Statistica Sinica **16**(2006), 255-266

TESTS WITH OPTIMAL AVERAGE POWER IN MULTIVARIATE ANALYSIS

Samuel S. Wu, Hongying Li and George Casella

University of Florida

Abstract: It is well known that in a general multi-parameter setting, there may not exist any unique best test. More importantly, unlike the univariate case, the power of different test procedures could vary remarkably. In this article we extend results of Hsu (1945) and introduce a new class of tests that have best average power for multivariate linear hypotheses. A simple method to implement the new tests is also provided.

Key words and phrases: Average power, multivariate linear hypotheses, multivariate location problem, Fisher's method of combining tests, U distribution.

1. Introduction

In this paper we derive tests for multivariate linear hypotheses that have best average power. We focus on the simplest special case of testing hypothesis about a one sample normal mean vector. Applications to two-sample problems, multivariate analysis of variance and linear regression are also discussed.

Assume that $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})'$, $1 \le i \le n$, are independent random samples from a p-variate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The problem is to test the hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus alternatives $H_a : \boldsymbol{\mu} \ne \mathbf{0}$. The T^2 test proposed by Hotelling (1931) is probably the best known test for this problem. The test statistic is defined as $T^2 = n\overline{\boldsymbol{Y}}'\mathbf{S}^{-1}\overline{\boldsymbol{Y}}$, where $\overline{\boldsymbol{Y}}$ is the sample mean and $\mathbf{S}$ is the sample covariance matrix. Hotelling's test is the likelihood ratio test and is uniformly most powerful (UMP) among all tests that are invariant under the group of nonsingular linear transformations (see Anderson (1984)). In fact of all tests of $\boldsymbol{\mu} = \mathbf{0}$ with power depending only on $n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, the T^2 test is UMP (Simaika (1941)). Also, the T^2 test is admissible. Further, it is minimax when $p = 2$, locally minimax as $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \to 0$ and asymptotically minimax as $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \to \infty$. (Pillai (1985) and the references therein). In addition, Kariya (1981) has shown that the test is UMP invariant under a broader class of distributions.

On the other hand, Hotelling's T^2 test can be viewed as combination of univariate t-tests. Through the union-intersection principle of test construction of Roy (1953), the T^2 test rejects the original multivariate hypothesis if and only if *all* univariate hypotheses $H_0(\mathbf{a}) : \mathbf{a}'\boldsymbol{\mu} = \mathbf{0}$ specified by varying elements of $\mathbf{a}$ are rejected. More specifically, the Hotelling's T^2 test statistic can be derived as the maximum of univariate t^2: $T^2 = \sup_{\mathbf{a}}\{n(\mathbf{a}'\overline{\mathbf{Y}})^{\mathbf{2}}/\mathbf{a}'\mathbf{S}\mathbf{a}\}$ (see Morrison (1990)). Therefore, the T^2 test can be considered as an extension of the Tippett (1931) method for combining finite number of tests to the combination of nonindependent univariate t-tests over infinite directions. However, Birnbaum (1954), Littell and Folks (1971), Goutis, Casella and Wells (1996) have shown that the method due to Fisher (1932) for combining probabilities is asymptotically most efficient among essentially all methods of combining independent tests. In addition, Wu (2003a) has shown that the power difference between the T^2 test and the Fisher's method could be as large as 0.5 in both directions for many alternatives. This motivates us to study tests that have best average power.

Hsu (1945) has proved optimal properties of the T^2 test that involves averaging the power over $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Although Theorem 3 in that paper also shows that there exist other exact tests which maximize average power weighted by certain functions of $\boldsymbol{\Sigma}$, Hsu dismissed those tests because of the great difficulty in numerical computation. In this article, we first derive tests that have best average power over a wide class of weighting schemes, which includes Hsu's results as special cases. Secondly, we provide an easy method to implement the new tests utilizing random samples from a p-variate unit ball. We also, for the first time, derive tests with optimal average power for multivariate linear hypotheses, where dimensions of the entire parameter space and the null space differ by more than one.

The article is organized as follows. In the next section we derive a statistical distribution on the unit ball, called the U distribution, from the multivariate normal distribution. A sampling method from the distribution is provided based on independent Beta random variables. Section 3 contains the main result about procedures that have best average power for testing one sample normal mean. General results for testing multivariate linear hypotheses are presented in Section 4 followed by a summary and some discussion in the final section.

2. The U Distribution

In this section, we define a statistical distribution on the p-variate unit ball ($\boldsymbol{x}'\boldsymbol{x} \le 1$), which we call the U distribution. It will be shown later to be critical for the optimal tests derived in this work.

Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})'$, $1 \le i \le n$, be independent and identically distributed random samples from a p-variate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. It is well known that the sample mean $\overline{\boldsymbol{Y}}$ is normally distributed and independent of $\boldsymbol{A} = \sum_{i=1}^n (\boldsymbol{Y}_i - \overline{\boldsymbol{Y}})(\boldsymbol{Y}_i - \overline{\boldsymbol{Y}})' \equiv (n-1)\mathbf{S}$, which has Wishart distribution $W(\boldsymbol{\Sigma}, n-1)$. Now, let $\boldsymbol{C} = \sum_{i=1}^n \boldsymbol{Y}_i\boldsymbol{Y}_i' = \boldsymbol{A} + n\overline{\boldsymbol{Y}}\,\overline{\boldsymbol{Y}}'$ and $\boldsymbol{X} = \sqrt{n}\boldsymbol{T}'^{-1}\overline{\boldsymbol{Y}}$, where $\boldsymbol{T}$ is the Cholesky decomposition of $\boldsymbol{C}$ such that $\boldsymbol{C} = \boldsymbol{T}'\boldsymbol{T}$. By the change of variable theorem, we can show the following about joint distribution of $\boldsymbol{C}$ and $\boldsymbol{X}$:

Theorem 1. (a) *$\boldsymbol{C}$ and $\boldsymbol{X}$ have joint density given by:*

$$k|\boldsymbol{\Sigma}|^{-\frac{n}{2}}|\boldsymbol{C}|^{\frac{n-p-1}{2}} \exp\Big(-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}(\boldsymbol{C} - 2\sqrt{n}\boldsymbol{T}'\boldsymbol{X}\boldsymbol{\mu}' + n\boldsymbol{\mu}\boldsymbol{\mu}')\Big) \\ (1-\boldsymbol{X}'\boldsymbol{X})^{(n-p-2)/2}, \tag{1}$$

where $k = 1/(2^{pn/2}\pi^{p(p+1)/4}\prod_{i=1}^{p}\Gamma[(n-i)/2])$, $\boldsymbol{C}$ is a positive definite symmetric matrix, and $\boldsymbol{X}$ is a p-variate vector such that $\boldsymbol{X}'\boldsymbol{X} \leq 1$.

(b) *If $\boldsymbol{\mu} = \mathbf{0}$ then $\boldsymbol{C}$ and $\boldsymbol{X}$ are independent, $\boldsymbol{C}$ has Wishart distribution $W(\boldsymbol{\Sigma}, n)$, and $\boldsymbol{X}$ has probability density $(\prod_{i=1}^{p}\Gamma[(n+1-i)/2]/\Gamma[(n-i)/2]/\sqrt{\pi})(1-\boldsymbol{X}'\boldsymbol{X})^{(n-p-2)/2}$, where $\boldsymbol{X}'\boldsymbol{X} \leq 1$.*

Note that, under the null hypothesis, the distribution of $\boldsymbol{X}$ does *not* depend on $\boldsymbol{\Sigma}$. We refer to the distribution of $\boldsymbol{X}$ defined in part (b) as the U distribution from a p-variate unit ball with $n-2$ degrees of freedom, denoted by $U_{p,n-2}$. The distribution is spherically symmetric, so we only need to focus on the positive quadrant in order to generate random samples from $U_{p,n-2}$. Under the polar coordinate system ($X_1 = r\cos\phi_1$, $X_j = r\sin\phi_1\cdots\sin\phi_{j-1}\cos\phi_j$ for $2 \leq j \leq p-1$, $X_p = r\sin\phi_1\cdots\sin\phi_{p-2}\sin\phi_{p-1}$), it is straightforward to show that the U distribution has density proportional to

$$(1-r^2)^{\frac{n-p-2}{2}}r^{p-1}|(\sin\phi_1)^{p-2}(\sin\phi_2)^{p-3}\cdots\sin\phi_{p-2}|.$$

Therefore, if we let $Z_j = (\sin\phi_j)^2$, $1 \leq j \leq p-1$ and $Z_p = r^2$, under the U distribution we have $Z_j \sim Beta((p-j)/2, 1/2), 1 \leq j \leq p-1$; $Z_p \sim Beta(p/2, (n-p)/2)$, and they are mutually independent. In summary, it is very easy to generate random samples from the U distribution based on independent Beta random variables.

3. The Main Result

3.1. Optimal tests

Let $\beta_\omega(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the power function of critical region ω for the rejection of $H_0 : \boldsymbol{\mu} = \mathbf{0}$. In this section we derive statistical tests that have best average power weighted by

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|l, \theta, m, \phi) = |\boldsymbol{\Sigma}|^{-\frac{l}{2}} \exp\Big\{-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}[\theta + m(\boldsymbol{\mu}-\phi)(\boldsymbol{\mu}-\phi)']\Big\}, \tag{2}$$

over $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \in (0, s)$, where $l \geq 0$, θ is a $p \times p$ positive definite matrix, $m \geq 0$, ϕ is a $p \times 1$ vector and $0 < s \leq \infty$. In other words, we consider the following average power:

$$\Gamma_\omega(l, \theta, m, \phi, s) = \int_{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \in (0,s)} \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma} | l, \theta, m, \phi)\, \beta_\omega(\boldsymbol{\mu}, \boldsymbol{\Sigma})\, d\boldsymbol{\mu}\, d\boldsymbol{\Sigma}. \tag{3}$$

If we let $l = m = 0$, then the weighting function reduces to $\exp\{-(1/2) tr \boldsymbol{\Sigma}^{-1} \theta\}$, which is the case studied in Hsu (1945). Theorem 3 in the Hsu paper showed that there exists a threshold function g such that rejection region $\overline{\boldsymbol{Y}}'(\boldsymbol{C}+\theta)^{-1}\overline{\boldsymbol{Y}} \geq g(\boldsymbol{C}, \theta)$ maximizes the average power $\Gamma_\omega(0, \theta, 0, \cdot, s)$ for any $s < \infty$. Hsu also pointed out that, given $\boldsymbol{C}$ and θ, the threshold $g(\boldsymbol{C}, \theta)$ can be obtained by solving the equation

$$\Big(\prod_{i=1}^{p} \frac{\Gamma[\frac{n+1-i}{2}]}{\sqrt{\pi}\Gamma[\frac{n-i}{2}]}\Big) \int_{\sum_{i=1}^{p}[\lambda_i/(1+\lambda_i)]x_i^2 \geq g} (1 - \boldsymbol{x}'\boldsymbol{x})^{\frac{n-p-2}{2}} \Pi d\boldsymbol{x} = \alpha,$$

where $\lambda_i, 1 \leq i \leq p$, are eigenvalues of matrix $\boldsymbol{C}\theta^{-1}$. However, the difficulty in evaluating the threshold at that time prohibited applications of the test.

In this article, we extend Hsu's Theorem 3 and derive tests with optimal average power defined in (3). Our main result is stated in the following theorem, whose proof can be found in a technical report (Wu (2003b)).

Theorem 2. *For any given $\boldsymbol{C} = \boldsymbol{T}'\boldsymbol{T}$, we let $g(\boldsymbol{C}, m, \phi, \theta)$ be the upper αth percentile of $(\sqrt{n}\boldsymbol{T}'\boldsymbol{U} + m\phi)'(\boldsymbol{C} + \theta + m\phi\phi')^{-1}(\sqrt{n}\boldsymbol{T}'\boldsymbol{U} + m\phi)$, where $\boldsymbol{U}$ is a random variable with the $U_{p,n-2}$ distribution. Then the region*

$$\omega_1 : (n\overline{\boldsymbol{Y}} + m\phi)'(\boldsymbol{C} + \theta + m\phi\phi')^{-1}(n\overline{\boldsymbol{Y}} + m\phi) \geq g(\boldsymbol{C}, m, \phi, \theta) \tag{4}$$

satisfies $\beta_\omega(\boldsymbol{0}, \boldsymbol{\Sigma}) = \alpha$ for all $\boldsymbol{\Sigma}$ and maximizes $\Gamma_\omega(l, \theta, m, \phi, s)$ whenever $m > 0$ or $s < \infty$.

We have shown in Section 2 that it is easy to generate random samples from the U distribution. Therefore the threshold function $g(\boldsymbol{C}, m, \phi, \theta)$ can be evaluated very quickly, usually only requiring about 1,000 samples from the $U_{p,n-2}$ based on our experience. More importantly, such a method also allows us to construct α level tests with data dependent θ and ϕ. The results are summarized in the following theorem.

Theorem 3. *Let $h(\boldsymbol{C}, \overline{\boldsymbol{Y}})$ be an arbitrary test statistic involving $\boldsymbol{C}$ and $\overline{\boldsymbol{Y}}$. For any given $\boldsymbol{C} = \boldsymbol{T}'\boldsymbol{T}$, if $g(\boldsymbol{C})$ is the upper αth percentile of $h(\boldsymbol{C}, \boldsymbol{T}'\boldsymbol{U}/\sqrt{n})$ where $\boldsymbol{U} \sim U_{p,n-2}$, then a rejection region defined by $\omega : h(\boldsymbol{C}, \overline{\boldsymbol{Y}}) \geq g(\boldsymbol{C})$ satisfies $\beta_\omega(\boldsymbol{0}, \boldsymbol{\Sigma}) = \alpha$ for all $\boldsymbol{\Sigma}$.*

Proof of the above theorem is straightforward after noting that $h(\boldsymbol{C}, \overline{\boldsymbol{Y}})$ and $h(\boldsymbol{C}, \boldsymbol{T}'\boldsymbol{U}/\sqrt{n})$ have the same probability distribution under the null hypothesis. It is worth noting that this result is a special case of the necessary and sufficient conditions of Simaika (1941) for a test to be level α for all covariance matrices.

3.2. Connection to Fisher's method of combining tests

Wu (2003a) compared the T^2 test with Fisher's combination of coordinate-wise tests, and also showed that Fisher's method is asymptotically equivalent to $S_1 = n\overline{\boldsymbol{Y}}' diag(\boldsymbol{S})^{-1}\overline{\boldsymbol{Y}}$, which is the sum of univariate t^2. This test can also be derived from (4) by choosing ϕ to be any vector in the same direction as $\overline{\boldsymbol{Y}}$, and $\theta = \tau diag(\boldsymbol{S})$ with $\tau \to \infty$.

It is obvious that Fisher's test statistic ignores the off-diagonal elements of the sample covariance matrix. Some intermediate approach is to shrink the off-diagonal elements. In other words we can base our tests on $n\overline{\boldsymbol{Y}}'[\boldsymbol{S} * V]^{-1}\overline{\boldsymbol{Y}}$ ($\boldsymbol{S} * V$ denotes the Hadamard product of matrices, with entries $s_{ij}v_{ij}$). This test is equivalent to the following in terms of $\boldsymbol{C}$ and $\overline{\boldsymbol{Y}}$: reject on

$$\omega_s : n(n-1)\overline{\boldsymbol{Y}}'[(\boldsymbol{C} - n\overline{\boldsymbol{Y}}\,\overline{\boldsymbol{Y}}') * V]^{-1}\overline{\boldsymbol{Y}} \geq g(\boldsymbol{C}, V), \tag{5}$$

where $g(\boldsymbol{C}, V)$ is the upper αth percentile of $(n-1)(\boldsymbol{T}'\boldsymbol{U})'[(\boldsymbol{C} - \boldsymbol{T}'\boldsymbol{U}\boldsymbol{U}'\boldsymbol{T}) * V]^{-1}(\boldsymbol{T}'\boldsymbol{U})$, $\boldsymbol{U} \sim U_{p,n-2}$. If V is the identity matrix, then the left hand side of the above test reduces to S_1, which only utilizes coordinate wise tests. We denote by S_2 the test defined by (5) with V chosen to be a autoregressive covariance matrix with correlation 0.3. In addition, S_3 is the test corresponding to a complete symmetric covariance V with correlation 0.3.

Figures 1 and 2 compare the power functions of S_1, S_2, S_3 with that of Hotelling's test T^2 and a pseudo-test $T^* = \overline{\boldsymbol{Y}}'\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{Y}}$ which uses the unknown covariance matrix $\boldsymbol{\Sigma}$. Assuming a multivariate normal distribution with constant correlation, the powers for testing zero mean are compared based on a random sample of size 20 for four different situations determined by (a) two mean shifts: $s_1 = (0.4, 0.4, \ldots, 0.4)'$ and $s_2 = 0.4p(1, 2^2, \ldots, p^2)'/(\sum_{i=1}^{p} i^2)$; (b) two covariance structures: compound symmetric (CS) and first order autoregressive (AR). Since all five tests are invariant for coordinate-wise linear transformations, the covariance matrices are assumed to have homogeneous variance 1. The first mean shift assumes that the effect sizes for all dimensions are equal, while there are big differences for the second case. We estimated the power of the five tests at the 0.05 level based on 10,000 replications, with the threshold $g(\boldsymbol{C}, V)$ evaluated based on 2,000 random samples from the U distribution in each case.

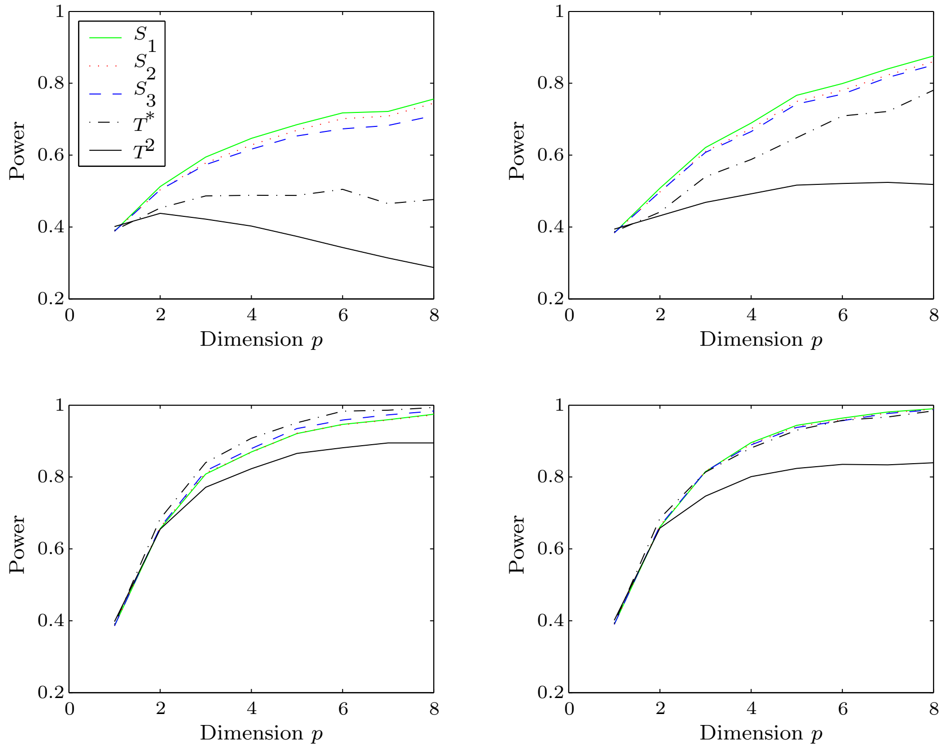

Figure 1. A comparison of power functions of the Hotelling's T^2, pseudo-test T^* and the new tests S_1, S_2, S_3 when the correlation is fixed at $\rho = 0.3$. Each plot shows the relationship between the number of dimension p and the power to reject the null hypothesis of zero mean based on a multivariate normal sample of size 20. The plots in the top row are for the alternative $\boldsymbol{\mu} = (0.4, 0.4, \ldots, 0.4)'$, while the bottom plots correspond to $\boldsymbol{\mu} \propto (1, 2^2, \ldots, p^2)'$ standardized to a 0.4 average. The plots in the left column are for a compound symmetric covariance, while those in the right column are for an auto-regressive covariance. In all four plots, the Hotelling's T^2 performs worse than the new tests except when dimension $p = 1$. And, as the number of dimension p increases, the differences in the power functions also increase. In addition, the new tests are better than the pseudo-test T^* for the first type of mean shift illustrated in the top row.

3.3. Other weight functions

The weight function π in (2) is proportional to the joint density of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distributed as $\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N(\phi, \boldsymbol{\Sigma}/m)$ and $\boldsymbol{\Sigma}^{-1} \sim W(\theta^{-1}, l+p)$. This normal-inverse-Wishart weight function is the conjugate prior distribution for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the multivariate normal model with unknown mean and variance. Let T_a and T_b be the optimal tests corresponding to $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|3, I, 1, s_1)$ and $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|3, I, 1, s_2)$, respectively. Table 1 compares these procedures with Hotelling's T^2 and Fisher method S_1 on average power when $p = 2$. The seven different weight functions are: two conjugate priors; the Jeffreys' prior; three reference priors; and one noninformative prior. Chang and Eves (1990) have shown that π_4 and π_5 are reference priors on the orbits parameterized by the correlation matrix and the eigenvalues of the covariance matrix (λ_1 and λ_2), respectively, and that π_6 is the reference prior on the orbits parameterized by the noncentrality parameter when $\mathbf{\Sigma}$ is restricted to $\sigma^2 I$. The noninformative prior π_7 satisfies Stein's sufficient condition for accurate frequentist coverage when $\mathbf{\Sigma} = I$ (Tibshirani (1989)). We averaged the power function over the region $\{(\boldsymbol{\mu}, \mathbf{\Sigma}) : \boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu} \in (0, 1.5)\}$ for the

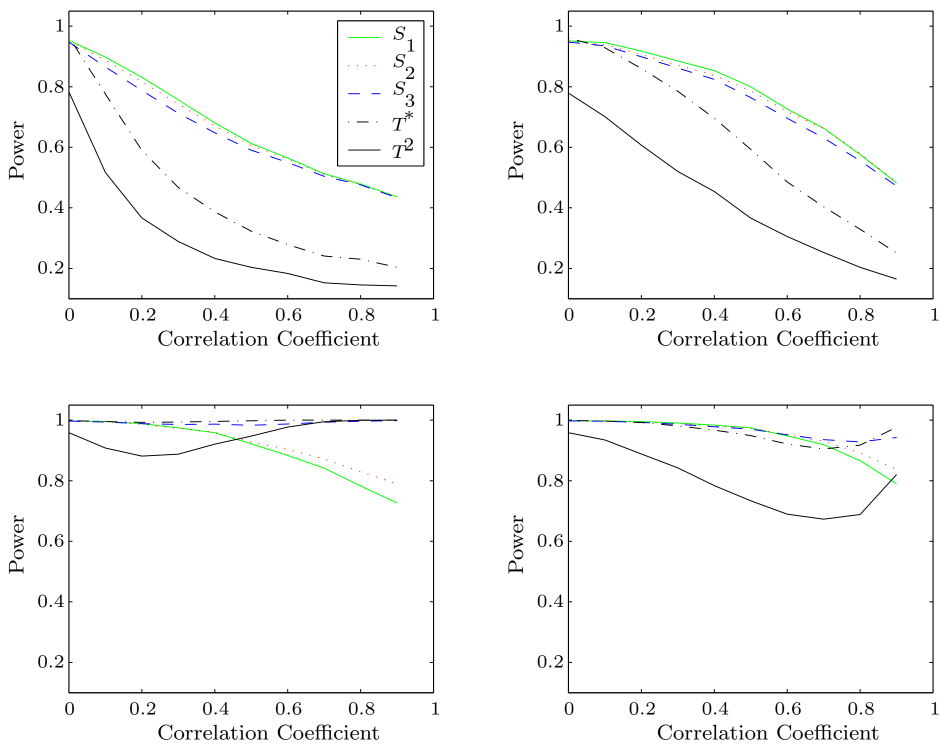

Figure 2. A comparison of power functions of the Hotelling's T^2, pseudo-test T^* and the new tests S_1, S_2, S_3 when the dimension is fixed at $p = 8$. Each plot shows the relationship between the correlation ρ and the power to reject the null hypothesis of zero mean, based on a multivariate normal sample of size 20. The mean shifts and the covariance structures are the same as in Figure 1 for the four plots. The Hotelling's T^2 performs worst except when the covariance is auto-regressive with large correlation. Once again, the "shrinkage" tests S_1, S_2, S_3 are better than the pseudo-test T^* for the first type of mean shift.

first two cases, and over the compact region $\{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : |\mu_i|/\sqrt{\sigma_{ii}} < 1, \sigma_{ii} < 1, |\sigma_{12}|/\sqrt{\sigma_{11}\sigma_{22}} < 0.9\}$ for the five improper priors, because all four tests have good power outside the two regions. The table showes that T_a and T_b have larger average powers for the conjugate priors and are nearly same as the Hotelling's test T^2 for the reference prior π_6, but T^2 is best in other cases.

Table 1. Comparison of four tests on average power weighted by seven different functions.

Weight functions	T^2	S_1	T_a	T_b
$\pi_1 = \|\boldsymbol{\Sigma}\|^{-\frac{3}{2}} \exp\left\{-(1/2)tr\boldsymbol{\Sigma}^{-1}[I + (\boldsymbol{\mu} - s_1)(\boldsymbol{\mu} - s_1)']\right\}$	0.72	0.71	0.74	0.74
$\pi_2 = \|\boldsymbol{\Sigma}\|^{-\frac{3}{2}} \exp\left\{-(1/2)tr\boldsymbol{\Sigma}^{-1}[I + (\boldsymbol{\mu} - s_2)(\boldsymbol{\mu} - s_2)']\right\}$	0.72	0.71	0.74	0.75
$\pi_3 = \|\boldsymbol{\Sigma}\|^{-\frac{3}{2}}$	0.80	0.72	0.73	0.72
$\pi_4 = \|\boldsymbol{\Sigma}\|^{-\frac{3}{2}} \|I + \boldsymbol{\Sigma} * \boldsymbol{\Sigma}^{-1}\|^{-\frac{1}{2}}$	0.79	0.73	0.72	0.71
$\pi_5 = \|\boldsymbol{\Sigma}\|^{-1} \|\lambda_1 - \lambda_2\|^{-1}$	0.78	0.74	0.75	0.75
$\pi_6 = \|\boldsymbol{\Sigma}\|^{-\frac{3}{4}} [4\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^2]^{-\frac{1}{2}}$	0.55	0.52	0.55	0.54
$\pi_7 = \|\boldsymbol{\Sigma}\|^{-\frac{3}{2}} (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{-1/2}$	0.57	0.52	0.52	0.52

4. Results for Multivariate Linear Hypotheses

Consider n independent multivariate normal vectors $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})'$, $1 \leq i \leq n$, with means $E(Y_{ij}) = \mu_{ij}$ and common covariance matrix $\boldsymbol{\Sigma}$. A *multivariate linear hypothesis* is defined in terms of two linear subspaces Π_Ω and Π_ω of n-dimensional space having dimensions $s < n$ and $0 \leq s - r \leq s$ (Lehmann (1986, p.453)). It is assumed known that for all $j = 1, \ldots, p$, the vectors $(\mu_{1j}, \ldots, \mu_{nj})'$ lie in the subspace Π_Ω; the hypothesis to be tested specifies that they lie in the subspace Π_ω. It is well known that hypothesis tests for two sample normal means, multivariate analysis of variance and multivariate linear regression are all special cases.

Assume that $(e_1, \ldots, e_r)$, $(e_{r+1}, \ldots, e_s)$ and $(e_{s+1}, \ldots, e_n)$ are orthogonal bases for subspaces $\Pi_\Omega \backslash \Pi_\omega, \Pi_\omega$ and $\Pi_\Omega^\perp$, respectively; let $\boldsymbol{D} = [e_1, \ldots, e_n], \boldsymbol{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n]$ and $\boldsymbol{Z} = \boldsymbol{Y}\boldsymbol{D}$. It is easy to see that, under the above transformation, our problem is reduced to the following canonical form: $\boldsymbol{Z}_i, 1 \leq i \leq n$, *are independently distributed according to a p-variate normal distributions with common covariance matrix* $\boldsymbol{\Sigma}$*. The means of* $\boldsymbol{Z}_{s+1}, \ldots, \boldsymbol{Z}_n$ *are zero, and the hypothesis to be tested is that the means of* $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_r$ *are zero.* Thus the joint distribution of the transformed observations is

$$(2\pi)^{-\frac{pn}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2} tr\boldsymbol{\Sigma}^{-1} [\sum_{i=1}^{s} (\boldsymbol{Z}_i - \boldsymbol{\mu}_i)(\boldsymbol{Z}_i - \boldsymbol{\mu}_i)' + \sum_{i=s+1}^{n} \boldsymbol{Z}_i \boldsymbol{Z}_i'] \Big\}. \quad (6)$$

It is clear that the $\boldsymbol{Z}_i, r < i < s$, can be of no use, and that the only useful quantities supplied by the $\boldsymbol{Z}_{s+1}, \dots, \boldsymbol{Z}_n$ are the statistics $\boldsymbol{A} = \sum_{i=s+1}^n \boldsymbol{Z}_i \boldsymbol{Z}_i'$.

Now let $\beta_\omega(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_r, \boldsymbol{\Sigma})$ be the power corresponding to the critical region ω, which is assumed to be a function of $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_r$ and $\boldsymbol{A}$, for testing the multivariate linear hypothesis in its canonical form. We consider maximizing the average power

$$\Gamma_\omega = \int_{\Omega(s)} \pi(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_r, \boldsymbol{\Sigma} | l, \theta, m_i, \phi_i, S, 1 \le i \le r)\, \beta_\omega(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}) \Pi d\boldsymbol{\mu}_i \, d\boldsymbol{\Sigma}, \tag{7}$$

where $\Omega(s) = \{(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}) : \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i \in (0, s),\ 1 \le i \le r\}$ and the weighting function is given by

$$\pi = |\boldsymbol{\Sigma}|^{-\frac{l}{2}} \exp\Big\{ -\frac{1}{2} tr \boldsymbol{\Sigma}^{-1} [\theta + \sum_{i=1}^r m_i (\boldsymbol{\mu}_i - \phi_i)(\boldsymbol{\mu}_i - \phi_i)'] \Big\}.$$

The optimal tests are provided by the following theorem, whose proof is found in Wu (2003b).

Theorem 4. *Let* $\boldsymbol{C} = \sum_{i=1}^r \boldsymbol{Z}_i \boldsymbol{Z}_i' + \sum_{i=s+1}^n \boldsymbol{Z}_i \boldsymbol{Z}_i'$ *and* $\mathcal{W} = [(\boldsymbol{Z}_1 + m_1\phi_1)/\sqrt{m_1+1}, \dots, (\boldsymbol{Z}_r + m_r\phi_r)/\sqrt{m_r+1}]$. *Then there is a function* $g(\boldsymbol{C}, \theta, m_i, \phi_i, 1 \le i \le r)$ *such that the rejection region*

$$\omega_2 : \det\Big(I_r - \mathcal{W}'(\boldsymbol{C} + \theta + \sum_{i=1}^r m_i \phi_i \phi_i')^{-1} \mathcal{W} \Big) \le g(\boldsymbol{C}, \theta, m_i, \phi_i, 1 \le i \le r) \tag{8}$$

satisfies $\beta_\omega(\boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{\Sigma}) = \alpha$ *and maximizes* $\Gamma_\omega(l, \theta, m_i, \phi_i, s, 1 \le i \le r)$, *provided* (1) $r = 1$, *all* $m_i > 0$ *or* $s < \infty$; *or* (2) $r \ge 2$, *all* $m_i > 0$ *AND* $s = \infty$.

Previously, this result was only known for the case $r = 1$ and $\pi = \exp\{-(1/2) tr \boldsymbol{\Sigma}^{-1} \theta\}$. For $r \ge 2$, our result only applies to the case when the power is averaged over the entire parameter space.

For a two sample location problem, we observe $\boldsymbol{Y}_i$, $1 \le i \le n_1$, from $N_p(\boldsymbol{\eta}_1, \boldsymbol{\Sigma})$ and $\boldsymbol{Y}_i, (n_1+1) \le i \le n = n_1 + n_2$, from $N_p(\boldsymbol{\eta}_2, \boldsymbol{\Sigma})$. We consider the hypothesis $H_0 : \boldsymbol{\eta}_1 = \boldsymbol{\eta}_2$. In this case we also have $r = 1$ because $\Pi_\Omega = span \begin{pmatrix} \boldsymbol{1}_{n_1} & \boldsymbol{0}_{n_1} \\ \boldsymbol{0}_{n_2} & \boldsymbol{1}_{n_2} \end{pmatrix}$ is a linear space of dimension $s = 2$ and $\Pi_\omega = span\{\boldsymbol{1}_n\}$ has dimension $s - r = 1$. It is straightforward to check that $\boldsymbol{Z}_1 = \sqrt{n_1 n_2 / n}(\overline{\boldsymbol{Y}}^{(1)} - \overline{\boldsymbol{Y}}^{(2)})$, the standardized version of the difference between the two sample means. Furthermore, we have $\boldsymbol{A} = \sum_{i=3}^n \boldsymbol{Z}_i \boldsymbol{Z}_i' = (n-2)\boldsymbol{S}$, where

$$\boldsymbol{S} = \frac{1}{n-2} \Big(\sum_{i=1}^{n_1} (\boldsymbol{Y}_i - \overline{\boldsymbol{Y}}^{(1)})(\boldsymbol{Y}_i - \overline{\boldsymbol{Y}}^{(1)})' + \sum_{i=n_1+1}^{n} (\boldsymbol{Y}_i - \overline{\boldsymbol{Y}}^{(2)})(\boldsymbol{Y}_i - \overline{\boldsymbol{Y}}^{(2)})' \Big).$$

Obviously $\boldsymbol{Z}_1$ is $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}_1 = \sqrt{n_1 n_2/n}(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2)$. Applying Theorem 4 to this special two sample problem, we have tests that have optimal average power weighted by

$$\pi(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}|l, \theta, m, \phi) = |\boldsymbol{\Sigma}|^{-\frac{l}{2}} \exp\Big\{-\frac{1}{2} tr \boldsymbol{\Sigma}^{-1}[\theta + m(\boldsymbol{\mu}_1 - \phi)(\boldsymbol{\mu}_1 - \phi)']\Big\}. \tag{9}$$

Corollary 1. *Given $\boldsymbol{Y}_i, 1 \le i \le n_1$, from $N_p(\boldsymbol{\eta}_1, \boldsymbol{\Sigma})$ and $\boldsymbol{Y}_i, (n_1+1) \le i \le n = n_1 + n_2$, from $N_p(\boldsymbol{\eta}_2, \boldsymbol{\Sigma})$, if $\boldsymbol{Z}_1 = \sqrt{n_1 n_2/n}(\overline{\boldsymbol{Y}}^{(1)} - \overline{\boldsymbol{Y}}^{(2)})$ and $\boldsymbol{C} = (n-2)\boldsymbol{S} + \boldsymbol{Z}_1 \boldsymbol{Z}_1'$, then among all rejection regions (for the hypothesis $H_0 : \boldsymbol{\eta}_1 = \boldsymbol{\eta}_2$) that are functions of $\boldsymbol{Z}_1$ and $\boldsymbol{S}$ and satisfy $\beta_\omega(\boldsymbol{0}, \boldsymbol{\Sigma}) = \alpha$ for all $\boldsymbol{\Sigma}$,*

$$\omega_3 : (\boldsymbol{Z}_1 + m\phi)'(\boldsymbol{C} + \theta + m\phi\phi')^{-1}(\boldsymbol{Z}_1 + m\phi) \ge g(\boldsymbol{C}, m, \phi, \theta) \tag{10}$$

has the maximum average power weighted by π defined in (9) whenever $m > 0$ or $s < \infty$, where $g(\boldsymbol{C}, m, \phi, \theta)$ is the upper αth percentile of $(\boldsymbol{T}'\boldsymbol{U} + m\phi)'(\boldsymbol{C} + \theta + m\phi\phi')^{-1}(\boldsymbol{T}'\boldsymbol{U} + m\phi)$, $\boldsymbol{T}$ satisfies $\boldsymbol{C} = \boldsymbol{T}'\boldsymbol{T}$, and $\boldsymbol{U} \sim U_{p,n-3}$.

5. Summary and Discussion

We have introduced a new class of procedures that have best average power for testing multivariate linear hypotheses. Our results extend Hsu (1945) not only in the choice of the weighting functions, but also in the linear hypotheses to be tested. Implementation of the new tests is provided by using random samples from the U distribution. Furthermore, "shrinkage" tests constructed from the optimal procedures compete very well against Hotelling's T^2, especially as p increases.

It is well known that, in the multiple normal means problem, the usual least squares estimator may be inadmissible (See Stein (1956) and Brown (1990)). The James-Stein shrinkage estimate dominates the usual sample mean in simultaneously estimating three (or more) parameters. Our Theorem 2 shows that, if we shrink the sample mean $\overline{\boldsymbol{Y}}$ toward any given vector ϕ and the sample covariance toward any given positive definite matrix θ, the resulting quadratic form yields a test with best average power, hence is admissible. Such tests have not been used before because of the difficulty in the computation of the rejection threshold. It can be very evaluated now based on random samples from the U distribution. We note that the permutation method, which is slightly different from our approach, is an alternative way to compute the threshold.

From another point of view, for fixed n and increasing p, the estimation of the covariance matrix becomes less accurate, and the new tests that shrink the off-diagonal elements prevail over Hotelling's T^2. We suspect that the optimal amount of shrinkage depends on the relative magnitudes of p and n.

The tests proposed here are only for multivariate normal data. However, their rejection regions can be easily implemented using the permutation method for other heavy-tailed distributions. As Wu (2003a) has shown empirically in such cases, Fisher's combination of coordinate-wise nonparametric tests can also outperform Hotelling-type tests. Whether the same optimal property in average power will hold for multivariate location tests with other probability distributions needs further investigation.

References

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

Birnbaum, A. (1954). Combining independent tests of significance. *J. Amer. Statist. Assoc.* **49**, 559-574.

Brown, L. D. (1990). An ancillarity paradox which appears in multiple linear regression. *Ann. Statist.* **18**, 471-493.

Chang, T. and D. Eaves (1990). Reference priors for the orbit in a group model. *Ann. Statist.* **18**, 1595-1614.

Fisher, R. A. (1932). *Statistical Methods for Research Workers.* 4th edition. Oliver and Boyd, London.

Goutis, C., G. Casella and M. T. Wells (1996). Assessing evidence in multiple hypotheses. *J. Amer. Statist. Assoc.* **91**, 1268-1277.

Hotelling, H. (1931). The genreralization of Student's ratio. *Ann. Math. Statist.* **2**, 360-378.

Hsu, P. L. (1945). On the Power functions for the E^2-test and the T^2-test. *Ann. Math. Statist.* **16**, 278-286.

Kariya, T. (1981). A robustness property of Hotelling's T^2-test. *Ann. Statist.* **9**, 211-214.

Lehmaun, E. L. (1986). *Testing Statistical Hypothesis.* 2nd edition. Wiley, New York.

Littell, R. C. and Folks, J. L. (1971). Asymptotic optimality of Fisher's method of combining independent tests. *J. Amer. Statist. Assoc.* **66**, 802-806.

Morrison, D. F. (1990). *Multivariate Statistical Methods.* 3rd edition. McGraw Hill, New York.

Pillai, K. C. S. (1985). Hotelling's T^2. In *Encyclopedia of Statistical Sciences* **6** (Edition by S. Kotz, N. L. Johnson and C. B. Read), 669-673. Wiley, New York.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24**, 220-238.

Simaika, J. B. (1941). On an optimum property of two important statistical tests. *Biometrika.* **32**, 70-80.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. 3rd Berkeley Symp. Math. Statist. Probab.* **1**, 197-206. Univ. California Press.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika.* **76**, 604-608.

Tippett, L. H. C. (1931). *The Methods of Statistics.* Williams and Norgate, London.

Wu, S. S. (2003a). Combining univariate tests for multivariate location problem. Technical Report, Department of Statistics, University of Florida.

Wu, S. S. (2003b). Tests with optimal average power in multivariate analysis. Technical Report, Department of Statistics, University of Florida.

Division of Biostatistics, University of Florida, Gainesville, Florida 32610, U.S.A.
E-mail: samwu@biostat.ufl.edu

Department of Statistics, University of Florida, Gainesville, Florida 32610, U.S.A.
E-mail: hli@stat.ufl.edu

Department of Statistics, University of Florida, Gainesville, Florida 32610, U.S.A.
E-mail: casella@stat.ufl.edu

(Received February 2004; accepted November 2004)