PRESCOTT, P. (1975). An approximate test for outliers in linear models. Technometrics, 17, 129-132.

RADHAKNISHNAN, S. (1974). Selection of variables in multiple regression. Ph.D. dissertation, University of Houston, Texas.

ROY, S.N. (1957). Some Aspects of Multivariate Analysis. Wiley.

SCHUMANN,D. and BRADLEY, R. (1957). The comparison of the sensitivities of similar experiments. Theory. Ann. Math. Statist., 28, 902-920.

SPJØTVOLLE, E. (1977). Alternatives to plotting C_p in multiple regression. Biometrika, 64, 1-8.

STEFANSKY, W. (1971). Rejecting outliers by maximum normed residual. Ann. Math. Statist., 42, 35-45.

STEFANSKY, W. (1972). Rejecting outliers in factorial designs. Technometrics, 14, 469-479.

STEIN, C. (1960). Multiple regression. In Contributions to Probability and Statistics. Ed. I. Olkin et al. Stanford University Press.

STEIN, C.M. (1962). Confidence sets for the mean of multivariate normal distribution. J. Roy. Statist. Soc., B, 24, 265-296.

SUBROHMANIAN,K(1976). Recent trends in multivariate normal distribution theory on the Zonal polynomials and other functions of matrix argument. Sankhyā, 38, 221-258.

SUGIYAMA, T. (1967). Distribution of the largest latent root and the smallest latent root of the generalized B statistic and F statistic in multivariate analysis. Ann. Math. Statist. 38, 1152-1159.

THEIL, H. (1971). Principles of Econometrics. Wiley.

THEOBALD, C.M. (1974). Generalization of mean square error applied to ridge regression. J. Roy. Statist. Soc., B, 36, 103-106.

THOMPSON, M. (1978). Selection of variables in multiple regression. Part I. A review and evaluation. International Statistical Review, 46, 1-19.

THOMPSON, M. (1978). Selection of variables in multiple regression. Part II.. Chosen procedures, computation and examples. International Statistical Review, 46, 29-146.

TIENTJEN, G.L. MOORE, R. and BECKMAN, R. (1973). Testing for a single outlier in simple linear regression. Technometrics, 15, 717-721.

TROSKIE, C.G. (1969). The generalised multiple correlation matrix.. S. Afr. Statist. Jour., 3, 109-122.

TROSKIE, C.G. (1971). The distributions of some test criteria in multivariate analysis. Ann. Math. Statist., 42, 1752-1757.

VINOD, H. (1978). Survey of ridge regression and related techniques for improvement over ordinary least squares. Review of Economics and Statistics, 1, 121-131.

WILKS, S.S. (1932). Certain generalizations in the analysis of variance. Biometrika, 24, 471-494.

FISHER, R.A. (1928). The general sampling distribution of the multiple correlation coefficient. Proc. Roy. Soc., A, 121, 654-673.