Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 64(2) 2010 219-228

Multivariate Directed Inference with Modified Hotelling's T-squared

John H. Carson, Jr.[1*] **and Arjun K. Gupta**[2]
[1]*The Shaw Group, Findlay OH, USA*
[2]*Bowling Green State University, Bowling Green OH, USA*

Received 10 February 2010; Revised 19 April 2010; Accepted 20 April 2010

SUMMARY

This paper presents an approach to the problem of multivariate directed inference, a generalization of one-side testing, in the setting of vector and matrix valued elliptically contoured (MEC) random variables. The T_+^2 statistic, a modification of Hotelling's T^2, is introduced. It gives a sensitive test of positivity in one or more components of a location vector, which is nonparametric over the MEC family. The T_+^2 statistic uses the positive part of the sample mean vector or of the difference between a sample mean vector and a reference vector. Other hypotheses, including order restrictions, may be tested by suitably transforming the data. The test is derived from the Generalized Likelihood Ratio Test and by the union-intersection principle. Principal properties and the null and power distributions are given.

Keywords: Multivariate analysis, Hotelling's T^2, One-sided testing, Elliptically contoured distributions, Union-intersection test, Generalized likelihood ratio test.

1. INTRODUCTION

The generalization of univariate statistical techniques to multivariate settings sometimes proceeds by analogy in a straightforward way. However, this has not been the case for the extension of univariate one-sided tests to the multivariate setting. The difficulty may be appreciated when one considers how many different meanings the term "one-sided" could have in p-dimensional space. This setting provides many opportunities for subtle differences in hypotheses to distinctly change the testing problem.

The authors recommend the term *directed inference* for the various extensions that can be made. As has been pointed out by Robertson, Wright and Dykstra (1988), these types of tests can be covered by the mantle of *order restricted inference*. The principal focus of this paper is the detection of positivity in one or more location components or a positive shift in location of one population with respect to another (or to a fixed standard) in at least one location component. This is important in many applications, including comparison of treatments to a control, combination of test statistics and detecting elevated levels of contaminants in the environment, particularly when they also occur naturally. More generally, these are orthant and order restricted hypotheses. In these cases, the null and alternative hypotheses can be described as sets of directions in the parameter space relative to some reference point. This underscores the appropriateness of the term *directed inference.*

Much previous work on directed multivariate inference, by Kudo (1963), Perlman (1969), Tang *et al.* (1989) and Fraser *et al.* (1991) has focused on applications to clinical trials and combinations of test statistics. They have explored hypotheses which are somewhat different than those which appear in this paper.

Another major difficulty with the study of directed inference is that it involves truncated random vectors, and the distributions of truncated random vectors are

*Corresponding author : John. H Carson, Jr.
E-mail address : john.carson@shawgrp.com

extremely complicated. The authors have found a reasonably nice solution to the directed inference problem which is nonparametric over a large class of distributions and asymptotically nonparametric in a wide variety of situations.

2. BACKGROUND

2.1 Matrix Variate Elliptically Contoured Distributions

Let $\mathbf{X} = (\mathbf{X}_1,\ldots, \mathbf{X}_n)$ be a $p \times n$ random matrix with characteristic function of the form

$$\psi_{\mathbf{X}}(\mathbf{T}) = \text{etr}(i\mathbf{T}'\mathbf{M})\ \Psi\ (\text{tr}(\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T})) \tag{1}$$

where $\mathbf{T}$ is a $p \times n$ matrix. Then $\mathbf{X}$ has a matrix variate elliptically contoured distribution with parameters $\mathbf{M}$, $\boldsymbol{\Sigma} \otimes \mathbf{I}_n$ and Ψ, written as

$$\mathbf{X} \sim E_{p,n}\ (\mathbf{M}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n, \Psi)$$

Consequently,

$$E(\mathbf{X}) = \mathbf{M}$$

and

$$\text{Var}(\mathbf{X}') = -2\Psi'(0)\ \boldsymbol{\Sigma} \otimes \mathbf{I}_n$$

where $\boldsymbol{\Sigma}$ is a $p \times p$ covariance matrix, $\otimes$ is the Kronecker product and $\mathbf{I}_n$ is the $n \times n$ identity matrix.

This defines a very large class of distributions that includes the matrix variate normal distribution and the matrix variate t-distributions-see Fang and Zhang (1990), Gupta and Varga (1993). We denote this class by $\mathcal{F}$. The family $\mathcal{F}$ is a generalization of the class of normally distributed random matrices and inherits many of its nice properties, particularly with respect to invariance properties and LR based tests.

The distribution of Hotelling's T^2 statistic and of other scale-equivariant statistics has been shown to be invariant over the subset of $\mathcal{F}$ with mean $\mathbf{0}$, by Fang and Zhang (1990) and Gupta and Varga (1993).

2.2 Decomposition of Hotelling's T^2

Let

$$\bar{\mathbf{X}} = n^{-1} \sum_{i=1}^{n} \mathbf{X}_i = n^{-1}\mathbf{X}\mathbf{1}_n \tag{2}$$

and

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})' = \mathbf{XDD'X'} = \mathbf{XDX'} \tag{3}$$

where $\mathbf{D} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n$, and $\mathbf{1}_n$ is the n-dimensional vector of ones. Note that $\mathbf{D}$ is symmetric and idempotent.

Let

$$T^2 = n\bar{\mathbf{X}}'\mathbf{S}^{-1}\bar{\mathbf{X}} = n^{-1}\mathbf{1}'_n\mathbf{X}'(\mathbf{XDX}')^{-1}\mathbf{X}\mathbf{1}_n$$

Then T^2 has the $\mathrm{T}^2_{p,n-1}$ distribution with noncentrality parameter $\delta^2 = \boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$. Write

$\bar{\mathbf{X}} = \bar{\mathbf{X}}_+ - \bar{\mathbf{X}}_-$, where $\left(\bar{\mathbf{X}}_+\right)_i = \bar{X}_i \vee 0$,

and $\left(\bar{\mathbf{X}}_-\right)_i = \bar{X}_i \wedge 0$ for $i = 1,\ldots, p$

Then we have

$$\begin{aligned}T^2 &= n\bar{\mathbf{X}}'\mathbf{S}^{-1}\bar{\mathbf{X}}\\ &= n(\bar{\mathbf{X}}_+ - \bar{\mathbf{X}}_-)'\mathbf{S}^{-1}(\bar{\mathbf{X}}_+ - \bar{\mathbf{X}}_-)\\ &= n\bar{\mathbf{X}}'_+\mathbf{S}^{-1}\bar{\mathbf{X}}_+ + n\bar{\mathbf{X}}'_-\mathbf{S}^{-1}\bar{\mathbf{X}}_- - 2n\bar{\mathbf{X}}'_+\mathbf{S}^{-1}\bar{\mathbf{X}}_-\\ &= T_+^2 + T_-^2 - 2\left\langle\bar{\mathbf{X}}_+, \bar{\mathbf{X}}_-\right\rangle_{[\mathbf{S}/n]^{-1}}\end{aligned}$$

where $\langle a, b\rangle_{[\mathbf{C}]}$ is the inner product with respect to the matrix $\mathbf{C}$. This has the form of the law of cosines in the random manifold generated by empirical Mahalanobis distance. It is also a decomposition of the total sum of squares.

3. DIRECTED INFERENCE USING T_+^2

3.1 Hypotheses

Let $\Theta = \mathbb{R}^p$, $\Theta_0 = \overline{\mathbb{R}^p_-}$ and $\Theta_1 = \Theta \backslash \Theta_0 = \mathbb{R}^p \backslash \overline{\mathbb{R}^p_-}$,

where $\bar{A}$ denotes the closure of the set A. Let $E(\bar{\mathbf{X}}) = \boldsymbol{\mu}$. Then H_0 has $\mu_i \le 0\ \forall\ i$ and H_1 has $\mu_i > 0$ for at least one i. H_0 is an orthant hypothesis, since it has $\boldsymbol{\mu}$ in the closure of the negative orthant. Orthants are the generalization of the familiar quadrants of $\mathbb{R}^p$ to p-dimensional space. Note that H_0 and H_1 partition $\mathbb{R}^p$. We want to detect when at least one component of the mean vector is positive, and we do not care whether any of them are negative. Other authors have used directed hypothesis that do not partition $\mathbb{R}^p$, such as H'_0: $\boldsymbol{\mu} = \mathbf{0}$ versus H'_1 : $\boldsymbol{\mu} \ge \mathbf{0}$, $\boldsymbol{\mu} \ne \mathbf{0}$.

3.2 Technique

The basic idea behind the T_+^2 test is to start with a very good test, Hotelling's T^2 (see Anderson (1984)) and to modify it so that negative components of $\bar{\mathbf{X}}$ do not inflate the value of the test statistic; that is, so that it is invariant under coordinate projection of $\bar{\mathbf{X}}$ onto the positive orthant. The acceptance region of the α-level test is defined by the inequality

$$T_+^2 = n\,\bar{\mathbf{X}}'_+ \mathbf{S}^{-1}\,\bar{\mathbf{X}}_+ \le \mathrm{C}(n, p, \mathbf{R}, \alpha) \qquad (4)$$

where $\mathbf{R}$ is the population correlation matrix.

3.3 Derivation as a Generalized Likelihood Ratio Test

The derivation of the T_+^2 test as a Generalized Likelihood Ratio (GLR) Test is similar to the derivation of Hotelling's T^2 as a GLR test.

Theorem 1. The T_+^2 test is equivalent to the GLR test of H_0 versus H_1.

Proof. See Appendix A.

3.4 Some Results on Truncated Spherical and Elliptic Random Vectors

Let $\mathbb{N}$ denote the set of positive integers.

Lemma 1. Let $\mathcal{S}_k$ denote the surface of the k-dimensional unit sphere, for some $k \in \mathbb{N}$. Let $\mathcal{B}(\mathcal{S}_k)$ denote the Borel σ-algebra of subsets of $\mathcal{S}_k$. Let $\mathbf{B} \in \mathcal{B}(\mathcal{S}_k)$, and let $\mathbf{U}_\mathbf{B}^{(k)}$ be the random vector which has the uniform distribution on $\mathbf{B}$. Then $\mathbf{U}_\mathbf{B}^{(k)\prime}\mathbf{U}_\mathbf{B}^{(k)} = 1$, $\forall$ $\mathbf{B} \in \mathcal{B}(\mathcal{S}_k)$ and $\forall$ $k \in \mathbb{N}$.

Proof. The proof is omitted because it is elementary.

Lemma 2. Using the notation of the previous lemma, let $\mathbf{Y}_\mathbf{B} \overset{d}{=} R\mathbf{U}_\mathbf{B}^{(k)}$, where R has a distribution on $(0, \infty)$, and $\mathbf{B} \in \mathcal{B}(\mathcal{S}_k)$. Let $\mathbf{U}^{(k)}$ have the uniform distribution on $\mathcal{S}_k$, and let $R \overset{d}{=} R'$. Then $\mathbf{Y}'_\mathbf{B}\,\mathbf{Y}_\mathbf{B} \overset{d}{=} \mathbf{X}'\mathbf{X}$.

Proof. $\mathbf{Y}'\mathbf{Y} \overset{d}{=} (R\mathbf{U}_\mathbf{B}^{(k)})'(R\mathbf{U}_\mathbf{B}^{(k)}) = R^2\mathbf{U}_\mathbf{B}^{(k)\prime}\mathbf{U}_\mathbf{B}^{(k)} = R^2$, by the previous lemma. Similarly,

$$\mathbf{X}'\mathbf{X} \overset{d}{=} (R'\mathbf{U}^{(k)})'(R'\mathbf{U}^{(k)}) = (R')^2\mathbf{U}^{(k)\prime}\mathbf{U}^{(k)} = (R')^2$$

Then $R \overset{d}{=} R' \Rightarrow R^2 \overset{d}{=} (R')^2$. Note that this holds for any k.

Lemma 3. Using the notation of the previous lemmas, $\mathbf{Y} \overset{d}{=} R\mathbf{U}^{(k)}$, $\mathbf{Y}_\mathbf{B} \overset{d}{=} R^*\mathbf{U}_\mathbf{B}^{(k)}$ and $\mathbf{X}_\mathrm{i} \overset{d}{=} R_i\mathbf{U}_i^{(k)}$, $i = 1,\ldots,n$ where $B \in \mathcal{B}(\mathcal{S}_k)$. Let $R \overset{d}{=} R^* \overset{d}{=} R_i$, $i = 1,\ldots,n$ where R has a distribution on $(0, \infty)$. Let $\mathbf{A} = \mathbf{D}'\mathbf{D}$ be a $k \times k$ symmetric matrix. Then $\mathbf{Y}'\mathbf{A}\mathbf{Y} \overset{d}{=} \mathbf{Y}'_\mathbf{B}\mathbf{A}\mathbf{Y}_\mathbf{B}$.

Proof. By the previous lemma, $\mathbf{Y}'_\mathbf{B}\mathbf{Y}_\mathbf{B} \overset{d}{=} \mathbf{X}'\mathbf{X}$

Then

$$(\mathbf{D}\mathbf{Y}_\mathbf{B})'(\mathbf{D}\mathbf{Y}_\mathbf{B}) \overset{d}{=} (\mathbf{D}\mathbf{Y})'(\mathbf{D}\mathbf{Y}) \Rightarrow \mathbf{Y}'_\mathbf{B}\mathbf{A}\mathbf{Y}_\mathbf{B} \overset{d}{=} \mathbf{YAY}$$

as required.

3.5 Stochastic Representation

Let $\mathcal{O}_K$ denote the union of all of the $\binom{p}{K}$ orthants with K positive coordinates. The following theorem presents a stochastic representation of the T_+^2 statistic as a discrete mixture of Hotelling's T^2 statistics.

Theorem 2. The T_+^2 statistic has a stochastic representation as a T^2 statistic on a K-dimensional space, where K is a discrete random variable with distribution

$$\Pr(K = k) = \Pr(\mathcal{O}_k)$$

Proof. We have

$$\begin{aligned} T_+^2 &= n\bar{\mathbf{X}}'_+\mathbf{S}^{-1}\bar{\mathbf{X}}_+ \qquad (5)\\ &= n[\mathbf{B}\bar{\mathbf{X}}]'\mathbf{S}^{-1}[\mathbf{B}\bar{\mathbf{X}}] = n\bar{\mathbf{X}}'_\mathbf{B}\mathbf{S}_\mathbf{B}^{-1}\bar{\mathbf{X}}_\mathbf{B}\end{aligned}$$

where

$$\mathbf{B} = \mathrm{diag}(b_1, \ldots, b_p), \text{ and}$$

$$b_i = \mathbf{I}\{\bar{\mathbf{X}}_i > 0\},\ i = 1, \ldots, p \qquad (6)$$

Here $\mathbf{I}(A)$ is the indicator of the set A, $\bar{\mathbf{X}}_\mathbf{B}$ is the K-dimensional subvector consisting of the positive components of the vector $\bar{\mathbf{X}}$, and $\mathbf{S}_\mathbf{B}^{-1}$ is the $K \times K$ submatrix formed by deleting the rows and columns of $\mathbf{S}^{-1}$ corresponding to the nonpositive elements of $\bar{\mathbf{X}}$.

There is a 1-1 correspondence between each observed value of the matrix $\mathbf{B}$ and the orthant into which the associated observation of $\bar{\mathbf{X}}$ falls. Let this orthant be denoted $\mathcal{O}^\mathbf{B}$. Clearly, $\mathbf{B}$ is a random matrix, $K = \mathrm{tr}(\mathbf{B})$, and, therefore, K is a random variable.

Note that $\mathbf{B}$ is symmetric and idempotent. It is the matrix of the orthogonal projection of $\bar{\mathbf{X}}$ onto the

K-dimensional subspace spanned by the positive coordinates of $\bar{\mathbf{X}}$. The probability that K of the coordinates of $\bar{\mathbf{X}}$ are positive equals the probability measure of the union of the $\binom{p}{K}$ orthants which have K positive axes. If we observe $\mathbf{X}$ and calculate $\bar{\mathbf{X}}$, then $\mathbf{B}$ is fixed, and $K = \text{tr}(\mathbf{B})$ is fixed, at say $K = k$. Then T_+^2 calculated from this sample is calculated as a T^2 statistic on a k-dimensional space.

Lemma 4. Let $\mathbf{X}$ have a distribution in the family $\mathcal{F}$, as defined by equation (1) with $\mathbf{M} = \mathbf{0}$. Then for fixed $\mathbf{\Sigma}$ the distributions of the random matrix $\mathbf{B}$ and the random variable $K = \text{tr}(\mathbf{B})$, as defined in the proof of Theorem 2, are *invariant* over the family $\mathcal{F}$. This distribution is a function only of the common correlation matrix, $\mathbf{\Phi}$, of the rows of $\mathbf{X}$.

Proof. In this case, the distribution of K, $\Pr(K = k)$ $\Pr(\mathcal{O}_k)$, is a function of the central orthant probabilities. This is so because $E(\bar{\mathbf{X}}) = \mathbf{0}$. Since the limits of integration in evaluating these probabilities are either from 0 to ∞ or from ∞ to 0, central orthant probabilities are independent of the coordinate scaling.

Now $\bar{\mathbf{X}}$ has a vector variate elliptically contoured distribution with location $\mathbf{0}$ and dispersion matrix proportional to $\frac{1}{n}\mathbf{\Sigma}$.

To show that the distribution of K is invariant over $\mathcal{F}$, consider the stochastic representation of $\bar{\mathbf{X}}$,

$$\bar{\mathbf{X}} \overset{d}{=} R\left[\frac{1}{n}\mathbf{\Sigma}\right]^{1/2} \mathbf{U}$$

where the rank of $\mathbf{\Sigma}$ is p, R is a continuous positive random variable, and $\mathbf{U}$ has the uniform distribution on the surface of the unit sphere in $\mathbb{R}^p$ (see Fang and Zhang (1990) for a detailed development). Since $\bar{\mathbf{X}}$ has a density, $\Pr(\bar{\mathbf{X}} = \mathbf{0}) = 0$.

Then, since $\Pr[\mathbf{U} \in \mathcal{O}^{\mathbf{B}}] = 2^{-p}$, $(\forall \mathbf{B})$, since $R > 0$ (wp1), and since $\Pr[\bar{\mathbf{X}} \in \mathcal{O}^{\mathbf{B}}]$ is independent of the scaling of the components of $\bar{\mathbf{X}}$, we conclude that for any random matrix $\mathbf{X}$, with distribution in $\mathcal{F}$ and having $\mathbf{M} = \mathbf{0}$, the orthant probabilities for the associated random variable $\bar{\mathbf{X}}$, are a function only of the common correlation matrix, $\mathbf{\Phi}$, of the rows of $\mathbf{X}$. Hence, the distribution of $\mathbf{B}$ is a function of $\mathbf{\Phi}$ alone.

Now, $K = \text{tr}(\mathbf{B})$. Then, the distribution of K is a function only of the central orthant probabilities, which in turn depend only on the correlation matrix $\mathbf{\Phi}$ and are invariant over $\mathcal{F}$.

3.6 Null Distribution of $\mathbf{T}_+^2$

Theorem 3. The null distribution of T_+^2 for the family $\mathcal{F}$ under the point null hypothesis, $\mathbf{H}_0^n : \boldsymbol{\mu} = \mathbf{0}$, is given by

$$\Pr\left(T_+^2 \le t\right) = \Pr(\mathcal{O}_0) + \sum_{k=1}^{p} \Pr(\mathcal{O}_k) \Pr\left(T_{p,n-1}^2 \le t\right) \qquad t \ge 0 \quad (7)$$

where $T_{p,n-1}^2$ represents a random variable with the $T_{p,n-1}^2$ distribution.

Proof. The proof is given in Appendix A.

After estimating the mixing weights by simulation or numerical integration, the null distribution of T_+^2 may be computed by using a mixture of T^2 distributions. Optionally, unconditional simulation may be used. For analysis with real data, the bootstrap may be preferred.

3.7 Union-Intersection Construction

Theorem 4. The T_+^2 test can be constructed by using Roy's union-intersection principle.

Proof. The hypotheses can be written as

$$\text{H}_0 = \bigcap_{i=1}^{p} \text{H}_{0i},\ \text{H}_{0i} : \mu_i \le 0$$

$$\text{H}_1 = \bigcup_{i=1}^{p} \text{H}_{1i},\ \text{H}_{1i} : \mu_i > 0$$

Note that H_1 can be written as

$$\text{H}_1 = \bigcup_{k=1}^{p} \bigcup_{\{i_1,\ldots i_k\}} \text{H}_{1,\{i_1,\ldots i_k\}} \qquad (8)$$

with

$$\text{H}_{1,\{i_1,\ldots i_k\}} : \mu_{i1} > 0, \ldots, \mu_{ik} > 0 \qquad (9)$$

where $\{i_1, \ldots i_k\}$ ranges over all subsets of the integers 1 to p of size k. Likewise, H_0 can be expressed as

$$\text{H}_0 = \bigcap_{k=1}^{p} \bigcap_{\{i_1,\ldots,i_k\}} \{\text{H}_{0,\{i_1,\ldots,i_k\}}, s, i_k\} \qquad (10)$$

with

$$\mathrm{H}_{0,\{i_1,\ldots,i_k\}} : \mu_{i1} \le 0, \ldots, \mu_{ik} \le 0 \tag{11}$$

In order to reject H_0, we must reject, for some value of k, any of the hypotheses $\mathrm{H}_{0,\{i_1,\ldots,i_k\}}$. But this is equivalent to testing the hypothesis corresponding to the positive components of $\bar{\mathbf{X}}$, since inclusion of simultaneous testing of means corresponding to the negative components of $\bar{\mathbf{X}}$ would reduce the probability of rejection.

From equation (4), we see that T_+^2 is the squared Mahalanobis distance from **0**, which corresponds to the point null hypothesis $\boldsymbol{\mu} = \mathbf{0}$, to $\bar{\mathbf{X}}_+$. Clearly, the likelihood of the point null is strictly decreasing as this sample distance increases. Correspondingly, we reject the null hypothesis for large values of the test statistic. This demonstrates the union-intersection construction.

4. PRINCIPAL PROPERTIES OF T_+^2 TEST

4.1 Consistency and Admissibility

Theorem 5. The T_+^2 test is both consistent and admissible.

Proof. Nandi (1965) has shown that unionintersection tests are consistent if the component tests are and that they are admissible if the component tests are. The component tests in the union-intersection construction which we have demonstrated are T^2 tests for $K > 1$ and t-tests for $K = 1$. In the space $\mathbb{R}_+^p$, where the null hypothesis is identified with the origin, the only tests of location which may exist are one-sided. The component tests, from which the T_+^2 test is constructed by union-intersection, are both consistent and admissible. Hence, tests based on the T_+^2 statistic are both consistent and admissible.

4.2 Unbiasedness

Theorem 6. T_+^2 based tests are unbiased.

Proof. The distribution of T_+^2 has an atom at zero because the mapping $\mathbf{X} \mapsto \mathbf{X}_+$ maps the negative orthant onto the origin. The mass assigned to this atom will be of the order of 2^{-p}. By elliptic symmetry, the probability measure of the negative orthant equals that of the positive orthant. Therefore, for $p > 1$ both of them will be strictly less than 0.5. Then the distribution of T_+^2 is continuous at the boundary of any reasonable critical region.

4.3 Acceptance Region

Theorem 7. The acceptance region of T_+^2 based tests is a bounded, closed and convex set in the relative topology induced by the mapping from $\mathbf{X} \mapsto \mathbf{X}_+$.

Proof. For a fixed constant $C > 0$, the set

$$\{\mathbf{u} \in \mathbb{R}_+^p : \mathbf{u}'\mathbf{A}^{-1}\mathbf{u} \le C, \mathbf{A} > 0 \text{ fixed}\}$$

is the intersection of a p-dimensional ellipsoid, say $\mathcal{B}$, centered at **0**, and the positive orthant $\mathbb{R}_+^p$.

Then

$$\begin{aligned} &\mathbf{u} \in \mathbb{R}_+^p : \mathbf{u}'\mathbf{S}^{-1}\mathbf{u} \le C\} \\ &= \{\mathbf{u} \in \mathbb{R}^p : \mathbf{u}'\mathbf{S}^{-1}\mathbf{u} \le C\} \cap \mathbb{R}_+^p \\ &= \mathcal{B} \cap \mathbb{R}_+^p \end{aligned}$$

Note that this set includes cases in which components of $\mathbf{u}_+$ are zero, which correspond to $K < p$. They lie on the boundary of the region. Note also that the origin, which is the only atom, does not lie on the boundary between the acceptance and rejection regions. This boundary is the set $(\partial\mathcal{B}) \cap \mathbb{R}_+^P$, which does not include the origin.

Then

$$\Pr((\partial\mathcal{B}) \cap \mathbb{R}_+^p) = 0$$

The acceptance region $\mathcal{B} \cap \mathbb{R}_+^p$ is the intersection of two closed, convex sets and is, therefore, closed and convex. Since $\mathcal{B}$ is bounded, the acceptance region is bounded also.

4.4 Symmetry of T_+^2 and T_-^2 Statistics

Theorem 8. T_+^2 and T_-^2 are identically distributed under the point null hypothesis $\boldsymbol{\mu} = \mathbf{0}$.

Proof. With $\boldsymbol{\mu} = \mathbf{0}$, $\bar{\mathbf{X}}$ is elliptically symmetric about **0**. Elliptical symmetry about the origin implies symmetry about the origin in each coordinate. Hence, $\bar{\mathbf{X}} \overset{d}{=} -\bar{\mathbf{X}}$. Then we have,

$$T_+^2 = n\bar{\mathbf{X}}_+' \mathbf{S}^{-1}\bar{\mathbf{X}}_+ \overset{d}{=} n(-\bar{\mathbf{X}})_+' \mathbf{S}^{-1}(-\bar{\mathbf{X}})_+$$
$$= n\bar{\mathbf{X}}_-' S^{-1}\bar{\mathbf{X}}_- = T_-^2 \quad (12)$$

4.5 Noncentral Distribution of T_+^2

Let **X** have the distribution given in equation (1). Let $\bar{\mathbf{X}}$ and **S** be calculated from **X** as in equations (2) and (3). Let $E\bar{\mathbf{X}} = \mu$ with at least one component of μ positive.

Theorem 9. The non-null distribution of T_+^2 is a discrete mixture of a random variable degenerate at 0 and $2^p - 1$ random variables, each of which is distributed as $\frac{n-k}{a_n k}$ times a $GF_{k,n-k}(\delta_{\mathbf{B}}^2, f)$ random variable, where $GF_{k,n-k}(\delta_{\mathbf{B}}^2, f)$ is a generalized non-central F distribution with degrees of freedom k and n-k and noncentrality $\delta_{\mathbf{B}}^2$. **B** is the random matrix of equation (6), and f is the function in the definition, equation (1), of the class of elliptically contoured matrix distributions.

Proof. See Appendix.

See Fang and Zhang (1990, section 2.9.3) for definition and properties of the generalized noncentral F-distribution. This distribution is not invariant over $\mathcal{F}$. This creates obvious difficulties for the evaluation of power. The situation is further complicated by the fact that the mixing weights in this case are functions of both mean and covariance as well as the form of the density function, since the symmetry of the central orthant probabilities is destroyed. For these two reasons, the non-null distribution of T_+^2 is best approximated by simulation.

5. CONCLUSION

An important approach to directed inference for multivariate data, the T_+^2 statistic, has been presented. The null distributions, robustness and power of procedures based on the T_+^2 statistic (and other important properties) have been studied. Much remains to be done in the way of studying the properties of these procedures and how they may be applied to more complex problems of directed inference.

REFERENCES

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, New York.

Fang, K.T. and Zhang, Y.T. (1990). *Generalized Multivariate Analysis*. Springer-Verlag, New York.

Fraser, D.A.S., Guttman, I. and Srivastava, M.S. (1991). Conditional inference for treatment and error in multivariate analysis *Biometrika*, **78**, 565-572.

Gupta, A.K. and Varga, T. (1991). Rank of a quadratic form in an elliptically contoured matrix random variable. *Statist. Probab. Lett.*, **12**, 131-134.

Gupta, A.K. and Varga, T. (1993). *Elliptically Contoured Models in Statistics*. Kluwer Academic Publisher, Dordrecht, NE.

Gupta, A.K. and Varga, T. (1994a). A new class of elliptically contoured distributions. *J. Italian Statist. Soc.*, **3**, 255-270.

Gupta, A.K. and Varga, T. (1994b). Moments and other expected values for matrix variate elliptically contoured distributions. *Statistica*, **54**, 361-373.

Gupta, A.K. and Varga, T. (1995a). Normal mixture representations for matrix variate elliptically contoured distributions. *Sankhya*, **A57**, 68-78.

Gupta, A.K. and Varga, T. (1995b). Some inference problems for matrix variate elliptically contoured distributions. *Statistics*, **26**, 219-229.

Kudo, A. (1963). A multivariate analogue of the onesided test. *Biometrika*, **50**, 403-418.

Nandi, H.K. (1965). On some properties of Roy's union-intersection tests. *Cal. Stat. Assoc. Bull.*, **14**, 9-13.

Perlman, M.D. (1969). One-sided testing problems in multivariate analysis. *Ann. Math. Statist.*, **40**, 549-567.

Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Inference*. John Wiley and Sons, New York.

Tang, D.I., Gnecco, C. and Geller, N.L. (1989). Likelihood ratio test for normal means with non-negative components. *Biometrika*, **76**, 577-583.

APPENDIX A: PROOFS

Proof of Theorem 1

For the problem under consideration, the likelihood function is

$$L(\mu, \Sigma) = c(n, p) \left|\Sigma\right|^{\frac{n}{2}} f(\mathrm{tr}(\mathbf{X} - \mathbf{M})' \Sigma^{-1} (\mathbf{X} - \mathbf{M}))$$

$$= c(n, p) \left|\Sigma\right|^{\frac{n}{2}} f\left(\sum_{i=1}^{n} (\mathbf{X}_i - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{X}_i - \boldsymbol{\mu}) \right)$$

$$= c(n, p) \left|\Sigma\right|^{\frac{n}{2}} f\left[\sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})' \Sigma^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) + n(\bar{\mathbf{X}} - \boldsymbol{\mu})' \Sigma^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})) \right] \quad \text{(A.1)}$$

$$= c(n, p) \left|\Sigma\right|^{\frac{n}{2}} f\left[\mathrm{tr}\, \Sigma^{-1} \mathbf{A} + n(\bar{\mathbf{X}} - \boldsymbol{\mu})' \Sigma^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}) \right]$$

where $\mathbf{A} = \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$

Since f is decreasing, the unrestricted likelihood is maximized by

$$\hat{\boldsymbol{\mu}}_{\Omega} = \bar{\mathbf{X}}$$

By a theorem of Fang and Zhang (1990), we have

$$\hat{\Sigma}_{\Omega} = \lambda_{\max}(n, f) \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$$

$$= \lambda_{\max}(n, f)\mathbf{A}$$

where $\lambda_{\max}(n, f)$ is a constant depending on n and f only. Note that $\lambda_{\max}(n, f)$ need not be specified, since it divides out in the likelihood ratio.

The likelihood function is maximized over the restricted parameter space (where $\boldsymbol{\mu} \leq \mathbf{0}$) by

$\hat{\boldsymbol{\mu}}_{\omega} = \min(\mathbf{0}, \bar{\mathbf{X}}) = \bar{\mathbf{X}}_{-}$. Then $\bar{\mathbf{X}} - \hat{\boldsymbol{\mu}}_{\omega} = \bar{\mathbf{X}}_{+}$

The concentrated likelihood function on the restricted parameter space is

$$L(\hat{\boldsymbol{\mu}}_{\omega}, \Sigma) = c(n, p) \left|\Sigma\right|^{\frac{n}{2}} f\left(\mathrm{tr}\Sigma^{-1} \mathbf{A} + (\bar{\mathbf{X}} - \hat{\boldsymbol{\mu}}_{\omega})' \Sigma^{-1} (\bar{\mathbf{X}} - \hat{\boldsymbol{\mu}}_{\omega}) \right)$$

$$= c(n, p) \left|\Sigma\right|^{\frac{n}{2}} f\left(\mathrm{tr}\Sigma^{-1} \mathbf{A} + \bar{\mathbf{X}}_{+}' \Sigma^{-1} \bar{\mathbf{X}}_{+} \right) \quad \text{(A.2)}$$

$$= c(n, p) \left|\Sigma\right|^{\frac{n}{2}} f\left(\mathrm{tr}\Sigma^{-1} (\mathbf{A} + \bar{\mathbf{X}}_{+} \bar{\mathbf{X}}_{+}') \right)$$

$$= c(n, p) \left|\Sigma\right|^{\frac{n}{2}} f\left(\mathrm{tr}\Sigma^{-1} \mathbf{A}^{*} \right)$$

As above, we have $\hat{\Sigma}_{\omega} = \lambda_{\max}(n, f)\mathbf{A}^{*}$, and we have (see Fang and Zhang 1990 Lemma 5.2.1)

$$L(\hat{\boldsymbol{\mu}}_{\Omega}, \hat{\Sigma}_{\Omega}) = c(n, p) \,|\, \lambda_{\max}(n, f)\, \mathbf{A} |^{\frac{n}{2}} \times f(p/\lambda_{\max}(n, f)), \text{ and}$$

$$L(\hat{\boldsymbol{\mu}}_{\omega}, \hat{\Sigma}_{\omega}) = c(n, p) \,|\, \lambda_{\max}(n, f)\, \mathbf{A}^{*} |^{\frac{n}{2}} \times f(p/\lambda_{\max}(n, f))$$

Hence

$$\lambda = \frac{L(\hat{\boldsymbol{\mu}}_{\omega}, \hat{\Sigma}_{\omega})}{L(\hat{\boldsymbol{\mu}}_{\Omega}, \hat{\Sigma}_{\Omega})} = \left(\frac{|\mathbf{A}|}{|\mathbf{A}^{*}|} \right)^{\frac{1}{2}}, \text{ and}$$

$$\lambda^{\frac{2}{n}} = \frac{|\mathbf{A}|}{|\mathbf{A}^{*}|} = \frac{|\mathbf{A}|}{|\mathbf{A} + \bar{\mathbf{X}}_{+} \bar{\mathbf{X}}_{+}'|}$$

$$= \frac{|\mathbf{A}|}{|\mathbf{A}||\mathbf{I} + \mathbf{A}^{-1} \bar{\mathbf{X}}_{+} \bar{\mathbf{X}}_{+}'|} = \frac{1}{1 + \bar{\mathbf{X}}_{+}' \mathbf{S}^{-1} \bar{\mathbf{X}}_{+}/(n-1)}$$

Clearly then, λ is a strictly decreasing function of T_{+}^{2}. Therefore based on the likelihood principle, we reject H_0 for large values of T_{+}^{2}, which coincides with the procedure outlined in Section 3.2.

Proof of Theorem 3

In this proof the notation $E_k(\xi, \Lambda, f)\, f$ is used to denote the distribution of a k-dimensional random vector which has a density on $\mathbb{R}^k$ of the form $C(k)\, f((\mathbf{X} - \xi)' \Lambda^{-1} (\xi - \mathbf{X}))$.

From equation (4), we note that, conditional on $K = k$, T_{+}^{2} is a squared generalized distance in a

k-dimensional subspace. We will show that it is, in fact, squared empirical Mahalanobis distance with the expected probability distribution.

Consider an arbitrary permutation and partitioning of $\bar{\mathbf{X}}$ into k and p-k dimensional subvectors $\bar{\mathbf{X}}^* = (\bar{\mathbf{X}}_{(1)}^{*\prime}, \bar{\mathbf{X}}_{(2)}^{*\prime})'$, with corresponding permutation and partitioning for $\mathbf{S}$ and $\mathbf{\Sigma}$. Under $\mathbf{H}_0'$, $\bar{\mathbf{X}}^* \sim E_p(\mathbf{0}, n^{-1}\Sigma^*, f)$. The random variable

$$V_{(1)} = \bar{\mathbf{X}}_{(1)}^* - [n^{-1}\mathbf{S}_{12}^*][n^{-1}\mathbf{S}_{22}^*]^{-1}\bar{\mathbf{X}}_{(2)}^*$$

$$= \bar{\mathbf{X}}_{(1)}^* - \mathbf{S}_{12}^* \mathbf{S}_{22}^{*-1} \bar{\mathbf{X}}_{(2)}^* \quad \text{(A.3)}$$

has an E_k $(\mathbf{0}, n^{-1}\ \mathbf{\Sigma}_{11.2}^*, f^*)$ distribution, where

$$\mathbf{\Sigma}_{11.2}^* = \mathbf{\Sigma}_{11}^* - \mathbf{\Sigma}_{12}^*\mathbf{\Sigma}_{22}^{*-1}\mathbf{\Sigma}_{21}^* \quad \text{(A.4)}$$

This is, of course, exactly analogous to the normal case. The proof of this may be found in Fang and Zhang (1990), Section 2.6.4. $V_{(1)}$ is the orthogonal projection (in the generalized metric defined by Σ^*) of $\bar{\mathbf{X}}^*$ onto the k-dimensional subspace spanned by $\bar{\mathbf{X}}_{(1)}^*$.

Let $\mathbf{P_B}$ be a random $p \times p$ permutation matrix which permutes $\bar{\mathbf{X}}$ so that it may be partitioned into positive and non-positive subvectors. Clearly, $\mathbf{P_B}$ is a function of the random matrix $\mathbf{B}$ defined by equation (6). In this new coordinate system

$$\bar{\mathbf{X}}^* = \mathbf{P_B}\bar{\mathbf{X}} = \begin{pmatrix} \bar{\mathbf{X}}_{(1)}^* \\ \bar{\mathbf{X}}_{(2)}^* \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{X}}_{\mathbf{B}} \\ \bar{\mathbf{X}}_{(2)}^* \end{pmatrix} \text{ and}$$

$$\mathbf{S}^* = \mathbf{P_B}\mathbf{S}\mathbf{P_B'} \begin{pmatrix} \mathbf{S}_{11}^* & \mathbf{S}_{12}^* \\ \mathbf{S}_{21}^* & \mathbf{S}_{22}^* \end{pmatrix}$$

where $\bar{\mathbf{X}}_{\mathbf{B}}$ is as defined in the proof of Theorem 2. Furthermore,

$$\bar{\mathbf{X}}_{\mathbf{B}} = \bar{\mathbf{X}}_{(1)}^* - \boldsymbol{\mu}_{(1)\omega}^* - \mathbf{S}_{12}^*\mathbf{S}_{22}^{*-1}\left(\mathbf{X}_{(2)}^* - \boldsymbol{\mu}_{(2)\omega}^*\right)$$

$$= \bar{\mathbf{X}}_{(1)}^* - \mathbf{S}_{12}^*\mathbf{S}_{22}^{*-1}\left(\mathbf{X}^{*(2)} - \mathbf{X}_{(2)}^*\right)$$

$$= \bar{\mathbf{X}}_{(1)}^* \quad \text{(A.5)}$$

with $\mathbf{B}$ as defined in equation (6). Then $\bar{\mathbf{X}}_{\mathbf{B}}$ is the predicted value of $\bar{\mathbf{X}}^{*(1)}$ based on the restricted MLE $\boldsymbol{\mu}_{\omega}^*$. It is also a restricted generalized least squares linear predictor of $\bar{\mathbf{X}}_{\mathbf{B}}$.

Now, let us consider a representation of a vector variate elliptically contoured distribution centered at the origin as a discrete mixture of orthant truncated densities with elliptic contours. The contours for the mixture weighted densities match those on adjacent orthants since the parent distribution is continuous. Orthogonal projections map ellipsoids into ellipsoids. Since the orthogonal projection from a specific orthant with, say k, positive axes, $\bar{\mathbf{X}} \mapsto \bar{\mathbf{X}}_{\mathbf{B}}$, is onto a k-dimensional linear set, albeit truncated, the induced density also has elliptic contours on the lower dimensional set. The obvious exceptions are the negative orthant, which maps into the origin, and the positive orthant, which remains invariant. Note also that each orthant with k ($<p$) positive axes is mapped onto a distinct k-dimensional boundary set of $\mathbb{R}_+^p$. From the previous arguments, we conclude that the truncated density on each of these boundary sets has elliptic contours determined in each case by $n^{-1}\Sigma_{11.2}^*$ and f^*, which are determined by the specific projection and the parent density, respectively.

Now we have

$$T_+^2 = n\bar{\mathbf{X}}_+'\mathbf{S}^{-1}\bar{\mathbf{X}}_+ = n(\mathbf{B}\bar{\mathbf{X}}')\,\mathbf{S}^{-1}(\mathbf{B}\bar{\mathbf{X}})$$

$$= n\bar{\mathbf{X}}'\,\mathbf{B}(\mathbf{P_B'}\mathbf{P_B})\mathbf{S}^{-1}(\mathbf{P_B'}\mathbf{P_B})\mathbf{B}\bar{\mathbf{X}}$$

(orthogonality of permutation matrices)

$$= \bar{\mathbf{X}}'\,\mathbf{B}\mathbf{P_B'}(\mathbf{P_B}\mathbf{S}^{-1}\mathbf{P_B'})\mathbf{P_B}\mathbf{B}\bar{\mathbf{X}}$$

$$= n(\mathbf{P_B}\mathbf{B}\bar{\mathbf{X}})'\,((\mathbf{P_B'})^{-1}\mathbf{S}^{-1}\mathbf{P_B'})\,(\mathbf{P_B}\mathbf{B}\bar{\mathbf{X}})$$

(since $\mathbf{B}$ is diagonal and $\mathbf{P}' = \mathbf{P_B^{-1}}$)

$$= n\bar{\mathbf{X}}_+^{*\prime}(\mathbf{P_B}\mathbf{S}\mathbf{P_B'})^{-1}\bar{\mathbf{X}}_+^* = n\bar{\mathbf{X}}_+^{*\prime}\mathbf{S}^{*-1}\bar{\mathbf{X}}_+^* \quad \text{(A.6)}$$

$$= n\begin{pmatrix} \bar{\mathbf{X}}_{\mathbf{B}} \\ \mathbf{0} \end{pmatrix}' \mathbf{S}^{*-1} \begin{pmatrix} \bar{\mathbf{X}}_{\mathbf{B}} \\ \mathbf{0} \end{pmatrix}$$

$$= n\begin{pmatrix}\bar{\mathbf{X}}_{\mathbf{B}} \\ \mathbf{0}\end{pmatrix}'\begin{pmatrix}\mathbf{S}_{\mathbf{B}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\bar{\mathbf{X}}_{\mathbf{B}} \\ \mathbf{0}\end{pmatrix}$$

$$= n\bar{\mathbf{X}}'_{\mathbf{B}}\mathbf{S}_{\mathbf{B}}^{-1}\bar{\mathbf{X}}_{\mathbf{B}} \text{ (see equation (6)).}$$

By a theorem from Anderson (1984, Theorem A.3.3), we see that

$$\mathbf{S}_{\mathbf{B}}^{-1} = \mathbf{S}_{11.2}^{*-1} = (\mathbf{S}_{11}^{*} - \mathbf{S}_{12}^{*}\mathbf{S}_{22}^{*-1}\mathbf{S}_{21}^{*})^{-1} \qquad \text{(A.7)}$$

Let $K = k > 0$ and $\mathbf{DD}' = \Sigma_{11.2}$. Since $\bar{\mathbf{X}}_{\mathbf{B}}$ has all k entries positive, under the point null hypothesis $\mathbf{H}_0^n : \mu = 0$,

$$n^{1/2}\bar{\mathbf{X}}_{\mathbf{B}} \overset{d}{=} R^*\mathbf{DU}_C^{(k)}$$

where $C = S_k \cap \mathbb{R}_+^k$. Also

$$\mathbf{S}_{11.2}^{*} \overset{d}{=} (n-1)^{-1}\sum_{i=1}^{n-1}\mathbf{Z}_i\mathbf{Z}'_i$$

where $\mathbf{Z}_i \overset{d}{=} R_i\mathbf{DU}_i^{(k)}$ and $\mathbf{U}_i^{(k)} \sim \mathrm{U}(S_k)$ for $i = 1,\ldots, n-1$. In this representation the variables R^*, R_1, ..., R_{n-1}, $\mathbf{U}_C^{(k)}, \mathbf{U}_1^{(k)}, \ldots, \mathbf{U}_{n-1}^{(k)}$ are all independent.

From equation (A.7) above

$$n\bar{\mathbf{X}}'_{\mathbf{B}}\mathbf{S}_{\mathbf{B}}^{-1}\bar{\mathbf{X}}_{\mathbf{B}}$$

$$= \left[n^{1/2}\bar{\mathbf{X}}_{\mathbf{B}}\right]'\mathbf{S}_{11.2}^{*-1}\left[n^{1/2}\bar{\mathbf{X}}_{\mathbf{B}}\right]$$

$$\overset{d}{=} \left(R^*\mathbf{DU}_C^{(k)}\right)'\left[\frac{1}{n-1}\sum_{i=1}^{n-1}(R_i\mathbf{DU}_i^{(k)})(R_i\mathbf{DU}_i^{(k)})'\right]^{-1} \times \left(R^*\mathbf{DU}_C^{(k)}\right)$$

$$= \left(R^*\mathbf{U}_C^{(k)}\right)'\mathbf{D}'\left[\frac{1}{n-1}\mathbf{D}\sum_{i=1}^{n-1}(R_i\mathbf{U}_i^{(k)})(R_i\mathbf{U}_i^{(k)})'\mathbf{D}'\right]^{-1} \times \mathbf{D}\left(R^*\mathbf{U}_C^{(k)}\right)$$

$$= \left(R^*\mathbf{U}_C^{(k)}\right)'\mathbf{D}'\mathbf{D}'^{-1}\left[\frac{1}{n-1}\sum_{i=1}^{n-1}(R_i\mathbf{U}_i^{(k)})(R_i\mathbf{U}_i^{(k)})'\right]^{-1} \times \mathbf{D}^{-1}\mathbf{D}\left(R^*\mathbf{U}_C^{(k)}\right)$$

$$= \left(R^*\mathbf{U}_C^{(k)}\right)'\left[\frac{1}{n-1}\sum_{i=1}^{n-1}(R_i\mathbf{U}_i^{(k)})(R_i\mathbf{U}_i^{(k)})'\right]^{-1} \times \left(R^*\mathbf{U}_C^{(k)}\right)$$

By Lemma 1, this has the same distribution as it would if $\mathbf{U}_C^{(k)}$ were replaced with $\mathbf{U}^{(k)}$, uniform on $\mathcal{S}_k$. By the invariance properties shown in Lemma 6, together with Fang and Zhang (1990) Theorems 5.1.1(a) and 5.2.1, this is just the central distribution of Hotelling's T^2 with parameters k and $n - 1$, which is invariant over the ECD class.

Then conditioning on K and for $t \geq 0$,

$$\Pr(T_+^2 \leq t) = \Pr(K=0) + \sum_{k=1}^{p}\Pr(K=k)\Pr(T_{k,n-1}^2 \leq t)$$

$$= \Pr(O_0) + \sum_{k=1}^{p}\Pr(O_k)\Pr(T_{k,n-1}^2 \leq t)$$

Proof of Theorem 9

Since $\bar{\mathbf{X}}$ is in this case no longer centered at $\underline{0}$, the associated orthant probabilities are noncentral and are no longer independent of the infinite dimensional parameter f. Let them be denoted as $\Pr(O_{\mu,\Sigma,f}^{\mathbf{B}})$, where the matrix $\mathbf{B}$ indicates the orthant. The noncentrality for orthant probabilities is not simply a function of the standard noncentrality $\delta^2 = \mu'\Sigma^{-1}\mu$.

Another complication is that for each of the $2^p - 1$ distinct nonzero values that the projection matrix $\mathbf{B}$ takes, $\bar{\mathbf{X}}_{\mathbf{B}}$ has noncentrality

$$\delta_{\mathbf{B}}^2 = [\mathbf{B}\mu]'\Sigma^{-1}[\mathbf{B}\mu]$$

$$= \mu'_{\mathbf{B}}\Sigma_{\mathbf{B}}^{-1}\mu_{\mathbf{B}} \qquad \text{(A.8)}$$

$$= \boldsymbol{\mu}^{*(1)'} \Sigma_{11.2}^{*-1} \boldsymbol{\mu}_{(1)}^{*}$$

Depending on the orthant in which the alternative lies, $\delta_{\mathbf{B}}^2$ can take on from 1 to $2^p - 1$ distinct nonzero values. If μ_+ is zero except for the j^{th} coordinate, then $\delta_{\mathbf{B}}^2$ equals 0 if **B** does not select the j^{th} coordinate and $\boldsymbol{\mu}_j^2 / \sigma_{11.2}^{jj}$ otherwise. For alternatives in the positive orthant, δ_B^2 can take on 2^p distinct values, one of which is zero.

Fang and Zhang (sections 2.9.3 and 5.2.4) show that the noncentral distribution of Hotelling's T^2 for members of the ECD family which have a density depends on the form of the density function *f*. They define the generalized noncentral F-distribution and show that in the noncentral case, $\frac{k}{n-k}$ times a T^2 statistic has a generalized noncentral F-distribution with parameters k, $n - k$ and f and noncentrality δ^2. In this case, it has a density which is given by Fang and Zhang (1990), Theorem 2.9.5

$$\frac{2\pi^{(n-1)/2}}{\Gamma(k/2)\Gamma((n-k)/2)} \frac{k}{n-k} \left(\frac{k}{n-k} t\right)^{(k-2)/2}$$

$$\times \left(1 + \frac{k}{n-k} t\right)^{-n/2} \int_0^{\pi} \int_0^{\infty} \sin^{k-2}(\theta)\, y^{n-2}$$

$$\times f(y^2 - 2\,(kt/((n-1)t))^{1/2}\, y \cos\theta + \delta^2)\,d\theta\, dy$$

for $t > 0$. We will denote this density as $g(t\,|\,\delta^2, f)$. Then the noncentral distribution of T_+^2 may be written as

$$\Pr(T_+^2 \le t \mid \mu, \Sigma, f) = \Pr(O_{\boldsymbol{\mu},\Sigma,f}^{0})$$

$$+ \sum_{\mathbf{B}} \Pr(O_{\boldsymbol{\mu},\Sigma,f}^{\mathbf{B}}) \int_0^{\frac{tk}{n-k}} g(u \mid \delta_{\mathbf{B}}^2, f)\, du \tag{A.9}$$

where the summation is taken over all nonzero values of the matrix **B**, $k = \text{tr}\mathbf{B}$, and O^0 is the negative orthant.