[1]

On the limiting distributions of multivariate depth-based rank sum statistics and related tests

By Yijun Zuo[2] and Xuming He[3]

Michigan State University and University of Illinois

A depth-based rank sum statistic for multivariate data introduced by Liu and Singh (1993) as an extension of the Wilcoxon rank sum statistic for univariate data has been used in multivariate rank tests in quality control and in experimental studies. Those applications, however, are based on a conjectured limiting distribution given in Liu and Singh (1993). The present paper proves the conjecture under general regularity conditions and therefore validates various applications of the rank sum statistic in the literature. The paper also shows that the corresponding rank sum tests can be more powerful than Hotelling's T^2 test and some commonly used multivariate rank tests in detecting location-scale changes in multivariate distributions.

1. Introduction. The key idea of data depth is to provide a center-outward ordering of multivariate observations. Points deep inside a data cloud get high depth and those on the outskirts get lower depth. The depth of a point decreases when

[1]Received July, 2003, revised June, 2004 and July 2005

AMS 2000 subject classifications. Primary 62G20; secondary 62G10, 62H10, 62H15

Key words and phrases. Data depth, limiting distribution, multivariate data, rank sum statistic, efficiency, two-sample problem.

[2]Research partially supported by NSF Grants DMS-0071976 and DMS-0134628.

[3]Research partially supported by NSF Grant DMS-0102411, NSA Grant H98230-04-1-0034, and NIH Grant R01 DC005603.

the point moves away from the center of the data cloud. Applications of depth-induced ordering are numerous. For example, Liu and Singh (1993) generalized, via data depth, the Wilcoxon rank sum statistic to the multivariate setting. Earlier generalizations of the statistic are due to, e.g., Puri and Sen (1971), Brown and Hettmansperger (1987) and Randles and Peters (1990). More recent ones include Choi and Marden (1997), Hettmansperger et al (1998), and Topchii et al (2003). A special version of the Liu-Singh depth-based rank sum statistic (with a reference sample) inherits the distribution-free property of the Wilcoxon rank sum statistic. The statistic discussed in this paper, like most other generalizations, is only asymptotically distribution-free under the null hypothesis. For its applications in quality control and experimental studies to detect quality deterioration and treatment effects, we refer to Liu and Singh (1993) and Liu (1992, 1995). These applications relied on a conjectured limiting distribution, provided by Liu and Singh (1993), of the depth-based rank sum statistic. Rousson (2002) made an attempt to prove the conjecture but did not handle the differentiability of the depth functionals for a rigorous treatment. The first objective of the present paper is to fill this mathematical gap by providing regularity conditions for the limiting distribution to hold, and by verifying those conditions for some commonly used depth functions. The empirical process theory, and in particular, a generalized Dvoretzk-Kiefer-Wolfowitz theorem in the multivariate setting turns out to be very useful here.

Our second objective is to investigate the power behavior of the test based on the Liu-Singh rank sum statistic. The test can outperform Hotelling's T^2 test and some other existing multivariate tests in detecting location-scale changes for a wide range of distributions. In particular, it is very powerful for detecting scale changes in the alternative, for which Hotelling's T^2 test is not even consistent.

Section 2 presents the Liu-Singh depth-based rank sum statistic and a theorem of asymptotic normality. Technical proofs of the main theorem and auxiliary lemmas are given in Section 3. The theorem is applied to several commonly used depth functions in Section 4, whereas Section 5 is devoted to a study of the power properties of the rank sum test. Concluding remarks in Section 6 ends the paper.

2. Liu-Singh statistic and its limiting distribution Let $X \sim F$ and $Y \sim G$ be two independent random variables in R^d. Let $D(y; H)$ be a depth function of a given distribution H in $\mathbb{R}^d$ evaluated at point y. Liu and Singh (1993) introduced $R(y; F) = P_F(X : D(X; F) \le D(y; F))$ to measure the relative outlyingness of y with respect to F and defined a *quality index*

$$Q(F, G) := \int R(y; F)dG(y) = P\{D(X; F) \le D(Y; F)|X \sim F, Y \sim G\}. \tag{2.1}$$

Since $R(y; F)$ is the fraction of F population that is "not as deep" as the point y, $Q(F, G)$ is the average fraction over all $y \in G$. As pointed out by Proposition 3.1 of Liu and Singh (1993), $R(Y; F) \sim U[0, 1]$ and consequently $Q(F, G) = 1/2$ when $Y \sim G = F$ and $D(X; F)$ has a continuous distribution. Thus, the index $Q(F, G)$ can be used to detect a treatment effect or quality deterioration. The Liu-Singh *depth-based rank sum statistic*

$$Q(F_m, G_n) := \int R(y; F_m)dG_n(y) = \frac{1}{n}\sum_{j=1}^{n} R(Y_j; F_m) \tag{2.2}$$

is a two sample estimator of $Q(F, G)$ based on the empirical distributions F_m and G_n. Under the null hypothesis $F = G$ (e.g., no treatment effect or quality deterioration), Liu and Singh (1993) proved in *one* dimension $d = 1$

$$\Big((1/m + 1/n)/12\Big)^{-1/2}\Big(Q(F_m, G_n) - 1/2\Big) \xrightarrow{d} N(0, 1), \tag{2.3}$$

and in *higher* dimensions, they proved the same for the Mahalanobis depth under the existence of the fourth moments, and conjectured that the same limiting distribution holds for general depth functions and in the general multivariate setting. In the next section we prove this conjecture under some regularity conditions, and generalize the result to the case of $F \neq G$ in order to perform a power study.

We first list assumptions that are needed for the main result. They will be verified in this and later sections for some commonly used depth functions. Assume without loss of generality that $m \leq n$ hereafter. Let F_m be the empirical version of F and $D(\cdot;\cdot)$ be a given depth function with $0 \leq D(x;H) \leq 1$ for any point x and distribution H in $\mathbb{R}^d$.

A1: $P\{y_1 \leq D(Y;F) \leq y_2\} \leq C|y_2 - y_1|$ for some C and any $y_1, y_2 \in [0,1]$.

A2: $\sup_{x\in\mathbb{R}^d} |D(x;F_m) - D(x;F)| = o(1)$, almost surely as $m \to \infty$.

A3: $E\left(\sup_{x\in\mathbb{R}^d} |D(x;F_m) - D(x;F)|\right) = O(m^{-1/2})$.

A4: $E\Big(\sum_i p_{iX}(F_m)p_{iY}(F_m)\Big) = o(m^{-1/2})$ if there exist c_i such that $p_{iX}(F_m) > 0$ and $p_{iY}(F_m) > 0$ for $p_{iZ}(F_m) := P(D(Z;F_m) = c_i \mid F_m)$, $i = 1, 2, \cdots$.

Assumption (A1) is the Lipschitz continuity of the distribution of $D(Y;F)$ and can be extended to a more general case with $|x_2 - x_1|$ replaced by $|x_2 - x_1|^\alpha$ for some $\alpha > 0$, if (A3) is also replaced by $E\left(\sup_{x\in\mathbb{R}^d} |D(x;F_m) - D(x;F)|\right)^\alpha = O(m^{-\alpha/2})$. The following main result of the paper still holds true.

THEOREM 1. *Let $X \sim F$ and $Y \sim G$ be independent, and $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be independent samples from F and G, respectively. Under* (A1)–(A4)*,*

$$\Big(\sigma_{GF}^2/m + \sigma_{FG}^2/n\Big)^{-1/2}\Big(Q(F_m, G_n) - Q(F,G)\Big) \xrightarrow{d} N(0,1), \quad \text{as } m \to \infty,$$

where

$$\sigma_{FG}^2 = \int P^2(D(X;F) \leq D(y,F))dG(y) - Q^2(F,G),$$
$$\sigma_{GF}^2 = \int P^2(D(x;F) \leq D(Y,F))dF(x) - Q^2(F,G).$$

Assumption (A2) in the theorem is satisfied by most depth functions such as the Mahalenobis, projection, simplicial, and halfspace depth functions; see Zuo (2003), Liu (1990), and Massé (1999) for related discussions. Assumptions (A3)-(A4) also hold true for many of the commonly used depth functions. Verifications can be technically challenging and are deferred to Section 4.

Remark 1 Under the null hypothesis $F = G$, it is readily seen that $Q(F,G) = 1/2$ and $\sigma^2_{GF} = \sigma^2_{FG} = 1/12$ in the theorem.

Remark 2 Note that (A1)-(A4) and consequently the theorem hold true for not only common depth functions that induce a center-outward ordering in R^d but also other functions that can induce a general (not necessarily center-outward) ordering in R^d. For example, if we define a function $D(x, F) = F(x)$ in R^1, then the corresponding Liu-Singh statistic is equivalent to the Wilcoxon rank sum statistic.

3. Proofs of the main result and auxiliary lemmas To prove the main theorem, we need the following auxiliary lemmas. Some proofs are skipped. For the sake of convenience, we write, for any distribution functions H, F_1 and F_2 in $\mathbb{R}^d$, points x and y in R^d, and a given (affine invariant) depth function $D(\cdot;\cdot)$,

$$I(x, y, H) = I\{D(x; H) \le D(y; H)\}, \quad I(x, y, F_1, F_2) = I(x, y, F_1) - I(x, y, F_2).$$

Lemma 1. *Let F_m and G_n be the empirical distributions based on independent samples of sizes m and n from distributions F and G, respectively. Then*

(i) $\iint I(x, y, F) d(G_n(y) - G(y)) d(F_m(x) - F(x)) = O_p(1/\sqrt{mn})$,

(ii) $\iint I(x, y, F_m, F) d(F_m(x) - F(x)) dG(y) = o_p(1/\sqrt{m})$ *under* (A1)-(A2), *and*

(iii) $\iint I(x, y, F_m, F) dF_m(x) d(G_n - G)(y) = O_p(m^{-1/4} n^{-1/2})$ *under* (A1) and (A3).

PROOF OF LEMMA 1. We prove (iii). The proofs of (i)-(ii) are skipped. Let $I_{mn} := \iint I(x, y, F_m, F) dF_m(x) d(G_n - G)(y)$. Then

$$
\begin{aligned}
E(I_{mn})^2 &\le E\Big\{ \int \Big[\int I(x, y, F_m, F) d(G_n - G)(y) \Big]^2 dF_m(x) \Big\} \\
&= E\Big[\int I(X_1, y, F_m, F) d(G_n - G)(y) \Big]^2 \\
&= E\Big[E\Big\{ \Big(\frac{1}{n} \sum_{j=1}^{n} (I(X_1, Y_i, F_m, F) - E_Y I(X_1, Y, F_m, F)) \Big)^2 \Big| X_1, \cdots, X_m \Big\} \Big] \\
&\le \frac{1}{n} E\Big[E_Y \Big\{ (I(X_1, Y_1, F_m, F))^2 \Big| X_1, \cdots, X_m \Big\} \Big] \\
&= \frac{1}{n} E\Big[E_Y \Big\{ |I(X_1, Y_1, F_m, F)| \Big| X_1, \cdots, X_m \Big\} \Big].
\end{aligned}
$$

One can verify that

$$
|I(x, y, F_m, F)| \le I\Big(|D(x; F) - D(y; F)| \le 2 \sup_{x \in \mathbb{R}^d} |D(x; F_m) - D(x; F)| \Big).
$$

By (A1) and (A3), we have

$$
E(I_{mn})^2 \le \frac{4C}{n} E\Big(\sup_{x \in \mathbb{R}^d} |D(x; F_m) - D(x; F)| \Big) \;=\; O(1/(m^{1/2} n)).
$$

The desired result follows from Markov's inequality. □

LEMMA 2. *Assume that $X \sim F$ and $Y \sim G$ are independent. Then under* (A4), *we have* $\iint I(D(x; F_m) = D(y; F_m)) dF(x) dG(y) = o(m^{-1/2})$.

PROOF OF LEMMA 2. Let $I(F_m) = \iint I\big(D(x; F_m) = D(y; F_m)\big) dF(x) dG(y)$. Conditionally on $X_1, \cdots, X_m$ (or equivalently on F_m), we have

$$
\begin{aligned}
I(F_m) &= \int_{\{y:\, P(D(X; F_m) = D(y; F_m) \,|\, F_m) > 0\}} P\Big(D(X; F_m) = D(y; F_m) \;\Big|\; F_m \Big) dG(y) \\
&= \sum_i \int_{\{y:\, P(D(X; F_m) = D(y; F_m) = c_i \,|\, F_m) > 0\}} P\Big(D(X; F_m) = c_i \;\Big|\; F_m \Big) dG(y)
\end{aligned}
$$

$$
= \sum_i P\Big(D(X; F_m) = c_i \;\Big|\; F_m \Big) P\Big(D(Y; F_m) = c_i \;\Big|\; F_m \Big) \;=\; \sum_i p_{iX}(F_m) p_{iY}(F_m),
$$

where $0 \le c_i \le 1$ such that $P\big(D(X;F_m) = c_i \mid F_m\big) = P\big(D(Y;F_m) = c_i \mid F_m\big) > 0$. (Note that there are at most countably many such c_i's.) Taking expectation with respect to $X_1, \cdots, X_m$, the desired result follows immediately from (A4). $\square$

LEMMA 3. *let $X \sim F$ and $Y \sim G$ be independent and $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_m$ be independent samples from F and G, respectively. Under* (A1)-(A4)

$$Q(F_m, G_n) - Q(F, G_n) = \int\int I(x, y, F) dG(y) d(F_m(x) - F(x)) + o_p(m^{-1/2}),$$

and consequently $\sqrt{m}\,(Q(F_m, G_n) - Q(F, G_n)) \xrightarrow{d} N(0, \sigma^2_{GF})$.

PROOF OF LEMMA 3. It suffices to consider the case $F = G$. First we observe

$$\begin{aligned}
Q(F_m, G_n) - Q(F, G_n) &= \int R(y; F_m) dG_n(y) - \int R(y; F) dG_n(y) \\
&= \int\int I(x, y, F_m) dF_m(x) dG_n(y) - \int\int I(x, y, F) dF(x) dG_n(y) \\
&= \int\int [I(x, y, F_m) - I(x, y, F)] dF_m(x) dG_n(y) \\
&\quad + \int\int I(x, y, F) d(G_n(y) - G(y)) d(F_m(x) - F(x)) \\
&\quad + \int\int I(x, y, F) dG(y) d(F_m(x) - F(x)).
\end{aligned}$$

Call the last three terms I_{mn1}, I_{mn2}, and I_{m3}, respectively. From Lemma 1 it follows immediately that $\sqrt{m}\, I_{mn2} = o_p(1)$. By a standard central limit theorem, we have

$$\sqrt{m}\, I_{m3} \xrightarrow{d} N(0, \sigma^2_{GF}). \tag{3.4}$$

We now show that $\sqrt{m}\, I_{mn1} = o_p(1)$. Observe that

$$\begin{aligned}
I_{mn1} &= \int\int I(x, y, F_m, F) dF_m(x) d(G_n - G)(y) + \int\int I(x, y, F_m, F) dF_m(x) dG(y) \\
&= \int\int I(x, y, F_m, F) dF(x) dG(y) + o_p(1/\sqrt{m})
\end{aligned}$$

by Lemma 1 and the given condition. It is readily seen that

$$\begin{aligned}
&\int\!\!\int I(x,y,F_m,F)dF(x)dG(y)\\
&= \int\!\!\int I(x,y,F_m)dF(x)dG(y) - \int\!\!\int I(x,y,F)dF(x)dG(y)\\
&= \frac{1}{2}\int\!\!\int [I(D(x,F_m) \le D(y;F_m)) + I(D(x,F_m) \ge D(y;F_m))]dF(x)dG(y) - \frac{1}{2}\\
&= \frac{1}{2}\int\!\!\int I(D(x,F_m) = D(y;F_m))dF(x)dG(y) = \ o(m^{-1/2}),
\end{aligned}$$

by Lemma 2. The desired result follows immediately. □

PROOF OF THEOREM 1. By Lemma 3 we have

$$\begin{aligned}
Q(F_m,G_n) - Q(F,G) \ &= \ (Q(F_m,G_n) - Q(F,G_n)) + (Q(F,G_n) - Q(F,G))\\
&= \int\!\!\int I(x,y,F)dG(y)d(F_m(x) - F(x)) + \int\!\!\int I(x,y,F)dF(x)d(G_n(y) - G(y))\\
&\quad + o_p(m^{-1/2}).
\end{aligned}$$

The independence of F_m and G_n and the central limit theorem give the result. □

4. Applications and Examples This section verifies (A3)-(A4) (and (A2)) for several common depth functions. Mahalanobis, halfspace, and projection depth functions are selected for illustration. The findings here and in Section 2 ensure the validity of Theorem 1 for these depth functions.

EXAMPLE 1 – Mahalanobis depth (MHD). The depth of a point x is defined as

$$MHD\,(x;F) = 1/(1 + (x - \mu(F))'\Sigma^{-1}(F)(x - \mu(F))), \quad x \in \mathbb{R}^d,$$

where $\mu(F)$ and $\Sigma(F)$ are location and covariance measures of a given distribution F; see Liu and Singh (1993) and Zuo and Serfling (2000a). Clearly both $MHD(x;F)$ and $MHD(x;F_m)$ vanish at infinity as $\|x\| \to \infty$, where F_m is the empirical version of F based on $X_1, \cdots, X_m$ and $\mu(F_m)$ and $\Sigma(F_m)$ are strongly consistent estimators of $\mu(F)$ and $\Sigma(F)$, respectively. Hence

$$\sup_{x \in \mathbb{R}^d} |MHD(x; F_m) - MHD(x; F)| = |MHD(x_m; F_m) - MHD(x_m; F)|,$$

by the continuity of $MHD(x; F)$ and $MHD(x; F_m)$ in x for some $x_m = x(F_m, F) \in \mathbb{R}^d$ and $\|x_m\| \leq M < \infty$ for some $M > 0$ and all large m. Write for simplicity μ and Σ for $\mu(F)$ and $\Sigma(F)$ and μ_m and Σ_m for $\mu(F_m)$ and $\Sigma(F_m)$, respectively. Then

$$\begin{aligned} &|MHD(x_m; F_m) - MHD(x_m; F)| \\ &= \frac{|(\mu_m - \mu)'\Sigma_m^{-1}(\mu_m + \mu - 2x_m) + (x_m - \mu)'(\Sigma_m^{-1} - \Sigma^{-1})(x_m - \mu)|}{(1 + \|\Sigma_m^{-1/2}(x_m - \mu_m)\|^2)(1 + \|\Sigma^{-1/2}(x_m - \mu)\|^2)}. \end{aligned}$$

This, in conjunction with the strong consistency of μ_m and Σ_m, yields (A2).

Hölder's inequality and expectations of quadratic forms (p.13 of Seber (1977)) yield (A3) if conditions (i) and (ii) below are met. (A4) holds trivially if (iii) holds.
(i) μ_m and Σ_m are strongly consistent estimators of μ and Σ, respectively,
(ii) $E\,(\mu_m - \mu)_i = O(m^{-1/2}),\ E\,(\Sigma_m^{-1} - \Sigma^{-1})_{jk} = O(m^{-1/2})$, $1 \leq i,\ j,\ k \leq d$, where the subscripts i and jk denote the elements of a vector and a matrix respectively,
(iii) The probability mass of X over any ellipsoid is 0.

Corollary 1. *Assume that conditions* (i), (ii) *and* (iii) *hold and the distribution of MHD*$(Y; F)$ *is Lipschitz continuous. Then Theorem* 1 *holds for MHD.* □

Example 2 – Halfspace depth (HD). Tukey (1975) suggested this depth as

$$HD\,(x; F) = \inf\{P(H_x):\ H_x \text{ closed halfspace with } x \text{ on its boundary}\}, \quad x \in \mathbb{R}^d,$$

where P is the probability measure corresponding to F. (A2) follows immediately [see, e.g., pages 1816-1817 of Donoho and Gasko (1992)]. Let $\mathcal{H}$ be the set of all closed halfspaces and P_m be the empirical probability measure of P. Define

$$D_m(\mathcal{H}) := m^{1/2}\,\|P_m - P\|_{\mathcal{H}} := \sup_{H\in\mathcal{H}} m^{1/2}|P_m(H) - P(H)|.$$

Note that $\mathcal{H}$ is a permissible class of sets with polynomial discrimination [see Section II. 4 of Pollard (1984) for definitions and arguments]. Let $S(\mathcal{H})$ be the degree of the corresponding polynomial. Then by a generalized Dvoretzk-Kiefer-Wolfowitz theorem [see Alexander (1984) and Massart (1983, 1986); also see Section 6.5 of Dudley (1999)], we have for any $\epsilon > 0$ that $P(D_m(\mathcal{H}) > M) \le Ke^{-(2-\epsilon)M^2}$ for some large enough constant $K = K(\epsilon, S(\mathcal{H}))$. This yields immediately (A3).

To verify (A4), we consider the case $F = G$ for simplicity. We first note that $HD(X; F_m)$ for given F_m is discrete and can take at most $O(m)$ values $c_i = i/m$ for $i = 0, 1, \cdots, m$. Let F be continuous. We first consider the univariate case. Let

$$A_0 = \mathbb{R}^1 - \cap H_m, \quad A_i = \cap H_{m-i+1} - \cap H_{m-i}, \quad A_{k+1} = \cap H_{m-k} - \cap\emptyset,$$

with $1 \le i \le k$ and $k = \lfloor (m-1)/2 \rfloor$, where H_i is any closed half-line containing exactly some i points of $X_1, \cdots, X_m$. It follows that for $0 \le i \le k$,

$$\begin{aligned} P(HD(X; F_m) = c_i \mid F_m) &= P(A_i) \\ &= [F(X_{(i+1)}) - F(X_{(i)})] + [F(X_{(m-i+1)}) - F(X_{(m-i)})] \\ P(HD(X; F_m) = c_{k+1} \mid F_m) &= P(A_{k+1}) = [F(X_{(m-k)}) - F(X_{(k+1)})], \end{aligned}$$

where $-\infty =: X_{(0)} \le X_{(1)} \le \cdots \le X_{(m)} \le X_{(m+1)} := \infty$ are order statistics.

On the other hand, $X_{(i)}$ and $F^{-1}(U_{(i)})$ are equal in distribution ($\stackrel{d}{=}$), where $0 =: U_{(0)} \le U_{(1)} \le \cdots \le U_{(m)} \le U_{(m+1)} := 1$ are order statistics based on a sample from the uniform distribution on $[0, 1]$. Let $D_i = F(X_{(i+1)}) - F(X_{(i)})$, $i = 0, \cdots, m$. The D_i's have the same distribution and

$$E(D_i) = \frac{1}{m+1}, \quad E(D_i^2) = \frac{2}{(m+1)(m+2)}, \quad E(D_iD_j) = \frac{1}{(m+1)(m+2)}.$$

Hence for $0 \le i \le k$, $E((P(HD(X;F_m) = c_i \mid F_m))^2) = 6/((m+1)(m+2))$, and $E((P(HD(X;F_m) = c_{k+1} \mid F_m))^2) = O(m^{-2})$. Thus (A4) follows immediately.

Now let us treat the multivariate case. Let $X_1, \cdots, X_m$ be given. Denote by H_i any closed halfspace containing exactly i points of $X_1, \cdots, X_m$. Define sets

$$A_0 = \mathbb{R}^d - \cap H_m,\ A_1 = \cap H_m - \cap H_{m-1}, \cdots, A_{m-k} = \cap H_{k+1} - \cap H_k, A_{m-k+1} = \cap H_k$$

with $(m-k+1)/m = \max_{x \in \mathbb{R}^d} HD(x;F_m) \le 1$. Then it is not difficult to see that

$$HD(x;F_m) = i/m, \quad \text{for } x \in A_i,\ i = 0, 1, \cdots, m-k, m-k+1.$$

Now let $p_i = P(HD(X;F_m) = c_i \mid F_m)$ with $c_i = i/m$. Then for any $0 \le i \le m-k+1$ $p_i = P(X \in A_i) = P(\cap H_{m-i+1}) - P(\cap H_{m-i})$, with $H_{m+1} = \mathbb{R}^d$ and $H_{k-1} = \emptyset$. Now treating p_i as random variables based on the random variables $X_1, \cdots, X_m$, by symmetry and the uniform spacings results used for the univariate case above we conclude that the p_i's have the same distribution for $i = 0, \cdots, m-k$ and

$$E(p_i) = O(m^{-1}), \quad E(p_i^2) = O(m^{-2}), \quad i = 0, \cdots, m-k+1.$$

Assumption (A4) follows in a straightforward fashion. Thus we have

Corollary 2. *Assume that F is continuous and the distribution of $HD(Y;F)$ is Lipschitz continuous. Then Theorem 1 holds true for HD.* □

Example 3 – Projection depth (PD). Stahel (1981) and Donoho (1982) defined the outlyingness of a point $x \in \mathbb{R}^d$ with respect to F in $\mathbb{R}^d$ as

$$O(x;F) = \sup_{u \in S^{d-1}} |u'x - \mu(F_u)|/\sigma(F_u),$$

where $S^{d-1} = \{u : \|u\| = 1\}$, $\mu(\cdot)$ and $\sigma(\cdot)$ are univariate location and scale estimators such that $\mu(aZ+b) = a\mu(Z)+b$ and $\sigma(aZ+b) = |a|\sigma(Z)$ for any scalars

$a, b \in \mathbb{R}^1$ and random variable $Z \in \mathbb{R}^1$, and $u'X \sim F_u$ with $X \sim F$. The projection depth of x with respect to F is then defined as [see Liu (1992) and Zuo (2003)]

$$PD(x; F) = 1/(1 + O(x; F)).$$

Under the following conditions on μ and σ,

(C1): $\sup_{u \in S^{d-1}} \mu(F_u) < \infty, \quad 0 < \inf_{u \in S^{d-1}} \sigma(F_u) \le \sup_{u \in S^{d-1}} \sigma(F_u) < \infty;$

(C2): $\sup_{u \in S^{d-1}} |\mu(F_{mu}) - \mu(F_u)| = o(1),\ \sup_{u \in S^{d-1}} |\sigma(F_{mu}) - \sigma(F_u)| = o(1),\ a.s.$

(C3): $E \sup_{\|u\|=1} |\mu(F_{mu}) - \mu(F_u)| = O(m^{-\frac{1}{2}}),\ E \sup_{\|u\|=1} |\sigma(F_{mu}) - \sigma(F_u)| = O(m^{-\frac{1}{2}}),$

where F_{mu} is the empirical distribution based on $u'X_1, \cdots, u'X_m$ and $X_1, \cdots, X_m$ is a sample from F, Assumption (A2) holds true by Theorem 2.3 of Zuo (2003) and (A3) follows from (C3) and the fact that for any $x \in \mathbb{R}^d$ and some constant $C > 0$

$$\begin{aligned} |PD(x; F_m) - PD(x; F)| &\le \sup_{u \in S^{d-1}} \frac{O(x; F)|\sigma(F_{mu}) - \sigma(F_u)| + |\mu(F_{mu}) - \mu(F_u)|}{(1 + O(x; F_m))(1 + O(x; F))\sigma(F_{mu})} \\ &\le C \sup_{u \in S^{d-1}} \{|\sigma(F_{mu}) - \sigma(F_u)| + |\mu(F_{mu}) - \mu(F_u)|\}. \end{aligned}$$

(C1)-(C3) is true for general smooth M-estimators of μ and σ [see Huber (1981)] and rather general distribution functions F. If we consider the median (Med) and the median absolute deviation (MAD), then (C3) holds under the following condition

(C4): F_u has a continuous density f_u around points $\mu(F_u) + \{0, \pm\sigma(F_u)\}$ such that $\inf_{\|u\|=1} f_u(\mu(F_u)) > 0$, $\inf_{\|u\|=1}(f_u(\mu(F_u) + \sigma(F_u)) + f_u(\mu(F_u) - \sigma(F_u))) > 0$,

To verify this, it suffices to establish just the first part of (C3) for μ=Med. Observe

$$\begin{aligned} F_u^{-1}(1/2 - \|F_{mu} - F_u\|_\infty) - F_u^{-1}(1/2) &\le \mu(F_{mu}) - \mu(F_u) \\ &\le F_u^{-1}(1/2 + \|F_{mu} - F_u\|_\infty) - F_u^{-1}(1/2), \end{aligned}$$

for any u and sufficiently large m. Hence

$$|\mu(F_{mu}) - \mu(F_u)| \le 2\|F_{mu} - F_u\|_\infty / \inf_{u \in S^{d-1}} f_u(\mu(F_u)) := C\|F_{mu} - F_u\|_\infty,$$

by (C4). Clearly, $\mu(F_{mu})$ is continuous in u. From (C4) and Lemma 5.1 and Theorem 3.3 of Zuo (2003), it follows that $\mu(F_u)$ is also continuous in u. Therefore

$$\begin{aligned} P\big(\sqrt{m}\sup_{u\in S^{d-1}}|\mu(F_{mu})-\mu(F_u)|>t\big) &\le P\big(\|F_{mu_0}-F_{u_0}\|_\infty > \big(t^2/(mC^2)\big)^{1/2}\big) \\ &\le 2e^{-2t^2/C^2}, \quad \text{for any } t>0, \end{aligned}$$

where unit vector u_0 may depend on m. Hence the first part of (C3) follows.

Assumption (A4) holds for PD since $P(PD(X;F_m)=c \mid F_m)=0$ for most commonly used (μ,σ) and F. First the continuity of $\mu(F_{mu})$ and $\sigma(F_{mu})$ in u gives

$$P\big(PD(X;F_m)=c \bigm| F_m\big) = P\big((u_X'X-\mu(F_{mu_X}))/\sigma(F_{mu_X})=(1-c)/c \bigm| F_m\big),$$

for some unit vector u_X depending on X. This probability is 0 for most F and (μ,σ). For example, if (μ,σ) =(mean, standard deviation), then

$$P\big(PD(X;F_m)=c \bigm| F_m\big) = P\big(\|S_m^{-1/2}(X-\bar{X}_m)\|=(1-c)/c \bigm| F_m\big),$$

where $S_m=\frac{1}{m-1}\sum_{i=1}^m (X_i-\bar{X}_m)(X_i-\bar{X}_m)'$, which is 0 as long as the mass of F on any ellipsoids is 0. Thus

Corollary 3. *Assume that* (C1)-(C3) *hold,* $P(PD(X;F_m)=c \mid F_m)=0$ *for any* $c\ge 0$*, and* $PD(Y;F)$ *satisfies* (A1)*. Then Theorem* 1 *holds for PD.* □

5. Power properties of the Liu-Singh multivariate rank sum test

Large sample properties A main application of the Liu-Singh multivariate rank-sum statistic is to test the following hypotheses:

$$H_0: F=G, \quad \text{versus} \quad H_1: F\ne G. \tag{5.5}$$

By Theorem 1, a large sample test based on the Liu-Singh rank sum statistic $Q(F_m,G_n)$ rejects H_0 at (an asymptotic) significance level α when

$$|Q(F_m,G_n)-1/2| > z_{1-\alpha/2}\,((1/m+1/n)/12)^{1/2}, \tag{5.6}$$

where $\Phi(z_r) = r$ for $0 < r < 1$ and normal c.d.f $\Phi(\cdot)$. The test is *affine invariant*, and is *distribution-free* in the asymptotic sense under the null hypothesis. Here we focus on the asymptotic power properties of the test. By Theorem 1, the (asymptotic) power function of the depth-based rank sum test with an asymptotic significance level α is

$$\beta_Q(F,G) = 1 - \Phi\Big(\frac{1/2 - Q(F,G) + z_{1-\alpha/2}\,\sqrt{(1/m+1/n)/12}}{\sqrt{\sigma^2_{GF}/m + \sigma^2_{FG}/n}}\Big) + \Phi\Big(\frac{1/2 - Q(F,G) - z_{1-\alpha/2}\,\sqrt{(1/m+1/n)/12}}{\sqrt{\sigma^2_{GF}/m + \sigma^2_{FG}/n}}\Big). \tag{5.7}$$

The asymptotic power function indicates that the test is *consistent* for *all* alternative distributions G such that $Q(F,G) \neq 1/2$. Before studying the behavior of $\beta_Q(F,G)$, let's take a look at its key component $Q(F,G)$, the so-called *quality index* in Liu and Singh (1993). For convenience, consider a normal family and $d = 2$. Assume without loss of generality (w.l.o.g.) that $F = N_2((0,0)', I_2)$ and consider $G = N_2(\mu, \Sigma)$, where I_2 is a 2×2 identity matrix. It can be shown

$$Q(F,G) = \big(|S|/|\Sigma|\big)^{1/2}\ \exp\big(-\mu'(\Sigma^{-1} - \Sigma^{-1} S\, \Sigma^{-1})\mu/2\big),$$

for *any* affine invariant depth functions, where $S = (I_2 + \Sigma^{-1})^{-1}$. In the case $\mu = (u,u)'$ and $\Sigma = \sigma^2 I_2$, write $Q(u,\sigma^2)$ for $Q(F,G)$. Then

$$Q(u,\sigma^2) := Q(F,G) = \exp(-u^2/(1+\sigma^2))/(1+\sigma^2).$$

Its behavior is revealed in Figure 1. It increases to its maximum value $(1+\sigma^2)^{-1}$ (or $\exp(-u^2)$) as $u \to 0$ for a fixed σ^2 (or as $\sigma^2 \to 0$ for a fixed u). When $u = 0$ and $\sigma^2 = 1$, $Q(F,G)$, as expected, is $1/2$, and it is $< 1/2$ when there is a dilution in the distribution ($\sigma^2 > 1$). Note that Liu and Singh (1993) also discussed $Q(u,\sigma^2)$. The results here are more accurate than their Table 1.

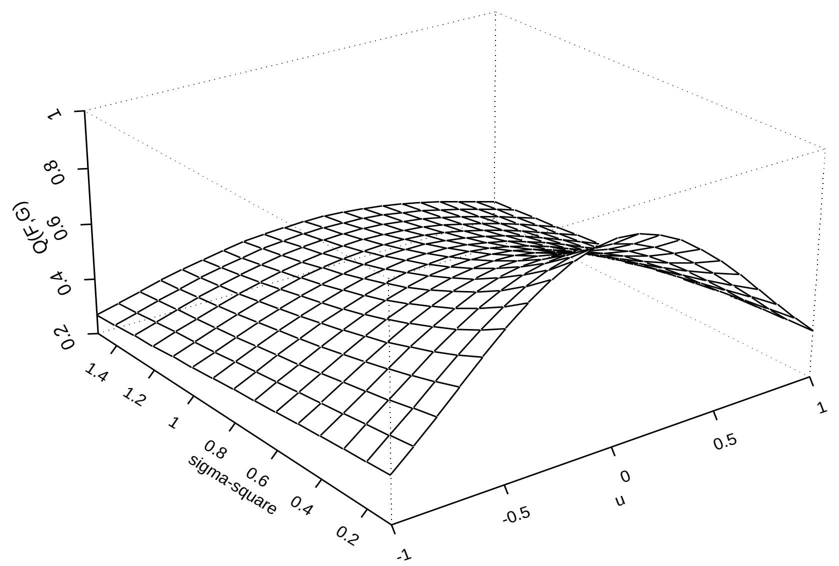

FIG. 1. *The behavior of* $Q(F,G)$ *with* $F = N_2((0,0)', I_2)$ *and* $G = N_2((u,u)', \sigma^2 I_2)$.

A popular large sample test for hypotheses (5.5) is based on Hotelling's T^2 statistic that rejects H_0 if

$$(\bar{X}_m - \bar{Y}_n)'((1/m + 1/n)S_{pooled})^{-1}(\bar{X}_m - \bar{Y}_n) > \chi^2_{1-\alpha}(d) \tag{5.8}$$

where $S_{pooled} = ((m-1)S_X + (n-1)S_Y)/(m+n-2)$, $\bar{X}_m$ and $\bar{Y}_n$ and S_X and S_Y are sample means and covariance matrices, and $\chi^2_r(d)$ is the rth quantile of the chi-square distribution with d degrees of freedom. The power function of the test is

$$\beta_{T^2}(F,G) = P\big((\bar{X}_m - \bar{Y}_n)'((1/m + 1/n)S_{pooled})^{-1}(\bar{X}_m - \bar{Y}_n) > \chi^2_{1-\alpha}(d)\big). \tag{5.9}$$

We also consider a multivariate rank sum test based on the Oja objective function in Hettmansperger et al (1998). The Oja test statistic, O, has the following null distribution with $N = m + n$ and $\lambda = n/N$,

$$O := (N\lambda(1-\lambda))^{-1} T_N' B_N^{-1} T_N \xrightarrow{d} \chi^2(d), \tag{5.10}$$

where $T_N = \sum_{k=1}^N a_k R_N(z_k)$, $R_N(z) = \frac{d!(N-d)!}{N!} \sum_{p \in P} S_p(z) n_p$, $a_k = (1-\lambda)I(k > m) - \lambda I(k < m)$, $z_k \in \{X_1, \cdots, X_m, Y_1, \cdots, Y_n\}$, $B_N = \frac{1}{N-1} \sum_k R_N(z_k) R_N'(z_k)$,

$P = \{p = (i_1, \cdots, i_d) : 1 \le i_1 < \cdots < i_d \le N\}$, $S_p(z) = \mathrm{sign}(n_{0p} + z' n_p)$, and

$$\det\begin{pmatrix} 1 & 1 & \cdots 1 & 1 \\ z_{i_1} & z_{i_2} & \cdots z_{i_d} & z \end{pmatrix} = n_{0p} + z' n_p,$$

where n_{0p} and $n_{jp}, j = 1, \cdots, d$, are the co-factors according to the column $(1, z')'$. The power function of this rank test with an asymptotic significance level α is

$$\beta_O(F, G) = P\big((N\lambda(1-\lambda))^{-1} T_N' B_N^{-1} T_N > \chi^2_{1-\alpha}(d)\big). \tag{5.11}$$

The asymptotic relative efficiency (ARE) of this test in Pitman's sense is discussed in the literature; see, e.g., Möttönen et al (1998). At the bivariate normal model, it is 0.937 relative to T^2.

In the following we study the behavior of β_Q, β_O, and β_{T^2}. To facilitate our discussion, we assume that $\alpha = 0.05$, $m = n$, $d = 2$, and G is normal or mixed (contaminated) normal, shrinking to the null distribution $F = N_2((0,0)', I_2)$. Note that the asymptotic power of the depth-based rank sum test, called the Q test from now on, is invariant in the choice of the depth function.

For pure location shift models $Y \sim G = N_2((u,u)', I_2)$, Hotelling's T^2 based test, called T^2 from now on, is the most powerful, followed by the Oja rank test, to be called the O test, and then followed by the Q test. All these tests are consistent at any fixed alternative. Furthermore, we note that when the dimension d gets larger, the asymptotic powers of these tests move closer.

On the other hand, for pure scale change models $G = N_2((0,0)', \sigma^2 I_2)$, the Q test is much more powerful than the other test. In fact, for these models, the T^2 test has trivial asymptotic power α at all alternatives. Figure 2, a plot of the power functions β_{T^2}, β_O and β_Q, clearly reveals the superiority of the Q test. The O test performs just slightly better than T^2.

In the following we consider a location shift with contamination, a scale change with contamination, and a simultaneous location and scale change as alternatives.

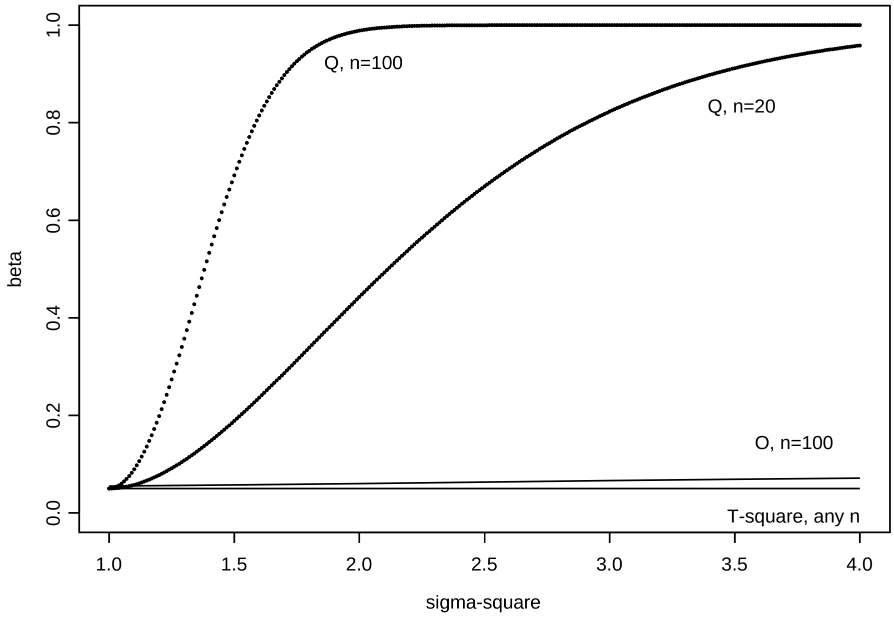

FIG. 2. $\beta_{T^2}(F,G)$ *and* $\beta_Q(F,G)$ *with* $F = N_2((0,0)', I_2)$, $G = N_2((0,0)', \sigma^2 I_2)$.

The contamination amount ϵ is set to be 10%. The asymptotic power calculations for T^2 and Q are based on the limiting distributions of the test statistics under the alternatives. Since the limiting distribution is not available for the O test (except for pure location shift models), we use Monte Carlo to estimate the powers.

For $G = (1-\epsilon)N_2((u,u)', I_2) + \epsilon N_2((0,0)', (1+10u\sigma^2)I_2)$, the contaminated location shift models with $u \to 0$, the (*asymptotic*) power function $\beta_{T^2}(F,G)$ is $P(Z_{2a} \geq \chi^2_{0.95}(2))$, where Z_{2a} has a non-central chi-square distribution with 2 degrees of freedom and non-centrality parameter $n(1-\epsilon)^2u^2/(1+5\epsilon u\sigma^2+\epsilon(1-\epsilon)u^2)$. Since the derivation of this result is quite tedious, we skip the details. Comparisons of β_{T^2}, β_O, and β_Q are listed in Table 1, which clearly reveals that T^2 becomes less powerful than Q when a pure location shift model is 10% contaminated. For large μ, O is more powerful than Q since the underlying model is mainly a location shift.

For $G = 0.9N_2((0,0)', \sigma^2 I_2) + 0.1N_2((u,u)', I_2)$, the contaminated scale change models with $\sigma^2 \to 1$ and $u \to 0$, the (*asymptotic*) power function β_{T^2} is equal to

$P(Z_{2b} \geq \chi^2_{0.95}(2))$, where Z_{2b} has the non-central chi-square distribution with 2 degrees of freedom and the non-centrality parameter $2n\epsilon^2u^2/(1+\epsilon+(1-\epsilon)\sigma^2+2\epsilon(1-\epsilon)u^2)$. Table 2 reveals the superiority of Q in detecting scale changes over T^2 and O, even when the model has a 10% contamination.

TABLE 1. The (asymptotic) power of tests based on T^2, O, and Q

u		0.0	0.15	0.20	0.25	0.30	0.35
$G = 0.9N_2((u,u)', I_2) + 0.1N_2((0,0)', (1+10u\sigma^2)I_2), \sigma = 4$							
n=100	β_{T^2}	0.050	0.117	0.155	0.196	0.239	0.284
	β_Q	0.051	0.245	0.286	0.307	0.381	0.443
	β_O	0.046	0.157	0.296	0.423	0.558	0.687
n=200	β_{T^2}	0.050	0.193	0.273	0.357	0.441	0.521
	β_Q	0.051	0.430	0.508	0.549	0.659	0.746
	β_O	0.056	0.342	0.546	0.712	0.881	0.941

TABLE 2. The (asymptotic) power of tests based on T^2, O, and Q

σ^2		1.0	1.2	1.4	1.6	1.8	2.0
$G = 0.9N_2((0,0)', \sigma^2 I_2) + 0.1N_2((u,u)', I_2), u = \sigma - 1$							
n=100	β_{T^2}	0.0500	0.0505	0.0522	0.0540	0.0565	0.0589
	β_Q	0.0510	0.1807	0.4300	0.7343	0.8911	0.9627
	β_O	0.0480	0.0540	0.0570	0.0640	0.0680	0.0700
n=200	β_{T^2}	0.0500	0.0513	0.0543	0.0583	0.0631	0.0679
	β_Q	0.0515	0.2986	0.7405	0.9505	0.9944	0.9996
	β_O	0.0520	0.0590	0.0630	0.0850	0.1120	0.1390

For $G = N_2((u,u)', \sigma^2 I_2)$, the simultaneous location and scale change models with $(u, \sigma^2) \to (0,1)$, the (*asymptotic*) power function β_{T^2} is $P(Z_{2c} \geq \chi^2_{0.95}(2))$, where Z_{2c} has the non-central chi-square distribution with 2 degrees of freedom and the non-centrality parameter $2nu^2/(1+\sigma^2)$. Table 3 reveals that Q can be more powerful than T^2 and O when there are simultaneous location and scale changes. Here we selected $(\sigma-1)/u = 1$. Our empirical evidence indicates that the superiority of Q holds as long as $(\sigma-1)/u$ is close to or greater than 1, that is, as long as the change in scale is not much less than that in location. Also note that T^2 in this model is more powerful than O.

TABLE 3. The (asymptotic) power of tests based on T^2, Q and O

u		0.0	0.15	0.20	0.25	0.30	0.35
		$G = N_2((u,u)', \sigma^2 I_2),\ \sigma = u+1$					
n=100	β_{T^2}	0.0500	0.2194	0.3482	0.4931	0.6345	0.7547
	β_Q	0.0506	0.4368	0.6622	0.8394	0.9409	0.9827
	β_O	0.0460	0.2180	0.3240	0.4300	0.5730	0.7080
n=200	β_{T^2}	0.0500	0.4039	0.6249	0.8050	0.9161	0.9698
	β_Q	0.0488	0.7247	0.9219	0.9875	0.9989	0.9999
	β_O	0.0560	0.3570	0.5690	0.7550	0.8820	0.9440

SMALL SAMPLE PROPERTIES To check the small sample power behavior of Q, we now examine the empirical behavior of the test based on $Q(F_m, G_n)$ and compare it with those of T^2 and O. We focus on the relative frequencies of rejecting H_0 of (5.5) at $\alpha = 0.05$ based on the tests (5.6), (5.8) and (5.10) and 1000 samples from F and G at the sample size $m = n = 25$. The projection depth with $(\mu, \sigma) =$ (Med, MAD) is selected in our simulation studies, and some results are given in Table 4. Again we skip the pure location shift or scale change models, in which cases, T^2 and Q perform best, respectively. Our Monte Carlo studies confirm the validity of the (asymptotic) power properties of Q at small samples.

Table 4. Observed relative frequency of rejecting H_0

u		0.0	0.15	0.20	0.25	0.30	0.35
$G = 0.9N_2((u,u)', I_2) + 0.1N_2((0,0)', (1+10u\sigma^2)I_2), \sigma = 4$							
n=25	β_{T^2}	0.058	0.083	0.108	0.142	0.151	0.189
	β_Q	0.057	0.154	0.156	0.170	0.203	0.216
	β_O	0.047	0.084	0.116	0.152	0.201	0.254
σ^2		1.0	1.2	1.4	1.6	1.8	2.0
$G = 0.9N_2((0,0)', \sigma^2 I_2) + 0.1N_2((u,u)', I_2), u = \sigma - 1$							
n=25	β_{T^2}	0.059	0.063	0.059	0.073	0.061	0.067
	β_Q	0.063	0.145	0.243	0.377	0.469	0.581
	β_O	0.051	0.058	0.041	0.053	0.043	0.055
u		0.0	0.15	0.20	0.25	0.30	0.35
$G = N_2((u,u)', \sigma^2 I_2),\ \sigma = u + 1$							
n=25	β_{T^2}	0.069	0.113	0.147	0.183	0.220	0.269
	β_Q	0.060	0.245	0.324	0.418	0.498	0.587
	β_O	0.044	0.082	0.089	0.127	0.197	0.221

6. Concluding remarks This paper proves the conjectured limiting distribution of the Liu-Singh multivariate rank sum statistic under some regularity conditions, which are verified for several commonly used depth functions. The asymptotic results in the paper are established for *general* depth structures and for *general* distributions F and G. The Q test requires neither the existence of a covariance matrix nor the symmetry of F and G. This is not always the case for Hotelling's T^2 test and other multivariate generalizations of Wilcoxon's rank sum test.

The paper also studies the power behavior of the rank sum test at large and small samples. Although the discussion focuses on the normal and mixed normal models and $d = 2$, what we learned from these investigations are typical for $d > 2$ and for many non-Gaussian models. Our investigations also indicate that the conclusions drawn from our two-sample problems are valid for one-sample problems.

The Liu-Singh rank sum statistic plays an important role in detecting scale changes similar to what Hotelling's T^2 does in detecting location shifts of distributions. When there is a scale change in F, the depths of almost all points y from G decrease or increase together, and consequently $Q(F, G)$ is very sensitive to the change. This explains why Q is so powerful in detecting small scale changes. On the other hand, when there is a small shift in location, the depths of some points y from G increase whereas those of the others decrease, and consequently $Q(F, G)$ will not be so sensitive to a small shift in location. Unlike the T^2 test for scale change alternatives, the Q test *is* consistent for location shift alternatives, nevertheless.

Finally, we briefly address the computing issue. Hotelling's T^2 statistic is clearly the easiest to compute. The computational complexity for the O test is $O(n^{d+1})$, while the complexity for the Q test based on the projection depth (PD) is $O(n^{d+2})$. Indeed, the projection depth can be computed exactly by considering $O(n^d)$ directions that are perpendicular to a hyperplane determined by d data points; see Zuo (2004) for a related discussion. The exact computing is of course time-consuming. In our simulation study, we employed approximate algorithms (see www.stt.msu.edu/˜zuo/table4.txt), which consider a large number (but not all n choosing d) directions in computing Q and O.

Acknowledgments. The authors thank two referees, an Associate Editor and Co-Editors Jianqing Fan and John Marden for their constructive comments and useful suggestions. They are grateful to Hengjian Cui, James Hannan, Hira Koul, Regina Liu, Hannu Oja and Ronald Randles for useful discussions.

REFERENCES

Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041-1067.

Brown, B. M. and Hettmansperger, T. P. (1987). Affine invariant rank methods in the bivariate location model. *J. Roy. Statist. Soci. Ser. B 49* 301-310.

Choi, K. and Marden, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *J. Amer. Statist. Assoc.* **92** 1581-1590.

Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Dept. Statistics, Harvard University.

Donoho, D. L., and Gasko, M. (1992). Breakdown properties of multivariate location parameters and dispersion matrices. *Ann. Statist.* **20** 1803-1827.

Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

Hettmansperger, T. P. , Möttönen, J. , and Oja, H. (1998). Affine invariant multivariate rank test for several samples. *Statist. Sicina* **8** 765-800.

Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414.

Liu, R. Y. (1992). Data Depth and Multivariate Rank Tests. In *L_1-Statistical Analysis and Related Methods* (Y. Dodge, ed.) 279-294.

Liu, R. Y. (1995). Control charts for multivariate processes. *J. Amer. Statist. Assoc.* **90** 1380-1387.

Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252-260.

Massé, J. C. (1999). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli* **10** 1-23.

Massart, P. (1983). Vitesses de convergence dans le théorème central limite pour des processus empiriques. *C. R. Acad. Sci. Paris Sér. I* **296** 937-940.

Massart, P. (1986). Rates of convergence in the central limit theorem for empirical processes. *Ann. Inst. H. Poincaré Probab.-Statist.* **22** 381-423.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* **18** 1269-1283.

Möttönen, J., Hettmansperger, T. P., Oja, H., and Tienari, J. (1998). On the efficiency of affine invariant multivariate rank tests. *J. Multivariate Anal.* **66** 118-132.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* John Wiley, New York.

Pollard, D. (1990). *Empirical Processes: Theorem and Applications.* Institute of Mathematical Statistics, Hayward, California.

Randles, R. H. and D. Peters (1990). Multivariate rank tests for the two-sample location problem. *Commn. Statist. Theory Methods* **19** 4225-4238.

Rousson, V. (2002). On distribution-free tests for the multivariate two-sample location -scale model. *J. Multivariate Anal.* **80** 43-57.

Seber, G. A. F. (1977). *Linear Regression Analysis.* John Wiley & Sons, New York.

Stahel, W. A. (1981). Breakdown of covariance estimators. Research Report 31, Fachgruppe für Statistik, ETH, Zürich.

Topchii, A., Tyurin, Y., and Oja, H. (2003). Inference based on the affine invariant multivariate Mann-Whitney-Wilcoxon statistic. *J. Nonparam. Statist.* **14** 403-414.

Tukey, J. W. (1975). Mathematics and picturing data. *Proc. Intern. Congr. Math. Vancouver 1974* **2** 523–531.

Zuo, Y. (2004). Projection Based Affine Equivariant Multivariate Location Estimators with the Best Possible Finite Sample Breakdown Point. *Statistica Sinica* **14** 1199-1208.

Zuo, Y. (2003). Projection based depth functions and associated medians. *Ann. Statist.* **31** 1460-1490.

Zuo, Y. and Serfling, R. (2000a). General notions of statistical depth function. *Ann. Statist.* **28**(2) 461-482.

Zuo, Y. and Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.* **28**(2) 483-499.

Department of Statistics and Probability
Michigan State University
East Lansing, MI 48824
Email: zuo@msu.edu

Department of Statistics
University of Illinois
Champaign, Illinois 61820
Email: he@stat.uiuc.edu