



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

Michael D. Perlman  
*University of Washington, Seattle, USA*

---

## 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 preliminary evidence suggesting that in some cases, no power may be lost, in fact may be gained, over a wide range of alternatives.

JEL Classification C12; C52.

*Keywords and phrases.* Multivariate normal distribution, hotelling's  $T^2$  test, student's  $t^2$ , variable selection, test for additional information, Mahalanobis distance.

---

## 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  $\chi_m^2(\lambda)$  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  $\chi_m^2(\lambda)/\chi_n^2$  (also denoted by  $f_{m,n}(\lambda)$ ), where the numerator and denominator are independent chi-square random variables and  $\chi_n^2 \equiv \chi_n^2(0)$ .

The upper  $\alpha$ -quantile of  $f_{m,n} \equiv f_{m,n}(0)$  is denoted by  $f_{m,n}^\alpha$ , so that

$$\Pr[f_{m,n} > f_{m,n}^\alpha] = \alpha. \quad (1)$$

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

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

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

$$c_{m,n;k}^\alpha := \Pr[f_{m+2k,n} > f_{m,n}^\alpha]; \quad (4)$$

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

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

$$b_{m,n} := \frac{f_{m,n}}{f_{m,n} + 1} \equiv \frac{\chi_m^2}{\chi_m^2 + \chi_n^2}, \quad (5)$$

whose probability density function (pdf) is given by

$$\phi_{m,n}(b) \equiv \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} b^{\frac{m}{2}-1} (1-b)^{\frac{n}{2}-1}, \quad 0 < b < 1. \quad (6)$$

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

$$b_{m,n}^\alpha = 1 - b_{n,m;\alpha}. \quad (7)$$

Later we shall need the following relation, obtained from Eqs. 4, 5, 6 and 7:

$$c_{m,n;k}^\alpha = \Pr[b_{n,m+2k} < b_{n,m;\alpha}]. \quad (8)$$

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

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

with  $\Sigma$  unknown is invariant under the group action  $X_i \rightarrow AX_i, i = 1, \dots, N$ , where  $A \in GL(p)$ , the group of all nonsingular  $p \times p$  matrices. A 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

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

is a maximal invariant parameter. Therefore the uniformly most powerful invariant size- $\alpha$  test rejects  $H_0$  if  $T^2 > f_{p, N-p; \alpha}$ , with power function  $\pi_\alpha(\Lambda; p, N - p)$ ; cf. Anderson, 2003 Theorem 5.6.1).<sup>1</sup>

It is informative to express  $\Lambda$  in terms of scale-free parameters, that is,

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

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

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

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

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

with  $R$  unknown.

*1.3 The  $T^2$  Variable-selection Problem* Denote the components of  $\bar{X}$  by  $\bar{X}_j, j = 1, \dots, p$ , and those of  $S$  by  $s_{jk}, j, k = 1, \dots, p$ . Let  $\Omega_p$  be the collection of all nonempty subsets of the index set  $I := \{1, \dots, p\}$ . For  $\omega \in \Omega_p$  denote the  $\omega$ -subvector of  $\bar{X}$  by  $\bar{X}_\omega$ , the  $\omega$ -submatrix of  $S$  by  $S_\omega$ , and similarly define  $\mu_\omega, \gamma_\omega, \Sigma_\omega$ , and  $R_\omega$ . The  $T^2$ -statistic based on  $(\bar{X}_\omega, S_\omega)$  is given by

$$T_\omega^2 \equiv N\bar{X}_\omega' S_\omega^{-1} \bar{X}_\omega \sim f_{|\omega|, N-|\omega|}(\Lambda_\omega), \tag{16}$$

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

---

<sup>1</sup> Among the larger class of all tests, invariant or non-invariant, the  $T^2$  test is admissible [Schaafsma (1982)], proper Bayes (Kiefer and Schwartz, 1965), and locally and asymptotically minimax for small and large values of  $\Lambda$ , respectively (Giri and Keifer, 1964).

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

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

This  $T_\omega^2$ -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 under which  $\{\mu \mid \mu_{I \setminus \omega} = 0\}$  has prior probability 1; cf. Kiefer and Schwartz (1965), Marden and Perlman (1980).

This paper addresses the feasibility of finding a parsimonious subset  $\omega$  such that the  $T_\omega^2$ -test maintains high power over a substantial portion of the alternative  $K$ . Because  $(\gamma, R)$  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  $(\gamma, R)$ ; see the Appendix. At worst, all  $2^p - 1$  nonempty subsets  $\omega$  must be considered.

Recently I consulted on such a variable-selection problem. The investigator, a research and development engineer, 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 product 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 wished to find a more readily interpretable measure of difference, namely a  $T_\omega^2$ -statistic based on a very small subset  $\omega$  of the 20 variables, hopefully with  $|\omega| = 1$  or 2.

Such a desire is not atypical of investigators presented with a multivariate data analysis. This led me to wonder how much power would be lost by restricting variable selection to small subsets  $\omega$ , for example to single variables or pairs of variables.

In fact some power might be gained. It is well known (e.g., Das Gupta and Perlman, 1974) that  $\pi_\alpha(\Lambda_\omega; |\omega|, N - |\omega|)$  is decreasing in  $|\omega|$  while increasing in  $\Lambda_\omega$ . Might the decreasing effect outweigh the increasing effect over a significant portion of the sample space? If so then restricting attention to small variable subsets might be desirable.

To state this more precisely, define

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

Thus  $\hat{\omega}_\alpha(\gamma, R)$  is the (not necessarily unique) subset  $\omega$  of variables that maximizes the power of the size- $\alpha$   $T_\omega^2$ -test to detect the alternative  $(\gamma, R)$  if the actual value of  $(\gamma, R)$  were revealed by an oracle. Whereas the admissibility of the overall size- $\alpha$   $T^2$ -test dictates that its power cannot be everywhere dominated by that of the size- $\alpha$   $T_\omega^2$ -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  $(\gamma, R)$ ? If so, then might one, with some confidence, limit variable selection to consideration of single variables (univariate  $t^2$ -tests) or pairs of variables (bivariate  $T^2$ -tests), as an alternative to simply applying the overall ( $p$ -variate)  $T^2$ -test?

Of course, corrections for multiple testing must be considered for any variable-selection procedure before definitive conclusions can be drawn and procedures implemented, see §5 for a brief example. However, restriction to small variable subsets has another desirable property: there are relatively few such subsets compared  $2^p - 1$  the number of all nonempty subsets of  $\Omega_p$ , greatly reducing any correction factor. For example, when  $p = 20$  as in the consulting problem cited above, there are 20 univariate subsets,  $\binom{20}{2} = 190$  bivariate subsets, compared to  $2^{20} - 1 = 1,048,575$  total nonempty subsets.

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. This report presents preliminary evidence indicating that limitation of variable selection to low-dimensional tests may not be entirely inappropriate.

In Sections 2, 3, and 4, several examples are considered where tractable algebraic expressions for the asymptotic ( $\Lambda_\omega \rightarrow \infty$ ), local ( $\Lambda_\omega \rightarrow 0$ ), 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  $T_\omega^2$  and  $T^2$ . These examples include both sparse and non-sparse mean-vector configurations, and the results may be the first that are based on algebraically-explicit power function comparisons of the low-dimensional and full-dimensional tests.

Examples 2.1 and 3.1 treat only the simplest possible case: the bivariate case ( $p = 2$ ) with  $N = 3$ .<sup>2</sup> 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  $(\gamma, R)$  in these regions.

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

---

<sup>2</sup> However, Giri et al. (1963) also began their study of Hotelling's  $T^2$  test by considering only this simplest case  $p = 2$ ,  $N = 3$ .

and vague condition that  $\alpha$  be sufficiently small, with “sufficiently small” determined by the value of the unknown noncentrality parameter – see Eqs. 126 and 131.

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  $T_\omega^2$ -tests ( $|\omega| = 2$ ) are compared to the power of the overall ( $p$ -variate)  $T^2$  test, again only for asymptotic and local alternatives and only for very special configurations of  $\gamma$  and  $R$ . In these cases, admittedly highly restrictive, the bivariate  $T_\omega^2$ -tests dominate the  $p$ -variate  $T^2$ -test 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 selection might be limited to small variable subsets  $\omega$ .

Together, the preliminary findings in this paper indicate the feasibility and potential benefit of limiting variable selection to small subsets, in particular to univariate or bivariate subsets. Further study will be needed to implement this approach to variable selection and to confirm its efficacy. See §5 and the Appendix for related comments.

## 2 Some Asymptotic Power Comparisons

The power function of the  $T_\omega^2$ -test is

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

(recall (16) and (17)). It follows from eqn. (3.4) in Marden and Perlman (1980) that as  $\Lambda_\omega \rightarrow \infty$ ,

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

Thus for two subsets  $\omega, \omega'$  with  $\omega \subset \omega'$ , there exists  $\Lambda_{|\omega|, |\omega'|, N; \alpha}^* > 0$  such that

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

Therefore power comparisons of  $T_\omega^2$  and  $T_{\omega'}^2$  for distant alternatives<sup>3</sup> 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

---

<sup>3</sup> We do not claim to know the value of  $\Lambda_{|\omega|, |\omega'|, N; \alpha}^*$ , even approximately.

limited scope<sup>4</sup> they begin to suggest that variable subset selection sometimes can be 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

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

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

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

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

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

**Example 2.1** In the bivariate case  $p = 2$ , abbreviate the singleton subsets  $\{1\}$  and  $\{2\}$  of  $\Omega_2$  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- $\alpha$   $t^2$ -tests to the power  $\pi_\alpha(\Lambda; 2, 1)$  of the overall (bivariate) size- $\alpha$   $T^2$ -test for distant alternatives.

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

$$\eta = \frac{\gamma_2}{\gamma_1}, \quad \rho = \rho_{12}, \quad (26)$$

where  $-1 < \rho < 1$ , so by Eqs. 13 and 17,

$$\Lambda_1 = N\gamma_1^2, \quad \Lambda_2 = N\eta^2\gamma_1^2, \quad \Lambda = N\left(\frac{1-2\eta\rho+\eta^2}{1-\rho^2}\right)\gamma_1^2. \quad (27)$$

Without loss of generality we can assume that  $|\gamma_1| \geq |\gamma_2|$ , so  $0 \leq \eta^2 \leq 1$  and

$$\max(\Lambda_1, \Lambda_2) = N\gamma_1^2. \quad (28)$$

The alternative hypotheses  $K$  can be represented as

$$K = \{(\gamma_1, \eta, \rho) \mid |\gamma_1| > 0, |\eta| \leq 1, |\rho| < 1\}, \quad (29)$$

while  $\omega_\alpha(\gamma, R)$  can be re-expressed as  $\omega_\alpha(\gamma_1, \eta, \rho)$ .

---

<sup>4</sup> But see Footnote 2.

Because  $\max(\Lambda_1, \Lambda_2) \leq \Lambda$ , it follows from Eqs. 22, 27, and 28 that

$$\gamma_1^2 > \max\left(\frac{1}{N}\Lambda_{1,2,N;\alpha}^*, Q_{1,2,N;\alpha}\left(\frac{1-2\eta\rho+\eta^2}{1-\rho^2}\right)\gamma_1^2\right) \quad (30)$$

$$\implies \max(\pi_\alpha(\Lambda_1; 1, 2), \pi_\alpha(\Lambda_2; 1, 2)) > \pi_\alpha(\Lambda; 2, 1) \quad (31)$$

$$\implies |\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1. \quad (32)$$

In the simplest case  $N = 3$ , Eq. 25 yields the explicit expression

$$Q_{1,2,3;\alpha} = \frac{b_{1,2;\alpha}}{b_{2,1;\alpha}} = \frac{\alpha}{2-\alpha}, \quad (33)$$

while the inequality in Eq. 30 is equivalent to

$$1 > \max\left(\frac{\Lambda_{1,2,3;\alpha}^*}{3\gamma_1^2}, Q_{1,2,3;\alpha}\left(\frac{1-2\eta\rho+\eta^2}{1-\rho^2}\right)\right). \quad (34)$$

Note that

$$1 > Q_{1,2,3;\alpha}\left(\frac{1-2\eta\rho+\eta^2}{1-\rho^2}\right) \quad (35)$$

$$\iff 0 > \rho^2 - 2Q_{1,2,3;\alpha}\eta\rho + Q_{1,2,3;\alpha}(1 + \eta^2) - 1 =: h_{\alpha,\eta}(\rho). \quad (36)$$

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

$$h_{\alpha,\eta}(-1) = Q_{1,2,3;\alpha}(1 + \eta)^2 \geq 0,$$

$$h_{\alpha,\eta}(1) = Q_{1,2,3;\alpha}(1 - \eta)^2 \geq 0,$$

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

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

$$\hat{\rho}_{\alpha,\eta}^\pm = Q_{1,2,3;\alpha}\eta \pm \sqrt{(1 - Q_{1,2,3;\alpha})(1 - Q_{1,2,3;\alpha}\eta^2)}; \quad (37)$$

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

It follows that if  $\alpha \leq \frac{2}{3}$  then for sufficiently large  $\gamma_1^2$ , i.e.,  $\gamma_1^2 \geq \frac{1}{3}\Lambda_{1,2,3;\alpha}^*$ ,

$$\rho \in (\hat{\rho}_{\alpha,\eta}^-, \hat{\rho}_{\alpha,\eta}^+) \implies |\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1, \quad (38)$$

that is, at least one of the two univariate  $t^2$ -tests is more powerful than the overall (bivariate)  $T^2$ -test. Specifically, when  $\gamma_1^2 \geq \frac{1}{3}\Lambda_{1,2,3;\alpha}^*$ ,  $|\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| =$



1 in the  $(\eta, \rho)$ -regions of the parameter space indicated in Table 1. From this it is seen that for  $p = 2$ ,  $N = 3$ , and the common (small) values of  $\alpha$ , the bivariate size- $\alpha$   $T^2$ -test is dominated by at least one of the two univariate size- $\alpha$   $t^2$ -tests for most of the distant alternative hypothesis  $K$ , i.e., for sufficiently large  $\gamma_1^2$ . In fact, for most cases this domination occurs over almost the entire range  $(-1, 1)$  of  $\rho$ .  $\square$

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

$$\{\pi(\Lambda_\omega; 2, p) \mid \omega \in \Omega_p, |\omega| = 2\} \quad \text{and} \quad \pi(\Lambda; p, 2). \quad (39)$$

Table 1: Take  $p = 2$ ,  $N = 3$ , and  $\gamma_1^2$  sufficiently large, that is,  $\gamma_1^2 \geq \frac{1}{3}\Lambda_{1,2,3;\alpha}^*$

$\alpha$	$Q_{1,2,3;\alpha}$	$\eta$	$(\hat{\rho}_{\alpha,\eta}^-, \hat{\rho}_{\alpha,\eta}^+)$
.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	$(-1, 1)$

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- $\alpha$   $t^2$ -tests dominates that of the bivariate size- $\alpha$   $T^2$ -test. (Note that  $\hat{\rho}_{\alpha,-\eta}^\pm = -\hat{\rho}_{\alpha,\eta}^\mp$ )

From Eq. 22,

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

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- $\alpha$   $T^2$ -tests will be more powerful than the  $p$ -variate size- $\alpha$   $T^2$ -test<sup>5</sup> provided that

$$\Lambda^{(2)} > Q_{2,p,p+2;\alpha}\Lambda. \quad (40)$$

From Eq. 25 we obtain the explicit expression

$$Q_{2,p,p+2;\alpha} = \frac{b_{2,p;\alpha}}{b_{p,2;\alpha}} = \frac{1-(1-\alpha)^{\frac{2}{p}}}{\alpha^{\frac{2}{p}}}. \quad (41)$$

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

$$\lim_{p \rightarrow \infty} Q_{2,p,p+2;\alpha} = 0, \quad (42)$$

$$\lim_{p \rightarrow \infty} U_{p;\alpha} = \frac{1}{-\log(1-\alpha)} \quad (> 1 \text{ for } \alpha < \frac{e-1}{e} = .6321); \quad (43)$$

$$\lim_{p \rightarrow \infty} (p-1)U_{p;\alpha} = \infty. \quad (44)$$

Table 2 shows that  $Q_{2,p,p+2;\alpha}$  decreases rapidly to 0 as  $p \rightarrow \infty$ , which suggests that Eq. 40 might hold over substantial regions of the alternative hypothesis  $K$ . We proceed to exhibit several such regions.

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

$$R_\rho := (1-\rho)I_p + \rho\mathbf{1}_p\mathbf{1}'_p, \quad (45)$$

where  $\mathbf{1}_p = (1, \dots, 1)'$ :  $p \times 1$  and the allowable range of  $\rho$  is  $(-\frac{1}{p-1}, 1)$ . Then

$$R_\rho^{-1} = \frac{1}{1-\rho} \left[ I_p - \frac{\rho\mathbf{1}_p\mathbf{1}'_p}{1+\rho(p-1)} \right], \quad (46)$$

$$\mathbf{1}'_p R_\rho^{-1} \mathbf{1}_p = \frac{p}{1+\rho(p-1)}, \quad (47)$$

---

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

ON THE FEASIBILITY OF PARSIMONIOUS VARIABLE SELECTION...

Table 2: Take  $p \geq 3$ ,  $N = p + 2$ , and  $\delta^2$  sufficiently large, that is,  $\delta^2 \geq (p + 2)^{-1} \Lambda_{2,p,p+2;\alpha}^*$

$p$	$\alpha$	$Q_{2,p,p+2;\alpha}$	$U_{p;\alpha}$	$\tilde{\psi}_{p;\alpha}^-$	$(\tilde{\rho}_{p;\alpha}^-, \tilde{\rho}_{p;\alpha}^+)$	$-\frac{1}{p-1}$
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 <sup>2</sup> 504	39.652	-.109	(-.111, .980)	
20	.20	.0259	3.858	-.039	(-.0525, .954)	-.0526
	.10	.0132	7.579	-.046	(-.0526, .976)	
	.05	.0 <sup>2</sup> 690	14.486	-.049	(-.0526, .988)	
	.01	.0 <sup>2</sup> 159	62.810	-.052	(-.0526, .997)	
40	.20	.0120	4.158	-.020	(-.0256, .977)	-.0256
	.10	.0 <sup>2</sup> 590	8.481	-.023	(-.0256, .989)	
	.05	.0 <sup>2</sup> 298	16.806	-.024	(-.0256, .994)	
	.01	.0 <sup>3</sup> 632	79.055	-.025	(-.0256, .999)	
$\infty$	.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)	

In Case 1, if  $\sim \psi_{p;\alpha}^- < \rho < 1$  then the power of all the  $\binom{p}{2}$  bivariate size- $\alpha$   $T^2$ -tests dominates that of the  $p$ -variate size- $\alpha$   $T^2$ -test. In Case 2, if  $\tilde{\rho}_{p;\alpha}^- < \rho < \sim \rho_{p;\alpha}^+$  then the power of at least one of the  $\binom{p}{2}$  bivariate size- $\alpha$   $T^2$ -tests dominates that of the  $p$ -variate size- $\alpha$   $T^2$ -test

By symmetry, all bivariate tests have the same power, and by Eqs. 13 and 17,

$$\Lambda = \frac{p(p+2)\delta^2}{1+(p-1)\rho}, \quad \Lambda^{(2)} = \frac{2(p+2)\delta^2}{1+\rho}. \quad (48)$$

Thus  $\Lambda^{(2)} \geq \Lambda_{2,p,p+2;\alpha}^*$  holds for all allowable  $\rho$  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 Eq. 40 is equivalent to each of the inequalities

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

$$[(p-1)U_{p;\alpha} - 1]\rho > 1 - U_{p;\alpha}. \quad (49)$$

Because  $(p-1)U_{p;\alpha} > 1$  for common (small) values of  $\alpha$  (see Eq. 44 and Table 2), in such cases Eq. 49 is equivalent to

$$\rho > \frac{1-U_{p;\alpha}}{(p-1)U_{p;\alpha}-1} =: \tilde{\psi}_{p;\alpha}^- \quad (50)$$

Table 2 shows that in Case 1,  $\tilde{\psi}_{p;\alpha}^-$  is close to the lower limit of the allowable range  $(-\frac{1}{p-1}, 1)$  for  $\rho$ . Thus by Eq. 50, all of the bivariate size- $\alpha$   $T^2$ -tests are more powerful than the  $p$ -variate size- $\alpha$   $T^2$ -test for most of the distant alternative hypothesis specified in Case 1, i.e., for sufficiently large  $\delta^2 (\geq (p+2)^{-1}\Lambda_{2,p,p+2;\alpha}^*)$ .

**Case 2 ( $\gamma$  sparse):**  $\gamma_i = \gamma_j =: \delta$  for some  $\{i, j\} \subset \{1, \dots, p\}$ ,  $\gamma_k = 0$  for  $k \neq i, j$ , and  $R$  has the intraclass form  $R_\rho$  in Eq. 45 By Eqs. 13 and 17,

$$\Lambda = \frac{2(p+2)\delta^2[1+(p-3)\rho]}{(1-\rho)[1+(p-1)\rho]}, \quad \Lambda^{(2)} = \frac{2(p+2)\delta^2}{1+\rho}, \quad (51)$$

so again  $\Lambda^{(2)} \geq \Lambda_{2,p,p+2;\alpha}^*$  holds for all allowable  $\rho$  if  $\delta^2 \geq (p+2)^{-1}\Lambda_{2,p,p+2;\alpha}^*$ . Abbreviating  $Q_{2,p,p+2;\alpha}$  by  $Q$ , Eq. 40 is equivalent to each of the inequalities

$$\begin{aligned} \frac{1}{1+\rho} &> \frac{Q[1+(p-3)\rho]}{(1-\rho)[1+(p-1)\rho]}, \\ 0 &> [(p-1) + (p-3)Q]\rho^2 - (p-2)(1-Q)\rho - (1-Q) =: h_{p;\alpha}(\rho). \end{aligned}$$

Since  $h_{p;\alpha}(0) = Q - 1 < 0$  for common (small) values of  $\alpha$  (see Eqs. 42-43 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 allowable range  $(-\frac{1}{p-1}, 1)$  for  $\rho$ . Thus at least one of the bivariate size- $\alpha$   $T^2$ -tests is more powerful than the  $p$ -variate size- $\alpha$   $T^2$ -test for most of the distant alternative hypothesis specified in Case 2, i.e., for sufficiently large  $\delta^2 (\geq (p+2)^{-1}\Lambda_{2,p,p+2;\alpha}^*)$ .

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

$$\Lambda^{(2)} = \frac{2(p+2)\delta^2}{1-\rho}, \quad \Lambda = \frac{2(p+2)\delta^2}{1+\rho}. \quad (52)$$

ON THE FEASIBILITY OF PARSIMONIOUS VARIABLE SELECTION...

Thus  $\Lambda^{(2)} \geq \Lambda_{2,p,p+2;\alpha}^*$  again holds for all allowable  $\rho$  if  $\delta^2 \geq (p+2)^{-1}\Lambda_{2,p,p+2;\alpha}^*$ , while Eq. 40 is equivalent to  $1 > Q_{2,p,p+2;\alpha}$ , which holds for most  $p, \alpha$  (see Eqs. 42-43 and Table 2). Again at least one of the bivariate size- $\alpha$   $T^2$ -tests will be more powerful than the  $p$ -variate size- $\alpha$   $T^2$ -test for the *entire* distant alternative hypothesis in Case 3, i.e., for sufficiently large  $\delta^2 (\geq (p+2)^{-1}\Lambda_{2,p,p+2;\alpha}^*)$ .

**Case 4:  $p = 2l$  is even,  $\gamma_i = \delta$  for  $l$  indices in  $\{1, \dots, p\}$ ,  $\gamma_i = -\delta$  for the remaining  $l$  indices, and  $R$  has the intraclass form  $R_\rho$**

By Eqs. 13 and 17,

$$\Lambda^{(2)} = \frac{2(p+2)\delta^2}{1-|\rho|}, \quad \Lambda = \frac{p(p+2)\delta^2}{(1-\rho)}. \quad (53)$$

Thus  $\Lambda^{(2)} \geq \Lambda_{2,p,p+2;\alpha}^*$  again holds for all allowable  $\rho$  if  $\delta^2 \geq (p+2)^{-1}\Lambda_{2,p,p+2;\alpha}^*$ , while Eq. 40 is equivalent to the inequality

$$U_{p;\alpha} > \frac{1-|\rho|}{1-\rho}. \quad (54)$$

Because  $\frac{1-|\rho|}{1-\rho} \leq 1$ , while  $U_{p;\alpha} > 1$  for holds for most  $p, \alpha$  (see Eq. 43 and Table 2), at least one of the bivariate size- $\alpha$   $T^2$ -tests is more powerful than the  $p$ -varisate size- $\alpha$   $T^2$ -test over the *entire* distant alternative hypothesis in Case 4, i.e., for sufficiently large  $\delta^2 (\geq (p+2)^{-1}\Lambda_{2,p,p+2;\alpha}^*)$

### 3 Some Local Power Comparisons

From Eqs. 2-4, as  $\Lambda_\omega \downarrow 0$  the power function  $\pi_\alpha(\Lambda_\omega) \equiv \pi_\alpha(\Lambda_\omega; |\omega|, N - |\omega|)$  of the  $T_\omega^2$ -test satisfies

$$\pi_\alpha(\Lambda_\omega) = e^{-\frac{\Lambda_\omega}{2}} \left[ \alpha + \frac{\Lambda_\omega}{2} c_{|\omega|, N-|\omega|; 1}^\alpha + O(\Lambda_\omega^2) \right] \quad (55)$$

$$= \alpha + \frac{\Lambda_\omega}{2} (c_{|\omega|, N-|\omega|; 1}^\alpha - \alpha) + O(\Lambda_\omega^2). \quad (56)$$

Thus for two subsets  $\omega, \omega'$  with  $\omega \subset \omega'$ , there exists  $\Lambda_{|\omega|, |\omega'|, N; \alpha}^{**} > 0$  such that

$$\Lambda_{\omega'} < \min \left( \Lambda_{|\omega|, |\omega'|, N; \alpha}^{**}, Z_{|\omega|, |\omega'|, N; \alpha} \Lambda_\omega \right) \implies \pi_\alpha(\Lambda_\omega) > \pi_\alpha(\Lambda_{\omega'}), \quad (57)$$

where, from (2.2) and (2.3) in Das Gupta and Perlman (1974),

$$Z_{|\omega|,|\omega'|,N;\alpha} := \frac{c_{|\omega|,N-|\omega|;1}^\alpha}{c_{|\omega'|,N-|\omega'|;1}^\alpha} > 1. \quad (58)$$

Therefore power comparisons of  $T_\omega^2$  and  $T_{\omega'}^2$  for local alternatives<sup>6</sup> require determination of the lower tail probabilities  $c_{m,n;k}^\alpha$ , which in turn require the lower quantiles  $b_{n,m;\alpha}$  (see Eq. 8).

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 from Eqs. 27 and 57 that

$$\left(\frac{1-2\eta\rho+\eta^2}{1-\rho^2}\right)\gamma_1^2 < \min\left(\frac{1}{N}\Lambda_{1,2,N;\alpha}^{**}, Z_{1,2,N;\alpha}\gamma_1^2\right) \quad (59)$$

$$\implies \max(\pi_\alpha(\Lambda_1; 1, 2), \pi_\alpha(\Lambda_2; 1, 2)) > \pi_\alpha(\Lambda; 2, 1), \quad (60)$$

$$\implies |\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1. \quad (61)$$

In the simplest case  $N = 3$ , it follows from Eqs. 8, 6, and 25 that

$$Z_{1,2,3;\alpha} = \frac{c_{1,2;1}^\alpha}{c_{2,1;1}^\alpha} = \frac{1-(1-\alpha)^3-\alpha}{\frac{3}{2}(\alpha-\frac{1}{3}\alpha^3)-\alpha} = \frac{2(2-\alpha)}{1+\alpha}. \quad (62)$$

First note that in Eq. 59,

$$\frac{1-2\eta\rho+\eta^2}{1-\rho^2} < Z_{1,2,3;\alpha} \quad (63)$$

$$\iff h_{\alpha,\eta}(\rho) := Z_{1,2,3;\alpha}\rho^2 - 2\eta\rho + \eta^2 + 1 - Z_{1,2,3;\alpha} < 0. \quad (64)$$

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

$$h_{\alpha,\eta}(-1) = (1 + \eta)^2 \geq 0,$$

$$h_{\alpha,\eta}(1) = (1 - \eta)^2 \geq 0,$$

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

---

<sup>6</sup> We do not claim to know the values of  $\Lambda_{|\omega|,|\omega'|,N;\alpha}^{**}$ , even approximately.

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

$$\check{\rho}_{\alpha,\eta}^{\pm} = \frac{\eta}{Z_{1,2,3;\alpha}} \pm \frac{1}{Z_{1,2,3;\alpha}} \sqrt{(Z_{1,2,3;\alpha} - 1)(Z_{1,2,3;\alpha} - \eta^2)}; \quad (65)$$

again  $\check{\rho}_{\alpha,-\eta}^{\pm} = -\check{\rho}_{\alpha,\eta}^{\mp}$ . Thus, if  $\alpha \leq \frac{1}{2}$  and  $\rho \in (\check{\rho}_{\alpha,\eta}^-, \check{\rho}_{\alpha,\eta}^+)$  then Eq. 63 must hold.

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

$$0 < \gamma_1^2 < \left( \frac{1-\rho^2}{1-2\eta\rho+\eta^2} \right) \frac{\Lambda_{1,2,3;\alpha}^{**}}{3}. \quad (66)$$

Because

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

for fixed  $\eta$ , Eq. 66 will be satisfied provided that

$$\rho \in (\check{\rho}_{\alpha,\eta}^-, \check{\rho}_{\alpha,\eta}^+), \quad (67)$$

$$\check{m}_{\alpha,\eta} := \max(|\check{\rho}_{\alpha,\eta}^-|, |\check{\rho}_{\alpha,\eta}^+|) < 1, \quad (68)$$

$$\gamma_1^2 < \frac{1}{6}(1 - m_{\alpha,\eta})\Lambda_{1,2,3;\alpha}^{**}. \quad (69)$$

It is straightforward to show that Eq. 68 holds for  $|\eta| < 1$  but not for  $|\eta| = 1$ .

Thus, if  $\alpha \leq \frac{1}{2}$ ,  $|\eta| < 1$ , and Eqs. 67, 68, and 69 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  $(\eta, \rho)$ -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  $p = 2$ ,  $N = 3$ , and the common (small) values of  $\alpha$ , the bivariate size- $\alpha$   $T^2$ -test is dominated by at least one of the two univariate size- $\alpha$   $t^2$ -tests over much of the local alternative hypothesis space. Compared to Table 1, this effect seems somewhat less than for distant alternatives.

**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- $\alpha$   $T^2$ -tests and the  $p$ -variate size- $\alpha$   $T^2$ -test for local alternatives, which requires comparison of the powers

$$\{\pi(\Lambda_{\omega}; 2, p) \mid \omega \in \Omega_p, |\omega| = 2\} \quad \text{and} \quad \pi(\Lambda; p, 2). \quad (70)$$

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- $\alpha$   $t^2$ -tests dominates that of the bivariate size- $\alpha$   $T^2$ -test

$\alpha$	$Z_{1,2,3;\alpha}$	$\eta$	$(\check{\rho}_{\alpha, \eta}^-, \check{\rho}_{\alpha, \eta}^+)$	$1 - \check{m}_{\alpha, \eta}$
.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

(Note that  $\check{\rho}_{\alpha, -\eta}^\pm = -\check{\rho}_{\alpha, \eta}^\mp$ )

From Eq. 57,

$$\Lambda < \min(\Lambda_{2,p,p+2;\alpha}^{**}, Z_{2,p,p+2;\alpha}\Lambda^{(2)}) \tag{71}$$

$$\implies \max\{\pi(\Lambda_\omega; 2, p) \mid \omega \in \Omega_p, |\omega| = 2\} > \pi(\Lambda; p, 2). \tag{72}$$

Therefore for sufficiently small values of  $\Lambda$ , namely  $\Lambda \leq \Lambda_{2,p,p+2;\alpha}^{**}$ , at least one of the bivariate size- $\alpha$   $T^2$ -tests will be more powerful than the  $p$ -variate size- $\alpha$   $T^2$ -test<sup>7</sup> whenever

$$\Lambda < Z_{2,p,p+2;\alpha}\Lambda^{(2)}. \tag{73}$$

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



From Eqs. 58, 8-6, 25, and some algebra, the explicit expression

$$Z_{2,p,p+2;\alpha} = \frac{c_{2,p;1}^\alpha - \alpha}{c_{p,2;1}^\alpha - \alpha} = \frac{p\alpha[1-\alpha^{\frac{2}{p}}]}{2(1-\alpha)[1-(1-\alpha)^{\frac{2}{p}}]} \quad (74)$$

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

$$\lim_{p \rightarrow \infty} Z_{2,p,p+2;\alpha} = \infty, \quad (75)$$

$$\lim_{p \rightarrow \infty} V_{p;\alpha} = \frac{\alpha \log \alpha}{(1-\alpha) \log(1-\alpha)} \quad (> 1 \text{ for } \alpha < \frac{1}{2}), \quad (76)$$

$$\lim_{p \rightarrow \infty} (p-1)V_{p;\alpha} = \infty. \quad (77)$$

Table 4 shows that  $Z_{2,p,p+2;\alpha}$  increases rapidly to  $\infty$  as  $p \rightarrow \infty$ , which suggests that Eq. 73 might hold over substantial regions of the alternative hypothesis. Several such regions are now exhibited.

**Case 1:**  $\gamma_1 = \dots = \gamma_p =: \delta$  and  $R$  has the intraclass form Eq. 45

Here  $-\frac{1}{p-1} < \rho < 1$  and as in Eq. 48,

$$\Lambda^{(2)} = \frac{2(p+2)\delta^2}{1+\rho}, \quad \Lambda = \frac{p(p+2)\delta^2}{1+(p-1)\rho}. \quad (78)$$

Here Eq. 73 is equivalent to each of the inequalities

$$\begin{aligned} V_{p;\alpha} &> \frac{1+\rho}{1+(p-1)\rho}; \\ V_{p;\alpha} - 1 &> -[(p-1)V_{p;\alpha} - 1]\rho. \end{aligned} \quad (79)$$

Because  $(p-1)V_{p;\alpha} > 1$  for common (small) values of  $\alpha$  (see Eq. 77 and Table 4), in such cases Eq. 79 in turn is equivalent to

$$\check{\psi}_{p;\alpha}^- := -\frac{V_{p;\alpha}-1}{(p-1)V_{p;\alpha}-1} < \rho < -1. \quad (80)$$

To conclude that Eqs. 71-72 holds,  $\delta^2$  must be sufficiently small, i.e.,

$$0 < \delta^2 < \left[ \frac{1+(p-1)\rho}{p(p+2)} \right] \Lambda_{2,p,p+2;\alpha}^{**}. \quad (81)$$

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

$$\frac{1+(p-1)\rho}{p(p+2)} > \frac{1+(p-1)\check{\psi}_{p;\alpha}^-}{p(p+2)} = \frac{p-2}{p(p+2)[(p-1)V_{p;\alpha}-1]} := \check{m}_{p,\alpha}. \quad (82)$$

Therefore Eq. 81 will be satisfied provided that

$$\rho > \check{\psi}_{p;\alpha}^- \quad \text{and} \quad \delta^2 < \check{m}_{p,\alpha} \Lambda_{2,p,p+2;\alpha}^{**}. \quad (83)$$

If  $p$  is large and  $\alpha$  is small, Table 4 shows that in Case 1,  $\check{\psi}_{p;\alpha}^-$  is close to the lower limit of the allowable range  $(-\frac{1}{p-1}, 1)$  for  $\rho$ . Then by Eq. 83, at least one of the bivariate size- $\alpha$   $T^2$ -tests will be more powerful than the  $p$ -variate size- $\alpha$   $T^2$ -test for most of the local alternative hypothesis covered by Case 1, i.e., for  $\delta^2 < \check{m}_{p,\alpha} \Lambda_{2,p,p+2;\alpha}^{**}$ .

**Case 2 ( $\gamma$  sparse):**  $\gamma_i = \gamma_j \equiv \delta$  for some  $\{i, j\} \subset \{1, \dots, p\}$ ,  $\gamma_k = 0$  for  $k \neq i, j$ , and  $R$  has the intraclass form  $R_\rho$  in Eq. 45 As in 51,

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

Abbreviating  $Z_{2,p,p+2;\alpha}$  by  $Z$ , Eq. 73 is equivalent to each of the inequalities

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

Since  $h_{p;\alpha}(0) = 1 - Z < 0$  (cf. Eq. 58),  $h_{p;\alpha}(\rho)$  has real roots  $\check{\rho}_{p;\alpha}^- < 0 < \check{\rho}_{p;\alpha}^+$  (found numerically). Therefore  $0 > h_{p;\alpha}(\rho)$  for  $\check{\rho}_{p;\alpha}^- < \rho < \check{\rho}_{p;\alpha}^+$ .

To conclude that Eqs. 71-72 holds,  $\delta^2$  must be sufficiently small, i.e.,

$$0 < \delta^2 < \left\{ \frac{(1-\rho)[1+(p-1)\rho]}{2(p+2)[1+(p-3)\rho]} \right\} \Lambda_{2,p,p+2;\alpha}^{**}. \quad (84)$$

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

$$\frac{(1-\rho)[1+(p-1)\rho]}{2(p+2)[1+(p-3)\rho]} > \frac{(1-\check{\rho}_{p;\alpha}^+)[1+(p-1)\check{\rho}_{p;\alpha}^+]}{2(p+2)[1+(p-3)\check{\rho}_{p;\alpha}^+]} := \check{m}'_{p,\alpha}. \quad (85)$$

Therefore Eq. 84 will be satisfied provided that

$$\rho \in (\check{\rho}_{p;\alpha}^-, \check{\rho}_{p;\alpha}^+) \quad \text{and} \quad \delta^2 < \check{m}'_{p,\alpha} \Lambda_{2,p,p+2;\alpha}^{**}. \quad (86)$$

Table 4 shows that in Case 2, the interval  $(\check{\rho}_{p;\alpha}^-, \check{\rho}_{p;\alpha}^+)$  covers almost all of the allowable range  $(-\frac{1}{p-1}, 1)$  for  $\rho$ . Thus at least one of the bivariate size- $\alpha$   $T^2$ -tests will be more powerful than the  $p$ -variate size- $\alpha$   $T^2$ -test for

ON THE FEASIBILITY OF PARSIMONIOUS VARIABLE SELECTION...

most of the local alternative hypothesis determined by Case 2, i.e., for  $\delta^2 < \check{m}'_{p,\alpha} \Lambda_{2,p,p+2;\alpha}^{**}$ .

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

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

Here Eq. 73 is equivalent to  $Z_{2,p,p+2;\alpha} > 1$ , which holds for all  $p, \alpha$  (see Eq. 58).

To conclude that Eqs. 71-72 holds,  $\delta^2$  must be sufficiently small, i.e.,

$$\delta^2 < \left(\frac{1-\rho}{2(p+2)}\right) \Lambda_{2,p,p+2;\alpha}^{**} \quad (87)$$

for all  $\rho \in (-\frac{1}{p-1}, 1)$ . This requires that  $\rho$  be bounded below 1, that is,  $\rho < 1 - \epsilon$  for some  $\epsilon > 0$ , whence Eq. 87 will be satisfied if

$$\delta^2 < \left(\frac{\epsilon}{2(p+2)}\right) \Lambda_{2,p,p+2;\alpha}^{**}. \quad (88)$$

Thus at least one of the bivariate size- $\alpha$   $T^2$ -tests is more powerful than the  $p$ -variate size- $\alpha$   $T^2$ -test if  $\rho < 1 - \epsilon$ , which covers almost all of the local region Eq. 88 in the alternative hypothesis determined by Case 3.

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

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

Here Eq. 73 is equivalent to the inequality

$$V_{p;\alpha} > \frac{1-|\rho|}{1-\rho}. \quad (89)$$

Because  $\frac{1-|\rho|}{1-\rho} \leq 1$ , while  $V_{p;\alpha} > 1$  for holds for most  $p, \alpha$  (see Eq. 76 and Table 4), Eq. 89 is satisfied for most  $p, \alpha$ .

To conclude that Eqs. 71-72 holds,  $\delta^2$  must be sufficiently small, i.e.,

$$\delta^2 < \left[\frac{1-\rho}{p(p+2)}\right] \Lambda_{2,p,p+2;\alpha}^{**} \quad (90)$$

for all  $\rho \in (-\frac{1}{p-1}, 1)$ . This again requires that  $\rho$  be bounded below 1, that is,  $\rho < 1 - \epsilon$  for some  $\epsilon > 0$ , whence Eq. 90 will be satisfied if

$$\delta^2 < \left[ \frac{\epsilon}{p(p+2)} \right] \Lambda_{2,p,p+2;\alpha}^{**} \quad (91)$$

Thus at least one of the bivariate size- $\alpha$   $T^2$ -tests is more powerful than the  $p$ -variate size- $\alpha$   $T^2$ -test if  $\rho < 1 - \epsilon$ , which covers almost all of the local region Eq. 91 in the alternative hypothesis determined by Case 4.  $\square$

**Remark 3.3** The mean-vector and covariance matrix configurations in Examples 2.1 and 2.2 are the same as those in Examples 3.1 and 3.2 respectively, so the results for distant alternatives in the former can be compared to those for local alternatives in the latter. Because both sets of results support the feasibility of limiting variable selection to small subsets, this suggests that this feasibility may extend to intermediate alternatives as well. Furthermore, both sparse cases (Cases 2 and 3) and non-sparse cases (Cases 1 and 4) exhibit this feasibility in §2 and §3. Of course, more extensive comparisons will be needed to confirm this conclusion.

#### 4 Some Exact Power Comparisons for the Bivariate Case

The results in Sections 2 and 3 compare the power of the overall ( $p$ -variate)  $T^2$ -test with those of univariate or bivariate  $T^2$ -tests based on the original variates. However these power comparisons are asymptotic or local, and are relevant only for noncentrality parameters  $\Lambda$  that approach  $\infty$  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  $\Lambda$ . 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  $\lambda > 0$ , there exists  $\alpha_l^*(\lambda) \in (0, 1)$  such that

$$0 < \alpha < \alpha_l^*(\lambda) \implies \pi_\alpha(\lambda; 1, 2l) > \pi_\alpha\left(\left(\frac{4l}{2l-1}\right)\lambda; 2, 2l-1\right), \quad (92)$$

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

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  $\lambda$  then we cannot conclude that, at least for small  $\alpha$ , one or both of the two univariate size- $\alpha$   $t^2$ -tests dominate the bivariate size- $\alpha$   $T^2$

Table 4: Suppose that  $p \geq 3$ ,  $N = p + 2$ , and  $\delta^2$  is sufficiently small, that is,  $\delta^2 < \check{m}_{p,\alpha} \Lambda_{2,p,p+2;\alpha}^{**}$  ( $\delta^2 < \check{m}'_{p,\alpha} \Lambda_{2,p,p+2;\alpha}^{**}$ )

$p$	$\alpha$	$Z_{2,p,p+2;\alpha}$	$V_{p;\alpha}$	$\check{\psi}_{p;\alpha}^-$	$\check{m}_{p,\alpha}$	$(\check{\rho}_{p;\alpha}^-, \check{\rho}_{p;\alpha}^+)$	$\check{m}'_{p,\alpha}$	$-\frac{1}{p-1}$
4	.20	2.618	1.309	-1.06	.0285	(-.282, .648)	.0524	-.333
	.10	2.961	1.480	-1.40	.0242	(-.289, .686)	.0475	
	.05	3.228	1.614	-1.60	.0217	(-.295, .710)	.0442	
	.01	3.627	1.814	-1.83	.0118	(-.298, .741)	.0400	
10	.20	7.882	1.576	-.044	.02506	(-.108, .815)	.02958	-.111
	.10	9.832	1.966	-.058	.02399	(-.109, .849)	.02783	
	.05	11.621	2.324	-.066	.02335	(-.109, .871)	.02669	
	.01	15.138	3.028	-.077	.02254	(-.110, .899)	.02630	
20	.20	16.842	1.684	-.022	.02132	(-.0523, .898)	.02257	-.0526
	.10	21.804	2.180	-.029	.02101	(-.0524, .921)	.02199	
	.05	26.630	2.663	-.034	.03825	(-.0524, .935)	.02164	
	.01	37.109	3.711	-.039	.03588	(-.0525, .953)	.02119	
40	.20	34.844	1.742	-.011	.03338	(-.0256, .947)	.02105	-.0256
	.10	45.995	2.300	-.015	.03255	(-.0256, .959)	.03514	
	.05	57.168	2.858	-.017	.03205	(-.0256, .967)	.03414	
	.01	90.567	4.528	-.020	.03129	(-.0256, .979)	.03263	
$\infty$	.20	$\infty$	1.803	0	0	[0, 1)	0	0
	.10	$\infty$	2.429	0	0	[0, 1)	0	
	.05	$\infty$	3.074	0	0	[0, 1)	0	
	.01	$\infty$	4.628	0	0	[0, 1)	0	

In Case 1 (Case 2) if  $\check{\psi}_{p;\alpha}^- < \rho < 1$  ( $\check{\rho}_{p;\alpha}^- < \rho < \check{\rho}_{p;\alpha}^+$ ) then the power of at least one of the  $\binom{p}{2}$  bivariate size- $\alpha$   $T^2$ -tests dominates that of the  $p$ -variate size- $\alpha$   $T^2$ -test. Note that the allowable range of  $\rho$  is  $(-\frac{1}{p-1}, 1)$

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  $\lambda$ , i.e.,  $\alpha_l^*(\lambda) = \alpha_l^*$ .  $\square$

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  $\alpha_l^*$ .

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

*Proof* By Eq. 3,

$$\pi_\alpha(\lambda; 1, 2l) > \pi_\alpha(2(1 + \delta)\lambda; 2, 2l - 1) \quad (93)$$

if and only if

$$s_{\delta, \lambda}^{(l)}(\alpha) := e^{(\frac{1}{2} + \delta)\lambda} \sum_{k=0}^{\infty} \left(\frac{\lambda}{2}\right)^k \frac{1}{k!} c_{1, 2l; k}^\alpha > \sum_{k=0}^{\infty} \frac{(1 + \delta)^k \lambda^k}{k!} c_{2, 2l - 1; k}^\alpha =: t_{\delta, \lambda}^{(l)}(\alpha). \quad (94)$$

From Eqs. 8 and 6 we find that

$$c_{1, 2l; k}^\alpha = \Pr[b_{2l, 1 + 2k} < b_{2l, 1; \alpha}] \quad (95)$$

$$= \frac{\Gamma(l + \frac{1}{2} + k)}{\Gamma(l)\Gamma(\frac{1}{2} + k)} \int_0^{b_{2l, 1; \alpha}} b^{l-1} (1 - b)^{k - \frac{1}{2}} db; \quad (96)$$

$$\alpha = \Pr[b_{2l, 1} < b_{2l, 1; \alpha}] \quad (97)$$

$$= \frac{\Gamma(l + \frac{1}{2})}{\Gamma(l)\Gamma(\frac{1}{2})} \int_0^{b_{2l, 1; \alpha}} b^{l-1} (1 - b)^{-\frac{1}{2}} db. \quad (98)$$

Set  $u = 1 - b$  in Eqs. 97-98, then differentiate with respect to  $\alpha$  to obtain

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

$$\frac{d}{d\alpha} c_{1, 2l; k}^\alpha = \frac{\Gamma(l + \frac{1}{2} + k)\Gamma(\frac{1}{2})}{\Gamma(l + \frac{1}{2})\Gamma(\frac{1}{2} + k)} (1 - b_{2l, 1; \alpha})^k \quad (100)$$

$$\frac{d}{d\alpha} s_{\delta, \lambda}^{(l)}(\alpha) = e^{(\frac{1}{2} + \delta)\lambda} \sum_{k=0}^{\infty} \left(\frac{\lambda}{2}\right)^k \frac{1}{k!} \frac{\Gamma(l + \frac{1}{2} + k)\Gamma(\frac{1}{2})}{\Gamma(l + \frac{1}{2})\Gamma(\frac{1}{2} + k)} (1 - b_{2l, 1; \alpha})^k \quad (101)$$

$$= \left[ \sum_{k=0}^{\infty} \frac{(\frac{1}{2} + \delta)^k \lambda^k}{k!} \right] \left[ \sum_{k=0}^{\infty} \left(\frac{\lambda}{2}\right)^k \frac{1}{k!} \frac{\Gamma(l + \frac{1}{2} + k)\Gamma(\frac{1}{2})}{\Gamma(l + \frac{1}{2})\Gamma(\frac{1}{2} + k)} (1 - b_{2l, 1; \alpha})^k \right] \quad (102)$$

$$= \sum_{k=0}^{\infty} \frac{(\frac{1}{2} + \delta)^k}{k!} \left[ \sum_{r=0}^k \binom{k}{r} \frac{1}{(1 + 2\delta)^r} \frac{\Gamma(l + \frac{1}{2} + r)\Gamma(\frac{1}{2})}{\Gamma(l + \frac{1}{2})\Gamma(\frac{1}{2} + r)} (1 - b_{2l, 1; \alpha})^r \right] \lambda^k. \quad (103)$$

ON THE FEASIBILITY OF PARSIMONIOUS VARIABLE SELECTION...

Next,

$$\alpha = \Pr[b_{2l-1,2} < b_{2l-1,2;\alpha}] \quad (104)$$

$$= \frac{\Gamma(l+\frac{1}{2})}{\Gamma(l-\frac{1}{2})\Gamma(1)} \int_0^{b_{2l-1,2;\alpha}} b^{l-\frac{3}{2}} db = b_{2l-1,2;\alpha}^{l-\frac{1}{2}}, \quad (105)$$

$$b_{2l-1,2;\alpha} = \alpha^{\frac{2}{2l-1}}; \quad (106)$$

$$c_{2,2l-1;k}^\alpha = \Pr[b_{2l-1,2+2k} < \alpha^{\frac{2}{2l-1}}] \quad (107)$$

$$= \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(l-\frac{1}{2})\Gamma(1+k)} \int_0^{\alpha^{\frac{2l}{2l-1}}} b^{l-\frac{3}{2}}(1-b)^k db \quad (108)$$

$$= \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(l+\frac{1}{2})\Gamma(1+k)} \int_0^\alpha (1-w^{\frac{2l}{2l-1}})^k dw; \quad (109)$$

$$\frac{d}{d\alpha} c_{2,2l-1;k}^\alpha = \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(l+\frac{1}{2})\Gamma(1+k)} (1-\alpha^{\frac{2l}{2l-1}})^k; \quad (110)$$

$$\frac{d}{d\alpha} t_{\delta,\lambda}^{(l)}(\alpha) = \sum_{k=0}^\infty \frac{(1+\delta)^k \lambda^k}{k!} \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(l+\frac{1}{2})\Gamma(1+k)} (1-\alpha^{\frac{2l}{2l-1}})^k. \quad (111)$$

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  $k \geq 0$ ,

$$\frac{(\frac{1}{2}+\delta)^k}{(1+\delta)^k (1-\alpha^{\frac{2l}{2l-1}})^k} \sum_{r=0}^k \binom{k}{r} \frac{(1-b_{2l,1;\alpha})^r}{(1+2\delta)^r} \frac{\Gamma(l+\frac{1}{2}+r)}{\Gamma(\frac{1}{2}+r)} \geq \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}, \quad (112)$$

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  $k \geq 0$ ,

$$\frac{(\frac{1}{2}+\delta)^k}{(1+\delta)^k} \sum_{r=0}^k \binom{k}{r} \frac{1}{(1+2\delta)^r} \frac{\Gamma(l+\frac{1}{2}+r)}{\Gamma(\frac{1}{2}+r)} \geq \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}, \quad (113)$$

with strict inequality for at least one  $k$ . After some algebra, Eq. 113 can be written equivalently as

$$\mathbb{E} \left[ \frac{\Gamma(l+\frac{1}{2}+R_{k,\delta})}{\Gamma(\frac{1}{2}+R_{k,\delta})} \right] \geq \frac{\Gamma(l+\frac{1}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}, \quad (114)$$

where  $R_{k,\delta} \sim \text{Binomial}(k, \frac{1}{1+\delta})$ . Because  $R_{k,\delta}$  is (strictly) stochastically decreasing in  $\delta$  (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 Eq. 114  $\equiv$  Eq. 113 is (strictly) decreasing in  $\delta$  (for  $k \geq 1$ ).

For  $k = 0$  both sides of Eq. 113 =  $\frac{\Gamma(l+\frac{1}{2})}{\Gamma(\frac{1}{2})}$ . For  $k = 1$ , Eq. 113 is equivalent to the inequality

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

which is equivalent to  $\delta \leq \frac{1}{2l-1}$ . Therefore the sufficient condition Eq. 113 for

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

will be satisfied for all  $\delta \leq \frac{1}{2l-1}$  if Eq. 113 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 Eqs. 93-94 that Eq. 113, with strict inequality for some  $k \geq 2$ , is a sufficient condition that for each  $\lambda > 0$ , there exists  $\alpha_l^*(\lambda) \in (0, 1)$  such that

$$0 < \alpha < \alpha_l^*(\lambda) \implies \pi_\alpha(\lambda; 1, 2l) > \pi_\alpha\left(\left(\frac{4l}{2l-1}\right)\lambda; 2, 2l-1\right), \quad (115)$$

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

For the simplest case  $l = 1$  ( $N = 3$ ), Eq. 113 with  $\delta = \frac{1}{2l-1} = 1$  becomes

$$\left(\frac{3}{4}\right)^k \sum_{r=0}^k \binom{k}{r} \frac{1}{3^r} \left(\frac{1}{2} + r\right) \geq \frac{\Gamma(\frac{3}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}, \quad (116)$$

which by Eq. 114 can be reduced to the equivalent form

$$1 + \frac{k}{2} \geq \frac{\Gamma(\frac{3}{2}+k)}{\Gamma(\frac{3}{2})\Gamma(1+k)}. \quad (117)$$

It is straightforward to verify Eq. 117 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 Eq. 115 holds for  $l = 1$ ; that is, for each  $\lambda > 0$ , there exists  $\alpha_1^*(\lambda) \in (0, 1)$  such that

$$0 < \alpha < \alpha_1^*(\lambda) \implies \pi_\alpha(\lambda; 1, 2) > \pi_\alpha(4\lambda; 2, 1), \quad (118)$$

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

Next consider the case  $l = 2$  ( $N = 5$ ). With  $\delta = \frac{1}{2l-1} = \frac{1}{3}$ , Eq. 113 becomes

$$\left(\frac{5}{8}\right)^k \sum_{r=0}^k \binom{k}{r} \left(\frac{3}{5}\right)^r \left(\frac{3}{2} + r\right) \left(\frac{1}{2} + r\right) \geq \frac{\Gamma(\frac{5}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}, \quad (119)$$



which can be reduced to the equivalent form

$$\frac{3}{4} + \frac{9k}{8} + \frac{9k(k-1)}{64} \geq \frac{\Gamma(\frac{5}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)} = \left(\frac{3}{2} + k\right)\left(\frac{1}{2} + k\right) \frac{\Gamma(\frac{1}{2}+k)}{\Gamma(\frac{1}{2})\Gamma(1+k)}. \quad (120)$$

Interestingly, Eq. 120 holds with equality for  $k = 2$  as well as for 0 and 1. Rewrite Eq. 120 in the equivalent form

$$\frac{k(k-1)\cdots 2 \cdot 1}{(k-\frac{1}{2})(k-\frac{1}{2})\cdots \frac{3}{2} \cdot \frac{1}{2}} \geq \frac{48+128k+64k^2}{48+63k+9k^2}. \quad (121)$$

To verify (121) by induction on  $k$ , it suffices to show that for  $k \geq 2$ ,

$$\frac{k+1}{k+\frac{1}{2}} \frac{48+128k+64k^2}{48+63k+9k^2} \geq \frac{48+128(k+1)+64(k+1)^2}{48+63(k+1)+9(k+1)^2}. \quad (122)$$

After simplification, this is equivalent to the inequality

$$4k^3 + 4k^2 - 5k - 3 \geq 0, \quad (123)$$

which holds for all  $k \geq 1$ , with strict inequality for  $k + 1 \geq 3$ . Thus Eq. 115 holds for  $l = 2$ : for each  $\lambda > 0$ , there exists  $\alpha_1^*(\lambda) \in (0, 1)$  such that

$$0 < \alpha < \alpha_2^*(\lambda) \implies \pi_\alpha(\lambda; 1, 4) > \pi_\alpha\left(\frac{8}{3}\lambda; 2, 3\right), \quad (124)$$

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

**Example 4.1** s Return to the bivariate Example 2.1, where  $p = 2$  and

$$\Lambda_1 = N\gamma_1^2, \quad \Lambda_2 = N\eta^2\gamma_1^2, \quad \Lambda = N\left(\frac{1-2\eta\rho+\eta^2}{1-\rho^2}\right)\gamma_1^2; \quad (125)$$

(recall Eq. 27). For  $N = 3$  it follows from Eqs. 118 and 125 that for each  $\gamma_1^2 > 0$ ,

$$0 < \alpha < \alpha_1^*(3\gamma_1^2) \implies \pi_\alpha(3\gamma_1^2; 1, 2) > \pi_\alpha(4 \cdot 3\gamma_1^2; 2, 1). \quad (126)$$

Furthermore,

$$4 \cdot 3\gamma_1^2 > 3\left(\frac{1-2\eta\rho+\eta^2}{1-\rho^2}\right)\gamma_1^2 \iff 4 > \frac{1-2\eta\rho+\eta^2}{1-\rho^2} \quad (127)$$

$$\iff 0 > 4\rho^2 - 2\eta\rho + (\eta^2 - 3) := h_\eta(\rho). \quad (128)$$

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

$$\rho \in (\hat{\rho}_\eta^-, \hat{\rho}_\eta^+) \implies \max(\pi_\alpha(\Lambda_1; 1, 2), \pi_\alpha(\Lambda_2; 1, 2)) > \pi_\alpha(\Lambda; 2, 1) \quad (129)$$

$$\implies |\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1; \quad (130)$$

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  $(\eta, \rho)$ -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 Eqs. 124 and 125 that for each  $\gamma_1^2 > 0$ ,

$$0 < \alpha < \alpha_2^*(5\gamma_1^2) \implies \pi_\alpha(5\gamma_1^2; 1, 4) > \pi_\alpha\left(\frac{8}{3} \cdot 5\gamma_1^2; 2, 3\right). \quad (131)$$

Furthermore,

$$\frac{8}{3} \cdot 5\gamma_1^2 > 5\left(\frac{1-2\eta\rho+\eta^2}{1-\rho^2}\right)\gamma_1^2 \iff \frac{8}{3} > \frac{1-2\eta\rho+\eta^2}{1-\rho^2} \quad (132)$$

$$\iff 0 > \frac{8}{3}\rho^2 - 2\eta\rho + (\eta^2 - \frac{5}{3}) := \tilde{h}_\eta(\rho). \quad (133)$$

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

$$\rho \in (\tilde{\rho}_\eta^-, \tilde{\rho}_\eta^+) \implies \max(\pi_\alpha(\Lambda_1; 1, 4), \pi_\alpha(\Lambda_2; 1, 4)) > \pi_\alpha(\Lambda; 2, 3) \quad (134)$$

$$\implies |\hat{\omega}_\alpha(\gamma_1, \eta, \rho)| = 1; \quad (135)$$

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

Thus for  $p = 2$ ,  $N = 3$  or  $5$ , and sufficiently small  $\alpha$  (but depending on  $\gamma_1^2$ ), the bivariate size- $\alpha$   $T^2$ -test is dominated by at least one of the two univariate size- $\alpha$   $t^2$ -tests over a fairly large portion of the entire alternative hypothesis, comprising local, intermediate, and distant alternatives.

## 5 Concluding Remarks

For the purpose of encouraging future research, the questions raised in this report are stated formally as follows:

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- $\alpha$   $t^2$ -tests exceeds that of the bivariate size- $\alpha$   $T^2$ -test

$\eta$	$(\hat{\rho}_\eta^-, \hat{\rho}_\eta^+)$	$(\tilde{\rho}_\eta^-, \tilde{\rho}_\eta^+)$
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)

(Note that  $\hat{\rho}_{-\eta}^\pm = -\hat{\rho}_\eta^\mp$  and  $\tilde{\rho}_{-\eta}^\pm = -\tilde{\rho}_\eta^\mp$ )

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

$$A_\alpha(i) \equiv \{(\gamma, R) \mid |\hat{\omega}_\alpha(\gamma, R)| = i\}, \quad i = 1, \dots, p. \quad (136)$$

**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.

Finally, Example 5.1 illustrates the gain in power that ideally can be attained by variable selection limited to univariate subsets even after the crude Bonferonni correction for multiple testing is applied.

**Example 5.1** For  $p = 10$ ,  $N = 12, 22, 41$ , and  $\alpha = .05$ , Table 6 shows the gain in power obtained by the Bonferroni-corrected test  $T_\omega^2$  with the oracular

Table 6: The powers of the overall  $T^2$  test and optimal univariate test  $T_\omega^2$  under a most-favorable parameter configuration where their noncentralities coincide:  $\Lambda = \Lambda_\omega = 18$

$N$	$\pi_{.05}(\Lambda; p, N - p)$	$\pi_{.005}(\Lambda_\omega;  \hat{\omega} , N -  \hat{\omega} )$
12	0.134	0.756
22	0.548	0.849
41	0.707	0.890
121	0.801	0.914

Here  $p = 10$  and  $\alpha = .05$

subset  $\hat{\omega} \equiv \hat{\omega}_\alpha(\gamma, R)$  when  $|\hat{\omega}| = 1$  and  $\Lambda = \Lambda_{\hat{\omega}} = 18$ . The gain in power can be substantial unless  $N$  is very large. (The powers are from Tiku, 1967.)

*Acknowledgements.* This work owes much to the late Somesh Das Gupta, my colleague, teacher, and friend. I am also grateful to David Perlman for raising the questions addressed here and providing supporting data, and to an anonymous referee who provided many insightful comments and suggestions.

## Declarations

**Conflicts of interest** No funding was received to assist with the preparation of this manuscript. The author has no financial or proprietary interests in any material discussed in this article.

## References

- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis, 3rd edition*, Wiley & Sons, New York.
- 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. Stat. Assoc.* **69** 174-180.
- Giri, N. and Kiefer, J. (1964). Local and asymptotic minimax properties of multivariate tests. *Ann. Math. Stat.* **35** 21-35.
- Giri, N., Kiefer, J., and Stein, C. (1963). Minimax character of Hotelling's  $T^2$ -test in the simplest case. *Ann. Math. Stat.* **34** 1524-1535.
- 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. Stat.* **36** 747-770.
- Marden, J. and Perlman, M. D. (1980). Invariant tests for means with covariates. *Ann. Stat.* **8** 25-63.
- Stein, C. (1956). The admissibility of Hotelling's  $T^2$ -test. *Ann. Math. Stat.* **27** 616-623.
- Tiku, M. L. (1967). Tables of the power of the  $F$ -test. *J.Amer. Stat. Assoc.* **62** 525-539.

*Publisher's Note* Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## 6 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  $\Omega_p$ , in general  $\Lambda_\omega \leq \Lambda_{\omega'}$ . The question of whether the power of the  $T_{\omega'}$ -test exceeds that of the  $T_\omega$ -test for the testing problem Eq. 9 usually was formulated as the problem of *testing for additional information (TAI)*, namely, testing

$$\Lambda'_{\omega} = \Lambda_{\omega} \quad \text{vs.} \quad \Lambda'_{\omega} > \Lambda_{\omega} \quad (137)$$

based on a preliminary sample – see [Rao (1973)] §8c.4. This formulation was adopted by many researchers, even while citing the following result of Das Gupta and Perlman (1974) 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 Eq. 2) of the non-central  $f$ -test is strictly decreasing in  $m$  and strictly increasing in  $n$ .<sup>8</sup> Therefore for any integer  $1 \leq q \leq n - 1$  there exists a unique real number

$$g_\alpha(\lambda) := g_\alpha(\lambda; m, n, q) > 0 \quad (138)$$

such that

$$\pi_\alpha(\lambda; m, n) = \pi_\alpha(\lambda + g_\alpha(\lambda); m + q, n - q). \quad (139)$$

Here  $g_\alpha(0) = 0$  and  $g_\alpha(\lambda)$  is strictly increasing in  $\lambda$ ; cf. (Theorem 3.1, Das Gupta and Perlman, 1974). Thus the power is increased only if

$$\Lambda_{\omega'} > \Lambda_\omega + g_\alpha(\Lambda_\omega; |\omega|, N - |\omega|, |\omega' \setminus \omega|). \quad (140)$$

---

<sup>8</sup> In line 2 of the second column on p.179 of Das Gupta and Perlman (1974), “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, “ $m \rightarrow \infty$ ” should be “ $n \rightarrow \infty$ ”.

Therefore (Section 4, Das Gupta and Perlman, 1974) introduced the problem of *testing for increased power (TIP)*, namely, testing

$$H_1 : \Lambda_{\omega'} \leq \Lambda_{\omega} + g_{\alpha}(\Lambda_{|\omega|}; |\omega|, N - |\omega|, |\omega' \setminus \omega|)$$

vs.  $K_1 : \Lambda_{\omega'} > \Lambda_{\omega} + g_{\alpha}(\Lambda_{|\omega|}; |\omega|, N - |\omega|, |\omega' \setminus \omega|),$

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  $(\omega, \omega')$  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 A1** The relation Eq. 92 in Conjecture 4.1 can be stated equivalently in terms of  $g_{\alpha}$ :

$$0 < \alpha < \alpha_l^*(\lambda) \implies g_{\alpha}(\lambda; 1, 2l, 1) > \left(\frac{2l+1}{2l-1}\right)\lambda \quad \forall \lambda > 0, \quad (141)$$

with equality when  $\alpha = \alpha_l^*(\lambda)$ . Thus the relations Eqs. 118 and 124 in Proposition 4.3 also can be stated equivalently in terms of  $g_{\alpha}$ :

$$0 < \alpha < \alpha_1^*(\lambda) \implies g_{\alpha}(\lambda; 1, 2, 1) > 3\lambda \quad \forall \lambda > 0, \quad (142)$$

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

$$0 < \alpha < \alpha_2^*(\lambda) \implies g_{\alpha}(\lambda; 1, 4, 1) > \frac{5}{3}\lambda \quad \forall \lambda > 0, \quad (143)$$

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

## Additional References for the Appendix

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

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

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.

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.

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

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

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.

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

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

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

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.

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

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

MICHAEL D. PERLMAN  
DEPARTMENT OF STATISTICS,  
UNIVERSITY OF WASHINGTON, SEATTLE  
WA 98195, USA

E-mail: mdperlma@uw.edu.

Paper received: 19 April 2023