Journal of Multivariate Analysis **68**, 264–282 (1999)
Article ID jmva.1998.1791, available online at http://www.idealibrary.com on IDEAL®

Two-Stage Likelihood Ratio and Union–Intersection Tests for One-Sided Alternatives Multivariate Mean with Nuisance Dispersion Matrix

Pranab K. Sen

University of North Carolina at Chapel Hill
E-mail: pksen@bios.unc.edu

and

Ming-Tien Tsai*

Academia Sinica, Taipei, Taiwain, Republic of China
E-mail: mttsai@stat.sinica.edu.tw

Received May 21, 1997; revised June 3, 1998

For a multinormal distribution with an unknown dispersion matrix, union-intersection (UI) tests for the mean against one-sided alternatives are considered. The null distribution of the UI test statistic is derived and its power monotonicity properties are studied. A Stain-type two-stage procedure is proposed to eliminate some of the inherent drawbacks of such tests. Some comparisons are also made with some recently proposed alternative conditional likelihood ratio tests. © 1999 Academic Press

AMS 1991 subject classifications: 62H15, 62L10.

Key words and phrases: Bartlett adjustment; M-matrix; non-null distribution; orthogonal projection; positive orthant; power monotonicity; projected tests; similar tests; stopping time; unbiasedness.

1. INTRODUCTION

Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be i.i.d.r.v.s with the $N_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ d.f. (distribution function), where $\boldsymbol{\Sigma}$ is positive definite (p.d.), and consider the hypotheses

$$H_0: \boldsymbol{\theta} = \mathbf{0} \qquad \text{vs} \qquad H_1: \boldsymbol{\theta} \in \Gamma_0 = \{\boldsymbol{\theta} \in R^p; \boldsymbol{\theta} \geqslant \mathbf{0}, \|\boldsymbol{\theta}\| > 0\}. \tag{1.1}$$

* Research supported in part by the National Science Council of Republic of China, NSC 87-2118-M-001-001.

0047-259X/99 $30.00

Various tests for this hypothesis have been considered under varying generality on $\mathbf{\Sigma}$: (i) $\mathbf{\Sigma}$ completely specified; (ii) $\mathbf{\Sigma} = \sigma^2 \mathbf{\Sigma}_0$, $\mathbf{\Sigma}_0$ specified but σ^2 unknown; and (iii) $\mathbf{\Sigma}$ is p.d. but completely unknown. Likelihood ratio tests (LRT) in case (i), considered by a host of researchers, are known to have some desirable properties; we refer the reader to Shapiro [12] for a useful discussion. Roy's [11] union–intersection tests (UIT) are strong competitors to the LRT, and often they are isomorphic. The projected tests proposed by Cohen, Kemperman, and Sackrowitz [4] may also be viewed as UIT.

Treating $\mathbf{\Sigma}$ as a p.d. nuisance matrix, Perlman [9] considered LRTs for a comparatively more general class of alternatives. Unfortunately, the null distribution of the LRT depends on the unknown $\mathbf{\Sigma}$, so that it is not a similar test. He considered suitable upper and lower bounds for this null distribution that may be used to prescribe some approximate or conservative tests. However, such a test is not known to have the usual properties, unbiasedness, monotonicity of power, etc. leaving us wondering: On what theoretical ground(s) may a LRT be advocated for such one-sided alternatives in case (iii)? The union–intersection (UI) principle and sufficient statistics are often used to yield alternative test criteria. In the next section we consider a class of tests for H_0 vs H_1 in case (iii) wherein the heuristic union–intersection (UI) principle of Roy [11] is incorporated in formulating UITs which are quite comparable to LRTs. The null distribution of a UI-scores test statistic is also derived in this section. Power-monotonicity and consistency of LRT and UIT are studied in Section 3. The dependence of the distribution of the UIT or LRT on the unknown $\mathbf{\Sigma}$ has been a major concern for the adoption of these tests in practice. Recently, Wang and McDermott [18] have considered a conditional version of the LRT that has an exact size; however, the critical values for such a conditional test are themselves random (they depend on the observed covariance matrix), and for values of p ($\geqslant 4$) their proposed algorithm may require extensive numerical computations. The present authors are not totally clear on the claimed unbiasedness property of the conditional LRT, and a counterexample is cited later on. Our approach, formulated in Section 4, is to extend the classical Stein [15] two-stage procedure to provide a resolution to this problem, along with an isomorphism of the LRT and UIT. Such two-stage LRT/UIT tests are shown to be similar, unbiased, and to have monotonicity of power properties. The allied distributional problems are also addressed. The concluding section is devoted to some general remarks pertaining to this one-sided hypothesis testing problem in case (iii) with some emphasis on power properties.

2. WHITHER UI-TESTS?

For every n ($\geqslant 2$), we let

$$\bar{\mathbf{X}}_n = n^{-1} \sum_{i=1}^{n} \mathbf{X}_i \qquad \text{and} \qquad \mathbf{S}_n = \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n)'. \tag{2.1}$$

Note that $(\bar{\mathbf{X}}_n, \mathbf{S}_n)$ is (jointly) sufficient for $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, so that a test statistic may be based solely on this set. We formulate the UIT along the line of Roy [11]. For each $\mathbf{b} \in \Gamma_0$, we define

$$H_{0,\mathbf{b}}: \mathbf{b}'\boldsymbol{\theta} = 0 \qquad \text{and} \qquad H_{1,\mathbf{b}}: \mathbf{b}'\boldsymbol{\theta} > 0, \tag{2.2}$$

so that

$$H_0 = \bigcap_{\{\mathbf{b} \in \Gamma_0\}} H_{0,\mathbf{b}} \qquad \text{and} \qquad H_1 = \bigcup_{\{\mathbf{b} \in \Gamma_0\}} H_{1,\mathbf{b}}. \tag{2.3}$$

Now for a given $\mathbf{b} \in \Gamma_0$, a one-sided UMP similar test statistic for testing $H_{0,\mathbf{b}}$ vs $H_{1,\mathbf{b}}$ is given by

$$T_n(\mathbf{b}) = n^{1/2}(\mathbf{b}'\bar{\mathbf{X}}_n)/\{\mathbf{b}'\mathbf{S}_n\mathbf{b}\}^{1/2}, \tag{2.4}$$

and $H_{0,\mathbf{b}}$ is rejected whenever $T_n(\mathbf{b})$ exceeds a critical level, say, $c_{n,\alpha}$, where α ($0 < \alpha < 1$) stands for the level of significance of the test. From (2.3) and (2.4), we obtain that the UIT for testing H_0 vs H_1 is based on the test statistic

$$T_n^* = \sup\{T_n(\mathbf{b}): \mathbf{b} \in \Gamma_0\}. \tag{2.5}$$

Our task is to find a closed expression for T_n^* and to find its critical value, say $c_{n,\alpha}^*$, that satisfies

$$P\{T_n^* \geqslant c_{n,\alpha}^* \mid H_0\} = \alpha. \tag{2.6}$$

To obtain a closed expression for T_n^*, we proceed as follows. Let $P = \{1, ..., p\}$, and for every a: $\emptyset \subseteq a \subseteq P$, let a' be its complement and $|a|$ its cardinality. Thus there are 2^p such possible subsets a: $\emptyset \subseteq a \subseteq P$ and $0 \leqslant |a| \leqslant p$. For each a, we partition (following possible rearrangement) $\bar{\mathbf{X}}_n$ and $\mathbf{S}_n$ as

$$\bar{\mathbf{X}}_n = \begin{pmatrix} \bar{\mathbf{X}}_{na} \\ \bar{\mathbf{X}}_{na'} \end{pmatrix} \qquad \text{and} \qquad \mathbf{S}_n = \begin{pmatrix} \mathbf{S}_{naa} & \mathbf{S}_{naa'} \\ \mathbf{S}_{na'a} & \mathbf{S}_{na'a'} \end{pmatrix}, \tag{2.7}$$

and write

$$\bar{\mathbf{X}}_{na:\,a'} = \bar{\mathbf{X}}_{na} - \mathbf{S}_{naa'}\,\mathbf{S}_{na'a'}^{-1}\,\bar{\mathbf{X}}_{na'}, \tag{2.8}$$

$$\mathbf{S}_{naa:\,a'} = \mathbf{S}_{naa} - \mathbf{S}_{naa'}\,\mathbf{S}_{na'a'}^{-1}\,\mathbf{S}_{na'a}. \tag{2.9}$$

Further, let

$$I_{na} = 1\{\,\bar{\mathbf{X}}_{na:\,a'} > \mathbf{0},\ \mathbf{S}_{na'a'}^{-1}\,\bar{\mathbf{X}}_{na'} \leqslant 0\}, \tag{2.10}$$

for $\varnothing \subseteq a \subseteq P$, where $1\{\cdot\}$ denotes the indicator function. Then using the Kuhn–Tucker–Lagrange (KTL) point formula theorem (Hadley [5]), we obtain that

$$T_n^* = \sum_{\varnothing \subseteq a \subseteq P} \{n\,\bar{\mathbf{X}}'_{na:\,a'}\,\mathbf{S}_{naa:\,a'}^{-1}\,\bar{\mathbf{X}}_{na:\,a'}\}\,I_{na}. \tag{2.11}$$

Side by side, we may express the LRT statistic L_n^* as

$$L_n^* = \sum_{\varnothing \subseteq a \subseteq P} \left\{\frac{n\,\bar{\mathbf{X}}'_{na:\,a'}\,\mathbf{S}_{naa:\,a'}^{-1}\,\bar{\mathbf{X}}_{na:\,a'}}{1 + n\,\bar{\mathbf{X}}'_{na'}\,\mathbf{S}_{na'a'}^{-1}\,\bar{\mathbf{X}}_{na'}}\right\} I_{na} \tag{2.12}$$

(see Section 7 of Perlman [9]). Note that only one of the I_{na} is nonzero, so that both the LRT and the UIT involve the common (random) nonzero I_{na} while the accompanying statistics are somewhat different. It is clear from the above that

$$T_n^* \geqslant L_n^*, \tag{2.13}$$

with probability one. It follows from Theorem 7.4 of Perlman [9] that for every $c > 0$,

$$\begin{aligned} P\{L_n^* \geqslant c \mid H_0, \boldsymbol{\Sigma}\} &= P_{0,\boldsymbol{\Sigma}}\{L_n^* \geqslant c\} \\ &= \sum_{k=1}^{p} w(p, k; \boldsymbol{\Sigma})\, P\{\chi_k^2/\chi_{n-p}^2 \geqslant c\}, \end{aligned} \tag{2.14}$$

where the χ_m^2 are independent r.v.s, χ_m^2 has the central chi-square distribution with m ($\geqslant 0$) degrees of freedom (DF), $\chi_0^2 = 0$ with probability 1, and for each k ($= 0, 1, ..., p$),

$$w(p, k; \boldsymbol{\Sigma}) = \sum_{\{a \subseteq P:\ |a| = k\}} P\{\mathbf{Z}_{a:\,a'} > \mathbf{0}, \boldsymbol{\Sigma}_{a'a'}^{-1}\mathbf{Z}_{a'} \leqslant \mathbf{0}\}, \tag{2.15}$$

$\mathbf{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, and the partitioning as in (2.8)–(2.9) (with $\mathbf{S}_n$ replaced by $\boldsymbol{\Sigma}$). Note that the dependence of (2.14) on $\boldsymbol{\Sigma}$ is only through the $w(p, k; \boldsymbol{\Sigma})$. For this reason, Perlman [9] allowed $\boldsymbol{\Sigma}$ to vary over the entire class $\{\boldsymbol{\Sigma} > \mathbf{0}\}$ of p.d. matrices and obtain that for every $c > 0$,

$$\sup_{\{\boldsymbol{\Sigma} > 0\}} P_{\mathbf{0}, \boldsymbol{\Sigma}}\{L_n^* \geqslant c\} = \tfrac{1}{2}[P\{\chi_p^2/\chi_{n-p}^2 \geqslant c\} + P\{\chi_{p-1}^2/\chi_{n-p}^2 \geqslant c\}]. \tag{2.16}$$

if the right-hand side of (2.16) can be equated to α, yielding $c = d_{n,\alpha}^*$, we obtain that

$$P_{\mathbf{0}, \boldsymbol{\Sigma}}\{L_n^* \geqslant d_{n\alpha}^*\} \leqslant \alpha, \qquad \forall \boldsymbol{\Sigma} > \mathbf{0}, \tag{2.17}$$

and hence the conservative character of the LRT based on the critical value $d_{n,\alpha}^*$ (when $\boldsymbol{\Sigma}$ is fixed but unknown) becomes apparent from (2.17).

The derivation of the null distribution of the UIT-statistic T_n^* is a bit more complex and it involves convolution of some independent chi-square variables. Let χ_a^2 and χ_b^2 be independent chi-square r.v.s with a and b degrees of freedom respectively. Let then

$$G_{a,b}(u) = P\{\chi_a^2/\chi_b^2 \leqslant u\}, \qquad u \in R^+, \tag{2.18}$$

and let $\bar{G}_{a,b}(u) = 1 - G_{a,b}(u)$. Let

$$\bar{G}_{n,a,p}^*(u) = \int_0^\infty \bar{G}_{a,n-p}\left(\frac{u}{1+t}\right) dG_{p-a,n-p+a}(t), \qquad u \in R^+. \tag{2.19}$$

Thus, G^* is the convolution of the d.f.s of χ_a^2/χ_{n-p}^2 and $1 + \chi_{p-a}^2/\chi_{n-p+a}^2$. Then we have the following.

THEOREM 2.1. *For every* $c > 0$,

$$P_{0,\boldsymbol{\Sigma}}\{T_n^* \geqslant c\} = P\{T_n^* \geqslant c \mid H_0, \boldsymbol{\Sigma}\}$$
$$= \sum_{k=1}^{p} w(p, k; \boldsymbol{\Sigma})\, \bar{G}_{n,k,p}^*(c), \tag{2.20}$$

where the $w(p, k; \boldsymbol{\Sigma})$ *are defined by* (2.15).

Proof. For every a: $\varnothing \subseteq a \subseteq P$, we define $V_a = n\bar{\mathbf{X}}'_{na:a'}\mathbf{S}_{naa:a'}^{-1}\bar{\mathbf{X}}_{naa:a'}$, $\mathbf{W}_a = n^{1/2}\mathbf{S}_{na'a'}^{-1}\bar{\mathbf{X}}_{na'}$, $\xi_a = \mathbf{W}'_a\mathbf{S}_{na'a'}\mathbf{W}_a$, $U_a = (1+\xi_a)^{-1} V_a$, and $\Upsilon_a = (1+\xi_a)^{-1/2} \sqrt{n}\,\bar{\mathbf{X}}_{na:a'}$. Then defining the I_{na} as in (2.10) and proceeding as in the proof of Theorem 7.2 of Perlman [9], we obtain that for every a: $\varnothing \subseteq a \subseteq P$ and $c > 0$,

$$
\begin{aligned}
E\{I_{na} I(V_a \geqslant c) \mid H_0, \boldsymbol{\Sigma}\}
&= P\{V_a \geqslant c, \Upsilon_a > \mathbf{0}, \mathbf{W}_a \leqslant \mathbf{0} \mid H_0, \boldsymbol{\Sigma}\} \\
&= P\{U_a(1+\xi_a) \geqslant c, \Upsilon_a > \mathbf{0}, \mathbf{W}_a \leqslant \mathbf{0} \mid H_0, \Sigma\} \\
&= P\{U_a(1+\xi_a) \geqslant c \mid H_0, \boldsymbol{\Sigma}\}\ P\{\Upsilon_a > \mathbf{0} \mid H_0, \boldsymbol{\Sigma}\}\ P\{\mathbf{W}_a \leqslant \mathbf{0} \mid H_0, \boldsymbol{\Sigma}\} \\
&= P_{0, \boldsymbol{\Sigma}}\{U_a(1+\xi_a) \geqslant c\}\ P_{0, \boldsymbol{\Sigma}}\{\Upsilon > \mathbf{0}\}\ P_{0, \boldsymbol{\Sigma}}\{\mathbf{W}_a \leqslant \mathbf{0}\}. \qquad (2.21)
\end{aligned}
$$

Next, note that under H_0, U_a and ξ_a are independent, and $U_a \overset{\mathscr{D}}{=} \chi^2_{|a|}/\chi^2_{n-p}$ and $\xi_a \overset{\mathscr{D}}{=} \chi^2_{|a'|}/\chi^2_{n-|a'|}$, where $\overset{\mathscr{D}}{=}$ stands for the equality of distributions. Thus, by (2.18), (2.19), and (2.21), for every $c > 0$,

$$
P_{\mathbf{0}, \boldsymbol{\Sigma}}\{U_a(1+\xi_a) \geqslant c\} = \bar{G}^*_{n, a, p}(c). \qquad (2.22)
$$

Note that for different a: $\varnothing \subseteq a \subseteq P$, $|a'| = p - |a|$, and further the $\bar{G}^*_n$ are all independent of $\boldsymbol{\Sigma}$. Finally,

$$
\sum_{\{a \subseteq P:\ |a| = k\}} P_{0, \boldsymbol{\Sigma}}\{\Upsilon_a > \mathbf{0}\}\ P_{0, \boldsymbol{\Sigma}}\{\mathbf{W}_a \leqslant \mathbf{0}\} = w(p, k; \boldsymbol{\Sigma}), \qquad \forall k = 0, 1, ..., p. \qquad (2.23)
$$

Therefore, (2.20) follows from (2.21), (2.22), and (2.23).

Note that in (2.14) the coefficient of $w(p, k; \boldsymbol{\Sigma})$ is $\bar{G}_{k, n-p}(c)$, whereas in (2.20) it is the convolution $\bar{G}^*_{n, k, p}(c)$. In this sense (2.20) differs from (2.14). For $k = p$, $\chi^2_0 = 0$ with probability 1, so that $\bar{G}^*_{n, p, p}(c) = \bar{G}_{p, n-p}(c)$, $\forall c > 0$, but for $k < p$, they are not the same. Nevertheless, (2.20), like (2.14), depends on $\boldsymbol{\Sigma}$ (unknown) only through the $w(p, k; \boldsymbol{\Sigma})$, so that like the LRT, the UIT is also not a similar test. As in (2.16), we may proceed to maximize (2.20) over the class $\{\boldsymbol{\Sigma} > \mathbf{0}\}$ of p.d. $\boldsymbol{\Sigma}$. We may virtually repeat the proof of Theorem 6.2 of Perlman [9] but work with (2.11) instead of (2.12) and conclude that for every $c > 0$,

$$
\sup_{\{\boldsymbol{\Sigma} > \mathbf{0}\}} P_{\mathbf{0}, \boldsymbol{\Sigma}}\{T^*_n \geqslant c\} = \tfrac{1}{2}[\bar{G}_{p, n-p}(c) + \bar{G}^*_{n, p-1, p}(c)], \qquad (2.24)
$$

where the $\bar{G}^*_n$ are defined by (2.19). We equate the right-hand side of (2.24) to α, yielding $c = c^*_{n, \alpha}$, and obtain that

$$
P_{\mathbf{0}, \boldsymbol{\Sigma}}\{T^*_n \geqslant c^*_{n, \alpha}\} \leqslant \alpha, \qquad \forall \boldsymbol{\Sigma} > \mathbf{0}. \qquad (2.25)
$$

Like the LRT, the UIT based on the critical level $c^*_{n,\alpha}$ is conservative and both of them are biased. In passing, we may note that the first term on the right-hand side of (2.16) is also $\bar{G}_{p,n-p}(c)$, while by (2.19), $\bar{G}_{n,p-1,p}(c) \geqslant G_{p-1,n-p}(c)$, $\forall c>0$. Therefore $c^*_{n,\alpha} \geqslant d^*_{n,\alpha}$ for every α: $0<\alpha<1$.

Let $R_p^+ = \{\mathbf{x} \in R^p \mid \mathbf{x} \geqslant \mathbf{0}\}$ and $\pi_{\mathbf{A}}(\mathbf{x}; \Gamma)$ be the orthogonal projection of $\mathbf{x}$ onto Γ with respect to the inner product $\langle\ ,\ \rangle_{\mathbf{A}}$. Then, by (2.11), we have

$$T_n^* = \|\pi_{\mathbf{S}_n}(n^{1/2}\bar{\mathbf{X}}_n; R_p^+)\|^2_{\mathbf{S}_n}, \tag{2.26}$$

where $\|\mathbf{x}\|^2_{\mathbf{A}}$ is defined to be $\mathbf{x}'\mathbf{A}^{-1}\mathbf{x}$. For the LRT in (2.12), we have

$$L_n^* = \|\pi_{\mathbf{S}_n}(n^{1/2}\bar{\mathbf{X}}_n; R_p^+)\|^2_{\mathbf{S}_n} \{1 + \|n^{1/2}\bar{\mathbf{X}}_n - \pi_{\mathbf{S}_n}(n^{1/2}\bar{\mathbf{X}}_n; R_p^+)\|^2_{\mathbf{S}_n}\}^{-1}. \tag{2.27}$$

This calls our attention to the structural difference between the LRT and UIT. The UIT is a direct orthogonal projection onto R_p^+ while the LRT is not so when $|a|<p$; $T_n^* = L_n^*$ when $|a| = p$, otherwise $T_n^* > L_n^*$. Thus, if $\boldsymbol{\theta}$ belongs to an edge of R_p^+, $|a|<p$ with a positive probability, and hence the LRT may not perform as well as the UIT.

At this stage, we refer to a recent work by Wang and McDermott [18] who considered a conditional version of the LRT and developed an algorithm for the computation of the conditional critical values (which are themselves stochastic in nature). In this context, we emphasize that the same conditionality principle applies to the UIT as well. Their algorithm can be modified to suit the computation of the conditional critical values of the UIT, given the sample covariance matrix. Moreover, by virtue of the ordering of the UIT and LIT, that is, the fact (as mentioned above) that, for $|a|<p$, $T_n^* > L_n^*$ and they are equal when $|a| = p$, we expect the critical levels of the two tests to be close to each other, but that the UIT will have power superiority to the LRT, especially when θ belongs to an edge of $\mathbf{R}_p^+$. We shall make some further comments on this in Section 4.

3. CONSISTENCY AND MONOTONICITY

The consistency of the LRT has already been established by Wang and McDermott [18]. Their proof virtually goes over to the case of the UIT after noting that $T_n^* \geqslant L_n^*$. Therefore, we omit the details. The "monotonicity of power" property for the LRT for some one-sided alternatives has been discussed by Perlman [9, Lemma 8.2]. However, his treatment does not cover the general case of H_0 vs H_1 under our consideration. Hu and Wright [7] studied this property for the LRT for testing of H_0 vs H_1 but only for the case (ii) but not (iii). The following theorem, also established in a simpler manner than in Hu and Wright [7], provides the desired result for both the LRT and UIT.

THEOREM 3.1. *Let* $\boldsymbol{\theta}_i \in R_p^+$, $i=1,2$. *If* $\boldsymbol{\theta}_2-\boldsymbol{\theta}_1 \in R_p^+$, *i.e.*, $\boldsymbol{\theta}_2 \geqslant \boldsymbol{\theta}_1 \geqslant \mathbf{0}$, *then for every* $c>0$,

$$P_{\boldsymbol{\theta}_2, \boldsymbol{\Sigma}}\{T_n^* \geqslant c\} \geqslant P_{\boldsymbol{\theta}_1, \boldsymbol{\Sigma}}\{T_n^* \geqslant c\} \tag{3.1}$$

and

$$P_{\boldsymbol{\theta}_2, \boldsymbol{\Sigma}}\{L_n^* \geqslant c\} \geqslant P_{\boldsymbol{\theta}_1, \boldsymbol{\Sigma}}\{L_n^* \geqslant c\}. \tag{3.2}$$

Proof. Note that for any $\boldsymbol{\eta} \in R_p^+$, $\Gamma_{\boldsymbol{\eta}} = \{\mathbf{x} = \mathbf{u} + \boldsymbol{\eta};\, \mathbf{u} \in R_p^+\} \subseteq R_p^+$, and hence

$$\|\pi_{\mathbf{S}_n}(n^{1/2}\overline{\mathbf{X}}_n;\, R_p^+)\|_{\mathbf{S}_n}^2 \geqslant \|\pi_{\mathbf{S}_n}(n^{1/2}\overline{\mathbf{X}}_n;\, \Gamma_{\boldsymbol{\eta}})\|_{\mathbf{S}_n}^2, \tag{3.3}$$

$\forall \boldsymbol{\eta} \in R_p^+$ respectively. Therefore, for every $c>0$, letting $\boldsymbol{\eta}_i = n^{1/2}\boldsymbol{\theta}_i$, $i=1,2$, $\boldsymbol{\eta} = \boldsymbol{\eta}_2 - \boldsymbol{\eta}_1$,

$$\begin{aligned}
P_{\boldsymbol{\theta}_2, \boldsymbol{\Sigma}}\{T_n^* \geqslant c\} &= P_{\boldsymbol{\theta}_2, \boldsymbol{\Sigma}}\{\|\pi_{\mathbf{S}_n}(n^{1/2}\overline{\mathbf{X}}_n;\, R_p^+)\|_{\mathbf{S}_n}^2 \geqslant c\} \\
&\geqslant P_{\boldsymbol{\theta}_2, \boldsymbol{\Sigma}}\{\|\pi_{\mathbf{S}_n}(n^{1/2}\overline{\mathbf{X}}_n;\, \Gamma_{\boldsymbol{\eta}})\|_{\mathbf{S}_n}^2 \geqslant c\} \\
&= P_{\boldsymbol{\theta}_2, \boldsymbol{\Sigma}}\{\|\pi_{\mathbf{S}_n}(n^{1/2}\overline{\mathbf{X}}_n - \boldsymbol{\eta};\, R_p^+)\|_{\mathbf{S}_n}^2 \geqslant c\} \\
&= P_{\boldsymbol{\theta}_1, \boldsymbol{\Sigma}}\{\|\pi_{\mathbf{S}_n}(n^{1/2}\overline{\mathbf{X}}_n - \mathbf{0};\, R_p^+)\|_{\mathbf{S}_n}^2 \geqslant c\} \\
&= P_{\boldsymbol{\theta}_1, \boldsymbol{\Sigma}}\{T_n^* \geqslant c\}.
\end{aligned} \tag{3.4}$$

And (3.2) can then be proved similarly to (3.4) by noting the fact that

$$\begin{aligned}
&\|\pi_{\mathbf{S}_n}(n^{1/2}\overline{\mathbf{X}}_n;\, R_p^+)\|_{\mathbf{S}_n}^2 \{1 + \|n^{1/2}\overline{\mathbf{X}}_n - \pi_{\mathbf{S}_n}(n^{1/2}\overline{\mathbf{X}}_n;\, R_p^+)\|_{\mathbf{S}_n}^2\}^{-1} \\
&\quad \geqslant \|\pi_{\mathbf{S}_n}(n^{1/2}\overline{\mathbf{X}}_n;\, \Gamma_{\boldsymbol{\eta}})\|_{\mathbf{S}_n}^2 \{1 + \|n^{1/2}\overline{\mathbf{X}}_n - \pi_{\mathbf{S}_n}(n^{1/2}\overline{\mathbf{X}}_n;\, \Gamma_{\boldsymbol{\eta}})\|_{\mathbf{S}_n}^2\}^{-1}.
\end{aligned} \tag{3.5}$$

These results ensure that as $\boldsymbol{\theta}$ moves in the interior of R_p^+, the powers of both the LRT and UIT increase. Moreover, we may remark that for the case (i) the LRT and UIT are isomorphic and both statistics are exactly he same as in (2.11) with $\mathbf{S}_n$ being replaced by the known covariance matrix $\boldsymbol{\Sigma}$. For the case (ii), both LRT and UIT statistics remain the same form as in (2.26) and (2.27) respectively with $\mathbf{S}_n$ being replaced by $s_n^2\boldsymbol{\Sigma}_0$, where s_n^2 is the maximum likelihood estimator of σ^2. Hence by virtually repeating the proof of Theorem 3.2 the power monotonicity property of LRT and UIT for the problem H_0 vs H_1 is also true for cases (i) and (ii).

4. TWO-STAGE TESTS

The conservativeness of LRT and UIT studied in the preceding section stems mainly from the adoption of the upper bounds in (2.16) and (2.24).

Since the exact null distribution in (2.14) or (2.20) or their conditional counterparts (given $\mathbf{S}_n$) depend on the unknown $\boldsymbol{\Sigma}$, a fixed-sample size solution to unbiasedness and attainment of the exact size of the LRT and UIT may not exist. In order to attain the exact size, Wang and McDermott [18] proposed a conditional LRT. However, their proof for the unbiasedness of their conditional LRT in Proposition 5 may not be totally correct; for $\mu, \mu_0 \in \boldsymbol{\Theta}$ with $\boldsymbol{\Theta}$ being the positive orthant and $\boldsymbol{\Sigma}$ being positive definite, their claim that $\mu' \boldsymbol{\Sigma}^{-1} \mu_0 > 0$ may not be generally true. To stress this point, we let $\mu = (1/2, 1/10)'$, $\boldsymbol{\Sigma}^{-1} = (\sigma^{ij})$ with $\sigma^{11} = \sigma^{22} = 1$ and $\sigma^{12} = \sigma^{21} = -1/2$, and $\mu_0 = (1/10, 9/10)$, then obviously we have $\mu' \boldsymbol{\Sigma}^{-1} \mu_0 = -9/100 < 0$. Also note that $\boldsymbol{\Sigma}$ is unknown but fixed, while $\mathbf{S}_n$ is stochastic. If $\boldsymbol{\Sigma}$ is an M-matrix, the sample covariance matrix $\mathbf{S}_n$ might not be an M-matrix with a positive probability, and as such the conditional LRT by conditioning on $\mathbf{S}_n$ might not have good power performance. The two-stage procedure to be considered in this and the following sections is primarily motivated by a desire to eliminate this shortcoming.

For the univariate normal mean testing problem, Stein [15] considered a two-stage procedure which has a power function independent of the nuisance parameter σ^2. Chatterjee [2, 3] considered various multivariate extensions of the Stein procedure and exhibited the $\boldsymbol{\Sigma}$-freeness of the power function. In all these cases, the singe-sample tests are similar and unbiased. In the current context, the fact that LRT and UIT are not similar implies that they cannot be unbiased at a given size. Therefore, we proceed to adopt a suitable modification of the Stein procedure so as to induce similarity as well as the unbiasedness of LRT and UIT in a simple way. In the next section, we shall study the power function.

We start with a couple $(n_0, \mathbf{D})$ where n_0 $(>p)$ is the first-stage sample size and $\mathbf{D}$ is a given p.d. matrix. In view of the class of restricted alternatives (H_1), we may choose $\mathbf{D}^{-1}$ to be an M-matrix; we may even allow $\mathbf{D}$ to be diagonal with positive elements reflecting the relative importance of the components of $\boldsymbol{\theta}$ with respect to H_1. Based on the first-stage sample $(\mathbf{X}_1, ..., \mathbf{X}_{n_0})$, we compute $\bar{\mathbf{X}}_{n_0} = n_0^{-1} \sum_{i=1}^{n_0} \mathbf{X}_i$ and an estimator of $\boldsymbol{\Sigma}$ by

$$\mathbf{S}_{n_0} = (n_0 - 1)^{-1} \sum_{i=1}^{n_0} (\mathbf{X}_i - \bar{\mathbf{X}}_{n_0})(\mathbf{X}_i - \bar{\mathbf{X}}_{n_0})'. \tag{4.1}$$

Note that $\mathbf{S}_{n_0}$ is p.d. with probability one and is an unbiased estimator of $\boldsymbol{\Sigma}$. Then, we define a stopping variable N $(= N(n_0, \mathbf{D}))$ by letting

$$N = \max\{n_0 + p^2, [\operatorname{tr}(\mathbf{D}^{-1}\mathbf{S}_{n_0})] + 1\}, \tag{4.2}$$

where $[s]$ denotes the integer part of s (>0) and $\operatorname{tr}(\mathbf{A})$ stands for the trace $\mathbf{A}$. Thus, $N \geqslant n_0 + p^2$ with probability 1, it is properly defined for all $\mathbf{D}$ p.d., and for $\mathbf{D}_1 \geqslant \mathbf{D}_2$ (in the sense $\mathbf{D}_1 - \mathbf{D}_2$ is positive semidefinite),

$N(n_0, \mathbf{D}_1) \leqslant N(n_0, \mathbf{D}_2)$. Finally, note that for every $n \geqslant n_0 + p^2$, $[N = n]$ is $\mathbf{S}_{n_0}$-measurable and independent of $\bar{\mathbf{X}}_{n_0}$.

Having obtained N in (4.2), we draw $N - n_0$ additional observations $\mathbf{X}_{n_0+1}, ..., \mathbf{X}_N$ from the same distribution, and write

$$\tilde{\mathbf{X}}_N = (\bar{\mathbf{X}}_{n_0}, \mathbf{X}_{n_0+1}, ..., \mathbf{X}_N), \qquad N^0 = N - n_0 + 1, \tag{4.3}$$

$$\mathbf{1}'_{N^0} = (1, ..., 1) \qquad \text{and} \qquad \mathbf{L} = \operatorname{diag}(n_0^{-1}, 1, ..., 1). \tag{4.4}$$

Define then $\mathbf{A} = (\mathbf{A}'_1, ..., \mathbf{A}'_p)'$, where

$$\mathbf{A}_i = (\mathbf{a}_{i1}, ..., \mathbf{a}_{ip})', \qquad i = 1, ..., p, \tag{4.5}$$

and the $\mathbf{a}_{ij}$ are all N^0-vectors, chosen so that

$$(\mathbf{A}_1 \mathbf{1}, ..., \mathbf{A}_p \mathbf{1}) = \mathbf{I}_p \qquad \text{and} \qquad \mathbf{A}\mathbf{L}\mathbf{A}' = \mathbf{D} \otimes \mathbf{S}_{n_0}^{-1}, \tag{4.6}$$

where $\otimes$ stands for the Kronecker product. This is indeed feasible, as can be verified by using the results in Chatterjee [2, 3]. Next, we define $\mathbf{Z}_N = (Z_{N1}, ..., Z_{Np})'$ by letting

$$Z_{Ni} = \operatorname{tr}(\mathbf{A}_i \tilde{\mathbf{X}}'_N), \qquad i = 1, ..., p. \tag{4.7}$$

Now, given n_0 and $\mathbf{D}$, conditionally on $\mathbf{S}_{n_0}$, N is fixed and so are the $\mathbf{A}_i$. Thus, proceeding as in Chatterjee [2, 3] we conclude that given $\mathbf{S}_{n_0}$,

$$\mathbf{Z}_N \sim \mathcal{N}_p(\boldsymbol{\theta}, \zeta \mathbf{D}); \qquad \zeta = \operatorname{tr}(\mathbf{S}_{n_0}^{-1} \boldsymbol{\Sigma}). \tag{4.8}$$

Further note that for any symmetric p.d. $\mathbf{Q}$, $\zeta = \operatorname{tr}(\mathbf{Q}^{-1} \mathbf{S}_{n_0}^{-1} \mathbf{Q}^{-1} \mathbf{Q} \boldsymbol{\Sigma} \mathbf{Q})$, so that taking $\mathbf{Q} = \boldsymbol{\Sigma}^{-1/2}$, we conclude that under $\boldsymbol{\Sigma}$, ζ has the same distribution as $\zeta_0 = \operatorname{tr}(\mathbf{S}_{n_0}^{-1})$ under $\boldsymbol{\Sigma} = \mathbf{I}_p$. Thus, the distribution of ζ does not depend on $\boldsymbol{\Sigma}$.

Based on the pair $(\mathbf{Z}_N, \mathbf{D})$, in (2.10)–(2.11), replacing $\mathbf{S}_n$ by $\mathbf{D}$, we define the two-stage UIT statistic as

$$T_N^{*0} = \sum_{\varnothing \subseteq a \subseteq P} \{\mathbf{Z}'_{Na:a'} \mathbf{D}_{aa:a'}^{-1} \mathbf{Z}_{Na:a'}\} \, 1\{\mathbf{Z}_{Na:a'} > \mathbf{0}, \mathbf{D}_{a'a'}^{-1} \mathbf{Z}_{Na'} \leqslant \mathbf{0}\}. \tag{4.9}$$

Note that for the $\mathbf{Z}_{Na:a'}$ we use $\mathbf{D}$ instead of $\mathbf{S}_n$. Moreover, we define the $w(p, k, \mathbf{D})$ as in (2.15) with the unknown $\boldsymbol{\Sigma}$ replaced by the given $\mathbf{D}$, so that they are all independent of $\boldsymbol{\Sigma}$. Also, they are all scale-invariant, so that

$$w(p, k; \zeta \mathbf{D}) = w(p, k; \mathbf{D}), \qquad \forall 0 \leqslant k \leqslant p, \quad \zeta > 0. \tag{4.10}$$

We denote the distribution function (d.f.) of χ_k^2 by $G_k(x)$, $k \geqslant 0$, and let $\bar{G}_k(x) = 1 - G_k(x)$, $x \geqslant 0$, $k \geqslant 0$. Since $(n_0 - 1) \boldsymbol{\Sigma}^{-1} \mathbf{S}_{n_0}$ has the Wishart $(p, n_0 - 1, \mathbf{I}_p)$ d.f., we obtain that $(n_0 - 1)^{-1} \mathbf{S}_{n_0}^{-1} \boldsymbol{\Sigma}$ has the inverted Wishart

$(p, n_0-1, \mathbf{I}_p)$ d.f.; we may refer to Theorem 7.7.1 of Anderson [1]. Therefore, $\zeta = \operatorname{tr}(\mathbf{S}_{n_0}^{-1}\boldsymbol{\Sigma}) = (n_0-1)\operatorname{tr}(\mathbf{W}^*)$, where $\mathbf{W}^*$ has the inverted Wishart $(p, n_0-1, \mathbf{I}_p)$ d.f. We denote the d.f. of ζ by $\Psi_{n0}(y)$, $y \geqslant 0$, and note that Ψ_{n_0} is free from $\boldsymbol{\Sigma}$. Consider then the convolutions

$$\bar{G}^*_{k,p}(x) = \int_0^\infty \bar{G}_k(x/y)\, d\Psi_{n0}(y), \qquad x \in R^+, \quad k \geqslant 0. \tag{4.11}$$

Since none of G_k and Ψ_{n_0} depend on $\boldsymbol{\Sigma}$, we claim that the $\bar{G}^*_{k,p}$ are all free from $\boldsymbol{\Sigma}$. Our main result of this section is the following:

THEOREM 4.1. *For given* $(n_0, \mathbf{D})$, *for every* $c>0$,

$$\begin{aligned} P_{0,\boldsymbol{\Sigma}}\{T_N^{0*} \geqslant c\} &= P\{T_N^{0*} \geqslant c \mid H_0, \boldsymbol{\Sigma}\} \\ &= \sum_{k=1}^p w(p,k;\mathbf{D})\, \bar{G}^*_{k,p}(c), \end{aligned} \tag{4.12}$$

and hence, is independent of the unknown $\boldsymbol{\Sigma}$.

Proof. Note that given $\mathbf{S}_{n_0}$, N and ζ are also fixed, and $\zeta\mathbf{D}$ is a scalar multiplication of the given $\mathbf{D}$. Thus, we may proceed as in the proof of Theorem 2.1 and obtain that the conditional null distribution of T_N^{*0}, given $\mathbf{S}_{n_0}$ is given by

$$P_0\{T_N^{*0}/\zeta \geqslant c \mid \mathbf{S}_{n0}\} = \sum_{k=1}^p w(p,k,\mathbf{D})\, \bar{G}_k(c), \qquad c>0. \tag{4.13}$$

Note that the right-hand side of (4.13) does not depend on $\boldsymbol{\Sigma}$. Therefore, we have for every $c>0$,

$$\begin{aligned} P_{0,\boldsymbol{\Sigma}}\{T_N^{*0} \geqslant c\} &= E[P_{0,\boldsymbol{\Sigma}}\{T_N^{*0}/\zeta \geqslant c/\zeta \mid \mathbf{S}_{n0}\}] \\ &= \int_0^\infty P_0\{T_N^{*0}/y \geqslant c/y \mid \mathbf{S}_{n0}\}\, d\Psi_{n0}(y) \\ &= \sum_{k=1}^p w(p,k;\mathbf{D}) \int_0^\infty \bar{G}_k(c/y)\, d\Psi_{n0}(y) \\ &= \sum_{k=1}^p w(p,k;\mathbf{D})\, \bar{G}^*_{k,p}(c). \end{aligned} \tag{4.14}$$

Next, we adopt the proof of Theorem 3.2, we take $\mathbf{Z}_N$ and $\mathbf{D}$ instead of $\sqrt{n}\,\bar{\mathbf{X}}_n$ and $\mathbf{S}_n$, thus we obtain that for every $\boldsymbol{\theta}_2 \geqslant \boldsymbol{\theta}_1 \geqslant \mathbf{0}$ and $c>0$,

$$P_{\boldsymbol{\theta}_2,\boldsymbol{\Sigma}}\{T_N^{*0} \geqslant c \mid \zeta\} \geqslant P_{\boldsymbol{\theta}_1,\boldsymbol{\Sigma}}\{T_N^{*0} \geqslant c \mid \zeta\} \qquad \text{a.a. } \zeta, \tag{4.15}$$

and as Ψ_{n_0}, the d.f. of ζ, is free from $\mathbf{\Sigma}$ or $\boldsymbol{\theta}$s, integrating both sides of (4.15) with respect to Ψ_{n_0}, we have

$$P_{\boldsymbol{\theta}_2, \mathbf{\Sigma}}\{T_N^{*0} \geqslant c\} \geqslant P_{\boldsymbol{\theta}_1, \mathbf{\Sigma}}\{T_N^{*0} \geqslant c\}, \qquad \forall c > 0, \quad \boldsymbol{\theta}_1 \geqslant \boldsymbol{\theta}_1 \geqslant \mathbf{0}. \quad (4.16)$$

Thus, the two-stage UIT has a monotone power property. On the other hand, equating (4.12) to α, we obtain a critical level c_α^{*0} which does not depend on $\mathbf{\Sigma}$. Thus, the two-stage UIT based on the rejection region $T_N^{*0} \geqslant c_\alpha^{*0}$ (and the stopping rule N in (4.2)) is a similar size-α test. Combining this similarity along with the power monotonicity property in (4.16), we obtain the following.

THEOREM 4.2. *The two-stage UIT is unbiased, similar, and has a monotone power property.*

We may note further that in (4.8) the dispersion matrix $\zeta\mathbf{D}$ corresponds to case (ii) (in Section 1), albeit in a conditional setup, given $\mathbf{S}_{n_0}$. As such, based on this conditional multinormal law, if we construct the LRT and denote the statistic by L_N^{*0}, it follows readily that

$$L_N^{*0} \equiv T_N^{*0}. \quad (4.17)$$

Combining Theorems 4.1 and 4.2, and (4.17), we have the following.

THEOREM 4.3. *The two-stage LRT and UIT are isomorphic and share the properties of similarity, unbiasedness and power monotonicity.*

Let us conclude this section with some pertinent remarks on the d.f.s $\bar{G}_{k,p}^*$ in (4.11), and for this we need to took into Ψ_{n_0} first. For $p = 2$, Ψ_{n_0} has a simple form and therefore for $\bar{G}_{1,p}^*$ and $\bar{G}_{2,p}^*$ workable forms are available in the literature. For $p \geqslant 3$, the situation is much more complex. For $\bar{G}_{p,p}^*$, $p \geqslant 3$, Chatterjee [2] has worked out some convenient approximations, and similar ones can be obtained for $\bar{G}_{k,p}^*$, $k \leqslant p$, $p \geqslant 3$. If n_0 is larger than $p + 2r$, for some positive integer r, then we may use the central moments (up to the order r) of ζ to provide another approximation in terms of $\bar{G}_{k+2j}$, $j \geqslant 0$. Toward this ends, we note first that if k is even $(= 2r$, say), then for $y > 0$, $\bar{G}_0(y) = 0$ and

$$\bar{G}_{2r}(y) = 2 \sum_{j=1}^{r} g_{2j}(y) = \sum_{j=0}^{r-1} \frac{1}{j!} \left(\frac{1}{2} y\right)^j e^{-y/2}. \quad (4.18)$$

For $k = 2r+1$, odd,

$$\bar{G}_{2r+1}(y) = \bar{G}_1(y) + 2\sum_{j=1}^{r} g_{2j+1}(y)$$
$$= \bar{G}_1(y) + \sum_{j=1}^{r} \frac{1}{\Gamma(j+1/2)} \left(\frac{1}{2}y\right)^{j-1/2} e^{-y/2}, \tag{4.19}$$

where

$$\bar{G}_1(y) = \frac{1}{\sqrt{2\pi}} \int_y^{\infty} x^{-1/2} e^{-(1/2)x}\, dx, \qquad y \geqslant 0. \tag{4.20}$$

Further, note that

$$E(\mathbf{S}_{n_0}^{-1} \mid \boldsymbol{\Sigma} = \mathbf{I}_p) = \frac{n_0 - 1}{n_0 - p - 2} \mathbf{I}_p, \qquad \forall n_0 > p+2, \tag{4.21}$$

so that

$$E(\zeta) = \frac{(n_0 - 1)\, p}{n_0 - p - 2} = \zeta^0, \qquad \text{say}, \tag{4.22}$$

where $\zeta^0 > p$. Also using the second order moment result on inverted Wishart matrix (viz., Siskind [13]), we obtain in the following some routine steps that for $n_0 > p+4$,

$$E(\zeta - \zeta^0)^2 = \operatorname{Var}(\zeta)$$
$$= \frac{p(n_0-1)^2}{(n_0-p-2)(n_0-p-4)} \left\{ \frac{n_0 p - p^2 - 3p + 2}{n_0 - p - 1} \right\}$$
$$- \frac{(n_0-1)^2\, p^2}{(n_0-p-2)^2}$$
$$= \frac{2p(n_0-1)^2\,(n_0-2)}{(n_0-p-2)^2\,(n_0-p-4)(n_0-p-1)}. \tag{4.23}$$

Higher order central moments of ζ can also be obtained from the higher order moments of inverted Wishart matrix (viz., von Rosen [10]), but they require even larger values of n_0, for which the first order approximation will be quite adequate. Hence we refrain from these additional computations.

It follows from (4.11) and (4.18) that for $k=2r$, $r \geqslant 1$, for every $c>0$,

$$\bar{G}^*_{k,\,p}(c) = \sum_{j=0}^{r-1} \frac{1}{j!} \int_0^\infty \left(\frac{c}{2y}\right)^j e^{-c/2y}\, d\Psi_{n_0}(y)$$
$$= \sum_{j=0}^{r-1} \frac{1}{j!} E[h_j(\zeta; c)], \tag{4.24}$$

where

$$h_j(\zeta; c) = \left(\frac{c}{2\zeta}\right)^j e^{-c/2\zeta}, \qquad \text{for} \quad j \geqslant 0, \quad \zeta > 0, \quad c > 0. \tag{4.25}$$

Next note that

$$(\partial/\partial\zeta)\, h_j(\zeta; c) = h_{j+1}(\zeta, c) \left[\frac{1}{\zeta} - \frac{2j}{c}\right], \qquad j = 0, 1, ..., \tag{4.26}$$

so that recursively

$$(\partial^2/\partial\zeta^2)\, h_j(\zeta; c) = -\frac{1}{\zeta^2} h_{j+1}(\zeta; c) + h_{j+2}(\zeta; c) \left[\frac{1}{\zeta} - \frac{2j}{c}\right]\left[\frac{1}{\zeta} - \frac{2(j+1)}{c}\right], \tag{4.27}$$

for $j = 0, 1, ...,$ and higher order derivatives can be obtained recursively. Moreover, using the harmonic mean-arithmetic mean inequality on the characteristic roots of $\mathbf{\Sigma}^{-1}\mathbf{S}_{n_0}$, we obtain that

$$\zeta^{-1} \leqslant p^{-2}\, \mathrm{tr}\, (\mathbf{\Sigma}^{-1}\mathbf{S}_{n_0})$$
$$= \{(n_0 - 1)\, p^2\}^{-1}\, \mathrm{tr}(\mathbf{W}) n \tag{4.28}$$

where $\mathbf{W}$ is a Wishart $(p, n_0 - 1, \mathbf{I}_p)$ matrix. Since the moment generating function of $\mathrm{tr}(\mathbf{W})$ exists for all $n_0 > p$, there is no problem with the convergence of the negative moments of ζ of any order. Thus, writing

$$a_j^0(c) = (\partial/\partial\zeta)\, h_j(\zeta; c)|_{\zeta = \zeta^0} \tag{4.29}$$

$$b_j^0(c) = (\partial^2/\partial\zeta^2)\, h_j(\zeta; c)|_{\zeta = \zeta^0}, \qquad j \geqslant 0, \tag{4.30}$$

we have a valid Taylor's expansion

$$h_j(\zeta; c) = h_j(\zeta^0; c) + (\zeta - \zeta^0)\, a_j^0(c) + \tfrac{1}{2}(\zeta - \zeta^0)^2\, b_j^0(c) + \cdots, \tag{4.31}$$

so that we consider the first order approximation

$$
\begin{aligned}
Eh_j(\zeta; c) &= h_j(\zeta^0; c) + \frac{1}{2}\, b_j^0(c)\, E(\zeta - \zeta^0)^2 + \cdots \\
&= h_j(\zeta^0; c) + \frac{1}{2}\, b_j^0(c)\, \frac{2p(n_0-1)^2\,(n_0-2)}{(n_0-p-2)^2\,(n_0-p-4)(n_0-p-1)} \\
&\quad + o\left(\frac{1}{n_0-p}\right).
\end{aligned} \tag{4.32}
$$

From (4.24), (4.27), and (4.32), we obtain that for $k = 2r$, $r \geqslant 1$, $c > 0$,

$$
\begin{aligned}
\bar{G}^*_{k,\,p}(c) &= \left\{1 - \frac{c(n_0-2)(n_0-p-2)}{2p^2(n_0-1)(n_0-p-1)(n_0-p-4)}\right\} \bar{G}_k\left(\frac{c(n_0-p-2)}{p(n_0-1)}\right) \\
&\quad + \frac{c^2(n_0-2)}{4p(n_0-p-1)(n_0-p-4)} \sum_{j=0}^{n-1} \frac{1}{j!}\left(\frac{c(n_0-p-2)}{p(n_0-1)}\right)^j \\
&\quad \times \left[\exp\left\{-\frac{c(n_0-p-2)}{2p(n_0-1)}\right\}\right]\left[\frac{n_0-p-2}{p(n_0-1)} - \frac{2j}{c}\right] \\
&\quad \times \left[\frac{n_0-p-2}{p(n_0-1)} - \frac{2(j+1)}{c}\right] + o\left(\frac{1}{n_0-p}\right).
\end{aligned} \tag{4.33}
$$

It is also clear from the above equation that $\bar{G}^*_{2r+2,\,p}(c)$ can be recursively computed form $\bar{G}^*_{2r,\,p}(c)$ by adding the additional contributions of the rth term in the right-hand side of (4.33) along with $2g_{2r+2}(c(n_0-p-2)/p(n_0-1))\{1 - c(n_0-2)(n_0-p-2)/2p^2(n_0-1)(n_0-p-1)(n_0-p-4)\}$ arising from the first term. A similar recursive scheme applies to $\bar{G}^*_{2r+1,\,p}(c)$ (by (4.19)), and hence it suffices to consider only the case of $\bar{G}^*_{1,\,p}(c)$. By a similar Taylor's expansion, we have

$$
\begin{aligned}
\bar{G}^*_{1,\,p}(c) &= \bar{G}_1\left(\frac{c(n_0-p-2)}{p(n_0-1)}\right) - \frac{(n_0-2)}{2p(n_0-p-4)(n_0-p-1)} \\
&\quad \times \left[\exp\left\{-\frac{c(n_0-p-2)}{2p(n_0-1)}\right\}\right] \frac{1}{\Gamma(1/2)}\left(\frac{c(n_0-p-2)}{p(n_0-1)}\right)^{1/2} \\
&\quad \times \left\{1 + 2\,\frac{c(n_0-p-2)}{p(n_0-1)}\right\} + o\left(\frac{1}{n_0-p}\right).
\end{aligned} \tag{4.34}
$$

For conformity with (4.33), we rewrite (4.34) as

$$\bar{G}^*_{1,p}(c)=\left\{1-\frac{c(n_0-2)(n_0-p-2)}{2p^2(n_0-1)(n_0-p-1)(n_0-p-4)}\right\}$$
$$\times\bar{G}_1\left(\frac{c(n_0-p-2)}{p(n_0-1)}\right)+\frac{(n_0-2)(n_0-p-2)}{2p(n_0-p-4)(n_0-p-1)}$$
$$\times\left[\frac{c}{(n_0-1)\,p}\,\bar{G}_1\left(\frac{c(n_0-p-2)}{p(n_0-1)}\right)-\frac{1}{\Gamma(1/2)}\right.$$
$$\times\left(\frac{c(n_0-p-2)}{p(n_0-1)}\right)^{1/2}\exp\left\{-\frac{c(n_0-p-2)}{2p(n_0-1)}\right\}$$
$$\left.\times\left\{1+\frac{2c(n_0-p-2)}{p(n_0-1)}\right\}\right]+o\left(\frac{1}{n_0-p}\right). \tag{4.35}$$

The similarity with the Bartlett adjustment for the LRT in the regular case (viz., Anderson [1, Chap. 8]) can be noticed. For moderate to large values of n_0 ($\gg p$), these approximations are quite adequate.

5. SOME GENERAL REMARKS

The stopping variable N, defined by (4.2), depends on the chosen pair $(n_0, \mathbf{D})$ and the stochastic matrix $\mathbf{S}_{n_0}$. We have commented on the choice of $\mathbf{D}^{-1}$ and advocated the use of an M-matrix. One of the main reasons for this recommendation is that in the conditional setup of (4.8), the dispersion matrix $\zeta\mathbf{D}$ becomes an (unknown) scalar multiple of the inverse of an M-matrix ($\mathbf{D}$), and then the corresponding two-stage LRT/UIT have convex acceptance regions. Ideally, n_0 should be chosen to be large compared to p. If this is the case, in the expressions for the $\bar{G}^*_{k,p}(c)$ considered in the preceding section, the remainder terms being typically $O((n_0-p)^{-2})$ are very small, and the first-order approximations are quite adequate in practice.

For the normal mean problem–global alternatives, it is known that the Stein [15] two-stage procedure is generally not fully efficient, in the sense that the ASN for this procedure is generally higher than the optimal value if $\boldsymbol{\Sigma}$ were known. A multistage procedure eliminates this problem to a greater extent (viz., Hall [6]). Alternatively, Mukhopadhyay [8] has shown that in an asymptotic setup (where $\boldsymbol{\theta}\to\mathbf{0}$), choosing n_0 large, but small compared to $\|\boldsymbol{\theta}\|^{-2}$, induces asymptotic efficiency of the Stein two-stage procedure. We like to discuss such results in the present study of one-sided alternatives. Motivated by this, we proceed as in Chatterjee [2, 3] and consider a sequence of stopping variables $\{N_d, d>0\}$, where

$$N_d = \max\{n_0 + p^2, [d^{-1}\operatorname{tr}(\mathbf{D}^{-1}\mathbf{S}_{n_0})] + 1\}, \qquad d>0. \tag{5.1}$$

Then N_d is properly defined for every $d>0$, it is nonincreasing in d (>0), and by definition

$$\lim_{d \downarrow 0} N_d = +\infty \qquad \text{almost surely (a.s.)} \tag{5.2}$$

In the Stein–Chatterjee case, one has the global alternatives $\boldsymbol{\theta} \neq \mathbf{0}$, and the choice of N_d was primarily made to make the power function of the test independent of the nuisance parameter $\boldsymbol{\Sigma}$. On the other hand, in Section 4, the two-stage procedure was primarily designed for establishing the similarity, unbiasedness, and monotonicity of power function of the LRT/UIT for such one-sided alternatives. Nevertheless, we may consider the entire class $\mathscr{C}^*$ of stopping rules

$$\mathscr{C} = \{N_d, d>0, \mathbf{D}^{-1} \text{ M-matrix}\}, \tag{5.3}$$

and for any member of this class we may show that the corresponding two-stage test (LRT/UIT) has the three desirable properties referred to earlier. For a given $\mathbf{D}$, using the monotonicity of N_d, it can be shown that the smaller the value of d (>0) chosen in (5.1), the better the corresponding test will be. However, varying $\mathbf{D}^{-1}$ over the class of all M-matrices, we have a much more complex situation, and a parallel statement is hard to make. The main complication arises due to the rather complex form of the non-null distribution of a two-stage LRT/UIT, and we shall make more comments on that later. Although the expected stopping time EN_d may be studied as in Chatterjee [2], the non-null distributions of different two-stage LRT/UITs belonging to this class $\mathscr{C}^*$, for one-sided alternatives, may not conform to a simple function of EN_d. Much of this complexity is due to the intractability of the power function of LRT/UITs for one-sided alternatives in a closed form (even for the one-sample setup). The independence of the three stochastic elements $\mathbf{Z}_{Na:\,a'}$, $\mathbf{D}^{-1}_{a'a'}\mathbf{Z}_{Na'}$, and $\mathbf{Z}_{Na:\,a'}\mathbf{D}^{-1}_{aa:\,a'}\mathbf{Z}_{Na:\,a'}$, $\varnothing \subseteq a \subseteq P$, does not generally hold when H_0 is not true. Thus, for the non-null distribution, parallel to (4.12), we may not have a closed form where the $w(p, k; \mathbf{D})$ are there and the $\bar{G}^*_{k,p}$ are replaced by appropriate non-central forms. In fact, even if this were true, the $w(p, k; \mathbf{D})$ would have depended on $\boldsymbol{\theta} \in R_p^+$ as under the alternative, $E\mathbf{Z}_N = \boldsymbol{\theta} \geqslant \mathbf{0}$. On top of that, the LRT/UIT for one-sided alternatives are generally not invariant under affine transformations, so that analogues of non-central Hotelling T^2 distributions may not exist for such UIT/LRTs. For the know $\mathbf{\Sigma}$ case, Tsai [16] proved that LRT/UIT are power-superior to that of Hotelling T^2 tests for some restricted alternatives, including the special case when $\mathbf{\Sigma}^{-1}$ is a known M-matrix, and the results may be adopted in this conditional setup as the (two-stage) LRT/UITs are shown to be the natural analogue of (two-stage) Hotelling T^2 tests. Since ζ is a positive random variable, the power of two-stage LRT/UIT should be at least as large as that of two-stage Hotelling T^2 test over the domain R_p^+ if $\mathbf{D}$ is in $\mathscr{M}$, although the opposite picture may hold for the complementary part $R_p \backslash R_p^+$.

Motivated by these factors, we may proceed as in Chatterjee [2, 3] and choose N_d corresponding to a given d (>0) such that the power of the two-stage Hotelling T^2 test based on N_d has a minimum power $(1-\beta)$ on an ellipsoidal contour on R_p. If we confine ourselves to the segment of that contour on R_p^+, then the two-stage LRT/UIT based on the same stopping rule N_d will not only be similar, unbiased, and a monotone power test, but also will be expected to have the same minimum power $1-\beta$ outside this contour in R_p^+. This provides a basis for the choice of d (>0) in (5.1).

ACKNOWLEDGMENT

We thank the reviewers for their very helpful comments.

REFERENCES

1. T. W. Anderson, "An Introduction to Multivariate Statistical Analysis," 2nd ed., Wiley, New York, 1984.
2. S. K. Chatterjee, On an extension of Stein's two-sample procedure to the multinormal problem, *Calcutta Statist. Assoc. Bull.* **8** (1959), 121–148.
3. S. K. Chatterjee, Sequential inference procedures of Stein's type for a class of multivariate regression problems, *Ann. Math. Statist.* **33** (1962), 1039–1064.
4. A. Cohen, J. H. B. Kemperman, and H. B. Sackrowitz, Projected tests for order restricted alternatives, *Ann. Statist.* **22** (1994), 1539–1546.
5. G. Hadley, "Non-linear and Dynamic Programming," Addison–Wesley, Reading, MA, 1964.
6. P. Hall, Asymptotic theory of triple sampling for sequential estimation of a mean, *Ann. Statist.* **9** (1981), 1229–1238.
7. X. Hu and F. T. Wright, Monotonicity properties of the power functions of likelihood radio tests for normal mean hypotheses constrained by linear space and a cone, *Ann. Statist.* **22** (1994), 1547–1554.
8. N. Mukhopadhyay, A consistent and asymptotically efficient two stage procedure to construct fixed width confidence intervals for the mean, *Metrika* **27** (1980), 281–284.
9. M. D. Perlman, One-sided problems in multivariate analysis, *Ann. Math. Statist.* **40** (1969), 549–567.

10. D. Von Rosen, Moments for the inverted Wishart distribution, *Scand. J. Statist.* **15** (1988), 97–109.
11. S. N. Roy, On a heuristic method of test construction and its use in multivariate analysis, *Ann. Math. Statist.* **24** (1953), 220–238.
12. A. Shapiro, Towards a unified theory of inequality constrained testing in multivariate analysis, *Internat. Statist. Rev.* **56** (1988), 49–62.
13. V. Siskind, Second moments of inverse Wishart-matrix elements, *Biometrika* **59** (1972), 690–691.
14. T. A. B. Snijders, "Asymptotic Optimality Theory for Testing Problems with Restricted Alternatives," Math. Centrum, Amsterdam, 1979.
15. C. Stein, A two-sample test for a linear hypothesis whose power function is independent of σ, *Ann. Math. Statist.* **16** (1945), 243–258.
16. M. T. Tsai, On the power superiority of likelihood ratio tests for restricted alternatives, *J. Multivariate Anal.* **42** (1992), 102–109.
17. M. T. Tsai, UI score tests for some restricted alternatives in exponential families, *J. Multivariate Anal.* **45** (1993), 305–323.
18. Y. Wang and M. P. McDermott, Conditional likelihood ratio test for a nonnegative normal mean vector, *J. Amer. Statist. Assoc.* **93** (1998), 380–386.