On the Feasibility of Parsimonious Variable Selection for Hotelling's T^2-test*

Michael D. Perlman†
University of Washington

October 8, 2019

Abstract

Hotelling's T^2-test for the mean of a multivariate normal distribution is one of the triumphs of classical multivariate analysis. It is uniformly most powerful among invariant tests, and admissible, proper Bayes, and locally and asymptotically minimax among all tests. Nonetheless, investigators often prefer non-invariant tests, especially those obtained by selecting only a small subset of variables from which the T^2-statistic is to be calculated, because such reduced statistics are more easily interpretable for their specific application. Thus it is relevant to ask the extent to which power is lost when variable selection is limited to very small subsets of variables, e.g. of size one (yielding univariate Student-t^2 tests) or size two (yielding bivariate T^2-tests). This study presents some evidence, admittedly fragmentary and incomplete, suggesting that in some cases no power may be lost over a wide range of alternatives.

*Key words: Multivariate normal distribution, mean vector, covariance matrix, hypothesis test, power function, Hotelling's T^2, Student's t^2, variable selection, parsimony, test for additional information.

†mdperlma@uw.edu.

This work is dedicated to the memory of Somesh Das Gupta, my colleague, teacher, and friend.

1. Introduction. This study is motivated by a re-examination of the variable-selection problem for Hotelling's T^2-test (closely related to variable selection for linear discriminant analysis). After some notational preliminaries in §1.1, Hotelling's T^2 is reviewed in §1.2. The variable-selection problem is described in §1.3, where the substance of this investigation is described.

1.1. The noncentral f-distribution. Let $\chi^2_m(\lambda)$ denote a noncentral chi-square random variable with m degrees of freedom and noncentrality parameter $\lambda > 0$. The noncentral $f_{m,n}(\lambda)$ distribution (nonnormalized) with m and n degrees of freedom and noncentrality parameter $\lambda > 0$ is the distribution of the ratio $\chi^2_m(\lambda)/\chi^2_n$ (also denoted by $f_{m,n}(\lambda)$), where the numerator and denominator are independent chi-square random variables and $\chi^2_n \equiv \chi^2_n(0)$. The upper α-quantile of $f_{m,n} \equiv f_{m,n}(0)$ is denoted by $f^\alpha_{m,n}$, so that

$$\Pr[f_{m,n} > f^\alpha_{m,n}] = \alpha. \tag{1}$$

The *noncentral $f_{m,n}$-test of size* $\alpha \geq 0$ for the problem of testing $\lambda = 0$ vs. $\lambda > 0$ has power function given by

$$\pi_\alpha(\lambda; m, n) = \Pr[f_{m,n}(\lambda) > f^\alpha_{m,n}] \tag{2}$$

$$= e^{-\frac{\lambda}{2}} \sum_{k=0}^{\infty} (\tfrac{\lambda}{2})^k \tfrac{1}{k!} c_{m,n;k;\alpha}, \tag{3}$$

$$c_{m,n;k;\alpha} \equiv \Pr[f_{m+2k,n} > f^\alpha_{m,n}]; \tag{4}$$

see Das Gupta and Perlman [DGP] (1974), eqn.(2.1). Clearly $\pi_\alpha(\lambda; m, n)$ is decreasing in α, with $\pi_0(\lambda; m, n) = 0$. Because $f_{m,n}(\lambda)$ has strictly monotone likelihood ratio in λ, $\pi_\alpha(\lambda; m, n)$ is strictly increasing in λ.

It will be convenient to work with the (central) *beta distribution* $b_{m,n}$:

$$b_{m,n} := \frac{f_{m,n}}{f_{m,n} + 1} \equiv \frac{\chi^2_m}{\chi^2_m + \chi^2_n}; \tag{5}$$

clearly, $b_{m,n} = 1 - b_{n,m}$. The upper and lower α-quantiles of $b_{m,n}$ are denoted by $b^\alpha_{m,n}$ and $b_{m,n;\alpha}$, respectively, so that

$$b^\alpha_{m,n} = 1 - b_{n,m;\alpha}. \tag{6}$$

Thus by (4),

$$c_{m,n;k;\alpha} = \Pr[b_{n,m+2k} < b_{n,m;\alpha}]. \tag{7}$$

The probability density function of $b_{m,n}$ is given by

$$\phi_{m,n}(b) \equiv \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} b^{\frac{m}{2}-1}(1-b)^{\frac{n}{2}-1}, \qquad 0 < b < 1. \tag{8}$$

1.2. Hotelling's T^2-test. Let $X_i : p \times 1$, $i = 1, \ldots, N$ $(N \geq p+1)$ be a random sample from the p-dimensional multivariate normal distribution $N_p(\mu, \Sigma)$, where μ $(p \times 1) \equiv (\mu_1, \ldots, \mu_p)' \in \mathbb{R}^p$ and Σ $(p \times p) \equiv (\sigma_{ij})$ is positive definite. The problem of testing

$$H_0 : \mu = 0 \quad \text{vs.} \quad K : \mu \neq 0 \tag{9}$$

with Σ unknown is invariant under the group action $X_i \to AX_i$, $i = 1, \ldots, N$, where $A \in GL(p)$, the group of all nonsingular $p \times p$ matrices. The maximal invariant statistic under $GL(p)$ is given by Hotelling's T^2 statistic:

$$T^2 := N\bar{X}'S^{-1}\bar{X}, \tag{10}$$

where $\bar{X} = \sum_{i=1}^N X_i$ and $S = \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})'$. Its distribution is

$$T^2 \sim f_{p,N-p}(\Lambda), \tag{11}$$

where

$$\Lambda = N\mu'\Sigma^{-1}\mu, \tag{12}$$

the maximal invariant parameter. Therefore the uniformly most powerful invariant size-α test rejects H_0 if $T^2 > f_{p,N-p;\alpha}$, with power function $\pi_\alpha(\Lambda; p, N-p)$; cf. [A] Theorem 5.6.1).[1]

It is useful to express Λ in terms of scale-free parameters, that is,

$$\Lambda = N\gamma' R^{-1}\gamma, \tag{13}$$

[1] Among the larger class of all tests, invariant or non-invariant, the T^2 test is admissible [S], proper Bayes [KS], and locally and asymptotically minimax for small and large values of Λ, respectively [GK].

where $R \equiv (\rho_{ij})$ is the $p \times p$ correlation matrix determined by Σ and

$$(14) \qquad \gamma\ (p \times 1) \equiv (\gamma_1, \ldots, \gamma_p)' := \left(\tfrac{\mu_1}{\sqrt{\sigma_{11}}}, \ldots, \tfrac{\mu_p}{\sqrt{\sigma_{pp}}}\right)'.$$

The testing problem (9) can be stated equivalently as that of testing

$$(15) \qquad H_0 : \gamma = 0 \quad \text{vs.} \quad K : \gamma \neq 0$$

with R unknown.

1.3. The T^2 variable-selection problem. Denote the components of $\bar{X}$ by $\bar{X}_j$, $j = 1, \ldots, p$, and those of S by s_{jk}, $j, k = 1, \ldots, p$. Let Ω_p be the collection of all nonempty subsets of $\{1, \ldots, p\}$. For $\omega \in \Omega_p$ denote the ω-subvector of $\bar{X}$ by $\bar{X}_\omega$, the ω-submatrix of S by S_ω, and similarly define γ_ω and R_ω. The T^2-statistic based on $(\bar{X}_\omega, S_\omega)$ is given by

$$(16) \qquad T^2_\omega \equiv N \bar{X}'_\omega S^{-1}_\omega \bar{X}_\omega \sim f_{|\omega|,\, N-|\omega|}(\Lambda_\omega),$$

$$(17) \qquad \Lambda_\omega \equiv \Lambda_\omega(\gamma, R) \;=\; N \gamma'_\omega R^{-1}_\omega \gamma_\omega.$$

($T^2_{\Omega_p} = T^2$, $\Lambda_{\Omega_p} = \Lambda \equiv \Lambda(\gamma, R)$.) The test that rejects H_0 if $T^2_\omega > f_{|\omega|, N-|\omega|; \alpha}$ has size α for H_0, and its power function is given by

$$(18) \qquad \pi_\alpha(\Lambda_\omega;\ |\omega|,\ N - |\omega|).$$

This T^2_ω-test is not invariant under $GL(p)$ but it is admissible for testing H_0 vs. K, being a unique proper Bayes test for a prior distribution that, under K, assigns mass 1 to $\{\mu \mid \mu_{\Omega_p \setminus \omega} = 0\}$; cf. [KS], [MP].

The T^2 variable-selection problem is that of choosing a parsimonious subset ω such that the T^2_ω-test maintains high power against alternatives deemed likely in some sense. Because (γ, R) is unknown, variable selection in practice is traditionally approached by forward and/or backward selection procedures based on a preliminary sample that yields estimates of (γ, R); see the Appendix. In general, all $2^p - 1$ nonempty subsets ω must be considered.

Recently I consulted on such a variable-selection problem. The investigator, a web-page designer, had observed 20 physiological variables (blood pressure, temperature, heart rate, etc.) on each of 100 subjects (the numbers are approximate). He wished to compare their responses to a new web-page design with their responses to the current design. The overall T^2-statistic, based on a linear combination of all 20 variables, indicated a significant difference between the two sets of responses. However, the client desired to find a more readily interpretable measure of difference, namely a T^2_ω-statistic based on a very small subset ω of the 20 variables, hopefully with $|\omega| = 1$ or 2. Such a desire is not atypical of investigators presented with a multivariate data analysis. Therefore it occurred to me to wonder just how much power would be lost by restricting variable selection to small subsets ω, for example to single variables or pairs of variables.

To state this more precisely, define

$$\hat{\omega}_\alpha(\gamma, R) = \arg\max_{\omega \in \Omega_p} \pi_\alpha(\Lambda_\omega(\gamma, R);\ |\omega|,\ N - |\omega|). \tag{19}$$

Thus $\hat{\omega}_\alpha(\gamma, R)$ is the (not necessarily unique) subset ω of variables that maximizes the power of the size-α T^2_ω-test to detect the alternative (γ, R) if the actual value of (γ, R) were revealed by an oracle. Whereas the admissibility of the overall size-α T^2-test dictates that its power cannot be dominated by that of the size-α T^2_ω-test when $\omega \neq \Omega_p$, might it happen that $|\hat{\omega}_\alpha(\gamma, R)|$ is small, perhaps 1 or 2, over a fairly wide range of parameter values (γ, R)? If so, then one might with some confidence limit variable selection to consideration of single variables (univariate t^2-tests) or pairs of variables (bivariate T^2-tests) – thus the overall p-variate T^2-test would be ruled out *a priori*.

Of course, such a radical suggestion flies in the face of 100 years of multivariate statistical theory, of which I have been but one of many proponents. Nonetheless, this report presents some evidence, admittedly fragentary and incomplete, indicating that this suggestion might not be entirely inappropriate in applications. In Sections 2, 3, and 4, a few very special cases are considered where tractable algebraic expressions for the asymptotic ($\Lambda_\omega \to \infty$), local ($\Lambda_\omega \to 0$), and/or exact values of $\pi_\alpha(\Lambda_\omega; |\omega|, N - |\omega|)$ are available. These in turn can be utilized to compare the powers of T^2_ω and T^2.

Examples 2.1 and 3.1 treat only the simplest possible case: the bivariate case ($p = 2$) with $N = 3$.[2] Here it is shown that $|\hat{\omega}_\alpha(\gamma, R)| = 1$ over large portions of the asymptotic and local regions of the alternative hypothesis K. This implies that the power of at least one of the two univariate Student t^2-tests ($|\omega| = 1$) exceeds that of the overall (bivariate) T^2-test for most alternatives (γ, R) in these regions.

[2] However, Giri, Kiefer, and Stein [GKS] also began their study of Hotelling's T^2 test by considering only this simplest case $p = 2$, $N = 3$.

In Example 4.4 this result is extended to the entire alternative hypothesis K, both for $N = 3$ and $N = 5$, but only under the highly restrictive and impracticably vague condition that α be sufficiently small, with "sufficiently small" determined by the value of the unknown noncentrality parameter – see (125) and (130).

Examples 2.2 and 3.2 go beyond the bivariate case. Here $p \geq 3$, $N = p+2$, and the powers of all possible bivariate T^2_ω-tests ($|\omega| = 2$) are compared to the power of the overall p-variate T^2 test, again only for asymptotic and local alternatives and, furthermore, only for very special configurations of γ and R. In these cases, admittedly highly restrictive, the bivariate T^2_ω-tests dominate the overall T^2 over a substantial portion of the alternative hypothesis K. This does not establish that $|\hat{\omega}_\alpha(\gamma, R)| = 2$ but again suggests that variable search might be limited to small variable subsets ω.

Of course, a much more comprehensive study is needed to confirm the efficacy of such an approach to variable selection. It is hoped that this report might encourage such an investigation.

2. Some asymptotic power comparisons. The power function of the T^2_ω-test is

$$\pi_\alpha(\Lambda_\omega) \equiv \pi_\alpha(\Lambda_\omega; |\omega|, N - |\omega|) \tag{20}$$

(recall (16)-(17)). It follows from eqn. (3.4) in [MP] that as $\Lambda_\omega \to \infty$,

$$\begin{aligned} \pi_\alpha(\Lambda_\omega) &\sim 1 - \exp[-\tfrac{\Lambda_\omega}{2}(f^\alpha_{|\omega|,N-|\omega|} + 1)^{-1}] \\ &= 1 - \exp[-\tfrac{\Lambda_\omega}{2} b_{N-|\omega|,|\omega|;\alpha}]. \end{aligned} \tag{21}$$

Thus for two subsets ω, ω' s.t. $\omega \subset \omega'$, $\exists\ \Lambda^*_{|\omega|,|\omega'|,N;\alpha} > 0$ s.t.

$$\Lambda_\omega > \max\left(\Lambda^*_{|\omega|,|\omega'|,N;\alpha},\ \tfrac{b_{N-|\omega'|,|\omega'|;\alpha}}{b_{N-|\omega|,|\omega|;\alpha}}\Lambda_{\omega'}\right) \ \Rightarrow\ \pi_\alpha(\Lambda_\omega) > \pi_\alpha(\Lambda_{\omega'}). \tag{22}$$

Therefore power comparisons of T^2_ω and $T^2_{\omega'}$ for distant alternatives[3] require determination of the lower quantiles $b_{n,m;\alpha}$. This can be done explicitly in Examples 2.1 and 2.2 below. Although these examples are of very limited scope[4] they begin to suggest that variable subset selection sometimes can be

[3] We do not claim to know the values of $\Lambda^*_{|\omega|,|\omega'|,N;\alpha}$, even approximately.

[4] But see Footnote 2.

limited to very small subsets $\omega \in \Omega_p$, e.g., singletons in the bivariate Example 2.1, or pairs (including singletons) in Example 2.2.

To simplify the notation, set

$$Q_{|\omega|,|\omega'|,N;\alpha} := \frac{b_{N-|\omega'|,|\omega'|;\alpha}}{b_{N-|\omega|,|\omega|;\alpha}} < 1. \tag{23}$$

The quantile $b_{n,m;\alpha}$ satisfies

$$\alpha = \frac{\Gamma\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \int_0^{b_{n,m;\alpha}} b^{\frac{n}{2}-1}(1-b)^{\frac{m}{2}-1} db \tag{24}$$

For the simple cases $n = 2$ or $m = 2$,

$$b_{2,m;\alpha} = 1 - (1-\alpha)^{\frac{2}{m}}, \quad b_{n,2;\alpha} = \alpha^{\frac{2}{n}}. \tag{25}$$

Example 2.1. In the bivariate case $p = 2$, abbreviate the singleton subsets $\{1\}$ and $\{2\}$ of Ω_2 by 1 and 2 respectively. We shall compare the powers $\pi_\alpha(\Lambda_1;\, 1,\, 2)$ and $\pi_\alpha(\Lambda_2;\, 1,\, 2)$ of the two univariate size-α t^2-tests to the power $\pi_\alpha(\Lambda;\, 2,\, 1)$ of the overall (bivariate) size-α T^2-test for distant alternatives.

Assume that $\gamma_1 \neq 0$ (recall (14)) and set

$$\eta = \tfrac{\gamma_2}{\gamma_1}, \qquad \rho = \rho_{12}, \tag{26}$$

where $-1 < \rho < 1$, so by (13) and (17),

$$\Lambda_1 = N\gamma_1^2, \quad \Lambda_2 = N\eta^2\gamma_1^2, \quad \Lambda = N\big(\tfrac{1-2\eta\rho+\eta^2}{1-\rho^2}\big)\gamma_1^2. \tag{27}$$

Without loss of generality we can assume that $|\gamma_1| \geq |\gamma_2|$, so $0 \leq \eta^2 \leq 1$ and $\max(\Lambda_1, \Lambda_2) = \gamma_1^2$. Thus the alternative hypotheses K can be represented as

$$K = \{(\gamma_1, \eta, \rho) \mid |\gamma_1| > 0,\, |\eta| \leq 1,\, |\rho| < 1\}, \tag{28}$$

while $\hat{\omega}_\alpha(\gamma, R)$ can be re-expressed as $\hat{\omega}_\alpha(\gamma_1, \eta, \rho)$.

Because $\max(\Lambda_1, \Lambda_2) \leq \Lambda$, it follows from (22) and (27) that

$$\gamma_1^2 > \max\Big(\tfrac{1}{N}\Lambda^*_{1,2,N;\alpha},\ Q_{1,2,N;\alpha}\big(\tfrac{1-2\eta\rho+\eta^2}{1-\rho^2}\big)\gamma_1^2\Big) \tag{29}$$

$$\Rightarrow\ \max(\pi_\alpha(\Lambda_1; 1, 2), \pi_\alpha(\Lambda_2; 1, 2)) > \pi_\alpha(\Lambda; 2, 1). \tag{30}$$

$$\Rightarrow |\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1. \tag{31}$$

In the simplest case $N = 3$, (25) yields the explicit expression

$$Q_{1,2,3;\alpha} = \tfrac{b_{1,2;\alpha}}{b_{2,1;\alpha}} = \tfrac{\alpha}{2-\alpha}, \textit{ while} \tag{32}$$

the inequality in (29) is equivalent to

$$1 > \max\left(\tfrac{\Lambda^*_{1,2,3;\alpha}}{3\gamma_1^2},\ Q_{1,2,3;\alpha}\left(\tfrac{1-2\eta\rho+\eta^2}{1-\rho^2}\right)\right). \tag{33}$$

Note that

$$1 > Q_{1,2,3;\alpha}\left(\tfrac{1-2\eta\rho+\eta^2}{1-\rho^2}\right) \tag{34}$$

$$\iff 0 > \rho^2 - 2Q_{1,2,3;\alpha}\eta\rho + Q_{1,2,3;\alpha}(1+\eta^2) - 1 =: h_{\alpha,\eta}(\rho). \tag{35}$$

The quadratic function $h_{\alpha,\eta}(\rho)$ $(-1 \le \rho \le 1)$ satisfies

$$\begin{aligned} h_{\alpha,\eta}(-1) &= Q_{1,2,3;\alpha}(1+\eta)^2 \ge 0, \\ h_{\alpha,\eta}(1) &= Q_{1,2,3;\alpha}(1-\eta)^2 \ge 0, \\ h_{\alpha,\eta}(0) &= Q_{1,2,3;\alpha}(1+\eta^2) - 1. \end{aligned}$$

It is easily seen that if $\alpha \le \frac{2}{3}$ then $Q_{1,2,3;\alpha} \le \frac{1}{2}$, so $h_{\alpha,\eta}(0) \le 0$ for all $\eta \in [-1,1]$. Thus if $\alpha \le \frac{2}{3}$ then $h_{\alpha,\eta}(\rho)$ must have one root in $[-1,0]$ and one root in $[0,1]$. The two roots are given by

$$\hat\rho^{\pm}_{\alpha,\eta} = Q_{1,2,3;\alpha}\eta \pm \sqrt{(1-Q_{1,2,3;\alpha})(1-Q_{1,2,3;\alpha}\eta^2)}; \tag{36}$$

note that $\hat\rho^{\pm}_{\alpha,-\eta} = -\hat\rho^{\mp}_{\alpha,\eta}$.

It follows that if $\alpha \le \frac{2}{3}$ then for sufficiently large γ_1^2, i.e., $\gamma_1^2 \ge \frac{1}{3}\Lambda^*_{1,2,3;\alpha}$,

$$\rho \in (\hat\rho^-_{\alpha,\eta},\ \hat\rho^+_{\alpha,\eta}) \Rightarrow |\hat\omega_\alpha(\gamma_1,\eta,\rho)| = 1, \tag{37}$$

that is, at least one of the two univariate t^2-tests is more powerful than the overall (bivariate) T^2-test. Specifically, when $\gamma_1^2 \ge \frac{1}{3}\Lambda^*_{1,2,3;\alpha}$, $|\hat\omega_\alpha(\gamma_1,\eta,\rho)| = 1$ in the (η,ρ)-regions of the parameter space indicated in Table 1. From this it is seen that for $p = 2$, $N = 3$, large γ_1^2, and the common (small) values of α, the overall size-α T^2-test is dominated by at least one of the two univariate size-α t^2-tests over a large portion of the alternative hypothesis K. □

α	$Q_{1,2,3;\alpha}$	η	$(\hat{\rho}^-_{\alpha,\eta},\ \hat{\rho}^+_{\alpha,\eta})$
.50	.333	1	(-.333, 1)
		.5	(-.615, .948)
		0	(-.816, .816)
.20	.111	1	(-.778, 1)
		.5	(-.874, .985)
		0	(-.943, .943)
.10	.0526	1	(-.895, 1)
		.5	(-.941, .993)
		0	(-.973, .973)
.05	.0256	1	(-.949, 1)
		.5	(-.971, .997)
		0	(-.987, .987)
.01	.00503	1	(-.990, 1)
		.5	(-.994, .999)
		0	(-.997, .997)
0+	0	1	(-1, 1)
		.5	(-1, 1)
		0	(-1, 1)

Table 1: For $p = 2$, $N = 3$, and sufficiently large $\gamma_1^2 \equiv \max(\Lambda_1, \Lambda_2)$ (i.e., $\gamma_1^2 \geq \frac{1}{3}\Lambda^*_{1,2,3;\alpha}$), if $\rho \in (\hat{\rho}^-_{\alpha,\eta},\ \hat{\rho}^+_{\alpha,\eta})$ then $|\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1$, i.e., the power of at least one of the two univariate size-α t^2-tests dominates that of the bivariate size-α T^2-test. (Note that $\hat{\rho}^\pm_{\alpha,-\eta} = -\hat{\rho}^\mp_{\alpha,\eta}$.)

Example 2.2. Suppose that $p \geq 3$ and $N = p + 2$. The powers of the $\binom{p}{2}$ bivariate size-α T^2-tests and the overall p-variate size-α T^2-test will be compared for distant alternatives, which requires comparison of the powers

$$\{\pi(\Lambda_\omega; 2, p) \mid \omega \in \Omega_p,\ |\omega| = 2\} \quad \text{and} \quad \pi(\Lambda; p,\ 2). \tag{38}$$

From (22),

$$\begin{aligned} \Lambda^{(2)} :&= \max\{\Lambda_\omega \mid \omega \in \Omega_p,\ |\omega| = 2\} > \max\left(\Lambda^*_{2,p,p+2;\alpha},\ Q_{2,p,p+2;\alpha}\Lambda\right) \\ &\Rightarrow\ \max\{\pi(\Lambda_\omega; 2, p) \mid \omega \in \Omega_p,\ |\omega| = 2\} > \pi(\Lambda; p,\ 2). \end{aligned}$$

Therefore for sufficiently large values of $\Lambda^{(2)}$, namely $\Lambda^{(2)} \geq \Lambda^*_{2,p,p+2;\alpha}$, at least one of the bivariate size-α T^2-tests will be more powerful than the overall (p-variate) size-α T^2-test[5] provided that

$$\Lambda^{(2)} > Q_{2,p,p+2;\alpha}\Lambda. \tag{39}$$

From (25) we obtain the explicit expression

$$Q_{2,p,p+2;\alpha} = \frac{b_{2,p;\alpha}}{b_{p,2;\alpha}} = \frac{1-(1-\alpha)^{\frac{2}{p}}}{\alpha^{\frac{2}{p}}}. \tag{40}$$

If we set $\nu_p = \frac{2}{p}$ and $U_{p;\alpha} = \frac{\nu_p}{Q_{2,p,p+2;\alpha}}$ then

$$\lim_{p\to\infty} Q_{2,p,p+2;\alpha} = 0, \tag{41}$$

$$\lim_{p\to\infty} U_{p;\alpha} = \frac{1}{-\log(1-\alpha)} \quad (> 1 \ \text{ for } \ \alpha < \tfrac{e-1}{e} = .6321); \tag{42}$$

$$\lim_{p\to\infty} (p-1)U_{p;\alpha} = \infty. \tag{43}$$

Table 2 shows that $Q_{2,p,p+2;\alpha}$ decreases rapidly to 0 as $p \to \infty$, which suggests that (39) might hold over substantial regions of the alternative hypothesis K. We proceed to exhibit several such regions.

Case 1: $\gamma_1 = \cdots = \gamma_p =: \delta$ and R has the intraclass form

$$R_\rho := (1-\rho)I_p + \rho\mathbf{e}_p\mathbf{e}_p'. \tag{44}$$

Here $-\frac{1}{p-1} < \rho < 1$ and $\mathbf{e}_p = (1,\ldots,1)' : p \times 1$. Then

$$R_\rho^{-1} = \frac{1}{1-\rho}\left[I_p - \frac{\rho\mathbf{e}_p\mathbf{e}_p'}{1+\rho(p-1)}\right], \tag{45}$$

$$\mathbf{e}_p' R_\rho^{-1}\mathbf{e}_p = \frac{p}{1+\rho(p-1)}, \tag{46}$$

so by (13) and (17),

$$\Lambda^{(2)} = \frac{2(p+2)\delta^2}{1+\rho}, \qquad \Lambda = \frac{p(p+2)\delta^2}{1+(p-1)\rho}. \tag{47}$$

Thus $\Lambda^{(2)} \geq \Lambda^*_{2,p,p+2;\alpha}$ holds for all feasible ρ if $\delta^2 \geq (p+2)^{-1}\Lambda^*_{2,p,p+2;\alpha}$. Also, if we set $\nu_p = \frac{2}{p}$ ($\leq \frac{2}{3}$) then (39) is equivalent to each of the inequalities

$$U_{p;\alpha} > \frac{1+\rho}{1+(p-1)\rho};$$

$$[(p-1)U_{p;\alpha} - 1]\,\rho > 1 - U_{p;\alpha}. \tag{48}$$

[5] Note that by itself this does not establish that $|\hat{\omega}_\alpha(\gamma, R)| = 2$.

p	α	$Q_{2,p,p+2;\alpha}$	$U_{p;\alpha}$	$\tilde{\psi}^-_{p;\alpha}$	$(\tilde{\rho}^-_{p;\alpha},\, \tilde{\rho}^+_{p;\alpha})$	$-\frac{1}{p-1}$
4	.20	.236	2.118	-.209	(-.304, .776)	-.333
	.10	.162	3.081	-.245	(-.314, .844)	
	.05	.113	4.415	-.279	(-.320, .890)	
	.01	.050	9.975	-.310	(-.328, .951)	
10	.20	.0602	3.321	-.080	(-.110, .907)	-.111
	.10	.0330	6.052	-.094	(-.110, .948)	
	.05	.0186	10.764	-.102	(-.111, .973)	
	.01	$.0^2504$	39.652	-.109	(-.111, .980)	
20	.20	.0259	3.858	-.039	(-.0525, .954)	-.0526
	.10	.0132	7.579	-.046	(-.0526, .976)	
	.05	$.0^2690$	14.486	-.049	(-.0526, .988)	
	.01	$.0^2159$	62.810	-.052	(-.0526, .997)	
40	.20	.0120	4.158	-.020	(-.0256, .977)	-.0256
	.10	$.0^2590$	8.481	-.023	(-.0256, .989)	
	.05	$.0^2298$	16.806	-.024	(-.0256, .994)	
	.01	$.0^3632$	79.055	-.025	(-.0256, .999)	
∞	.20	0	4.481	0	$[0, 1)$	0
	.10	0	9.491	0	$[0, 1)$	
	.05	0	19.496	0	$[0, 1)$	
	.01	0	99.499	0	$[0, 1)$	

Table 2: Take $p \geq 3$, $N = p + 2$, and δ^2 sufficiently large, that is, $\delta^2 \geq (p+2)^{-1}\Lambda^*_{2,p,p+2;\alpha}$. In Case 1 (Case 2) if $\tilde{\psi}^-_{p;\alpha} < \rho < 1$ ($\tilde{\rho}^-_{p;\alpha} < \rho < \tilde{\rho}^+_{p;\alpha}$) then the power of at least one of the $\binom{p}{2}$ bivariate size-α T^2-tests dominates that of the overall size-α T^2-test. Note that the feasible range of ρ is $\big(-\frac{1}{p-1}, 1\big)$.

Because $(p-1)U_{p;\alpha} > 1$ for common (small) values of α (see (43) and Table 2), in such cases (48) in turn is equivalent to

$$\rho > \tfrac{1-U_{p;\alpha}}{(p-1)U_{p;\alpha}-1} =: \tilde{\psi}^-_{p;\alpha}. \tag{49}$$

Table 2 shows that in Case 1, $\tilde{\psi}^-_{p;\alpha}$ is close to the lower limit of the feasible range $(-\frac{1}{p-1}, 1)$ for ρ. Thus by (49), if $\delta^2 \geq (p+2)^{-1}\Lambda^*_{2,p,p+2;\alpha}$ then at least one of the bivariate size-α T^2-tests will be more powerful than the overall (p-variate) size-α T^2-test for most of the region in the alternative hypothesis specified in Case 1.

Case 2: $\gamma_i = \gamma_j =: \delta$ *for some* $\{i,j\} \subset \{1,\ldots,p\}$, $\gamma_k = 0$ *for* $k \neq i,j$, *and* R *has the intraclass form* R_ρ *in* (44). By (13) and (17),

$$\Lambda^{(2)} = \tfrac{2(p+2)\delta^2}{1+\rho}, \qquad \Lambda = \tfrac{2(p+2)\delta^2[1+(p-3)\rho]}{(1-\rho)[1+(p-1)\rho]}, \tag{50}$$

so again $\Lambda^{(2)} \geq \Lambda^*_{2,p,p+2;\alpha}$ holds for all feasible ρ if $\delta^2 \geq (p+2)^{-1}\Lambda^*_{2,p,p+2;\alpha}$. Also, abbreviating $Q_{2,p,p+2;\alpha}$ by Q, (39) is equivalent to each of the inequalities

$$\tfrac{1}{1+\rho} > \tfrac{Q[1+(p-3)\rho]}{(1-\rho)[1+(p-1)\rho]};$$
$$0 > [(p-1)+(p-3)Q]\rho^2 - (p-2)(1-Q)\rho - (1-Q) =: h_{p;\alpha}(\rho).$$

Since $h_{p;\alpha}(0) = Q - 1 < 0$ for common (small) values of α (see (41)-(42) and Table 2), $h_{p;\alpha}(\rho)$ has two real roots $\tilde{\rho}^-_{p;\alpha} < 0 < \tilde{\rho}^+_{p;\alpha}$ (found numerically). Therefore $0 > h_{p;\alpha}(\rho)$ for $\tilde{\rho}^-_{p;\alpha} < \rho < \tilde{\rho}^+_{p;\alpha}$.

Table 2 shows that in Case 2, the interval $(\tilde{\rho}^-_{p;\alpha}, \tilde{\rho}^+_{p;\alpha})$ covers almost all of the feasible range $(-\frac{1}{p-1}, 1)$ for ρ. Thus if $\delta^2 \geq (p+2)^{-1}\Lambda^*_{2,p,p+2;\alpha}$ then at least one of the bivariate size-α T^2-tests will be more powerful than the overall p-variate size-α T^2-test for most of the region in the alternative hypothesis specified by Case 2.

Case 3: $\gamma_i = \delta$ *and* $\gamma_j = -\delta$ *for some* $\{i,j\} \subset \{1,\ldots,p\}$, $\gamma_k = 0$ *for* $k \neq i,j$, *and* R *has the intraclass form* R_ρ. By (13) and (17),

$$\Lambda^{(2)} = \tfrac{2(p+2)\delta^2}{1-\rho}, \qquad \Lambda = \tfrac{2(p+2)\delta^2}{1-\rho}. \tag{51}$$

Thus $\Lambda^{(2)} \geq \Lambda^*_{2,p,p+2;\alpha}$ again holds for all feasible ρ if $\delta^2 \geq (p+2)^{-1}\Lambda^*_{2,p,p+2;\alpha}$, while (39) is equivalent to $1 > Q_{2,p,p+2;\alpha}$, which holds for most p, α (see (41)-(42) and Table 2). Therefore if $\delta^2 \geq (p+2)^{-1}\Lambda^*_{2,p,p+2;\alpha}$ then at least one of the bivariate size-α T^2-tests will be more powerful than the overall p-variate size-α T^2-test for the *entire* region in the alternative hypothesis covered by Case 3.

Case 4: $p =: 2l$ *is even,* $\gamma_i = \delta$ *for* l *indices in* $\{1,\ldots,p\}$, $\gamma_i = -\delta$ *for the remaining* l *indices, and* R *has the intraclass form* R_ρ. By (13) and (17),

$$\Lambda^{(2)} = \tfrac{2(p+2)\delta^2}{1-|\rho|}, \qquad \Lambda = \tfrac{p(p+2)\delta^2}{(1-\rho)}. \tag{52}$$

Thus $\Lambda^{(2)} \geq \Lambda^*_{2,p,p+2;\alpha}$ again holds for all feasible ρ if $\delta^2 \geq (p+2)^{-1}\Lambda^*_{2-2,p+2;\alpha}$, while (39) is equivalent to the inequality

$$U_{p;\alpha} > \tfrac{1-|\rho|}{1-\rho}. \tag{53}$$

Because $\frac{1-|\rho|}{1-\rho} \leq 1$, while $U_{p;\alpha} > 1$ for holds for most p, α (see (42) and Table 2), we see that if $\delta^2 \geq (p+2)^{-1}\Lambda^*_{2,p,p+2;\alpha}$ then at least one of the bivariate size-α T^2-tests is more powerful than the overall p-variate size-α T^2-test over the *entire* region in the alternative hypothesis in Case 4. □

3. Some local power comparisons. From (2)-(4) and (7), as $\Lambda_\omega \downarrow 0$ the power function $\pi_\alpha(\Lambda_\omega) \equiv \pi_\alpha(\Lambda_\omega; |\omega|, N-|\omega|)$ of the T^2_ω-test satisfies

$$\pi_\alpha(\Lambda_\omega) = e^{-\frac{\Lambda_\omega}{2}}[\alpha + \tfrac{\Lambda_\omega}{2}c_{|\omega|,N-|\omega|;1;\alpha} + O(\Lambda^2_\omega)] \tag{54}$$

$$= \alpha + \tfrac{\Lambda_\omega}{2}(c_{|\omega|,N-|\omega|;1;\alpha} - \alpha) + O(\Lambda^2_\omega). \tag{55}$$

Thus for two subsets ω, ω' s.t. $\omega \subset \omega'$, $\exists\ \Lambda^{**}_{|\omega|,|\omega'|,N;\alpha} > 0$ s.t.

$$\Lambda_{\omega'} < \min\left(\Lambda^{**}_{|\omega|,|\omega'|,N;\alpha},\ Z_{|\omega|,|\omega'|,N;\alpha}\Lambda_\omega\right) \ \Rightarrow\ \pi_\alpha(\Lambda_\omega) > \pi_\alpha(\Lambda_{\omega'}), \tag{56}$$

where, from (2.2) and (2.3) in [DGP],

$$Z_{|\omega|,|\omega'|,N;\alpha} := \tfrac{c_{|\omega|,N-|\omega|;1;\alpha}-\alpha}{c_{|\omega'|,N-|\omega'|;1;\alpha}-\alpha} > 1. \tag{57}$$

Therefore power comparisons of T^2_ω and $T^2_{\omega'}$ for local alternatives[6] require determination of the lower tail probabilities $c_{m,n;k;\alpha}$ (see (7)), which in turn require the lower quantiles $b_{n,m;\alpha}$.

In parallel with Section 2, this is done explicitly in Examples 3.1 and 3.2. As in Examples 2.1 and 2.2 these examples begin to suggest that variable selection might be limited to very small subsets $\omega \in \Omega_p$, e.g., singletons in the bivariate Example 3.1, or pairs (plus singletons) in Example 3.2.

Example 3.1. As in Example 2.1 consider the bivariate case $p = 2$. Repeat the first two paragraphs from Example 2.1 verbatim, except replace "distant alternatives" by "local alternatives". Because $\max(\Lambda_1, \Lambda_2) \leq \Lambda$, it follows

[6] We do not claim to know the values of $\Lambda^{**}_{|\omega|,|\omega'|,N;\alpha}$, even approximately.

from (27) and (56) that

$$\left(\tfrac{1-2\eta\rho+\eta^2}{1-\rho^2}\right)\gamma_1^2 < \min\left(\tfrac{1}{N}\Lambda^{**}_{1,2,N;\alpha},\ Z_{1,2,N;\alpha}\gamma_1^2\right) \tag{58}$$

$$\Rightarrow\ \max(\pi_\alpha(\Lambda_1;1,2),\pi_\alpha(\Lambda_2;1,2)) > \pi_\alpha(\Lambda;2,1), \tag{59}$$

$$\Rightarrow |\hat{\omega}_\alpha(\gamma_1,\eta,\rho)| = 1. \tag{60}$$

In the simplest case $N = 3$, it follows from (7), (8), and (25) that

$$Z_{1,2,3;\alpha} = \tfrac{c_{1,2;1;\alpha}-\alpha}{c_{2,1;1;\alpha}-\alpha} = \tfrac{1-(1-\alpha)^3-\alpha}{\frac{3}{2}(\alpha-\frac{1}{3}\alpha^3)-\alpha} = \tfrac{2(2-\alpha)}{1+\alpha}. \tag{61}$$

First note that in (58),

$$\tfrac{1-2\eta\rho+\eta^2}{1-\rho^2} < Z_{1,2,3;\alpha} \tag{62}$$

$$\iff\ h_{\alpha,\eta}(\rho) := Z_{1,2,3;\alpha}\rho^2 - 2\eta\rho + \eta^2 + 1 - Z_{1,2,3;\alpha} < 0. \tag{63}$$

The quadratic function $h_{\alpha,\eta}(\rho)$ $(-1 \le \rho \le 1)$ satisfies

$$\begin{aligned} h_{\alpha,\eta}(-1) &= (1+\eta)^2 \ge 0,\\ h_{\alpha,\eta}(1) &= (1-\eta)^2 \ge 0,\\ h_{\alpha,\eta}(0) &= \eta^2 + 1 - Z_{1,2,3;\alpha}, \end{aligned}$$

It is easily seen that if $\alpha \le \frac{1}{2}$ then $Z_{1,2,3;\alpha} \ge 2$, so $h_{\alpha,\eta}(0) \le 0$ for all $\eta \in [-1,1]$. Therefore, if $\alpha \le \frac{1}{2}$ then $h_{\alpha,\eta}(\rho)$ must have one root in $[-1,0]$ and one root in $[0,1]$. The two roots are given by

$$\check{\rho}^{\pm}_{\alpha,\eta} = \tfrac{\eta}{Z_{1,2,3;\alpha}} \pm \tfrac{1}{Z_{1,2,3;\alpha}}\sqrt{(Z_{1,2,3;\alpha}-1)(Z_{1,2,3;\alpha}-\eta^2)}; \tag{64}$$

again $\check{\rho}^{\pm}_{\alpha,-\eta} = -\check{\rho}^{\mp}_{\alpha,\eta}$. Thus, if $\alpha \le \frac{1}{2}$ and $\rho \in (\check{\rho}^{-}_{\alpha,\eta},\ \check{\rho}^{+}_{\alpha,\eta})$ then (62) must hold.

To conclude that $|\hat{\omega}_\alpha(\gamma_1,\eta,\rho)| = 1$, γ_1^2 must be sufficiently small, i.e.,

$$0 < \gamma_1^2 < \left(\tfrac{1-\rho^2}{1-2\eta\rho+\eta^2}\right)\tfrac{\Lambda^{**}_{1,2,3;\alpha}}{3}. \tag{65}$$

Because

$$\min_{|\eta|\le 1}\left(\tfrac{1-\rho^2}{1-2\eta\rho+\eta^2}\right) = \tfrac{1}{2}(1-|\rho|),$$

for fixed η, (65) will be satisfied provided that

$$\rho \in (\check{\rho}^{-}_{\alpha,\eta},\ \check{\rho}^{+}_{\alpha,\eta}), \tag{66}$$

$$\check{m}_{\alpha,\eta} := \max(|\check{\rho}^{-}_{\alpha,\eta}|,\ |\check{\rho}^{+}_{\alpha,\eta}|) < 1, \tag{67}$$

$$\gamma_1^2 < \tfrac{1}{6}(1-m_{\alpha,\eta})\Lambda^{**}_{1,2,3;\alpha}. \tag{68}$$

It is straightforward to show that (67) holds for $|\eta| < 1$ but not for $|\eta| = 1$.

Thus, if $\alpha \leq \frac{1}{2}$, $|\eta| < 1$, and (66)-(68) are satisfied then $|\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1$, in which case at least one of the two univariate t^2-tests are more powerful than the overall (bivariate) T^2-test. This occurs in the (η, ρ)-regions of the parameter space indicated in Table 3, provided that $\gamma_1^2 < \frac{1}{6}(1 - \check{m}_{\alpha,\eta})\Lambda^{**}_{1,2,3;\alpha}$. Thus, for $p = 2$, $N = 3$, sufficiently small γ_1^2, and the common (small) values of α, the overall (bivariate) size-α T^2-test is dominated by at least one of the two univariate size-α t^2-tests over a large portion of the alternative hypothesis. □

α	$Z_{1,2,3;\alpha}$	η	$(\check{\rho}^-_{\alpha,\eta},\ \check{\rho}^+_{\alpha,\eta})$	$1 - \check{m}_{\alpha,\eta}$
.50	2	.9	(-.095, .9954)	.0046
		.5	(-.411, .911)	.089
		0	(-.707, .707)	.293
.20	3	.9	(-.398, .9976)	.0024
		.5	(-.615, .948)	.052
		0	(-.816, .816)	.184
.10	3.454	.9	(-.477, .9980)	.0020
		.5	(-.667, .957)	.043
		0	(-.843, .843)	.157
.05	3.714	.9	(-.514, .9982)	.0018
		.5	(-.691, .962)	.038
		0	(-.855, .855)	.145
.01	3.941	.9	(-.542, .9983)	.0017
		.5	(-.709, .963)	.037
		0	(-.864, .864)	.136
0+	4	.9	(-.548, .9984)	.0016
		.5	(-.714, .964)	.036
		0	(-.866, .866)	.134

Table 3: For $p = 2$, $N = 3$, and sufficiently small $\gamma_1^2 \equiv \max(\Lambda_1, \Lambda_2)$ (i.e., $\gamma_1^2 < \frac{1}{6}(1 - \check{m}_{\alpha,\eta})\Lambda^{**}_{1,2,3;\alpha}$), if $\rho \in (\check{\rho}^-_{\alpha,\eta},\ \check{\rho}^+_{\alpha,\eta})$ then $|\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1$, i.e., the power of at least one of the two univariate size-α t^2-tests dominates that of the bivariate size-α T^2-test. (Note that $\check{\rho}^\pm_{\alpha,-\eta} = -\check{\rho}^\mp_{\alpha,\eta}$.)

Example 3.2. Suppose that $p \geq 3$ and $N = p + 2 \geq 3$. We shall compare the powers of the $\binom{p}{2}$ bivariate size-α T^2-tests and the overall p-variate size-α T^2-test for local alternatives, which requires comparison of the powers

$$\{\pi(\Lambda_\omega; 2, p) \mid \omega \in \Omega_p,\ |\omega| = 2\} \quad \text{and} \quad \pi(\Lambda; p,\ 2). \tag{69}$$

From (56),

$$\Lambda < \min(\Lambda^{**}_{2,p,p+2;\alpha},\ Z_{2,p,p+2;\alpha}\Lambda^{(2)}) \tag{70}$$
$$\Rightarrow\ \max\{\pi(\Lambda_\omega; 2, p) \mid \omega \in \Omega_p,\ |\omega| = 2\} > \pi(\Lambda; p,\ 2). \tag{71}$$

Therefore for sufficiently small values of Λ, namely $\Lambda \leq \Lambda^{**}_{2,p,p+2;\alpha}$, at least one of the bivariate size-α T^2-tests will be more powerful than the overall p-variate size-α T^2-test[7] whenever

$$\Lambda < Z_{2,p,p+2;\alpha}\Lambda^{(2)}. \tag{72}$$

From (57), (7)-(8), (25), and some algebra, the explicit expression

$$Z_{2,p,p+2;\alpha} \equiv \frac{c_{2,p;1;\alpha} - \alpha}{c_{p,2;1;\alpha} - \alpha} = \frac{p\alpha[1-\alpha^{\frac{2}{p}}]}{2(1-\alpha)[1-(1-\alpha)^{\frac{2}{p}}]} \tag{73}$$

is obtained. Setting $\nu_p = \frac{2}{p}$ $(\leq \frac{2}{3})$ and $V_{p;\alpha} := \nu_p Z_{2,p,p+2;\alpha}$, we have

$$\lim_{p\to\infty} Z_{2,p,p+2;\alpha} = \infty, \tag{74}$$
$$\lim_{p\to\infty} V_{p;\alpha} = \frac{\alpha \log \alpha}{(1-\alpha)\log(1-\alpha)} \quad (> 1 \ \text{ for } \ \alpha < \tfrac{1}{2}), \tag{75}$$
$$\lim_{p\to\infty} (p-1)V_{p;\alpha} = \infty. \tag{76}$$

Table 4 shows that $Z_{2,p,p+2;\alpha}$ increases rapidly to ∞ as $p \to \infty$, which suggests that (72) might hold over substantial regions of the alternative hypothesis. Several such regions are now exhibited.

Case 1: $\gamma_1 = \cdots = \gamma_p =: \delta$ *and* R *has the intraclass form* (44). Here $-\frac{1}{p-1} < \rho < 1$ and as in (47),

$$\Lambda^{(2)} = \frac{2(p+2)\delta^2}{1+\rho}, \qquad \Lambda = \frac{p(p+2)\delta^2}{1+(p-1)\rho}. \tag{77}$$

---

[7] As in Example 2.2, this does not establish that $|\hat{\omega}_\alpha(\gamma, R)| = 2$.

Here (72) is equivalent to each of the inequalities

$$V_{p;\alpha} > \tfrac{1+\rho}{1+(p-1)\rho};$$

(78) $$V_{p;\alpha} - 1 > -\left[(p-1)V_{p;\alpha} - 1\right]\rho.$$

Because $(p-1)V_{p;\alpha} > 1$ for common (small) values of α (see (76) and Table 4), in such cases (78) in turn is equivalent to

(79) $$\breve{\psi}^{-}_{p;\alpha} := -\tfrac{V_{p;\alpha}-1}{(p-1)V_{p;\alpha}-1} < \rho < -1.$$

To conclude that (70)-(71) holds, δ^2 must be sufficiently small, i.e.,

(80) $$0 < \delta^2 < \left[\tfrac{1+(p-1)\rho}{p(p+2)}\right]\Lambda^{**}_{2,p,p+2;\alpha}.$$

However, $\rho > \breve{\psi}^{-}_{p;\alpha}$ implies that

(81) $$\tfrac{1+(p-1)\rho}{p(p+2)} > \tfrac{1+(p-1)\breve{\psi}^{-}_{p;\alpha}}{p(p+2)} = \tfrac{p-2}{p(p+2)[(p-1)V_{p;\alpha}-1]} := \breve{m}_{p,\alpha}.$$

Therefore (80) will be satisfied provided that

(82) $$\rho > \breve{\psi}^{-}_{p;\alpha} \quad \text{and} \quad \delta^2 < \breve{m}_{p,\alpha}\Lambda^{**}_{2,p,p+2;\alpha}.$$

If p is large and α is small, Table 4 shows that in Case 1, $\breve{\psi}^{-}_{p;\alpha}$ is close to the lower limit of the feasible range $(-\tfrac{1}{p-1}, 1)$ for ρ. Here, by (82), if $\delta^2 < \breve{m}_{p,\alpha}\Lambda^{**}_{2,p,p+2;\alpha}$ then at least one of the bivariate size-α T^2-tests will be more powerful than the overall (p-variate) size-α T^2-test for most of the region in the alternative hypothesis covered by Case 1.

Case 2: $\gamma_i = \gamma_j =: \delta$ for some $\{i, j\} \subset \{1, \ldots, p\}$, $\gamma_k = 0$ for $k \neq i, j$, and R has the intraclass form R_ρ in (44). As in (50),

$$\Lambda^{(2)} = \tfrac{2(p+2)\delta^2}{1+\rho}, \qquad \Lambda = \tfrac{2(p+2)\delta^2[1+(p-3)\rho]}{(1-\rho)[1+(p-1)\rho]},$$

Abbreviating $Z_{2,p,p+2;\alpha}$ by Z, (72) is equivalent to each of the inequalities

$$\tfrac{Z}{1+\rho} > \tfrac{[1+(p-3)\rho]}{(1-\rho)[1+(p-1)\rho]};$$
$$0 > [(p-1)Z + (p-3)]\rho^2 - (p-2)(Z-1)\rho - (Z-1) =: h_{p;\alpha}(\rho).$$

Since $h_{p;\alpha}(0) = 1 - Z < 0$ (cf. (57)), $h_{p;\alpha}(\rho)$ has real roots $\breve{\rho}^-_{p;\alpha} < 0 < \breve{\rho}^+_{p;\alpha}$ (found numerically). Therefore $0 > h_{p;\alpha}(\rho)$ for $\breve{\rho}^-_{p;\alpha} < \rho < \breve{\rho}^+_{p;\alpha}$.

To conclude that (70)-(71) holds, δ^2 must be sufficiently small, i.e.,

$$0 < \delta^2 < \left\{ \tfrac{(1-\rho)[1+(p-1)\rho]}{2(p+2)[1+(p-3)\rho]} \right\} \Lambda^{**}_{2,p,p+2;\alpha}. \tag{83}$$

Because $\frac{(1-\rho)[1+(p-1)\rho]}{1+(p-3)\rho}$ is decreasing in ρ, $\rho < \breve{\rho}^+_{p;\alpha}$ implies that

$$\tfrac{(1-\rho)[1+(p-1)\rho]}{2(p+2)[1+(p-3)\rho]} > \tfrac{(1-\breve{\rho}^+_{p;\alpha})[1+(p-1)\breve{\rho}^+_{p;\alpha}]}{2(p+2)[1+(p-3)\breve{\rho}^+_{p;\alpha}]} := \breve{m}'_{p,\alpha}. \tag{84}$$

Therefore (83) will be satisfied provided that

$$\rho \in (\breve{\rho}^-_{p;\alpha},\, \breve{\rho}^+_{p;\alpha}) \quad \text{and} \quad \delta^2 < \breve{m}'_{p,\alpha} \Lambda^{**}_{2,p,p+2;\alpha}. \tag{85}$$

Table 4 shows that in Case 2, the interval $(\breve{\rho}^-_{p;\alpha}, \breve{\rho}^+_{p;\alpha})$ covers almost all of the feasible range $(-\frac{1}{p-1}, 1)$ for ρ. Thus if $\delta^2 < \breve{m}'_{p,\alpha} \Lambda^{**}_{2,p,p+2;\alpha}$ then at least one of the bivariate size-α T^2-tests will be more powerful than the overall p-variate size-α T^2-test for most of the region in the alternative hypothesis determined by Case 2.

Case 3: $\gamma_i = \delta$ and $\gamma_j = -\delta$ for some $\{i, j\} \subset \{1, \ldots, p\}$, $\gamma_k = 0$ for $k \neq i, j$, and R has the intraclass form R_ρ. By (13) and (17),

$$\Lambda^{(2)} = \tfrac{2(p+2)\delta^2}{1-\rho}, \qquad \Lambda = \tfrac{2(p+2)\delta^2}{1-\rho}.$$

Here (72) is equivalent to $Z_{2,p,p+2;\alpha} > 1$, which holds for all p, α (see (57)).

To conclude that (70)-(71) holds, δ^2 must be sufficiently small, i.e.,

$$\delta^2 < \left(\tfrac{1-\rho}{2(p+2)}\right) \Lambda^{**}_{2,p,p+2;\alpha} \tag{86}$$

for all $\rho \in (-\frac{1}{p-1}, 1)$. This requires that ρ be bounded below 1, that is, $\rho < 1 - \epsilon$ for some $\epsilon > 0$, whence (86) will be satisfied if

$$\delta^2 < \left(\tfrac{\epsilon}{2(p+2)}\right) \Lambda^{**}_{2,p,p+2;\alpha}. \tag{87}$$

Therefore if (87) holds then at least one of the bivariate size-α T^2-tests will be more powerful than the overall p-variate size-α T^2-test if $\rho < 1 - \epsilon$, which covers almost all of the region in the alternative hypothesis determined by Case 3.

Case 4: $p =: 2l$ is even, $\gamma_i = \delta$ for l indices in $\{1, \ldots, p\}$, $\gamma_i = -\delta$ for the remaining l indices, and R has the intraclass form R_ρ. By (13) and (17),

$$\Lambda^{(2)} = \tfrac{2(p+2)\delta^2}{1-|\rho|}, \qquad \Lambda = \tfrac{p(p+2)\delta^2}{(1-\rho)}.$$

Here (72) is equivalent to the inequality

$$V_{p;\alpha} > \tfrac{1-|\rho|}{1-\rho}. \tag{88}$$

Because $\frac{1-|\rho|}{1-\rho} \leq 1$, while $V_{p;\alpha} > 1$ for holds for most p, α (see (75) and Table 4), (88) is satisfied for most p, α.

To conclude that (70)-(71) holds, δ^2 must be sufficiently small, i.e.,

$$\delta^2 < \left[\tfrac{1-\rho}{p(p+2)}\right]\Lambda^{**}_{2,p,p+2;\alpha} \tag{89}$$

for all $\rho \in (-\frac{1}{p-1}, 1)$. This again requires that ρ be bounded below 1, that is, $\rho < 1 - \epsilon$ for some $\epsilon > 0$, whence (89) will be satisfied if

$$\delta^2 < \left[\tfrac{\epsilon}{p(p+2)}\right]\Lambda^{**}_{2,p,p+2;\alpha} \tag{90}$$

Therefore if (90) holds then at least one of the bivariate size-α T^2-tests will be more powerful than the overall p-variate size-α T^2-test if $\rho < 1 - \epsilon$, which covers almost all of the region in the alternative hypothesis determined by Case 4. □

p	α	$Z_{2,p,p+2;\alpha}$	$V_{p;\alpha}$	$\breve{\psi}^{-}_{p;\alpha}$	$\breve{m}_{p,\alpha}$	$(\breve{\rho}^{-}_{p;\alpha}, \breve{\rho}^{+}_{p;\alpha})$	$\breve{m}'_{p,\alpha}$	$-\frac{1}{p-1}$
4	.20	2.618	1.309	-.106	.0285	(-.282, .648)	.0524	-.333
	.10	2.961	1.480	-.140	.0242	(-.289, .686)	.0475	
	.05	3.228	1.614	-.160	.0217	(-.295, .710)	.0442	
	.01	3.627	1.814	-.183	.0118	(-.298, .741)	.0400	
10	.20	7.882	1.576	-.044	$.0^2506$	(-.108, .815)	$.0^2958$	-.111
	.10	9.832	1.966	-.058	$.0^2399$	(-.109, .849)	$.0^2783$	
	.05	11.621	2.324	-.066	$.0^2335$	(-.109, .871)	$.0^2669$	
	.01	15.138	3.028	-.077	$.0^2254$	(-.110, .899)	$.0^2630$	
20	.20	16.842	1.684	-.022	$.0^2132$	(-.0523, .898)	$.0^2257$	-.0526
	.10	21.804	2.180	-.029	$.0^2101$	(-.0524, .921)	$.0^2199$	
	.05	26.630	2.663	-.034	$.0^3825$	(-.0524, .935)	$.0^2164$	
	.01	37.109	3.711	-.039	$.0^3588$	(-.0525, .953)	$.0^2119$	
40	.20	34.844	1.742	-.011	$.0^3338$	(-.0256, .947)	$.0^2105$	-.0256
	.10	45.995	2.300	-.015	$.0^3255$	(-.0256, .959)	$.0^3514$	
	.05	57.168	2.858	-.017	$.0^3205$	(-.0256, .967)	$.0^3414$	
	.01	90.567	4.528	-.020	$.0^3129$	(-.0256, .979)	$.0^3263$	
∞	.20	∞	1.803	0	0	$[0, 1)$	0	0
	.10	∞	2.429	0	0	$[0, 1)$	0	
	.05	∞	3.074	0	0	$[0, 1)$	0	
	.01	∞	4.628	0	0	$[0, 1)$	0	

Table 4: Suppose that $p \geq 3$, $N = p + 2$, and δ^2 is sufficiently small, that is, $\delta^2 < \breve{m}_{p,\alpha}\Lambda^{**}_{2,p,p+2;\alpha}$ ($\delta^2 < \breve{m}'_{p,\alpha}\Lambda^{**}_{2,p,p+2;\alpha}$). In Case 1 (Case 2) if $\breve{\psi}^{-}_{p;\alpha} < \rho < 1$ ($\breve{\rho}^{-}_{p;\alpha} < \rho < \breve{\rho}^{+}_{p;\alpha}$) then the power of at least one of the $\binom{p}{2}$ bivariate size-α T^2-tests dominates that of the overall p-variate size-α T^2-test. Note that the feasible range of ρ is $\left(-\frac{1}{p-1}, 1\right)$.

4. Some exact power comparisons for the bivariate case. The results in Sections 2 and 3 compare the power of the overall T^2-test with those of univariate or bivariate T^2-tests based on the original variates. However these power comparisons are asymptotic or local, relevant only for noncentrality parameters Λ that approach ∞ or 0. In this section we consider the bivariate case $p = 2$ and attempt to compare the exact power functions of the T^2-test and the two univariate t^2-tests for all values of Λ. Two conjectures are presented; the first of these is confirmed in Proposition 4.3 and applied in

Example 4.4 for only two simple cases.

Conjecture 4.1 (weak). Suppose that $p = 2$ and N is odd: $N = 2l + 1$. Then for each $\lambda > 0$, $\exists\, 0 < \alpha_l^*(\lambda) < 1$ such that

$$0 < \alpha < \alpha_l^*(\lambda) \quad \Rightarrow \quad \pi_\alpha(\lambda;\, 1,\, 2l) > \pi_\alpha\Big(\big(\tfrac{4l}{2l-1}\big)\lambda;\, 2,\, 2l-1\Big), \tag{91}$$

with equality when $\alpha = \alpha_l^*(\lambda)$. □

Conjecture 4.1 is established below for $l = 1, 2$ and we expect it to hold for all $l \geq 3$ as well. However, it is unsatisfactory in that if $\alpha_l^*(\lambda)$ depends nontrivially on λ then we cannot conclude that, at least for small α, one or both of the two univariate size-α t^2-tests dominate the bivariate size-α T^2 test in a large region of the alternative hypothesis. For this the following stronger result would be needed.

Conjecture 4.2 (strong). Conjecture 4.1 holds with $\alpha_l^*(\lambda)$ not depending on λ, i.e., $\alpha_l^*(\lambda) = \alpha_l^*$. □

At this time we do not have evidence either for or against Conjecture 4.2. If valid, it would be essential to determine or approximate the values of α_l^*.

Proposition 4.3. Conjecture 4.1 is valid for $l = 1$ and 2.

Proof. By (3),

$$\pi_\alpha(\lambda;\, 1,\, 2l) > \pi_\alpha(2(1+\delta)\lambda;\, 2,\, 2l-1) \tag{92}$$

if and only if

$$s_{\delta,\lambda}^{(l)}(\alpha) := e^{(\frac{1}{2}+\delta)\lambda} \textstyle\sum_{k=0}^\infty (\frac{\lambda}{2})^k \frac{1}{k!} c_{1,2l;k;\alpha} > \sum_{k=0}^\infty \frac{(1+\delta)^k \lambda^k}{k!} c_{2,2l-1;k;\alpha} =: t_{\delta,\lambda}^{(l)}(\alpha). \tag{93}$$

From (7) and (8) we find that

$$c_{1,2l;k;\alpha} = \Pr[b_{2l,1+2k} < b_{2l,1;\alpha}] \tag{94}$$

$$= \tfrac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(l)\Gamma(\frac{1}{2}+k)} \int_0^{b_{2l,1;\alpha}} b^{l-1}(1-b)^{k-\frac{1}{2}} db; \tag{95}$$

$$\alpha = \Pr[b_{2l,1} < b_{2l,1;\alpha}] \tag{96}$$

$$= \tfrac{\Gamma(l+\frac{1}{2})}{\Gamma(l)\Gamma(\frac{1}{2})} \int_0^{b_{2l,1;\alpha}} b^{l-1}(1-b)^{-\frac{1}{2}} db. \tag{97}$$

Set $u = 1 - b$ in (96)-(97), then differentiate w.r.to α to obtain

$$\frac{d}{d\alpha} b_{2l,1;\alpha} = \frac{\Gamma(l)\Gamma(\frac{1}{2})}{\Gamma(l+\frac{1}{2})} \frac{(1-b_{2l,1;\alpha})^{\frac{1}{2}}}{b_{2l,1;\alpha}^{l-1}}; \tag{98}$$

$$\frac{d}{d\alpha} c_{1,2l;k,\alpha} = \frac{\Gamma(l+\frac{1}{2}+k)\Gamma(\frac{1}{2})}{\Gamma(l+\frac{1}{2})\Gamma(\frac{1}{2}+k)}(1 - b_{2l,1;\alpha})^k \tag{99}$$

$$\frac{d}{d\alpha} s_{\delta,\lambda}^{(l)}(\alpha) = e^{(\frac{1}{2}+\delta)\lambda} \sum_{k=0}^{\infty} (\tfrac{\lambda}{2})^k \frac{1}{k!} \frac{\Gamma(l+\frac{1}{2}+k)\Gamma(\frac{1}{2})}{\Gamma(l+\frac{1}{2})\Gamma(\frac{1}{2}+k)} (1 - b_{2l,1;\alpha})^k \tag{100}$$

$$= \Big[\sum_{k=0}^{\infty} \frac{(\frac{1}{2}+\delta)^k \lambda^k}{k!} \Big] \Big[\sum_{k=0}^{\infty} (\tfrac{\lambda}{2})^k \frac{1}{k!} \frac{\Gamma(l+\frac{1}{2}+k)\Gamma(\frac{1}{2})}{\Gamma(l+\frac{1}{2})\Gamma(\frac{1}{2}+k)} (1 - b_{2l,1;\alpha})^k \Big] \tag{101}$$

$$= \sum_{k=0}^{\infty} \frac{(\frac{1}{2}+\delta)^k}{k!} \Big[\sum_{r=0}^{k} \binom{k}{r} \frac{1}{(1+2\delta)^r} \frac{\Gamma(l+\frac{1}{2}+r)\Gamma(\frac{1}{2})}{\Gamma(l+\frac{1}{2})\Gamma(\frac{1}{2}+r)} (1 - b_{2l,1;\alpha})^r \Big] \lambda^k. \tag{102}$$

Next,

$$\alpha = \Pr[b_{2l-1,2} < b_{2l-1,2;\alpha}] \tag{103}$$

$$= \frac{\Gamma(l+\frac{1}{2})}{\Gamma(l-\frac{1}{2})\Gamma(1)} \int_0^{b_{2l-1,2;\alpha}} b^{l-\frac{3}{2}} db = b_{2l-1,2;\alpha}^{l-\frac{1}{2}}, \tag{104}$$

$$b_{2l-1,2;\alpha} = \alpha^{\frac{2}{2l-1}}; \tag{105}$$

$$c_{2,2l-1;k;\alpha} = \Pr[b_{2l-1,2+2k} < \alpha^{\frac{2}{2l-1}}] \tag{106}$$

$$= \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(l-\frac{1}{2})\Gamma(1+k)} \int_0^{\alpha^{\frac{2l}{2l-1}}} b^{l-\frac{3}{2}} (1-b)^k db \tag{107}$$

$$= \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(l+\frac{1}{2})\Gamma(1+k)} \int_0^{\alpha} (1 - w^{\frac{2l}{2l-1}})^k dw; \tag{108}$$

$$\frac{d}{d\alpha} c_{2,2l-1;k,\alpha} = \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(l+\frac{1}{2})\Gamma(1+k)} (1 - \alpha^{\frac{2l}{2l-1}})^k; \tag{109}$$

$$\frac{d}{d\alpha} t_{\delta,\lambda}^{(l)}(\alpha) = \sum_{k=0}^{\infty} \frac{(1+\delta)^k \lambda^k}{k!} \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(l+\frac{1}{2})\Gamma(1+k)} (1 - \alpha^{\frac{2l}{2l-1}})^k. \tag{110}$$

Therefore a sufficient condition that $\frac{d}{d\alpha} s_{\delta,\lambda}^{(l)}(\alpha) > \frac{d}{d\alpha} t_{\delta,\lambda}^{(l)}(\alpha)$ is that for all $k \geq 0$,

$$\frac{(\frac{1}{2}+\delta)^k}{(1+\delta)^k (1-\alpha^{\frac{2l}{2l-1}})^k} \sum_{r=0}^{k} \binom{k}{r} \frac{(1-b_{2l,1;\alpha})^r}{(1+2\delta)^r} \frac{\Gamma(l+\frac{1}{2}+r)}{\Gamma(\frac{1}{2}+r)} \geq \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}, \tag{111}$$

with strict inequality for at least one k.

Thus for $\alpha = 0$, a sufficient condition that $\frac{d}{d\alpha} s_{\delta,\lambda}^{(l)}(\alpha = 0) > \frac{d}{d\alpha} t_{\delta,\lambda}^{(l)}(\alpha = 0)$ is that for all $k \geq 0$,

$$\frac{(\frac{1}{2}+\delta)^k}{(1+\delta)^k} \sum_{r=0}^{k} \binom{k}{r} \frac{1}{(1+2\delta)^r} \frac{\Gamma(l+\frac{1}{2}+r)}{\Gamma(\frac{1}{2}+r)} \geq \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}, \tag{112}$$

with strict inequality for at least one k. After some algebra, (112) can be written equivalently as

$$\mathrm{E}\left[\tfrac{\Gamma(l+\frac{1}{2}+R_{k,\delta})}{\Gamma(\frac{1}{2}+R_{k,\delta})}\right] \geq \tfrac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}, \tag{113}$$

where $R_{k,\delta} \sim \mathrm{Binomial}(k, \frac{1}{1+\delta})$. Because $R_{k,\delta}$ is (strictly) stochastically decreasing in δ (for $k \geq 1$) while $\frac{\Gamma(l+\frac{1}{2}+R_{k,\delta})}{\Gamma(\frac{1}{2}+R_{k,\delta})}$ is (strictly) increasing in $R_{k,\delta}$, the left side of (113) $\equiv$ (112) is (strictly) decreasing in δ (for $k \geq 1$).

For $k = 0$ both sides of (112) $= \frac{\Gamma(l+\frac{1}{2})}{\Gamma(\frac{1}{2})}$. For $k = 1$, (112) is equivalent to the inequality

$$\tfrac{(\frac{1}{2}+\delta)}{(1+\delta)}\left[\tfrac{\Gamma(l+\frac{1}{2})}{\Gamma(\frac{1}{2})} + \tfrac{1}{(1+2\delta)}\tfrac{\Gamma(l+\frac{3}{2})}{\Gamma(\frac{3}{2})}\right] \geq \tfrac{\Gamma(l+\frac{3}{2})}{\Gamma(\frac{1}{2})},$$

which is equivalent to $\delta \leq \frac{1}{2l-1}$. Therefore the sufficient condition (112) for

$$\tfrac{d}{d\alpha} s_{\delta,\lambda}^{(l)}(\alpha = 0) > \tfrac{d}{d\alpha} t_{\delta,\lambda}^{(l)}(\alpha = 0)$$

will be satisfied for all $\delta \leq \frac{1}{2l-1}$ if (112) holds for $\delta = \frac{1}{2l-1}$ for all $k \geq 2$, with strict inequality for at least one $k \geq 2$.

Because $s_{\delta,\lambda}^{(l)}(0) = t_{\delta,\lambda}^{(l)}(0) = 0$, it follows from (92)-(93) that (112), with strict inequality for some $k \geq 2$, is a sufficient condition that for each $\lambda > 0$, $\exists\, 0 < \alpha_l^*(\lambda) < 1$ such that

$$0 < \alpha < \alpha_l^*(\lambda) \quad \Rightarrow \quad \pi_\alpha(\lambda;\, 1,\, 2l) > \pi_\alpha\big(\big(\tfrac{4l}{2l-1}\big)\lambda;\, 2,\, 2l-1\big), \tag{114}$$

with equality when $\alpha = \alpha_l^*(\lambda)$.

For the simplest case $l = 1$ ($N = 3$), (112) with $\delta = \frac{1}{2l-1} = 1$ becomes

$$\big(\tfrac{3}{4}\big)^k \textstyle\sum_{r=0}^{k} \binom{k}{r} \frac{1}{3^r}(\frac{1}{2} + r) \geq \tfrac{\Gamma(\frac{3}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}, \tag{115}$$

which by (113) can be reduced to the equivalent form

$$1 + \tfrac{k}{2} \geq \tfrac{\Gamma(\frac{3}{2}+k)}{\Gamma(\frac{3}{2})\Gamma(1+k)}. \tag{116}$$

It is straightforward to verify (116) by induction on k, with strict inequality holding for large k because $\frac{\Gamma(\frac{3}{2}+k)}{\Gamma(1+k)} = O(k^{\frac{1}{2}})$. Therefore we conclude that (114) holds for $l = 1$; that is, for each $\lambda > 0$, $\exists\, 0 < \alpha_1^*(\lambda) < 1$ such that

$$0 < \alpha < \alpha_1^*(\lambda) \quad \Rightarrow \quad \pi_\alpha(\lambda;\, 1,\, 2) > \pi_\alpha(4\lambda;\, 2,\, 1), \tag{117}$$

with equality when $\alpha = \alpha_1^*(\lambda)$.

Next consider the case $l = 2$ ($N = 5$). With $\delta = \frac{1}{2l-1} = \frac{1}{3}$, (112) becomes

$$(118) \qquad \left(\tfrac{5}{8}\right)^k \sum_{r=0}^{k} \binom{k}{r} \left(\tfrac{3}{5}\right)^r (\tfrac{3}{2}+r)(\tfrac{1}{2}+r) \geq \frac{\Gamma(\frac{5}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)},$$

which can be reduced to the equivalent form

$$(119) \qquad \tfrac{3}{4} + \tfrac{9k}{8} + \tfrac{9k(k-1)}{64} \geq \frac{\Gamma(\frac{5}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)} = (\tfrac{3}{2}+k)(\tfrac{1}{2}+k)\frac{\Gamma(\frac{1}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}.$$

Interestingly, (119) holds with equality for $k = 2$ as well as for 0 and 1. Rewrite (119) in the equivalent form

$$(120) \qquad \frac{k(k-1)\cdots 2\cdot 1}{(k-\frac{1}{2})(k-\frac{1}{2})\cdots\frac{3}{2}\cdot\frac{1}{2}} \geq \frac{48+128k+64k^2}{48+63k+9k^2}.$$

To verify (120) by induction on k, it suffices to show that for $k \geq 2$,

$$(121) \qquad \frac{k+1}{k+\frac{1}{2}}\,\frac{48+128k+64k^2}{48+63k+9k^2} \geq \frac{48+128(k+1)+64(k+1)^2}{48+63(k+1)+9(k+1)^2}.$$

After simplification, this is equivalent to the inequality

$$(122) \qquad 4k^3 + 4k^2 - 5k - 3 \geq 0,$$

which holds for all $k \geq 1$, with strict inequality for $k+1 \geq 3$. Therefore we conclude that (114) holds for $l = 2$: for each $\lambda > 0$, $\exists\, 0 < \alpha_2^*(\lambda) < 1$ such that

$$(123) \qquad 0 < \alpha < \alpha_2^*(\lambda) \quad \Rightarrow \quad \pi_\alpha(\lambda;\, 1,\, 4) > \pi_\alpha(\tfrac{8}{3}\lambda;\, 2,\, 3),$$

with equality when $\alpha = \alpha_2^*(\lambda)$. □

Example 4.4. Return to the bivariate Example 2.1, where $p = 2$ and

$$(124) \qquad \Lambda_1 = N\gamma_1^2, \quad \Lambda_2 = N\eta^2\gamma_1^2, \quad \Lambda = N\left(\tfrac{1-2\eta\rho+\eta^2}{1-\rho^2}\right)\gamma_1^2;$$

(recall (27)). For $N = 3$ it follows from (117) and (124) that for each $\gamma_1^2 > 0$,

$$(125) \qquad 0 < \alpha < \alpha_1^*(3\gamma_1^2) \quad \Rightarrow \quad \pi_\alpha(3\gamma_1^2;\, 1,\, 2) > \pi_\alpha(4\cdot 3\gamma_1^2;\, 2,\, 1).$$

Furthermore,

$$(126)\qquad 4\cdot 3\gamma_1^2 > 3\big(\tfrac{1-2\eta\rho+\eta^2}{1-\rho^2}\big)\gamma_1^2 \iff 4 > \tfrac{1-2\eta\rho+\eta^2}{1-\rho^2}$$

$$(127)\qquad \iff 0 > 4\rho^2 - 2\eta\rho + (\eta^2-3) := h_\eta(\rho).$$

The two roots of $h_\eta(\rho)$ are $\hat{\rho}_\eta^\pm = \frac{\eta\pm\sqrt{12-3\eta^2}}{4}$; note that $\hat{\rho}_{-\eta}^\pm = -\hat{\rho}_\eta^\mp$. Some values appear in Table 5. Thus, if $\alpha < \alpha_1^*(3\gamma_1^2)$ then

$$(128)\qquad \rho \in (\hat{\rho}_\eta^-,\, \hat{\rho}_\eta^+) \Rightarrow \max(\pi_\alpha(\Lambda_1;\, 1, 2),\, \pi_\alpha(\Lambda_2;\, 1, 2)) > \pi_\alpha(\Lambda;\, 2, 1)$$

$$(129)\qquad \Rightarrow |\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1;$$

that is, at least one of the two univariate t^2-tests is more powerful than the overall bivariate T^2-test. This occurs in the (η, ρ)-regions of the parameter space indicated in Table 5, which constitute a substantial part of the alternative hypothesis.

Similarly, for $N = 5$ it follows from (123) and (124) that for each $\gamma_1^2 > 0$,

$$(130)\qquad 0 < \alpha < \alpha_2^*(5\gamma_1^2) \;\Rightarrow\; \pi_\alpha(5\gamma_1^2;\, 1,\, 4) > \pi_\alpha(\tfrac{8}{3}\cdot 5\gamma_1^2;\, 2,\, 3).$$

Furthermore,

$$(131)\qquad \tfrac{8}{3}\cdot 5\gamma_1^2 > 5\big(\tfrac{1-2\eta\rho+\eta^2}{1-\rho^2}\big)\gamma_1^2 \iff \tfrac{8}{3} > \tfrac{1-2\eta\rho+\eta^2}{1-\rho^2}$$

$$(132)\qquad \iff 0 > \tfrac{8}{3}\rho^2 - 2\eta\rho + (\eta^2 - \tfrac{5}{3}) := \tilde{h}_\eta(\rho).$$

The two roots of $\tilde{h}_\eta(\rho)$ are $\tilde{\rho}_\eta^\pm = \frac{3\eta\pm\sqrt{40-15\eta^2}}{8}$; note that $\tilde{\rho}_{-\eta}^\pm = -\tilde{\rho}_\eta^\mp$. Some values appear in Table 5. Thus, if $\alpha < \alpha_2^*(5\gamma_1^2)$ then

$$(133)\qquad \rho \in (\tilde{\rho}_\eta^-,\, \tilde{\rho}_\eta^+) \Rightarrow \max(\pi_\alpha(\Lambda_1;\, 1, 4),\, \pi_\alpha(\Lambda_2;\, 1, 4)) > \pi_\alpha(\Lambda;\, 2, 3)$$

$$(134)\qquad \Rightarrow |\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1;$$

that is, at least one of the two univariate t^2-tests is more powerful than the overall bivariate T^2-test. Again this occurs in the (η, ρ)-regions of the parameter space indicated in Table 5.

Thus for $p = 2$, $N = 3$ or 5, and sufficiently small α (but depending on γ_1^2), the overall size-α T^2-test is dominated by at least one of the two univariate size-α t^2-tests over a fairly large portion of the alternative hypothesis. □

η	$(\hat{\rho}_\eta^-, \hat{\rho}_\eta^+)$	$(\tilde{\rho}_\eta^-, \tilde{\rho}_\eta^+)$
1	(-.5, 1)	(-.25, 1)
.75	(-.615, .990)	(-.421, .983)
.5	(-.714, .964)	(-.566, .940)
.25	(-.797, .922)	(-.687, .875)
0	(-.866, .866)	(-.791, .791)

Table 5: For $p = 2$, $N = 3$ ($N = 5$), and $\alpha < \alpha_1^*(3\gamma_1^2)$ ($\alpha < \alpha_2^*(5\gamma_1^2)$), if $\rho \in (\hat{\rho}_\eta^-, \hat{\rho}_\eta^+)$ ($\rho \in (\tilde{\rho}_\eta^-, \tilde{\rho}_\eta^+)$) then $|\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1$, so the power of at least one of the two univariate size-α t^2-tests exceeds that of the bivariate size-α T^2-test. (Note that $\hat{\rho}_{-\eta}^\pm = -\hat{\rho}_\eta^\mp$ and $\tilde{\rho}_{-\eta}^\pm = -\tilde{\rho}_\eta^\mp$.)

5. Concluding remarks. For the purpose of stimulating future research, the questions raised in this report are stated formally as follows:

The Oracular Variable-Selection Problem (OVSP) is that of determining the function $\hat{\omega}_\alpha(\gamma, R)$, as defined in (19), and using this to determine the regions

(135) $$A_\alpha(i) \equiv \big\{(\gamma, R) \;\big|\, |\hat{\omega}_\alpha(\gamma, R)| = i\big\}, \quad i = 1, ..., p.$$

The Parsimonious Variable-Selection Problem (PVSP) asks if $A_\alpha(i)$ comprises a substantial portion of the alternative hypothesis K for small values of i, e.g., $i = 1, 2$.

If the answer to the PVSP is positive, then variable selection in some applied investigations can be limited to small, easily interpretable subsets of variables.

Acknowledgment. Special thanks are due to David Perlman for raising the questions addressed here and providing some supporting data.

References

[A] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis, 3rd edition,* Wiley & Sons, New York.

[DGP] Das Gupta, S. and Perlman, M. D. (1974). On the power of the noncentral F-test: effect of additional variates on Hotelling's T^2-test. *J.Amer.*

Statist. Assoc. **69** 174-180.

[GK] Giri, N. and Kiefer, J. (1964). Local and asymptotic minimax properties of multivariate tests. *Ann. Math. Statist.* **35** 21-35.

[GKS] Giri, N., Kiefer, J., and Stein, C. (1963). Minimax character of Hotelling's T^2-test in the simplest case. *Ann. Math. Statist.* **34** 1524-1535.

[KS] Kiefer, J. and Schwartz, R. (1965). Admissible Bayes character of T^2-, R^2, and other fully invariant tests for classical multivariate normal testing problems. *Ann. Math. Statist.* **36** 747-770.

[MP] Marden, J. and Perlman, M. D. (1980). Invariant tests for means with covariates. *Ann. Statist.* **8** 25-63.

[S] Stein, C. (1956). The admissibility of Hotelling's T^2-test. *Ann. Math. Statist.* **27** 616-623.

Appendix

A. Testing for additional information. Variable selection for the T^2-test and related linear discriminant analysis was thoroughly studied in the 1970s and 1980s, an era of limited computer power, and subsequently by several authors with greater ability to consider all-subsets methods; a list of references appears below. Almost all of these studies were based on testing for additional information (= increased Mahalanobis distance), as now described.

For any two nested subsets $\omega \subset \omega'$ in Ω_p, in general $\Lambda_\omega \le \Lambda_{\omega'}$. The question of whether the power of the $T_{\omega'}$-test exceeds that of the T_ω-test for the testing problem (9) usually was formulated as the problem of *testing for additional information (TAI),* namely, testing

$$\Lambda'_\omega = \Lambda_\omega \quad \text{vs.} \quad \Lambda'_\omega > \Lambda_\omega \tag{136}$$

based on a preliminary sample – see [R] §8c.4. This formulation was adopted by many researchers, even while citing the following result of [DGP] which implies that this standard formulation of TAI is inappropriate.

It was shown in [DGP, Theorem 2.1] that for fixed $\lambda > 0$, the power function $\pi_\alpha(\lambda; m, n)$ (recall (2)) of the non-central f-test is strictly decreasing in m and strictly increasing in n.[8] Therefore for any integer $1 \leq q \leq n-1$ there exists a unique real number

$$g_\alpha(\lambda) \equiv g_\alpha(\lambda; m, n, q) > 0 \tag{137}$$

such that

$$\pi_\alpha(\lambda; m, n) = \pi_\alpha(\lambda + g_\alpha(\lambda); m+q, n-q). \tag{138}$$

Here $g_\alpha(0) = 0$ and $g_\alpha(\lambda)$ is strictly increasing in λ; cf. [DGP, Theorem 3.1]. Thus the power is increased only if

$$\Lambda_{\omega'} > \Lambda_\omega + g_\alpha(\Lambda_\omega; |\omega|, N-|\omega|, |\omega'\backslash\omega|). \tag{139}$$

Therefore [DGP, Section 4] introduced the problem of *testing for increased power (TIP),* namely, testing

$$\begin{aligned} H_1 &: \Lambda_{\omega'} \leq \Lambda_\omega + g_\alpha(\Lambda_{|\omega|}; |\omega|, N-|\omega|, |\omega'\backslash\omega|) \\ \text{vs.}\quad K_1 &: \Lambda_{\omega'} > \Lambda_\omega + g_\alpha(\Lambda_{|\omega|}; |\omega|, N-|\omega|, |\omega'\backslash\omega|), \end{aligned}$$

and proposed several (approximate) tests.

This proposal was noted by subsequent authors but never implemented for variable selection, possibly because of difficulties in computing the functions $g_\alpha(\cdot)$, especially if many pairs (ω, ω') must be considered. However, if as suggested above, variable selection might be limited to very small subsets of variables in practical applications, then replacing the TAI by the TIP might be feasible.

Remark A.1. The relation (91) in Conjecture 4.1 can be stated equivalently in terms of g_α:

$$0 < \alpha < \alpha_l^*(\lambda) \quad \Rightarrow \quad g_\alpha(\lambda; 1, 2l, 1) > \left(\tfrac{2l+1}{2l-1}\right)\lambda \quad \forall\ \lambda > 0, \tag{140}$$

with equality when $\alpha = \alpha_l^*(\lambda)$. Thus the relations (117) and (123) in Proposition 4.3 also can be stated equivalently in terms of g_α:

$$0 < \alpha < \alpha_1^*(\lambda) \quad \Rightarrow \quad g_\alpha(\lambda; 1, 2, 1) > 3\lambda \quad \forall\ \lambda > 0, \tag{141}$$

[8] In line 2 of the second column on p.179 of [DGP], "conclude" should be"include". In the line following the third display in the second column on p.179, "j" should be"f". In Remark 4.1 on p.180, "increasing in m" should be "decreasing in m". In the next line, "$m \to \infty$" should be "$n \to \infty$".

with equality when $\alpha = \alpha_1^*(\lambda)$;

$$0 < \alpha < \alpha_2^*(\lambda) \quad \Rightarrow \quad g_\alpha(\lambda; 1, 4, 1) > \tfrac{5}{3}\lambda \quad \forall\ \lambda > 0, \tag{142}$$

with equality when $\alpha = \alpha_2^*(\lambda)$.

Additional References for the Appendix

[H1] Hand, D. J. (1981). *Discrimination and Classification,* Wiley & Sons, New York. [Chapter 6]

[H2] Hawkins, D. M. (1976). The subset problem in multivariate analysis of variance. *J. Royal Statist. Soc, Series B* **38** 132-139.

[JW] Jain, A. K. and Waller, W. G. (1978). On the optimal number of features in the classification of multivariate Gaussian data. *Pattern Recognition* **10** 103-109.

[JWT] Jiang, W., Wang, K., and Tsung, F. (2012). A variable-selection-based multivariate EWMA chart for process monitoring and diagnosis. *J. Quality Technology* **44** 209-230.

[McC] McCabe, G. P. Jr. (1975). Computations for variable selection in discriminant analysis. *Technometrics* **17** 259-263.

[McK] McKay, R. J. (1976). A graphical aid to selection of variables in two-group discriminant analysis. *Appl. Statist.* **27** 259-263.

[McL1] McLachlan, G. J. (1976). On the relationship between the F test and the overall error rate for variable selection in two-group discriminant function. *Biometrics* **36** 501-510.

[McL2] McLachlan, G. J. (1980). A criterion for selecting variables for the linear discriminant function. *Biometrics* **32** 529-534.

[McL3] McLachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition,* Wiley & Sons, New York. [Chapter 12]

[Mu] Murray, G. D. (1977). A cautionary note on selection of variables in discriminant analysis. *Appl. Statist.* **26** 246-250.

[NT] Nobuo, S. and Takahisa, I. (2016). A variable selection method for detecting abnormality base on the T^2 test. *Comm. Statist. - Theory and Methods* **46** 501-510.

[R] Rao, C. R. (1973). *Linear Statistical Inference and its Application, 2nd edition,* Wiley & Sons, New York.

[S] Schaafsma, W. (1982). In *Handbook of Statistics, Vol. 2,* P. R. Krishnaiah and L. N. Kanal, eds., 857-881. North Holland, Amsterdam.