THE LINEAR STOCHASTIC ORDER AND DIRECTED INFERENCE FOR MULTIVARIATE ORDERED DISTRIBUTIONS

By Ori Davidov[1] and Shyamal Peddada[2]

University of Haifa and National Institute of Environmental Health Sciences

Researchers are often interested in drawing inferences regarding the order between two experimental groups on the basis of multivariate response data. Since standard multivariate methods are designed for two-sided alternatives, they may not be ideal for testing for order between two groups. In this article we introduce the notion of the linear stochastic order and investigate its properties. Statistical theory and methodology are developed to both estimate the direction which best separates two arbitrary ordered distributions and to test for order between the two groups. The new methodology generalizes Roy's classical largest root test to the nonparametric setting and is applicable to random vectors with discrete and/or continuous components. The proposed methodology is illustrated using data obtained from a 90-day pre-chronic rodent cancer bioassay study conducted by the National Toxicology Program (NTP).

1. Introduction. In a variety of applications researchers are interested in comparing two treatment groups on the basis of several, potentially dependent outcomes. For example, to evaluate if a chemical is a neuro-toxicant, toxicologists compare a treated group of animals with an untreated control group in terms of various correlated outcomes such as tail-pinch response, click response and gait score, etc.; cf. Moser (2000). The statistical problem of interest is to compare the multivariate distributions of the outcomes in the control and treatment groups. Moreover, the outcome distributions are expected to be ordered in some sense. The theory of stochastic order relations [Shaked and Shanthikumar (2007)] provides the theoretical foundation for such comparisons.

To fix ideas let $\mathbf{X}$ and $\mathbf{Y}$ be p-dimensional random variables (RVs); $\mathbf{X}$ is said to be smaller than $\mathbf{Y}$ in the multivariate stochastic order, denoted $\mathbf{X} \preceq_{\mathrm{st}} \mathbf{Y}$, provided $\mathbb{P}(\mathbf{X} \in U) \leq \mathbb{P}(\mathbf{Y} \in U)$ for all upper sets $U \in \mathbb{R}^p$ [Shaked and Shanthikumar (2007)]. If for some upper set the above inequality is sharp, we say that $\mathbf{X}$ is strictly smaller than $\mathbf{Y}$ (in the multivariate stochastic order) which we denote by

$\mathbf{X} \prec_{\text{st}} \mathbf{Y}$. Recall that a set $U \in \mathbb{R}^p$ is called an upper set if $\mathbf{u} \in U$ implies that $\mathbf{v} \in U$ whenever $\mathbf{u} \leq \mathbf{v}$, that is, if $u_i \leq v_i$, $i = 1, \ldots, p$. Note that comparing $\mathbf{X}$ and $\mathbf{Y}$ with respect to the multivariate stochastic order requires comparing their distributions over all upper sets in $\mathbb{R}^p$. This turns out to be a very high-dimensional problem. For example, if $\mathbf{X}$ and $\mathbf{Y}$ are multivariate binary RVs, then $\mathbf{X} \preceq_{\text{st}} \mathbf{Y}$ provided $\sum_{\mathbf{t} \in U} p_{\mathbf{X}}(\mathbf{t}) \leq \sum_{\mathbf{t} \in U} p_{\mathbf{Y}}(\mathbf{t})$ where $p_{\mathbf{X}}(\mathbf{t})$ and $p_{\mathbf{Y}}(\mathbf{t})$ are the corresponding probability mass functions. Here $U \in \mathcal{U}_p$ where $\mathcal{U}_p$ is the family of upper sets defined on the support of a p-dimensional multivariate binary RV. It turns out that the cardinality of $\mathcal{U}_p$, denoted by N_p, grows super-exponentially with p. In fact $N_1 = 1$, $N_2 = 4$, $N_3 = 18$, $N_4 = 166$, $N_5 = 7579$ and $N_6 = 7{,}828{,}352$. The values of N_7 and N_8 are also known, but N_9 is not. However, good approximations for N_p are available for all p; cf. Davidov and Peddada (2011). Obviously the number of upper sets for general multivariate RVs is much larger. Since in many applications p is large, it would seem that the analysis of high-dimensional stochastically ordered data is practically hopeless. As a consequence, the methodology for analyzing multivariate ordered data is underdeveloped. It is worth mentioning that Sampson and Whitaker (1989) as well as Lucas and Wright (1991) studied stochastically ordered bivariate multinomial distributions. They noted the difficulty of extending their methodology to high-dimensional data due to the large number of constraints that need to be imposed. Recently Davidov and Peddada (2011) proposed a framework for testing for order among K, p-dimensional, ordered multivariate binary distributions.

In this paper we address the dimensionality problem by considering an easy to understand stochastic order which we refer to as the linear stochastic order.

DEFINITION 1.1.   The RV $\mathbf{X}$ is said to be smaller than the RV $\mathbf{Y}$ in the (multivariate) linear stochastic order, denoted $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$, if for all $\mathbf{s} \in \mathbb{R}_+^p = \{\mathbf{s} : \mathbf{s} \geq 0\}$,

$$(1.1) \qquad\qquad \mathbf{s}^T \mathbf{X} \preceq_{\text{st}} \mathbf{s}^T \mathbf{Y},$$

where $\preceq_{\text{st}}$ in (1.1) denotes the usual (univariate) stochastic order.

Note that it is enough to limit (1.1) to all nonnegative real vectors satisfying $\|\mathbf{s}\| = 1$, and accordingly we denote by $\mathcal{S}_+^{p-1}$ the positive part of the unit sphere in $\mathbb{R}^p$. We call each $\mathbf{s} \in \mathcal{S}_+^{p-1}$ a "direction." In other words the RVs $\mathbf{X}$ and $\mathbf{Y}$ are ordered by the linear stochastic order if every nonnegative linear combination of their components is ordered by the usual (univariate) stochastic order. Thus instead of considering all upper sets in $\mathbb{R}^p$ we need for each $\mathbf{s} \in \mathcal{S}_+^{p-1}$ to consider only upper sets in $\mathbb{R}$. This is a substantial reduction in dimensionality. In fact we will show that only one value of $\mathbf{s}$ need be considered. Note that the linear stochastic order, like the multivariate stochastic order, is a generalization of the usual univariate stochastic order to multivariate data. Both of these orders indicate, in different ways, that one random vector is more likely than another to take on large values.

In this paper we develop the statistical theory and methodology for estimation and testing for linearly ordered multivariate distributions. For completeness we note that weaker notions of the linear stochastic order are discussed by Hu, Homem-de Mello and Mehrotra (2011) and applied to various optimization problems in queuing and finance.

Comparing linear combinations has a long history in statistics. For example, in Phase I clinical trials it is common to compare dose groups using an overall measure of toxicity. Typically, this quantity is an ad hoc weighted average of individual toxicities where the weights are often known as "severity weights;" cf. Bekele and Thall (2004) and Ivanova and Murphy (2009). This strategy of dimension reduction is not new in the statistical literature and has been used in classical multivariate analysis when comparing two or more multivariate normal populations. For example, using the union-intersection principle, the comparison of multivariate normal populations can be reduced to the comparison of all possible linear combinations of their mean vectors. This approach is the basis of Roy's classical largest root test [Roy (1953), Johnson and Wichern (1998)]. Our proposed test may be viewed as nonparametric generalization of the classical normal theory method described above with the exception that we limit consideration only to nonnegative linear combinations (rather than all possible linear combinations) since our main focus is to make comparisons in terms of stochastic order. We emphasize that the linear stochastic order will allow us to address the much broader problem of directional ordering for multivariate ordered data, that is, to find the direction which best separates two ordered distributions. Based on our survey of the literature, we are not aware of any methodology that addresses the problems investigated here.

This paper is organized in the following way. In Section 2 some probabilistic properties of the linear stochastic order are explored, and its relationships with other multivariate stochastic orders are clarified. In Section 3 we provide the background and motivation for directional inference under the linear stochastic order and develop estimation and testing procedure for independent as well as paired samples. In particular the estimator of the best separating direction is presented and its large sampling properties derived. We note that the problem of estimating the best separating direction is a nonsmooth optimization problem. The limiting distribution of the best separating direction is derived in a variety of settings. Tests for the linear stochastic order based on the best separating direction are also developed. One advantage of our approach is that it avoids the estimation of multivariate distributions subject to order restrictions. Simulation results, presented in Section 4, reveal that for large sample sizes the proposed estimator has negligible bias and mean squared error (MSE). The bias and MSE seem to depend on the true value of the best separating direction, the dependence structure and the dimension of the problem. Furthermore, the proposed test honors the nominal type I error rate and has sufficient power. In Section 5 we illustrate the methodology using data obtained from the National Toxicology Program (NTP). Concluding remarks and some open research problems are provided in Section 6. For convenience all

proofs are provided in the Appendix where additional concepts are defined when
needed.

2. Some properties of the linear stochastic order. We start by clarifying
the relationship between the linear stochastic order and the multivariate stochastic
order. First note that $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$ if and only if $\mathbb{P}(\mathbf{s}^T\mathbf{X} \geq t) \leq \mathbb{P}(\mathbf{s}^T\mathbf{Y} \geq t)$ for all
$(t, \mathbf{s}) \in \mathbb{R} \times \mathbb{R}_+^p$ which is equivalent to $\mathbb{P}(\mathbf{X} \in H) \leq \mathbb{P}(\mathbf{Y} \in H)$ for all $H \in \mathcal{H}$ where
$\mathcal{H}$ is the collection of all upper half-planes, that is, sets which are both half planes
and upper sets. Thus $\mathbf{X} \preceq_{\text{st}} \mathbf{Y} \Rightarrow \mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$. The converse does not hold in general.

EXAMPLE 2.1. Let $\mathbf{X}$ and $\mathbf{Y}$ be bivariate RVs such that $\mathbb{P}(\mathbf{X} = (1, 1)) =$
$\mathbb{P}(\mathbf{X} = (0, 1)) = \mathbb{P}(\mathbf{X} = (1, 0)) = 1/3$ and $\mathbb{P}(\mathbf{Y} = (3/4, 3/4)) = \mathbb{P}(\mathbf{Y} = (1, 2)) =$
$\mathbb{P}(\mathbf{Y} = (2, 1)) = 1/3$. It is easy to show that $\mathbf{X}$ is smaller than $\mathbf{Y}$ in the linear
stochastic order but not in the multivariate stochastic order.

The following theorem provides some closure results for the linear stochastic
order.

THEOREM 2.1. (i) *If* $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$, *then* $g(\mathbf{X}) \preceq_{l\text{-st}} g(\mathbf{Y})$ *for any affine increas-*
ing function; (ii) *if* $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$, *then* $\mathbf{X}_I \preceq_{l\text{-st}} \mathbf{Y}_I$ *for each subset* $I \in \{1, \ldots, p\}$;
(iii) *if* $\mathbf{X}|\mathbf{Z} = \mathbf{z} \preceq_{l\text{-st}} \mathbf{Y}|\mathbf{Z} = \mathbf{z}$ *for all* $\mathbf{z}$ *in the support of* $\mathbf{Z}$, *then* $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$; (iv) *if*
$\mathbf{X}_1, \ldots, \mathbf{X}_n$ *are independent RVs with dimensions* p_i *and similarly for* $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$
and if in addition $\mathbf{X}_i \preceq_{l\text{-st}} \mathbf{Y}_i$, *then* $(\mathbf{X}_1, \ldots, \mathbf{X}_n) \preceq_{l\text{-st}} (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$; (v) *finally, if*
$\mathbf{X}_n \to \mathbf{X}$ *and* $\mathbf{Y}_n \to \mathbf{Y}$ *where convergence can be in distribution, in probability or*
almost surely and if $\mathbf{X}_n \preceq_{l\text{-st}} \mathbf{Y}_n$ *for all* n, *then* $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$.

Theorem 2.1 shows that the linear stochastic order is closed under increasing
linear transformations, marginalization, mixtures, conjugations and convergence.
In particular parts (ii) and (iii) of Theorem 2.1 imply that if $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$, then $X_i \preceq_{\text{st}}$
Y_i and $X_i + X_j \preceq_{\text{st}} Y_i + Y_j$ for all i and j; that is, all marginals are ordered as are
all convolutions. Although the multivariate stochastic order is in general stronger
than the linear stochastic order, there are situation in which both orders coincide.

THEOREM 2.2. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be continuous elliptically distributed RVs sup-*
ported on $\mathbb{R}^p$ *with the same generator. Then* $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$ *if and only if* $\mathbf{X} \preceq_{\text{st}} \mathbf{Y}$.

Note that the elliptical family of distributions is large and includes the multi-
variate normal, multivariate t and the exponential power family; see Fang, Kots
and Ng (1989). Thus Theorem 2.2 shows that the multivariate stochastic order co-
incides with the linear stochastic order in the normal family. Incidentally, in the
proof of Theorem 2.2 we generalize the results of Ding and Zhang (2004) on mul-
tivariate stochastic ordering of elliptical RVs. Another interesting example is the
following:

THEOREM 2.3. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be multivariate binary RVs. Then* $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$ *is equivalent to* $\mathbf{X} \preceq_{\text{st}} \mathbf{Y}$ *if and only if* $p \leq 3$.

REMARK 2.1. In the proof of Theorem 2.2 distributional properties of the elliptical family play a major role. In contrast, Theorem 2.3 is a consequence of the geometry of the upper sets of multivariate binary RVs which turn out to be upper half planes if and only if $p \leq 3$.

We now explore the role of the dependence structure.

THEOREM 2.4. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *have the same copula. Then* $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$ *if and only if* $\mathbf{X} \preceq_{\text{st}} \mathbf{Y}$.

Theorem 2.4 establishes that if two RVs have the same dependence structure as quantified by their copula function [cf. Joe (1997)], then the linear and multivariate stochastic orders coincide. Such situations arise when the correlation structure among outcomes is not expected to vary with dose.

The orthant orders are also of interest in statistical applications. We say that $\mathbf{X}$ is smaller than $\mathbf{Y}$ in the upper orthant order, denoted $\mathbf{X} \preceq_{\text{uo}} \mathbf{Y}$, if $\mathbb{P}(\mathbf{X} \in O) \leq \mathbb{P}(\mathbf{Y} \in O)$ for all $O \in \mathcal{O}$ where $\mathcal{O}$ is the collection of upper orthants, that is, sets of the form $\{\mathbf{z} : \mathbf{z} \geq \mathbf{x}\}$ for some fixed $\mathbf{x} \in \mathbb{R}^p$. The lower orthant order is similarly defined; cf. Shaked and Shanthikumar (2007) or Davidov and Herman (2011). It is obvious that the orthant orders are weaker than the usual multivariate stochastic order, that is, $\mathbf{X} \preceq_{\text{st}} \mathbf{Y} \Rightarrow \mathbf{X} \preceq_{\text{uo}} \mathbf{Y}$ and $\mathbf{X} \preceq_{\text{lo}} \mathbf{Y}$. In general the linear stochastic order does not imply the upper (or lower) orthant order, nor is the converse true. However, as stated below, under some conditions on the copula functions, the linear stochastic order implies the upper (or lower) orthant order.

THEOREM 2.5. *If* $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$ *and* $C_{\mathbf{X}}(\mathbf{u}) \leq C_{\mathbf{Y}}(\mathbf{u})$ *for all* $\mathbf{u} \in [0, 1]^p$, *then* $\mathbf{X} \preceq_{\text{lo}} \mathbf{Y}$. *Similarly if* $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$ *and* $\bar{C}_{\mathbf{X}}(\mathbf{u}) \leq \bar{C}_{\mathbf{Y}}(\mathbf{u})$ *for all* $\mathbf{u} \in [0, 1]^p$, *then* $\mathbf{X} \preceq_{\text{uo}} \mathbf{Y}$.

Note that $C_{\mathbf{X}}(\mathbf{u})$ and $\bar{C}_{\mathbf{X}}(\mathbf{u})$ above are the copula and tail-copula functions for the RV $\mathbf{X}$ [cf. Joe (1997)] and are defined in the Appendix and similarly for $C_{\mathbf{Y}}(\mathbf{u})$ and $\bar{C}_{\mathbf{Y}}(\mathbf{u})$. Further note that the relations $C_{\mathbf{X}}(\mathbf{u}) \leq C_{\mathbf{Y}}(\mathbf{u})$ and/or $\bar{C}_{\mathbf{X}}(\mathbf{u}) \leq \bar{C}_{\mathbf{Y}}(\mathbf{u})$ indicate that the components of $\mathbf{Y}$ are more strongly dependent than the components of $\mathbf{X}$. This particular dependence ordering is known as positive quadrant dependence. It can be further shown that strong dependence and the linear stochastic order do not in general imply stochastic ordering.

Additional properties of the linear stochastic order as they relate to estimation and testing problems are given in Section 3.

3. Directional inference.

3.1. *Background and motivation.* There exists a long history of well-developed theory for comparing two or more multivariate normal (MVN) populations. Methods for assessing whether there are any differences between the populations [which differ? in which component(s)? and by how much?] have been addressed in the literature using a variety of simultaneous confidence intervals and multiple comparison methods; cf. Johnson and Wichern (1998). Of particular interest to us is Roy's largest root test. To fix ideas consider two multivariate normal random vectors $\mathbf{X}$ and $\mathbf{Y}$ with means $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, respectively, and a common variance matrix $\boldsymbol{\Sigma}$. Using the union-intersection principle Roy (1953) expressed the problem of testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\nu}$ as a collection of univariate testing problems, by showing that H_0 and H_1 are equivalent to $\bigcap_{\mathbf{s} \in \mathbb{R}^p} H_{0,\mathbf{s}}$ and $\bigcup_{\mathbf{s} \in \mathbb{R}^p} H_{1,\mathbf{s}}$ where $H_{0,\mathbf{s}} : \mathbf{s}^T \boldsymbol{\mu} = \mathbf{s}^T \boldsymbol{\nu}$ and $H_{1,\mathbf{s}} : \mathbf{s}^T \boldsymbol{\mu} \neq \mathbf{s}^T \boldsymbol{\nu}$. Implicitly Roy's test identifies the linear combination $\mathbf{s}_{\max}^T (\boldsymbol{\nu} - \boldsymbol{\mu})$ that corresponds to the largest "distance" between the mean vectors, that is, the direction which best separates their distributions. The resulting test, known as Roy's largest root test, is given by the largest eigenvalue of $\mathbf{BS}^{-1}$ where $\mathbf{B}$ is the matrix of between groups (or populations) sums of squares and cross products, and $\mathbf{S}$ is the usual unbiased estimator of $\boldsymbol{\Sigma}$. In the special case when there are only two populations, this test statistic is identical to Hotelling's T^2 statistic. From the simultaneous confidence intervals point of view, the critical values derived from the null distribution of this statistic can be used for constructing Scheffe's simultaneous confidence intervals for all possible linear combinations of the difference $(\boldsymbol{\mu} - \boldsymbol{\nu})$. Further note that the estimated direction corresponding to Roy's largest root test is $\mathbf{S}^{-1}(\bar{\mathbf{Y}} - \bar{\mathbf{X}})$ where $\bar{\mathbf{X}}$ and $\bar{\mathbf{Y}}$ are the respective sample means.

Our objective is to extend and generalize the classical multivariate method, described above, to nonnormal multivariate ordered data. Our approach will be nonparametric. Recall that comparing MVNs is done by considering the family of statistics $T_{n,m}(\mathbf{s}) = \mathbf{s}^T (\bar{\mathbf{Y}} - \bar{\mathbf{X}})$ for all $\mathbf{s} \in \mathbb{R}^p$. In the case of nonnormal populations, the population mean alone may not be enough to characterize the distribution. In such cases, it may not be sufficient to compare the means of the populations but one may have to compare entire distributions. One possible way of doing so is by considering rank statistics. Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ are independent random samples from two multivariate populations. Let

$$R_k(\mathbf{s}) = \sum_{i=1}^{n} \mathbb{I}_{(\mathbf{s}^T \mathbf{X}_i \leq \mathbf{s}^T \mathbf{X}_k)} + \sum_{j=1}^{m} \mathbb{I}_{(\mathbf{s}^T \mathbf{Y}_j \leq \mathbf{s}^T \mathbf{X}_k)}$$

be the rank of $\mathbf{s}^T \mathbf{X}_k$ in the combined sample $\mathbf{s}^T \mathbf{X}_1, \ldots, \mathbf{s}^T \mathbf{X}_n, \mathbf{s}_1^T \mathbf{Y}_1, \ldots, \mathbf{s}^T \mathbf{Y}_m$. For fixed $\mathbf{s} \in \mathbb{R}^p$ the distributions of $\mathbf{s}^T \mathbf{X}$ and $\mathbf{s}^T \mathbf{Y}$ can be compared using a rank test. For example, if we use $W_{n,m}(\mathbf{s}) = \sum_{i=1}^{n} R_i(\mathbf{s})$ our comparison is done in terms of Wilcoxon's rank sum statistics. It is well known that rank tests are well

suited for testing for univariate stochastic order [cf. Hájek, Šidák and Sen (1999), Davidov (2012)] where the restrictions that $\mathbf{s} \in \mathcal{S}_+^{p-1}$ must be made. Although any rank test can be used, the Mann–Whitney form of Wilcooxon's (WMW) statistic is particularly attractive in this application. Therefore in the rest of this paper we develop estimation and testing procedures for the linear stochastic order based on the family of statistics

$$(3.1) \qquad \Psi_{n,m}(\mathbf{s}) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbb{I}_{(\mathbf{s}^T \mathbf{X}_i \leq \mathbf{s}^T \mathbf{Y}_j)},$$

where $\mathbf{s}$ varies over $\mathcal{S}_+^{p-1}$. Note that (3.1) unbiasedly estimates

$$(3.2) \qquad \Psi(\mathbf{s}) = \mathbb{P}(\mathbf{s}^T \mathbf{X} \leq \mathbf{s}^T \mathbf{Y}).$$

The following result is somewhat surprising.

PROPOSITION 3.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be independent MVNs with means $\boldsymbol{\mu} \leq \boldsymbol{\nu}$ and common variance matrix $\boldsymbol{\Sigma}$. Then Roy's maximal separating direction $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\nu} - \boldsymbol{\mu})$ also maximizes $\mathbb{P}(\mathbf{s}^T \mathbf{X} \leq \mathbf{s}^T \mathbf{Y})$.*

Proposition 3.1 shows that the direction which separates the means, in the sense of Roy, also maximizes (3.2). Thus it provides further support for choosing (3.1) as our test statistic. Note that in general $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\nu} - \boldsymbol{\mu})$ may not belong to $\mathcal{S}_+^{p-1}$. Since we focus on the linear statistical order, we restrict ourselves to $\mathbf{s} \in \mathcal{S}_+^{p-1}$. Consequently we define $\mathbf{s}_{\max} := \arg \max_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \Psi(\mathbf{s})$ and refer to $\mathbf{s}_{\max}$ as the best separating direction. Further note that if $\mathbf{X}$ and $\mathbf{Y}$ are independent and continuous and if $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$, then $\Psi(\mathbf{s}) \geq 1/2$ for all $\mathbf{s} \in \mathcal{S}_+^{p-1}$. This simply means that $\mathbf{s}^T \mathbf{X}$ tends to be smaller than $\mathbf{s}^T \mathbf{Y}$ more than 50% of the time. Note that probabilities of type (3.2) were introduced by Pitman (1937) and further studied by Peddada (1985) for comparing estimators. Random variables satisfying such a condition are said to be ordered by the precedence order [Arcones, Kvam and Samaniego (2002)].

Once $\mathbf{s}_{\max}$ is estimated we can plug it into (3.1) to get a test statistic. Hence our test may be viewed as a natural generalization of Roy's largest root test from MVNs to arbitrary ordered distributions. However, unlike Roy's method, which does not explicitly estimate $\mathbf{s}_{\max}$, we do. On the other hand the proposed test does not require the computation of the inverse of the sample covariance matrix whereas Roy's test and Hotteling's T^2 test require such computations. Consequently, such tests cannot be used when $n < p$ whereas our test can be used in all such instances.

REMARK 3.1. In the above description $\mathbf{X}_i$ and $\mathbf{Y}_j$ are independent for all i and j and therefore the probability $\mathbb{P}(\mathbf{s}^T \mathbf{X}_i \leq \mathbf{s}^T \mathbf{Y}_j)$ is independent of both i and j. However, in many applications such as repeated measurement and crossover

designs, the data are a random sample of dependent pairs $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_N, \mathbf{Y}_N)$ for which $\mathbf{Z}_i = \mathbf{Y}_i - \mathbf{X}_i$ are i.i.d. For example, such a situation may arise when $\mathbf{Y}_i = \mathbf{Y}'_i + \boldsymbol{\epsilon}_i$ and $\mathbf{X}_i = \mathbf{X}'_i + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i$ are pair-specific random effects and the RVs $\mathbf{Y}'_i$ (as well as $\mathbf{X}'_i$) are i.i.d. In this situation $\mathbb{P}(\mathbf{s}^T \mathbf{X}_i \le \mathbf{s}^T \mathbf{Y}_i)$ is independent of i and $\mathbf{s}_{\max}$ is well defined. Moreover the objective function analogous to (3.1) is

$$(3.3) \qquad \Psi_N(\mathbf{s}) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{(\mathbf{s}^T \mathbf{X}_i \le \mathbf{s}^T \mathbf{Y}_i)}.$$

In the following we consider both sampling designs which we refer to as: (a) independent samples and (b) paired or dependent samples. Results are developed primarily for independent samples, but modification for paired samples are mentioned as appropriate.

3.2. *Estimating the best separating direction.* Consider first the case of independent samples, that is, $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ are random samples from the two populations. Rewrite (3.1) as

$$(3.4) \qquad \Psi_{n,m}(\mathbf{s}) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbb{I}_{(\mathbf{s}^T \mathbf{Z}_{ij} \ge 0)},$$

where $\mathbf{Z}_{ij} = \mathbf{Y}_j - \mathbf{X}_i$. The maximizer of (3.4) is denoted by $\hat{\mathbf{s}}_{\max}$, that is,

$$(3.5) \qquad \hat{\mathbf{s}}_{\max} = \arg \max_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \Psi_{n,m}(\mathbf{s}).$$

Finding (3.5) with $\mathbf{s} \in \mathcal{S}_+^{p-1}$ is a nonsmooth optimization problem. Consider first the situation where $p = 2$. In this case we maximize (3.4) subject to $\mathbf{s} \in \mathcal{S}_+^1 = \{(s_1, s_2) : s_1^2 + s_2^2 = 1, (s_1, s_2) \ge 0\}$. Geometrically $\mathcal{S}_+^1$ is a quarter circle spanning the first quadrant. Now let $\mathbf{Z} = (Z_1, Z_2)$, and without any loss of generality assume that $\|\mathbf{Z}\| = 1$. We examine the behavior of the function $\mathbb{I}_{(\mathbf{s}^T \mathbf{Z} \ge 0)}$ as a function of $\mathbf{s}$. Clearly if $\mathbf{Z} \ge 0$, that is, if $Z_1 \ge 0, Z_2 \ge 0$, then for all $\mathbf{s} \in \mathcal{S}_+^1$ we have $\mathbb{I}_{(\mathbf{s}^T \mathbf{Z} \ge 0)} = 1$. In other words any value of $\mathbf{s}$ on the arc $\mathcal{S}_+^1$ maximizes $\mathbb{I}_{(\mathbf{s}^T \mathbf{Z} \ge 0)}$. Similarly if $\mathbf{Z} < 0$ then for all $\mathbf{s} \in \mathcal{S}_+^1$ we have $\mathbb{I}_{(\mathbf{s}^T \mathbf{Z} \ge 0)} = 0$ and again the entire arc $\mathcal{S}_+^1$ maximizes $\mathbb{I}_{(\mathbf{s}^T \mathbf{Z} \ge 0)}$. Now let $Z_1 \ge 0$ and $Z_2 < 0$. It follows that $\mathbb{I}_{(\mathbf{s}^T \mathbf{Z} \ge 0)} = 1$ provided $\cos(\mathbf{s}^T \mathbf{Z}) \ge 0$. Thus $\mathbb{I}_{(\mathbf{s}^T \mathbf{Z} \ge 0)} = 1$ for all $\mathbf{s}$ on the arc $[0, \theta]$ for some θ. If $Z_1 < 0$ and let $Z_2 \ge 0$ the situation is reversed and $\mathbb{I}_{(\mathbf{s}^T \mathbf{Z} \ge 0)} = 1$ for all angles $\mathbf{s}$ on the arc $[\theta, \pi/2]$. The value of θ is given by (3.6). In other words each $\mathbf{Z}$ is mapped to an arc on $\mathcal{S}_+^1$ as described above. Now, the function (3.4) simply counts the number of arcs covering each $\mathbf{s} \in \mathcal{S}_+^1$. The maximizer of (3.4) lies in the region where the maximum number of arcs overlap. Clearly this implies that the maximizer of (3.4) is not unique. A quick way to find the maximizer is the following:

ALGORITHM 3.1. Let M denote the number of $\mathbf{Z}_{ij}$'s which belong to the second or fourth quadrant. Map

$$(3.6) \qquad \mathbf{Z}_{ij} \mapsto \theta_{ij} = \begin{cases} \pi/2 - \cos^{-1}(Z_{ij,1}), & \text{if } Z_{ij,1} \geq 0, Z_{ij,2} < 0, \\ \cos^{-1}(Z_{ij,1}) - \pi/2, & \text{if } Z_{ij,1} < 0, Z_{ij,2} \geq 0. \end{cases}$$

Relabel and order the resulting angles as $\theta_{[1]} < \cdots < \theta_{[M]}$. Also define $\theta_{[0]} = 0$ and $\theta_{[M+1]} = \pi/2$. Evaluate $\Psi_{n,m}(\mathbf{s}_{[i]})$ $i = 1, \ldots, M$ where $\mathbf{s}_{[i],1} = \cos(\theta_{[i]})$ and $\mathbf{s}_{[i],2} = \sin(\theta_{[i]})$. If a maximum is attained at $\theta_{[j]}$, then any value in $[\theta_{[j-1]}, \theta_{[j]}]$ or $[\theta_{[j]}, \theta_{[j+1]}]$ maximizes (3.1).

In light of the above discussion we can be easily prove the following:

PROPOSITION 3.2. *For $p = 2$ Algorithm 3.1 maximizes* (3.1).

In the general case, that is, for $p \geq 3$, each observation $\mathbf{Z}_{ij}$ is associated with a "slice" of $\mathcal{S}_+^{p-1}$. The boundaries of the slice are the intersection of $\mathcal{S}_+^{p-1}$ and some half-plane. Note that when $p = 2$ the slices are arcs. The shape of the slice depends on the quadrant to which $\mathbf{Z}_{ij}$ belongs. The maximizer of (3.1) is again the value of $\mathbf{s}$ which belongs to the largest number of slices. Although the geometry of the resulting optimization problem is easy to understand, we have not been able to devise a simple algorithm, which scales with p, based on the ideas above. However, we have found that (3.5) can be obtained by converting the data into polar coordinates and then using the Nelder–Mead algorithm which does not require the objective function to be differentiable. We emphasize that this maximization process results in a single maximizer of (3.1) and we do not attempt to find the entire set of maximizers. For completness we note that there are methods for optimizing (3.1) specifically designed for nonsmooth problems. For more details see Price, Reale and Robertson (2008) and Audet, Béchard and Le Digabel (2008) and the references therein for both algorithms and convergence results.

REMARK 3.2. It is clear that the estimation procedure for paired samples is the same as for independent samples.

3.3. *Large sample behavior.* We find three different asymptotic regimes for $\hat{\mathbf{s}}_{\max}$ depending on distributional assumptions and the sampling scheme (paired versus independent samples).

Note that the parameter space is the unit sphere not the usual Euclidian space. There are several ways of dealing with this irregularity, one of which is to re-express the last coordinate of $\mathbf{s} \in \mathcal{S}_+^{p-1}$ as $\mathbf{s}_p = \sqrt{1 - s_1^2 - \cdots - s_{p-1}^2}$ and consider the parameter space $\{\mathbf{s} \geq \mathbf{0} : s_1^2 + \cdots + s_{p-1}^2 \leq 1\}$ which is a compact subset of $\mathbb{R}^{p-1}$. Clearly these parameterizations are equivalent, and without any further

ambiguity we will denote them both by $\mathcal{S}_+^{p-1}$. Thus in the proofs below both views of $\mathcal{S}_+^{p-1}$ are used interchangeably as convenient.

We begin our discussion with independent samples assuming continuous distributions for both $\mathbf{X}$ and $\mathbf{Y}$.

THEOREM 3.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ have continuously differentiable densities. If $\Psi(\mathbf{s})$ is uniquely maximized by $\mathbf{s}_{\max} \in \text{interior}(\mathcal{S}_+^{p-1})$. Then $\hat{\mathbf{s}}_{\max}$, the maximizer of* (3.1), *is strongly consistent, that is, $\hat{\mathbf{s}}_{\max} \overset{a.s.}{\to} \mathbf{s}_{\max}$. Furthermore $\hat{\mathbf{s}}_{\max} = \mathbf{s}_{\max} + O_p(N^{-1/2})$ where $N = n + m$. Finally, if $n/(n+m) \to \lambda \in (0,1)$, then*

$$N^{1/2}(\hat{\mathbf{s}}_{\max} - \mathbf{s}_{\max}) \Rightarrow N(0, \boldsymbol{\Sigma}),$$

where the matrix $\boldsymbol{\Sigma}$ is defined in the body of the proof.

Although (3.1) is not continuous (nor differentiable) its U-statistic structure guarantees that it is "almost" so [i.e., it is continuous up to an $o_p(1/N)$ term], and therefore its maximizer converges at a $\sqrt{N}$ rate to a normal limit [Sherman (1993)]. We also note that it is difficult to estimate the asymptotic variance $\boldsymbol{\Sigma}$ directly since it depends on unknown functions ($\nabla \psi_j$ and $\nabla^2 \psi_j$ for $j = 1, 2$ are defined in the body of the proof). Nevertheless bootstrap variance estimates are easily derived.

REMARK 3.3. Note that if either $\mathbf{X}$ or $\mathbf{Y}$ are continuous RVs, then $\Psi(\mathbf{s})$ is continuous. This is a necessary condition for the uniqueness of $\mathbf{s}_{\max}$.

We have not been able to find general condition(s) for a unique maximizer for $\Psi(\mathbf{s})$, although important sufficient conditions can be found. For example:

PROPOSITION 3.3. *If $\mathbf{Z} = \mathbf{Y} - \mathbf{X}$ and there exist $\boldsymbol{\delta} = \boldsymbol{\nu} - \boldsymbol{\mu} \geq \mathbf{0}$ and $\boldsymbol{\Sigma}$ so the distribution of*

$$\frac{\mathbf{s}^T \mathbf{Z} - \mathbf{s}^T \boldsymbol{\delta}}{\sqrt{\mathbf{s}^T \boldsymbol{\Sigma} \mathbf{s}}}$$

is independent of $\mathbf{s}$, then the maximizer of $\Psi(\mathbf{s})$ is unique.

The condition above is satisfied by location scale families, and it may be convenient to think of $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}$ as the location and scale parameters for $\mathbf{Z}$. In general, however, $\Psi(\mathbf{s})$ may not have a unique maximum nor be continuous. For example, if both $\mathbf{X}$ and $\mathbf{Y}$ are discrete RVs, then $\Psi(\mathbf{s})$ is a step function. In such situations $\mathbf{s}_{\max}$ is set valued, and we may denote it by $\mathbf{S}_{\max}$. As we have seen earlier the maximizer of $\Psi_{n,m}(\mathbf{s})$ is always set valued (typically, however, we find only one maximizer). Consider, for example, the case where $\mathbb{P}(\mathbf{X} = (-1, -1)) = \mathbb{P}(\mathbf{X} = (1, 1)) = 1/2$ and let $\mathbb{P}(\mathbf{Y} = (-1, -1)) = 1/2 - \varepsilon$, and $\mathbb{P}(\mathbf{Y} = (1, 1)) = 1/2 + \varepsilon$ for some $\varepsilon > 0$.

It is clear that $\mathbf{X} \prec_{st} \mathbf{Y}$. Further note that $\mathbf{Z} \in \{(2, 2), (0, 0), (-2, -2)\}$, and it follows that $\Psi(\mathbf{s})$ is constant on $\mathcal{S}_+^1$ which implies that $\mathbf{S}_{max}$ coincides with $\mathcal{S}_+^1$. Similarly $\Psi_{n,m}(\mathbf{s})$ is constant on $\mathcal{S}_+^1$ and therefore $\hat{\mathbf{s}}_{max} \in \mathbf{S}_{max}$ for all n, m. This means that consistency is guaranteed and the limiting distribution is degenerate. More generally, we have:

THEOREM 3.2. *If* $\mathbf{X}$ *and* $\mathbf{Y}$ *have discrete distributions with finite support and* $\hat{\mathbf{s}}_{max}$ *is a maximizer of* (3.1), *then*

$$\mathbb{P}(\hat{\mathbf{s}}_{max} \notin \mathbf{S}_{max}) \leq C_1 \exp(-C_2 N)$$

for some positive constants C_1 and C_2.

Theorem 3.2 shows that the probability that a maximizer of $\Psi_{n,m}$ is not in $\mathbf{S}_{max}$ is exponentially small when the underlying distributions of $\mathbf{X}$ and $\mathbf{Y}$ are discrete. Hence $\hat{\mathbf{s}}_{max}$ is consistent and converges exponentially fast. In fact the proof of Theorem 3.2 implies that $\hat{\mathbf{S}}_{max} \to \mathbf{S}_{max}$; that is, the set of maximizers of $\Psi_{n,m}(\cdot)$ converges to the set of maximizers of $\Psi(\cdot)$; that is, $\rho_H(\hat{\mathbf{S}}_{max}, \mathbf{S}_{max}) \to 0$ where ρ_H is the Hausdroff metric defined on compact sets. A careful reading of the proof shows that Theorem 3.2 also holds under paired samples.

Finally, we consider the case of continuous RVs under paired samples. Then under the conditions of Theorem 3.1 and provided the density of $\mathbf{Z} = \mathbf{Y} - \mathbf{X}$ is bounded we have:

THEOREM 3.3. *Under the above mentioned conditions* $\hat{\mathbf{s}}_{max}$, *the maximizer of* (3.3), *is strongly consistent, that is,* $\hat{\mathbf{s}}_{max} \overset{a.s.}{\to} \mathbf{s}_{max}$, *converges at a cube root rate, that is,* $\hat{\mathbf{s}}_{max} = \mathbf{s}_{max} + O_p(N^{-1/3})$, *and*

$$N^{1/3}(\hat{\mathbf{s}}_{max} - \mathbf{s}_{max}) \Rightarrow \mathbf{W},$$

where $\mathbf{W}$ *has the distribution of the almost surely unique maximizer of the process* $\mathbf{s} \longmapsto -[Q(\mathbf{s}) + \mathbb{W}(\mathbf{s})]$ *on* $\mathcal{S}_+^{p-1}$ *where $Q(\mathbf{s})$ is a quadratic function and $\mathbb{W}(\mathbf{s})$ is a zero mean Gaussian process described in the body of the proof.*

Theorem 3.3 shows that in paired samples $\hat{\mathbf{s}}_{max}$ is consistent, but in contrast with Theorem 3.1 it converges at a cube-root rate to a nonnormal limit. The cube root rate is due to the discontinuous nature of the objective function (3.3). General results dealing with this kind of asymptotics for independent observations are given by Kim and Pollard (1990). The main difference between Theorems 3.1 and 3.3 is that the objective function (3.1) is smoothed by its U-statistic structure while (3.3) is not.

3.4. *A confidence set for* $\hat{\mathbf{s}}_{\max}$. Since the parameter space is the surface of a unit sphere it is natural to define the $(1 - \alpha) \times 100\%$ confidence set for $\mathbf{s}_{\max}$ centered at $\hat{\mathbf{s}}_{\max}$ by

$$\{\mathbf{s} \in \mathcal{S}_+^{p-1} : \hat{\mathbf{s}}_{\max}^T \mathbf{s} \le C_{\alpha,N}\},$$

where $C_{\alpha,N}$ satisfies $\mathbb{P}(\hat{\mathbf{s}}_{\max}^T \mathbf{s} \le C_{\alpha,N}) = 1 - \alpha$. For more details see Fisher and Hall (1989) or Peddada and Chang (1996). Hence the confidence set is the set of all $\mathbf{s} \in \mathcal{S}_+^{p-1}$ which have a small angle with $\hat{\mathbf{s}}_{\max}$. In theory one may appeal to Theorem 3.1 to derive the critical value for any $\alpha \in (0, 1)$. However the limit law in Theorem 3.1 requires knowledge of unknown parameters and functions. For this reason, we explore the bootstrap for estimating $C_{\alpha,N}$.

REMARK 3.4.   Since in the case of paired samples, the estimator converges at cube root rate rather than the square root rate, the standard bootstrap methodology may yield inaccurate coverage probabilities; see Abrevaya and Huang (2005) and Sen, Banerjee and Woodroofe (2010). For this reason we recommend the "M out of N" bootstrap methodology. For further discussion on the "M out of N" bootstrap methodology one may refer to Lee (1999), Delagdo, Rodriguez-Poo and Wolf (2001), Bickel and Sakov (2008).

3.5. *Testing for order.*   Consider first the case of independent samples where interest is in testing the hypothesis

(3.7) $H_0 : \mathbf{X} =_{\text{st}} \mathbf{Y}$ versus $H_1 : \mathbf{X} \prec_{\text{st}} \mathbf{Y}$.

Thus (3.7) tests whether the distributions of $\mathbf{X}$ and $\mathbf{Y}$ are equal or ordered (later on we briefly discuss testing $H_0 : \mathbf{X} \preceq_{\text{st}} \mathbf{Y}$ versus $H_1 : \mathbf{X} \nprec_{\text{st}} \mathbf{Y}$). In this section we propose a new test for detecting an ordering among two multivariate distributions based on the maximal separating direction. The test is based on the following observation:

THEOREM 3.4.   *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be independent and continuous RVs. If* $\mathbf{X} =_{\text{st}} \mathbf{Y}$, *then* $\mathbb{P}(\mathbf{s}^T\mathbf{X} \le \mathbf{s}^T\mathbf{Y}) = 1/2$ *for all* $\mathbf{s} \in \mathcal{S}_+^{p-1}$, *and if both* (i) $\mathbf{X} \preceq_{\text{st}} \mathbf{Y}$ *and* (ii) $\mathbb{P}(\mathbf{s}^T\mathbf{X} \le \mathbf{s}^T\mathbf{Y}) > 1/2$ *for some* $\mathbf{s} \in \mathcal{S}_+^{p-1}$ *hold, then* $\mathbf{X} \prec_{\text{st}} \mathbf{Y}$.

Theorem 3.4 says that if it is known a priori that $\{\mathbf{X} \preceq_{\text{st}} \mathbf{Y}\} = \{\mathbf{X} =_{\text{st}} \mathbf{Y}\} \cup \{\mathbf{X} \prec_{\text{st}} \mathbf{Y}\}$, that is, the RVs are either equal or ordered [which is exactly what (3.7) implies], then a strict linear stochastic ordering implies a strict ordering by the usual multivariate stochastic order. In particular under the alternative there must exist a direction $\mathbf{s} \in \mathcal{S}_+^{p-1}$ for which $\mathbf{s}^T\mathbf{X} \prec_{l\text{-st}} \mathbf{s}^T\mathbf{Y}$.

REMARK 3.5.   The assumption that $\mathbf{X} \preceq_{\text{st}} \mathbf{Y}$ is natural in applications such as environmental sciences where high exposures are associated with increased risk.

Nevertheless if the assumption that $\mathbf{X} \preceq_{\mathrm{st}} \mathbf{Y}$ is not warranted then the alternative hypothesis formulated in terms of the linear stochastic order actually tests whether there exists a $\mathbf{s} \in \mathcal{S}_+^{p-1}$ for which $\mathbb{P}(\mathbf{s}^T \mathbf{X} \leq \mathbf{s}^T \mathbf{Y}) > 1/2$. This amounts to a precedence (or Pitman) ordering between $\mathbf{s}^T \mathbf{X}$ and $\mathbf{s}^T \mathbf{Y}$.

REMARK 3.6. In the proof of Theorem 3.4 we use the fact that given that $\mathbf{X} \preceq_{\mathrm{st}} \mathbf{Y}$ we have $\mathbf{X} \prec_{\mathrm{st}} \mathbf{Y}$ provided $X_i \prec_{\mathrm{st}} Y_i$ for some $1 \leq i \leq p$. Note that if $X_i \prec_{\mathrm{st}} Y_i$, then $\mathbb{E}(X_i) < \mathbb{E}(Y_i)$. Thus it is possible to test (3.7) by comparing means (or any other monotone function of the data). Although such a test will be consistent it may lack power because tests based on means are often far from optimal when the data is not normally distributed. The WMW procedure, however, is known to have high power for a broad collection of underlying distributions.

Hence (3.7) can be reformulated in terms of the linear stochastic. In particular it justifies using the statistic

$$(3.8) \qquad S_{n,m} = N^{1/2}\big(\Psi_{n,m}(\hat{\mathbf{s}}_{\max}) - 1/2\big).$$

To the best of our knowledge this is the first general test for multivariate ordered distributions. In practice we first estimate $\hat{\mathbf{s}}_{\max}$ and then define $\hat{U}_i = \hat{\mathbf{s}}_{\max}^T \mathbf{X}_i$ and $\hat{V}_j = \hat{\mathbf{s}}_{\max}^T \mathbf{Y}_j$ where $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Hence (3.8) is nothing but a WMW test based on the $\hat{U}'$'s and $\hat{V}$'s. It is also a Kolmogorov–Smirnov type test.

The large sample distribution of (3.8) is given in the following.

THEOREM 3.5. *Suppose the null* (3.7) *holds. Let* $n, m \to \infty$ *and* $n/(n + m) \to \lambda \in (0, 1)$. *Then*

$$S_{n,m} \Rightarrow S = \sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \mathbb{G}(\mathbf{s}),$$

where $\mathbb{G}(\mathbf{s})$ *is a zero mean Gaussian process with covariance function given by* (A.20).

REMARK 3.7. Since $\hat{\mathbf{s}}_{\max} \xrightarrow{\mathrm{a.s.}} \mathbf{s}_{\max}$ by Slutzky's theorem the power of test (3.8) converges to the power of a WMW test comparing the samples $(\mathbf{s}_{\max}^T \mathbf{X}_1, \ldots, \mathbf{s}_{\max}^T \mathbf{X}_n)$ and $(\mathbf{s}_{\max}^T \mathbf{Y}_1, \ldots, \mathbf{s}_{\max}^T \mathbf{Y}_m)$. The "synthetic" test, assuming that $\mathbf{s}_{\max}$ is known, serves as a gold standard as verified by our simulation study.

REMARK 3.8. Furthermore, the power of the test under local alternatives, that is, when $\mathbf{Y} =_{\mathrm{st}} \mathbf{X} + N^{-1/2}\boldsymbol{\delta}$ and $N \to \infty$ is bounded by the power of the WMW test comparing the distributions of $\mathbf{s}_{\max}^T \mathbf{X}$ and $\mathbf{s}_{\max}^T \mathbf{Y} = \mathbf{s}_{\max}^T \mathbf{X} + N^{-1/2}\mathbf{s}_{\max}^T \boldsymbol{\delta}$.

Alternatives to the "sup" statistic (3.8) are the "integrated" statistics

$$
I_{n,m} = \int_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \big[N^{1/2}\big(\Psi_{n,m}(\mathbf{s}) - 1/2\big) \big] \, d\mathbf{s} \quad \text{and}
$$

(3.9)

$$
I_{n,m}^+ = \int_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \big[N^{1/2}\big(\Psi_{n,m}(\mathbf{s}) - 1/2\big) \big]_+ \, d\mathbf{s},
$$

where $[x]_+ = \max(0, x)$. It is clear that $I_{n,m} \Rightarrow N(0, \sigma^2)$ where

$$
\sigma^2 = \int_{\mathbf{u} \in \mathcal{S}_+^{p-1}} \int_{\mathbf{v} \in \mathcal{S}_+^{p-1}} C(\mathbf{u}, \mathbf{v}) \, d\mathbf{u} \, d\mathbf{v}
$$

and $C(\mathbf{u}, \mathbf{v})$, the covariance function of $\mathbb{G}$, is given by (A.20). Also

$$
I_{n,m}^+ \Rightarrow \int_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \big[\mathbb{G}(\mathbf{s}) \big]_+ \, d\mathbf{s}.
$$

This distribution does not have a closed form. The statistics $I_{n,m}$ and $I_{n,m}^+$ have univariate analogues; cf. Davidov and Herman (2012). Finally, we have the following theorem:

THEOREM 3.6.   *The tests* (3.8) *and* (3.9) *are consistent. Furthermore if* $\mathbf{X} \preceq_{l\text{-st}}$ $\mathbf{Y} \preceq_{l\text{-st}} \mathbf{Z}$, *then all three tests for* $H_0 : \mathbf{X} =_{\text{st}} \mathbf{Z}$ *versus* $H_1 : \mathbf{X} \prec_{\text{st}} \mathbf{Z}$ *are more powerful than the respective tests for* $H_0 : \mathbf{X} =_{\text{st}} \mathbf{Y}$ *versus* $H_1 : \mathbf{X} \prec_{\text{st}} \mathbf{Y}$.

Theorem 3.6 shows that the tests are consistent and that their power function is "monotone" in the linear stochastic order.

REMARK 3.9.   Qualitatively similar results are obtainable in the paired sampling case; the only difference being the limiting process. For example, it easy to see that the paired sample analogue of (3.8) satisfies

$$
N^{1/2}\big(\Psi_N(\hat{\mathbf{s}}_{\max}) - 1/2\big) \Rightarrow \sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \mathbb{Q}(\mathbf{s}),
$$

where $\mathbb{Q}(\mathbf{s})$ is the empirical process on $\mathcal{S}_+^{p-1}$ associated with (3.3). Analogues of $I_{n,m}$ and $I_{n,m}^+$ are similarly defined and analyzed. Tests for paired samples may be similarly implemented using bootstrap or permutation methods.

4. Simulations.   For simplicity of exposition, and motivated by the fact that the example we analyzed in this paper deals with independent samples, we limit our simulations to the case of independent samples.

4.1. *The distribution of* $\hat{\mathbf{s}}_{\max}$. We start by investigating the distribution of $\hat{\mathbf{s}}_{\max}$ by simulation. For simplicity we choose $p = 3$ and generated $\mathbf{X}_i$ $(i = 1, \ldots, n)$ distributed as $N_3(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{Y}_j$ $(j = 1, \ldots, m)$ distributed as $N_3(\boldsymbol{\delta}, \mathbf{\Sigma})$ where $\mathbf{\Sigma} = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$, $\mathbf{I}$ is the identity matrix and $\mathbf{J}$ is a square matrix of 1s. We simulated 1000 realizations of $\hat{\mathbf{s}}_{\max}$ for various sample sizes and correlation coefficients. To get a visual description of the density of $\hat{\mathbf{s}}_{\max}$, we provide a pair of plots for each configuration of ρ and sample size n. In Figure 1 we provide the joint density of the two-dimensional polar angles (θ, ϕ) of $\hat{\mathbf{s}}_{\max}$. There are four panels in Figure 1, corresponding to all combinations of $\rho = 0, 0.9$ and $n = 10, 100$. The mean vector $\boldsymbol{\delta}$ in this plot was taken to be $\boldsymbol{\delta} = (2, 2, 2)^T$. In Figure 2 we provide the density of the polar residual defined by $1 - \hat{\mathbf{s}}_{\max}^T \mathbf{s}_{\max}$. The four panels of Figure 2 correspond to all combinations of $\rho = 0, 0.9$ and $n = 10, 100$ and two patterns of $\boldsymbol{\delta}$, namely, $(2, 2, 2)^T$ and $(3, 2, 1)^T$. We see from Figure 1 that $\hat{\mathbf{s}}_{\max}$ converges to a unimodal, normal looking distribution as the sample size increases. Interestingly, from Figure 2 we see that the concentration of the distribution around the true parameter depends upon the values of $\boldsymbol{\delta}$ and ρ (which together determine $\mathbf{s}_{\max}$). If the components of the underlying random vector are exchangeable [e.g., $\boldsymbol{\delta} = (2, 2, 2)^T$], the residuals tend to concentrate more closely around zero [Figure 2(a) and (c)] compared to the case when they are not exchangeable [Figure 2(b) and (d)].

4.2. *Study design.* The simulation study consists of three parts. In the first part we evaluate the accuracy and precision of $\hat{\mathbf{s}}_{\max}$ by estimating its bias and mean squared error (MSE). In the second part we investigate the coverage probability of bootstrap confidence intervals. In the third part we estimate type I errors and powers of the proposed test $S_{n,m}$ as well as the integral tests $I_{n,m}$ and $I_{n,m}^+$.

To evaluate the bias and MSEs we generated $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim N_3(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_m \sim N_3(\boldsymbol{\delta}, \mathbf{\Sigma})$ where $n = m = 20$ or 100 observations. The common variance matrix is assumed to have intra-class correlation structure, that is, $\mathbf{\Sigma} = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$ where $\mathbf{I}$ is the identity matrix and $\mathbf{J}$ is a matrix of ones. Various patterns of the mean vectors $\boldsymbol{\delta}$ and correlation coefficient ρ were considered as described in Table 1.

We conducted extensive simulation studies to evaluate the performance of the bootstrap confidence intervals. In this paper we present a small sample of our study. We generated data from two 5-dimensional normal populations with means $\mathbf{0}$ and $\boldsymbol{\delta}$, respectively, and a common covariance $\mathbf{\Sigma} = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$. We considered 5 patterns of ρ and 2 patterns of sample sizes ($n = m = 20$ and 40). The nominal coverage probability was 0.95. Results are summarized in Table 2.

The goal of the third part of our simulation study is to evaluate the type I error and the power of the test (3.8). To evaluate the type I error three different baseline distributions for the two populations $\mathbf{X}$ and $\mathbf{Y}$ were employed as follows: (1) both distributed as $N(\mathbf{0}, \mathbf{\Sigma})$; (2) both distributed as $\pi N(\mathbf{0}, \mathbf{\Sigma}) + (1 - \pi)N(\boldsymbol{\delta}, \mathbf{\Sigma})$ with $\pi = 0.2$ or $\pi = 0.8$; and (3) both distributed as $\exp(\mathbf{Z}) = (\exp(Z_1), \ldots, \exp(Z_p))$

Figure 1(a):delta = (2,2,2),rho=0, n=10

Figure 1(b):delta = (2,2,2),rho=0, n=100

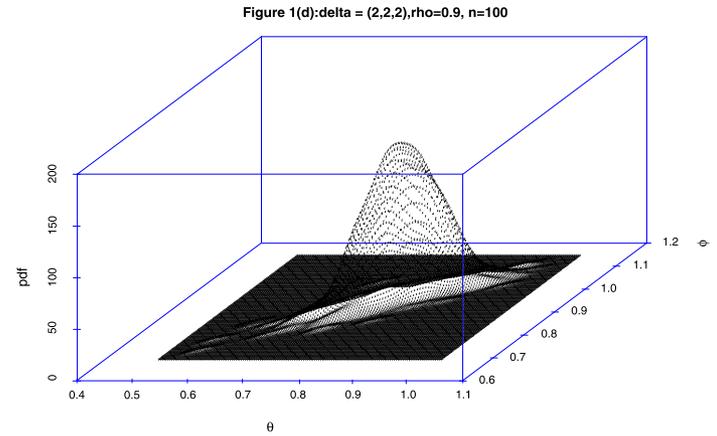**Figure 1(c):delta = (2,2,2),rho=0.9, n=10**

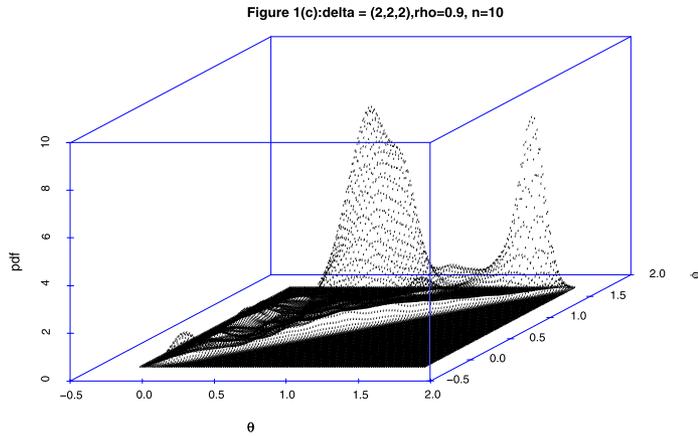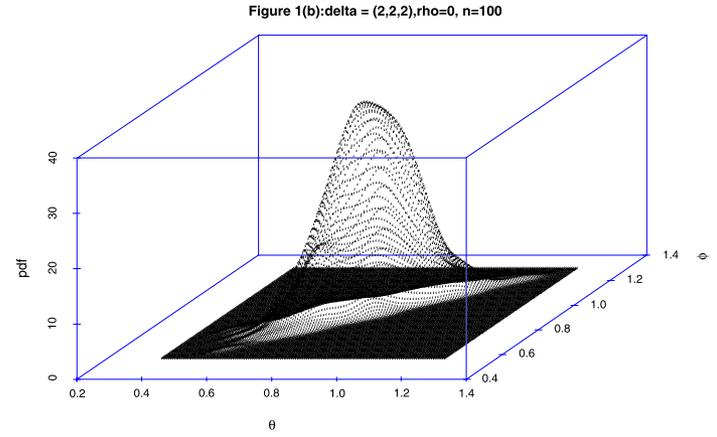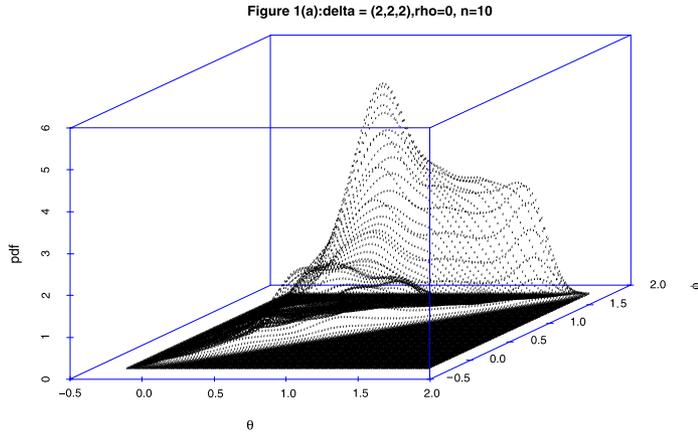**Figure 1(d):delta = (2,2,2),rho=0.9, n=100**

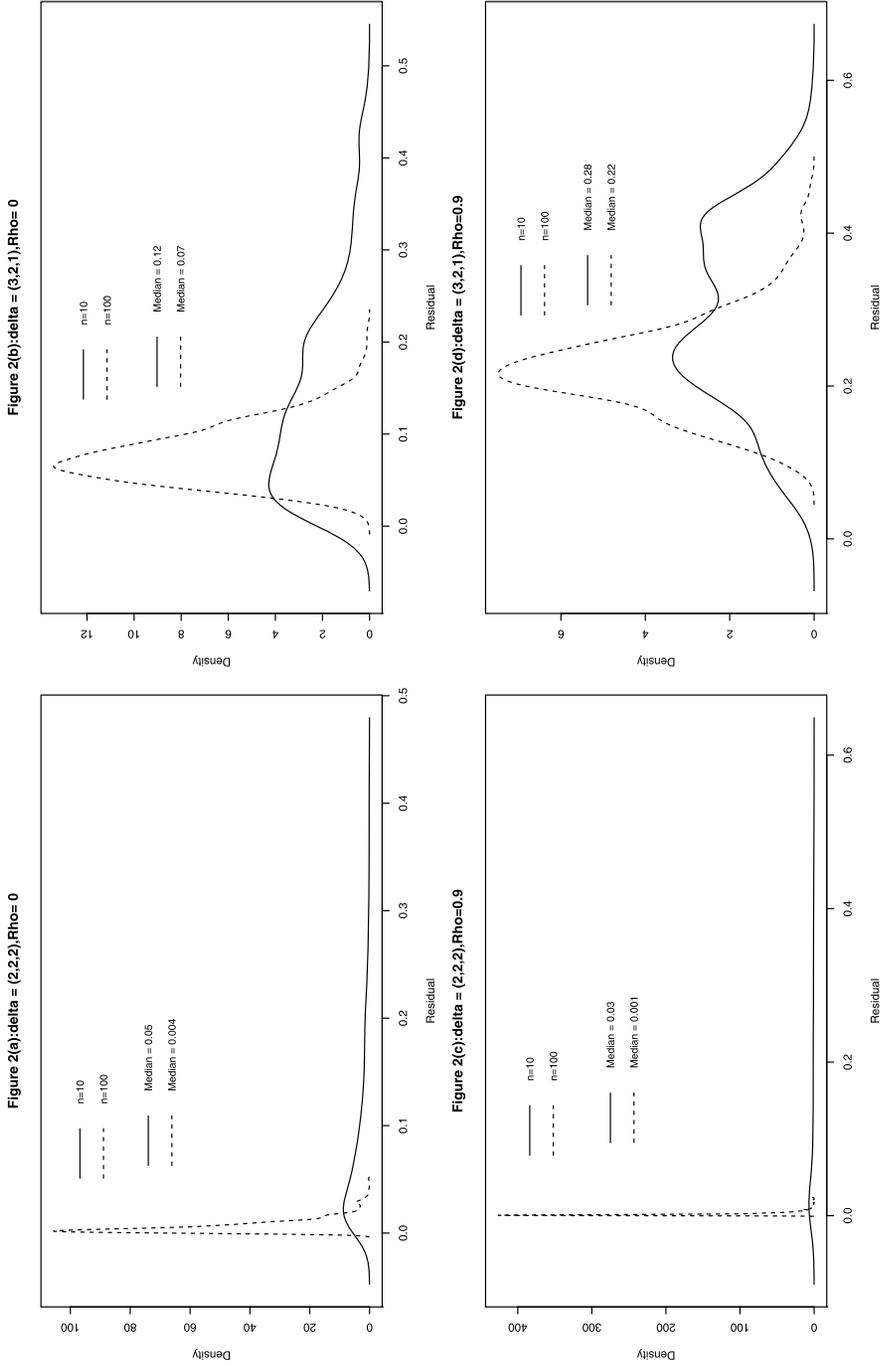FIG. 1. *Plot of* (*simulated*) *polar angles*.

F𝙸G. 2.   *Plot of (simulated) polar residuals.*

where $\mathbf{Z}$ follows a $N(\boldsymbol{\delta}, \boldsymbol{\Sigma})$. We refer to this distribution as the multivariate lognormal distribution. Throughout the variance matrix is assumed to have the intra-class structure described above. Various patterns of the mean vectors $\boldsymbol{\delta}$ and correlation coefficient ρ the dimension p were considered as described in Table 3. Sample sizes of $n = m = 15$ or 25 are reported.

Power comparisons were carried out for data generated from $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_m \sim N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ where $p = 3$ or 5 and a variety of patterns for $\boldsymbol{\delta}$ as described in Table 4. If Roy's maximal separating direction (cf. Proposition 3.1) was known then a "natural gold standard" would be the test based on $\Psi_{n,m}(\mathbf{s}_{\max})$. We shall refer to this test as the true maximal direction (TMD) test. Clearly the TMD test cannot be used in practice since it involves the unknown direction $\mathbf{s}_{\max}$. Nevertheless the TMD test provides an upper bound for the power of the proposed test which uses the estimated direction. Hence we compute the efficiency of the proposed test relative to TMD test. An additional test, referred to as the RMD test is also compared. The RMD test has the same form but uses Roy's maximal direction given by $\mathbf{S}^{-1}(\bar{\mathbf{Y}} - \bar{\mathbf{X}})$. As suggested by a reviewer we also evaluated the power of the two integral based tests, described in (3.9), which do not require the determination of the best separating direction.

Additionally, in Table 5 we evaluate the type I error and power of our test when $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_m \sim N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ and $n = m = p = 10$ and $n = m = 10$ and $p = 20$ (i.e., $p < n$ set up). Note that in neither of these cases the standard Hotteling's T^2 (or Roy's largest root test) can be computed whereas the proposed test can be calculated.

Simulation results reported in this paper are based on 1000 simulation runs. Confidence sets are calculated using 1000 bootstrap samples. The bootstrap critical values for estimating type I error were based on 500 bootstrap samples. Since the results between 100 bootstrap samples and 500 bootstrap samples did not differ by much, all powers were estimated using 100 bootstrap samples.

4.3. *Simulation results.* The Bias and MSEs for the patterns considered are summarized in Table 1. It is clear that the bias decreases with the sample size as do the MSEs. We observe that the bias tends to be smaller under independence and negative dependence compared with positive dependence. It also tends to be smaller when the data are exchangeable. Although results are not presented, we evaluated squared bias and MSE for larger values of p (e.g., $p = 5$, 10 and 20) and as expected the total squared bias and total MSE increased with the dimension p.

In Table 2 we summarize the estimated coverage probabilities of the bootstrap confidence intervals when $p = 5$. Our simulation study suggests that the proposed bootstrap methodology seems to perform better for larger sample sizes but rather poorly for smaller samples sizes.

Type I errors for different patterns considered in our simulation study are summarized in Table 3. Our simulation studies suggest that in every case the proposed bootstrap based test maintains the nominal level of 0.05. In general it is slightly

TABLE 1
Bias and MSE of $\hat{s}_{max}$

δ	ρ	Bias	MSE
		$n = m = 20$	
(1, 1, 1)	−0.25	0.001	0.072
(1, 1, 1)	0	0.004	0.129
(1, 1, 1)	0.25	0.009	0.187
(1, 1, 1)	0.50	0.012	0.216
(1, 1, 1)	0.90	0.010	0.203
(3, 2, 1)	−0.25	0.018	0.090
(3, 2, 1)	0	0.001	0.066
(3, 2, 1)	0.25	0.053	0.114
(3, 2, 1)	0.50	0.060	0.113
(3, 2, 1)	0.90	0.112	0.170
		$n = m = 100$	
(1, 1, 1)	−0.25	0.00009	0.014
(1, 1, 1)	0	0.00021	0.027
(1, 1, 1)	0.25	0.00045	0.041
(1, 1, 1)	0.50	0.00079	0.056
(1, 1, 1)	0.90	0.00050	0.044
(3, 2, 1)	−0.25	0.02400	0.039
(3, 2, 1)	0	0.00004	0.012
(3, 2, 1)	0.25	0.05200	0.065
(3, 2, 1)	0.50	0.06400	0.077
(3, 2, 1)	0.90	0.14100	0.158

conservative. The performance of the test is not affected by the shape of the underlying distribution. This is not surprising, owing to the nonparametric nature of the test. Furthermore, we evaluated the type I error of the proposed bootstrap test for testing the null hypothesis (3.7) for p as large as 20 with $n = m = 10$ and discovered that the proposed test attains the nominal level of 0.05 even $n \leq p$. See Table 4. As commented earlier in the paper, Hotelling's T^2 statistic cannot be applied here since the Wishart matrix is singular in this case. However, the proposed method is still applicable since the estimation of the best direction does not require the inversion of a matrix.

The power of tests (3.8) and (3.9) for various patterns considered in our simulation study are summarized in Table 5.

As expected, in every case the power of the TMD test is higher than that of $S_{n,m}$ test and the RMD test. The $S_{n,m}$ test is almost always more powerful than the RMD test. The relative efficiency of $S_{n,m}$ compared to the TMD test is quite high in most cases. When $n = m = 15$ the relative efficiency ranges between 65–95%. It is almost always above 90% when the sample size increases to 25 per group. In

TABLE 2
*Coverage probabilities for the bootstrap confidence intervals
for $p = 5$ normal data. Pattern $i = 1, 2$ corresponds to
$\delta_1 = (0.1, 0.25, 0.5, 0.75, 0.9)$ and $\delta_2 = (0.5, 0.5, 0.5, 0.5, 0.5)$*

Set up		Coverage probability	
Pattern	ρ	$n = m = 20$	$n = m = 40$
1	-0.25	0.981	0.971
1	0	0.913	0.918
1	0.25	0.916	0.933
1	0.50	0.971	0.969
1	0.90	0.993	0.989
2	-0.25	0.982	0.967
2	0	0.984	0.972
2	0.25	0.986	0.978
2	0.50	0.968	0.968
2	0.90	0.950	0.954

general the two integral tests had very similar power. They had larger power than $S_{n,m}$ when $\rho < 0$. As ρ increased, the power of $S_{n,m}$ improved relative to the two integral tests. Test (3.8) seems to perform better when the components of δ were unequal. We also note that when the integral tests outperform $S_{n,m}$ the difference is usually small, whereas the $S_{n,m}$ test can outperform the integral tests substantially. For example, observe pattern 2 where the powers of $S_{n,m}$ and $I_{n,m}$ are 0.93 and 0.97, respectively, when $\rho = -0.25$ and 0.63 versus 0.40 when $\rho = 0.90$.

5. Illustration. Prior to conducting a two-year rodent cancer bioassay to evaluate the toxicity/carcinogenicity of a chemical, the National Toxicology Program (NTP) routinely conducts a 90-day pre-chronic dose finding study. One of the goals of the 90-day study is to determine the maximum tolerated dose (MTD) that can be used in the two-year chronic exposure study. Accurate determination of the MTD is critical for the success of the two-year cancer bioassay. Cancer bioassays are typically very expensive and time consuming. Therefore their proper design, that is, choosing the correct dosing levels, is very important. When the highest dose used in the two-year study exceeds the MTD, a large proportion of animals in the high dose group(s) may die well before the end of the study, and the data from such group(s) cannot be used reliably. This results in inefficiency and wasted resources.

Typically the NTP uses the 90-day study to determine the MTD on the basis of a large number of correlated endpoints that provide information regarding toxicity. These include body weight, organ weights, clinical chemistry (red blood cell counts, cell volume, hemoglobin, hematocrit, lymphocytes, etc.), histopathology (lesions in various target organs), number of deaths and so forth. The dose response

TABLE 3
Type I errors for the proposed procedure with nominal level $\alpha = 0.05$.
Three types of distributions are considered: MVNs, MV-LogN
(multivariate lognormal) and Mix-MVN (mixtures of MVNs)

Set up			Type I error	
Distribution	p	ρ	$n = m = 15$	$n = m = 25$
MVNs	3	−0.25	0.041	0.037
MVNs	3	0.00	0.023	0.044
MVNs	3	0.25	0.037	0.033
MVNs	3	0.50	0.027	0.032
MVNs	3	0.90	0.031	0.036
MVNs	5	−0.25	0.035	0.035
MVNs	5	0.00	0.040	0.041
MVNs	5	0.25	0.045	0.032
MVNs	5	0.50	0.038	0.043
MVNs	5	0.90	0.044	0.031
MV-LogN	3	−0.25	0.025	0.040
MV-LogN	3	0.00	0.038	0.049
MV-LogN	3	0.25	0.025	0.027
MV-LogN	3	0.50	0.028	0.037
MV-LogN	3	0.90	0.026	0.034
MV-LogN	5	−0.25	0.026	0.039
MV-LogN	5	0.00	0.035	0.018
MV-LogN	5	0.25	0.039	0.039
MV-LogN	5	0.50	0.036	0.046
MV-LogN	5	0.90	0.034	0.042
Mix-MVNs	3	−0.25	0.032	0.040
Mix-MVNs	3	0.00	0.038	0.028
Mix-MVNs	3	0.25	0.039	0.032
Mix-MVNs	3	0.50	0.036	0.035
Mix-MVNs	3	0.90	0.041	0.028
Mix-MVNs	5	−0.25	0.042	0.035
Mix-MVNs	5	0.00	0.040	0.031
Mix-MVNs	5	−0.25	0.041	0.028
Mix-MVNs	5	0.50	0.034	0.040
Mix-MVNs	5	0.90	0.042	0.036

data is analyzed for each variable separately using Dunnett's or the Williams's test (or their nonparametric versions, Dunn's test and Shirley's test, resp.). NTP combines results from all such analyses qualitatively and uses other biological and toxicological information when making decisions regarding the highest dose for the two-year cancer bioassay. Analyzing correlated variables one at a time may result in loss of information. The proposed methodology provides a convenient

TABLE 4
Type I errors and power for some settings with $p \geq n$. Here $n = m = 10$ and δ_1
has components $1/2$ and δ_2 has components i/p

Type I error and power $p = 10$, $n = m = 10$			Type I error and power $p = 20$, $n = m = 10$		
δ	ρ	Type I error	δ	ρ	Type I error
0	0.00	0.054	**0**	0.00	0.081
0	0.25	0.051	**0**	0.25	0.050
0	0.50	0.028	**0**	0.50	0.046
0	0.90	0.038	**0**	0.90	0.048
Power			**Power**		
δ_1	0.00	0.83	δ_1	0.00	0.97
δ_1	0.25	0.48	δ_1	0.25	0.53
δ_1	0.50	0.26	δ_1	0.50	0.42
δ_1	0.90	0.20	δ_1	0.90	0.22
δ_2	0.00	0.98	δ_2	0.00	0.98
δ_2	0.25	0.80	δ_2	0.25	0.59
δ_2	0.50	0.67	δ_2	0.50	0.43
δ_2	0.90	0.71	δ_2	0.90	0.40

method to combine information from several outcome variables to make comparisons between groups.

We now illustrate our methodology by re-analyzing data obtained from a recent NTP study of the chemical Citral [NTP (2003)]. Citral is a flavoring agent that is widely used in a variety of food items. The NTP assigned a random sample of 10 male rats to the control group and 10 to the 1785 mg/kg dose group. Hematological and clinical chemistry measurements such as the number of platelets (in 1000 per l), urea nitrogen (UN) (in mg/dl), alkaline phosphatase (AP) (in IU/l) and bile acids (BA) (in mol/l) were recorded on each animal at the end of the study. The NTP performed univariate analysis on each of these variables and found no significant difference between the control and dose group except for the concentration of urea nitrogen which was increased in the high dose group. This increase was marginally significant at the 5% level and not at all after correcting for multiplicity. We applied the proposed methodology to compare the control with the high-dose group (1785 mg/kg) in terms of all nonnegative linear combinations of the above mentioned four variables. We test the null hypothesis of no difference between the control and the high-dose group against the alternative that the high-dose group is stochastically larger (in the above four variables) than the control group. The resulting p-value based on 10,000 bootstrap samples was 0.025, which is significant at a 5% level of significance. The estimated value of $\mathbf{s}_{\max}$ was $(0.074, 0.986, 0.012, 0.150)^T$ and the estimated 95% confidence region is given by

TABLE 5
Power comparisons of the two proposed test procedures with type I error of 0.050. *Here*
$\delta_1 = (0.1, 0.5, 0.9)$, $\delta_2 = (0.1, 0.25, 0.5, 0.75, 0.9)$, $\delta_3 = (0.5, 0.5, 0.5)$ *and*
$\delta_4 = (0.5, 0.5, 0.5, 0.5, 0.5)$

	Set up		**Directional tests**			**Integral tests**	
p	δ	ρ	$S_{n,m}$	**RMD test**	**TMD test**	$I_{n,m}$	$I_{n,m}^+$
3	δ_1	-0.25	0.79 (90%)	0.62 (71%)	0.88	0.89 (100%)	0.89 (100%)
3	δ_1	0.00	0.64 (82%)	0.45 (57%)	0.78	0.68 (87%)	0.68 (87%)
3	δ_1	0.25	0.53 (78%)	0.38 (56%)	0.68	0.54 (79%)	0.54 (79%)
3	δ_1	0.50	0.51 (73%)	0.41 (59%)	0.70	0.47 (67%)	0.47 (67%)
3	δ_1	0.90	0.62 (64%)	0.85 (99%)	0.97	0.40 (41%)	0.41 (42%)
5	δ_2	-0.25	0.93 (95%)	0.74 (76%)	0.98	0.97 (99%)	0.97 (99%)
5	δ_2	0.00	0.80 (87%)	0.56 (60%)	0.92	0.86 (93%)	0.86 (93%)
5	δ_2	0.25	0.59 (73%)	0.39 (47%)	0.81	0.66 (81%)	0.66 (81%)
5	δ_2	0.50	0.56 (67%)	0.42 (50%)	0.84	0.48 (57%)	0.48 (57%)
5	δ_2	0.90	0.63 (64%)	0.88 (89%)	0.99	0.40 (40%)	0.40 (40%)
3	δ_3	-0.25	0.74 (89%)	0.54 (64%)	0.83	0.83 (100%)	0.83 (100%)
3	δ_3	0.00	0.56 (87%)	0.34 (53%)	0.64	0.59 (92%)	0.59 (92%)
3	δ_3	0.25	0.42 (87%)	0.23 (48%)	0.49	0.46 (93%)	0.46 (93%)
3	δ_3	0.50	0.33 (86%)	0.15 (40%)	0.38	0.37 (97%)	0.37 (97%)
3	δ_3	0.90	0.27 (83%)	0.12 (38%)	0.32	0.27 (83%)	0.27 (83%)
5	δ_4	-0.25	0.92 (95%)	0.65 (68%)	0.96	0.95 (99%)	0.95 (99%)
5	δ_4	0.00	0.75 (90%)	0.43 (51%)	0.83	0.82 (99%)	0.82 (99%)
5	δ_4	0.25	0.49 (87%)	0.20 (35%)	0.57	0.60 (100%)	0.60 (100%)
5	δ_4	0.50	0.41 (90%)	0.16 (34%)	0.45	0.43 (100%)	0.43 (100%)
5	δ_4	0.90	0.29 (92%)	0.10 (32%)	0.31	0.33 (100%)	0.33 (100%)

The header row **Power and RE %** ($n = m = 15$) spans the table.

$\{\mathbf{s} \in \mathcal{S}_+^{p-1} : \hat{\mathbf{s}}_{\max}^T \mathbf{s} \le 0.93\}$. Hence the confidence set includes any $\mathbf{s}$ which is within $21.5°$ degrees of $\hat{\mathbf{s}}_{\max}$. This is a relatively large set due to the small sample sizes. Clearly our methodology appears to be sensitive to detect statistical differences which were not noted by NTP. Furthermore, our methodology allows us to infer that indeed 1785 mg/kg dose group is larger in the multivariate stochastic order than the control group. This is a much stronger conclusion than the simple ordering of their means. Thus we believe that the proposed framework and methodology for studying ordered distributions can serve as a useful tool in toxicology and is also applicable to a wide range of other problems as alluded to in this paper.

6. Concluding remarks and some open problems. In many applications, researchers are interested in comparing two experimental conditions, for example, a treatment and a control group, in terms of a multivariate response. In classical multivariate analysis one addresses such problems by comparing the mean vectors

	Set up		Directional tests			Integral tests	
p	δ	ρ	$S_{n,m}$	RMD test	TMD test	$I_{n,m}$	$I_{n,m}^{+}$

Power and RE % ($n = m = 25$)

p	δ	ρ	$S_{n,m}$	RMD test	TMD test	$I_{n,m}$	$I_{n,m}^{+}$
3	δ_1	−0.25	0.96 (98%)	0.90 (91%)	0.98	0.98 (100%)	0.98 (100%)
3	δ_1	0.00	0.85 (92%)	0.72 (78%)	0.92	0.85 (92%)	0.86 (92%)
3	δ_1	0.25	0.80 (88%)	0.69 (76%)	0.90	0.75 (83%)	0.75 (83%)
3	δ_1	0.50	0.75 (84%)	0.67 (75%)	0.89	0.66 (74%)	0.66 (74%)
3	δ_1	0.90	0.89 (89%)	0.98 (99%)	1.00	0.59 (59%)	0.61 (61%)
5	δ_2	−0.25	1.00 (100%)	0.98 (98%)	1.00	1.00 (100%)	1.00 (100%)
5	δ_2	0.00	0.96 (97%)	0.85 (86%)	0.99	0.98 (99%)	0.98 (99%)
5	δ_2	0.25	0.85 (88%)	0.74 (76%)	0.97	0.83 (86%)	0.83 (86%)
5	δ_2	0.50	0.81 (84%)	0.74 (77%)	0.96	0.70 (73%)	0.70 (73%)
5	δ_2	0.90	0.90 (90%)	0.99 (100%)	1.00	0.57 (57%)	0.58 (58%)
3	δ_3	−0.25	0.94 (96%)	0.85 (87%)	0.98	0.96 (98%)	0.96 (98%)
3	δ_3	0.00	0.75 (92%)	0.57 (69%)	0.82	0.79 (96%)	0.79 (96%)
3	δ_3	0.25	0.62 (89%)	0.39 (56%)	0.70	0.66 (94%)	0.66 (94%)
3	δ_3	0.50	0.54 (90%)	0.31 (52%)	0.60	0.55 (92%)	0.55 (92%)
3	δ_3	0.90	0.44 (90%)	0.20 (42%)	0.49	0.42 (86%)	0.42 (86%)
5	δ_4	−0.25	0.99 (99%)	0.94 (94%)	1.00	1.00 (100%)	1.00 (100%)
5	δ_4	0.00	0.94 (96%)	0.72 (74%)	0.97	0.97 (100%)	0.97 (100%)
5	δ_4	0.25	0.71 (90%)	0.41 (52%)	0.79	0.79 (100%)	0.79 (100%)
5	δ_4	0.50	0.58 (91%)	0.25 (39%)	0.63	0.64 (100%)	0.64 (100%)
5	δ_4	0.90	0.42 (87%)	0.18 (36%)	0.49	0.46 (94%)	0.46 (94%)

using Hotelling's T^2 statistic. The assumption of MVN, underlying Hotelling's T^2 test, may not hold in practice. Moreover if the data is not MVN, then the comparison of population means may not always provide complete information regarding the differences between the two experimental groups. Secondly, Hotelling's T^2 statistics are designed for two-sided alternatives and may not be ideal if a researcher is interested in one-sided, that is, ordered alternatives. Addressing such problems requires one to compare the two experimental groups nonparametrically in terms of the multivariate stochastic order. Such comparisons, however, are very high dimensional and not easy to perform.

In this article we circumvent this challenge by considering the notion of the linear stochastic order between two random vectors. The linear stochastic order is a "weak" generalization of the univariate stochastic order. The linear stochastic order is simple to interpret and has an intuitive appeal. Using this notion of ordering, we developed nonparametric directional inference procedures. Intuitively, the proposed methodology seeks to determine the direction that best separates two multivariate populations. Asymptotic properties of the estimated direction are derived.

Our test based on the best separating direction may be viewed as a generalization of Roy's classical largest root test for comparing several MVN populations. To the best of our knowledge this is the first general test for multivariate ordered distributions. Since in practice sample sizes are small, we use the bootstrap methodology for drawing inferences.

We illustrated the proposed methodology using a data obtained from a recent toxicity/carcinogenicity study conducted by the US National Toxicology Program (NTP) on the chemical Citral. A re-analysis of their 90-day data using our proposed methodology revealed a linear stochastic increase in platelets, urea nitrogen, alkaline phosphatase and bile acids in the high-dose group relative to the control group, which was not seen in the original univariate analysis conducted by the NTP. These findings suggest that the proposed methodology may have greater sensitivity than the commonly used univariate statistical procedures. Our methodology is sufficiently general since it is nonparametric and can be applied to discrete and/or continuous outcome variables. Furthermore, our methodology exploits the underlying dependence structure in the data, rather than analyzing one variable at a time.

We note that our example and some of our results pertain to continuous RVs. However, the methodology may be used, with appropriate modification (e.g., methods for dealing with ties) with discrete (or mixed) data with no problem. Although the focus of this paper has been the comparison of two multivariate vectors, in many applications, especially in dose response studies, researchers may be interested in determining trends (order) among several groups. Similarly to classical parametric order restricted inference literature, one could generalize the methodology developed in this paper to test for order restrictions among multiple populations. For example, one could extend the results to $K \geq 2$ RVs ordered by the simple ordering, that is, $\mathbf{X}_1 \prec_{l\text{-st}} \mathbf{X}_2 \prec_{l\text{-st}} \cdots \prec_{l\text{-st}} \mathbf{X}_K$ or to RVs ordered by the tree ordering, that is, $\mathbf{X}_1 \prec_{l\text{-st}} \mathbf{X}_j$ where $j = 2, \ldots, K$. As pointed out by a referee the hypotheses $H_0 : \mathbf{X} \preceq_{\text{st}} \mathbf{Y}$ versus $H_1 : \mathbf{X} \nprec_{\text{st}} \mathbf{Y}$ can also be formulated and tested using the approach described. First note that the null hypothesis implies $\Psi(\mathbf{s}) \geq 1/2$ for all $\mathbf{s} \in \mathcal{S}_+^{p-1}$. On the other hand under the alternative there is an $\mathbf{s} \in \mathcal{S}_+^{p-1}$ for which $\Psi(\mathbf{s}) < 1/2$. Thus a test may be based on the statistic

$$N^{1/2}\big(\Psi_{n,m}(\hat{\mathbf{s}}_{\min}) - 1/2\big),$$

where $\hat{\mathbf{s}}_{\min}$ is the value which minimizes $\Psi_{n,m}(\mathbf{s})$. It is also clear that the least favorable configuration occurs when $\Psi(\mathbf{s}) = 1/2$ for all $\mathbf{s} \in \mathcal{S}_+^{p-1}$ which is equivalent to $\mathbf{X} =_{\text{st}} \mathbf{Y}$.

We believe that the result obtained here may be useful beyond order restricted inference. Our simulation study suggests that our estimator of the best separating direction, that is, (3.5) may be useful even in the context of classical multivariate analysis where it may be viewed as a robust alternative to Roy's classical estimate. Finally we note that the linear stochastic order may be useful in a variety of other

statistical problems. For example, we believe that it provides a useful framework for linearly combining the results of several diagnostic markers. This is a well-known problem in the context of ROC curve analysis in diagnostic medicine.

APPENDIX: PROOFS

PROOF OF THEOREM 2.1. (i) Let $g : \mathbb{R}^p \to \mathbb{R}^n$ be an affine increasing function. Clearly $g(\mathbf{x}) = \mathbf{v} + \mathbf{M}\mathbf{x}$ for some n vector $\mathbf{v}$ and $n \times p$ matrix $\mathbf{M}$ with non-negative elements. Thus for any $\mathbf{u} \in \mathbb{R}_+^n$ we have $\mathbf{s} = \mathbf{M}^T \mathbf{u} \in \mathbb{R}_+^p$. Hence

$$\mathbf{u}^T g(\mathbf{X}) = \mathbf{u}^T (\mathbf{v} + \mathbf{M}\mathbf{X}) = \mathbf{u}^T \mathbf{v} + \mathbf{s}^T \mathbf{X} \preceq_{\mathrm{st}} \mathbf{u}^T \mathbf{v} + \mathbf{s}^T \mathbf{Y} = \mathbf{u}^T (\mathbf{v} + \mathbf{M}\mathbf{Y}) = \mathbf{u}^T g(\mathbf{Y})$$

as required where the inequality holds because $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$. (ii) Fix $I \in \{1, \ldots, p\}$. Let $\mathbf{X} = (\mathbf{X}_I, \mathbf{X}_{\bar{I}})$, $\mathbf{Y} = (\mathbf{Y}_I, \mathbf{Y}_{\bar{I}})$ where $\bar{I}$ is the complement of I in $\{1, \ldots, p\}$. Further define $\mathbf{s}^T = (\mathbf{s}_I^T, \mathbf{s}_{\bar{I}}^T)$ where $\mathbf{s} \in \mathbb{R}_+^p$, and set $\mathbf{s}_{\bar{I}}^T = 0$. It follows that for all $\mathbf{s}_I \in \mathbb{R}^{\dim(I)}$ we have

$$\mathbf{s}_I^T \mathbf{X}_I = \mathbf{s}^T \mathbf{X} \preceq_{\mathrm{st}} \mathbf{s}^T \mathbf{Y} = \mathbf{s}_I^T \mathbf{Y}_I$$

as required. (iii) Let $\phi : \mathbb{R} \to \mathbb{R}$ be any increasing function. Note that

$$\mathbb{E}(\phi(\mathbf{s}^T \mathbf{X})) = \mathbb{E}(\mathbb{E}(\phi(\mathbf{s}^T \mathbf{X}) | \mathbf{Z})) \le \mathbb{E}(\mathbb{E}(\phi(\mathbf{s}^T \mathbf{Y}) | \mathbf{Z})) = \mathbb{E}(\phi(\mathbf{s}^T \mathbf{Y})).$$

The inequality is a consequence of $\mathbf{X} | \mathbf{Z} = \mathbf{z} \preceq_{l\text{-st}} \mathbf{Y} | \mathbf{Z} = \mathbf{z}$. Since ϕ is arbitrary it follows that $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$ as required. (iv) Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, and define $\mathbf{Y}$ similarly. Let $\mathbf{s} \in \mathbb{R}_+^p$ where $p = p_1 + \cdots + p_n$. Now

$$\mathbf{s}^T \mathbf{X} = \mathbf{s}_1^T \mathbf{X}_1 + \cdots + \mathbf{s}_n^T \mathbf{X}_n \quad \text{and} \quad \mathbf{s}^T \mathbf{Y} = \mathbf{s}_1^T \mathbf{Y}_1 + \cdots + \mathbf{s}_n^T \mathbf{Y}_n$$

by assumption $\mathbf{s}_i^T \mathbf{X}_i \preceq_{\mathrm{st}} \mathbf{s}_i^T \mathbf{Y}_i$ for $i = 1, \ldots, n$. In addition $\mathbf{s}_i^T \mathbf{X}_i$ and $\mathbf{s}_j^T \mathbf{X}_j$ are independent for $i \ne j$. It follows from Theorem 1.A.3 in Shaked and Shanthikumar (2007) that $\mathbf{s}_1^T \mathbf{X}_1 + \cdots + \mathbf{s}_n^T \mathbf{X}_n \preceq_{\mathrm{st}} \mathbf{s}_1^T \mathbf{Y}_1 + \cdots + \mathbf{s}_n^T \mathbf{Y}_n$, that is, $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$ as required. (v) By assumption $\mathbf{X}_n \Rightarrow \mathbf{X}$ and $\mathbf{Y}_n \Rightarrow \mathbf{Y}$ where the symbol $\Rightarrow$ denotes convergence in distribution. By the continuous mapping theorem $\mathbf{s}^T \mathbf{X}_n \Rightarrow \mathbf{s}^T \mathbf{X}$ and $\mathbf{s}^T \mathbf{Y}_n \Rightarrow \mathbf{s}^T \mathbf{Y}$. It follows that

$$\text{(A.1)} \quad \mathbb{P}(\mathbf{s}^T \mathbf{X}_n \ge t) \to \mathbb{P}(\mathbf{s}^T \mathbf{X} \ge t) \quad \text{and} \quad \mathbb{P}(\mathbf{s}^T \mathbf{Y}_n \ge t) \to \mathbb{P}(\mathbf{s}^T \mathbf{Y} \ge t).$$

Moreover since $\mathbf{X}_n \preceq_{l\text{-st}} \mathbf{Y}_n$ we have

$$\text{(A.2)} \qquad \mathbb{P}(\mathbf{s}^T \mathbf{X}_n \ge t) \le \mathbb{P}(\mathbf{s}^T \mathbf{Y}_n \ge t) \qquad \text{for all } n \in \mathbb{N}.$$

Combining (A.1) and (A.2) we have $\mathbb{P}(\mathbf{s}^T \mathbf{X} \ge t) \le \mathbb{P}(\mathbf{s}^T \mathbf{Y} \ge t)$, that is, $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$ as required. $\square$

Before proving Theorem 2.2, we provide a definition and a preliminary lemma.

DEFINITION A.1. We say that the RV $\mathbf{X}$ has an elliptical distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and generator $\phi(\cdot)$, denoted $\mathbf{X} \sim E_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, if its characteristic function is given by $\exp(i\mathbf{t}^T \boldsymbol{\mu})\phi(\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t})$.

For this and other facts about elliptical distributions which we use in the proofs below, see Fang, Kots and Ng (1989).

LEMMA A.1. *Let $X \sim E_1(\mu, \sigma, \phi)$ and $Y \sim E_1(\mu', \sigma', \phi)$ be univariate elliptical RVs supported on $\mathbb{R}$. Then $X \preceq_{\mathrm{st}} Y$ if and only if $\mu \leq \mu'$ and $\sigma = \sigma'$.*

PROOF. Since X and Y have the same generator they have the stochastic representation:

$$(A.3) \qquad X =_{\mathrm{st}} \mu + \sigma R U \quad \text{and} \quad Y =_{\mathrm{st}} \mu' + \sigma' R U,$$

where R is a nonnegative RV, independent of the RV U, satisfying $\mathbb{P}(U = \pm 1) = 1/2$; cf. Fang, Kots and Ng (1989). It follows that RU is a symmetric RV supported on $\mathbb{R}$ with a strictly increasing DF which we denoted by F_0. Let F_X and F_Y denote the DFs of X and Y, respectively. Note that $X \preceq_{\mathrm{st}} Y$ if and only if $F_X(t) \geq F_Y(t)$ for all $t \in \mathbb{R}$, or equivalently by (A.3), if and only if

$$(A.4) \qquad F_0\left(\frac{t - \mu}{\sigma}\right) \geq F_0\left(\frac{t - \mu'}{\sigma'}\right)$$

for all $t \in \mathbb{R}$. It is obvious that (A.4) holds when $\mu \leq \mu'$ and $\sigma = \sigma'$, establishing sufficiency. Now assume that $X \preceq_{\mathrm{st}} Y$. Put $t = \mu$ in (A.4), and use the strict monotonicity of F_0 to get $0 \geq (\mu - \mu')/\sigma'$, that is, $\mu' \geq \mu$. Suppose now that $\sigma' > \sigma$. It follows from (A.4) and the the strict monotonicity of F_0 that $(t - \mu)/\sigma \geq (t - \mu')/\sigma'$ which is equivalent to $t \geq (\mu\sigma' - \mu'\sigma)/(\sigma' - \sigma)$. The latter, however, contradicts the fact that (A.4) holds for all $t \in \mathbb{R}$. A similar argument shows that $\sigma' < \sigma$ cannot hold; hence we must have $\sigma = \sigma'$ as required. $\square$

REMARK A.1. Note that Lemma A.1 may not hold for distributions with a finite support. For example, if $R \sim U(0, 1)$, then by (A.3) $X \sim U(\mu - \sigma, \mu + \sigma)$ and $Y \sim U(\mu' - \sigma', \mu' + \sigma')$. It is easily verified that in this case $X \preceq_{\mathrm{st}} Y$ if and only if $\Delta = \mu' - \mu \geq 0$ and $-\Delta \leq \sigma' - \sigma \leq \Delta$; that is, it is not required that $\sigma = \sigma'$. Hence the assumption that X and Y are supported on $\mathbb{R}$ is necessary.

We continue with the proof of Theorem 2.2.

PROOF OF THEOREM 2.2. Let $\mathbf{X}$ and $\mathbf{Y}$ be be $E_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ and $E_p(\boldsymbol{\mu}', \boldsymbol{\Sigma}', \phi)$ supported on $\mathbb{R}^p$. Suppose that $\mathbf{X} \preceq_{l\text{-}\mathrm{st}} \mathbf{Y}$. Choose $\mathbf{s} = \mathbf{e}_i$ where $\mathbf{e}_{ik} = 1$ if $i = k$ and 0 otherwise. It now follows from Definition 1.1 that $X_i \preceq_{\mathrm{st}} Y_i$. Since X_i and Y_i

are marginally elliptically distributed RVs with the same generator and supported on $\mathbb{R}$, then by Lemma A.1 we must have

(A.5) $\mu_i \leq \mu_i'$ and $\sigma_{ii} = \sigma_{ii}'$.

The latter holds, of course, for all $1 \leq i \leq p$. Choosing $\mathbf{s} = \mathbf{e}_i + \mathbf{e}_j$ we have $X_i + X_j \preceq_{\text{st}} Y_i + Y_j$. Note that $X_i + X_j$ and $Y_i + Y_j$ are supported on $\mathbb{R}$ and follow a univariate elliptical distribution with the same generator [Fang, Kots and Ng (1989)]. Applying Lemma A.1 again we find that

(A.6) $\mu_i + \mu_j \leq \mu_i' + \mu_j'$ and $\sigma_{ii} + \sigma_{jj} + 2\sigma_{ij} = \sigma_{ii}' + \sigma_{jj}' + 2\sigma_{ij}'$.

The latter holds, of course, for all $1 \leq i \neq j \leq p$. It is easy to see that equations (A.5) and (A.6) imply that $\boldsymbol{\mu} \leq \boldsymbol{\mu}'$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}'$. Recall [cf. Fang, Kots and Ng (1989)] that we may write $\mathbf{X} =_{\text{st}} \boldsymbol{\mu} + R\mathbf{S}\mathbf{U}$ and $\mathbf{Y} =_{\text{st}} \boldsymbol{\mu}' + R\mathbf{S}\mathbf{U}$ where $\boldsymbol{\Sigma} = S^T S$, $\mathbf{U}$ is a uniform RV on $\mathcal{S}_+^{p-1}$, and R is a nonnegative RV. Let S be an upper set in $\mathbb{R}^p$. Clearly the set $[S - \boldsymbol{\mu}] := \{\mathbf{x} - \boldsymbol{\mu} : x \in S\}$ is also an upper set and $[S - \boldsymbol{\mu}] \subseteq [S - \boldsymbol{\mu}']$ since $\boldsymbol{\mu} \leq \boldsymbol{\mu}'$. Now,

$$\mathbb{P}(\mathbf{X} \in S) = \mathbb{P}(\mathbf{X}_0 \in [S - \boldsymbol{\mu}]) \leq \mathbb{P}(\mathbf{X}_0 \in [S - \boldsymbol{\mu}']) = \mathbb{P}(\mathbf{Y} \in S),$$

where $\mathbf{X}_0 = R\mathbf{S}\mathbf{U}$, hence $\mathbf{X} \preceq_{\text{st}} \mathbf{Y}$. This proves the "if" part. The "only if" part follows immediately. □

PROOF OF THEOREM 2.3.   Let $\mathcal{X}_p = \{\mathbf{x} : (x_1, \ldots, x_p) \in \{0, 1\}^p\}$ denote the support of a p-dimensional multivariate binary (MVB) RV. By definition the relationship $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$ implies that for all $(t, \mathbf{s}) \in \mathbb{R}_+ \times \mathbb{R}_+^p$,

(A.7) $\mathbb{P}(\mathbf{s}^T\mathbf{X} > t) \leq \mathbb{P}(\mathbf{s}^T\mathbf{Y} > t)$.

Now note that

(A.8) $\mathbb{P}(\mathbf{s}^T\mathbf{X} > t) = \sum_{\mathbf{x} \in \mathcal{X}_p} f(\mathbf{x})\mathbb{I}_{(\mathbf{s}^T\mathbf{x} > t)}$ and $\mathbb{P}(\mathbf{s}^T\mathbf{Y} > t) = \sum_{\mathbf{x} \in \mathcal{X}_p} g(\mathbf{x})\mathbb{I}_{(\mathbf{s}^T\mathbf{x} > t)}$,

where f and g are the probability mass functions of $\mathbf{X}$ and $\mathbf{Y}$, respectively. Let U be an upper set on $\mathcal{X}_p$. It is well known [cf. Davey and Priestley (2002)] that U can be written as

(A.9) $U = \bigcup_{j \in J} U(\mathbf{x}_j),$

where $\mathbf{x}_j$ are the distinct minimal elements of U, and $U(\mathbf{x}_j) = \{\mathbf{x} : \mathbf{x} \geq \mathbf{x}_j\}$ are themselves upper sets [in fact $U(\mathbf{x}_j)$ is an upper orthant]. The set $\{\mathbf{x}_j : j \in J\}$ is often referred to as an anti-chain. Now observe that for any $\mathbf{s} \in \mathbb{R}_+^p$ the set $\{\mathbf{x} : \mathbf{s}^T\mathbf{x} > t\}$ is an upper set. Hence it must be of the form of (A.9) for some anti-chain $\{\mathbf{x}_j : j \in J\}$. Suppose now, that for some $U \in \mathcal{X}_p$ there is a vector $\mathbf{s}_U \in \mathbb{R}_+^p$

such that $U = \{\mathbf{x} : \mathbf{s}_U^T \mathbf{x} > t\}$ for some fixed $t > 0$. Then using (A.7) and (A.8) we have

$$\mathbb{P}(\mathbf{X} \in U) = \sum_{\mathbf{x} \in \{\mathbf{x} : \mathbf{s}_U^T \mathbf{x} > t\}} f(\mathbf{x}) = \mathbb{P}(\mathbf{s}_U^T \mathbf{X} > t) \leq \mathbb{P}(\mathbf{s}_U^T \mathbf{Y} > t)$$

$$= \sum_{\mathbf{x} \in \{\mathbf{x} : \mathbf{s}_U^T \mathbf{x} > t\}} g(\mathbf{x}) = \mathbb{P}(\mathbf{Y} \in U).$$

We will complete the proof by showing that for each upper set $U \in \mathcal{X}_p$, we can find a vector $\mathbf{s}_U$ for which $\mathbf{s}_U^T \mathbf{x} > t$ for $\mathbf{x} \in U$ and $\mathbf{s}_U^T \mathbf{x} \leq t$ for $\mathbf{x} \in U^c = \mathcal{X} \setminus U$ if and only if $p \leq 3$. To do so we will first solve the system of equations $\mathbf{s}^T \mathbf{x}_j = t$ for $j \in J$. This system can also be written as $\mathbf{X}\mathbf{s} = \mathbf{t}$ where

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_J \end{pmatrix}$$

is a $J \times p$ matrix whose rows are the member of the anti-chain defining U, and $\mathbf{t} = (t, \ldots, t)$ has dimension J. Clearly the elements of $\mathbf{X}$ are ones and zeros. If $J \leq p$, the matrix $\mathbf{X}$ is of full rank since its rows are linearly independent by the fact that they are an anti-chain. Hence a solution for $\mathbf{s}$ exists. With a bit of algebra, we can further show that a solution $\mathbf{s} \geq 0$ exists. This, of course, is trivially verified when $p \leq 3$. Now set $\mathbf{s}_U = \mathbf{s} + \boldsymbol{\varepsilon}$ for some $\boldsymbol{\varepsilon} \geq 0$. It is clear that we can choose $\boldsymbol{\varepsilon}$ small enough to guarantee that $\mathbf{s}_U^T \mathbf{x} > t$ if and only if $\mathbf{x} \in U$. Hence if $J \leq p$, upper set (A.9) can be mapped to a vector $\mathbf{s}_U$. However, the inequality $J \leq p$ for *all* upper sets $U \subset \mathcal{X}_p$ holds if and only if $p \leq 3$. This can be easily shown by enumerating all 18 upper sets belonging to $\mathcal{X}_3$ [cf. Davidov and Peddada (2011)] and noting that they have at most three minimal elements. Hence if $p \leq 3$, then $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y} \iff \mathbf{X} \preceq_{\text{st}} \mathbf{Y}$ as required.

Now let $p = 4$, and consider the upper set U generated by the anti-chain $\mathbf{x}_j$, $j = 1, \ldots, J$ where $\mathbf{x}_j$ are all the distinct permutations of the vector $(1, 1, 0, 0)$. Clearly $J = 6$. Note that although $J > p$, the system of equations $\mathbf{X}\mathbf{s} = \mathbf{t}$ is uniquely solved by $\mathbf{s}_*^T = (t/2, t/2, t/2, t/2)$. However, this solution coincides with the solution of the system $\mathbf{X}'\mathbf{s} = \mathbf{t}$ where $\mathbf{X}'$ is any matrix obtained from $\mathbf{X}$ by deleting any two (or just one) of its rows. Note that the rows of $\mathbf{X}'$ correspond to an upper set $U' \subset U$. This, in turn, implies that for any such U' one cannot find a vector $\mathbf{s}_{U'}$ satisfying $\mathbf{s}_{U'}^T \mathbf{x} > t$ if and only if $\mathbf{x} \in U'$ because the inequality will hold for all $\mathbf{x} \in U$. Thus U' does not define an upper half plane. This shows that the linear stochastic order and the multivariate stochastic order do not coincide when $p = 4$. A similar argument may be used for any $p \geq 5$. This completes the proof. $\quad\square$

We first define the term copula.

DEFINITION A.2. Let F be the DF of a p-dimensional RV with marginal DFs $F_1, \ldots, F_p$. The copula C associated with F is a DF such that

$$F(\mathbf{x}) = C(\mathbf{x}) = C(F_1(x_1), \ldots, F_p(x_p)).$$

It follows that the tail-copula $\bar{C}(\cdot)$ is nothing but the tail of the DF $C(\cdot)$.

PROOF OF THEOREM 2.4. Suppose that $\mathbf{X}$ and $\mathbf{Y}$ have the same copula. Let $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y}$. Choosing $\mathbf{s} = \mathbf{e}_i$ where $\mathbf{e}_{ik} = 1$ if $i = k$ and 0 otherwise, we find using the definition that $X_i \preceq_{st} Y_i$. The latter holds, of course, for all $1 \leq i \leq p$. Applying Theorem 6.B.14 in Shaked and Shanthikumar (2007), we find that $\mathbf{X} \preceq_{st} \mathbf{Y}$. The reverse direction is immediate. □

PROOF OF THEOREM 2.5. Note that for any $\mathbf{x} \in \mathbb{R}^p$ we have

$$F(\mathbf{x}) = C_{\mathbf{X}}(F_1(x_1), \ldots, F_p(x_p)) \geq C_{\mathbf{X}}(G_1(x_1), \ldots, G_p(x_p))$$
$$\geq C_{\mathbf{Y}}(G_1(x_1), \ldots, G_p(x_p)) = G(\mathbf{x}).$$

This means that $\mathbf{X} \preceq_{lo} \mathbf{Y}$. The other part of the theorem is proved similarly. □

PROOF OF PROPOSITION 3.1. Let $\mathbf{X}$ and $\mathbf{Y}$ be independent MVNs with means $\boldsymbol{\mu} \leq \boldsymbol{\nu}$ and common variance matrix $\boldsymbol{\Sigma}$. Clearly

$$\mathbb{P}(\mathbf{s}^T \mathbf{X} \leq \mathbf{s}^T \mathbf{Y}) = \Phi\left(-\frac{\mathbf{s}^T(\boldsymbol{\mu} - \boldsymbol{\nu})}{\sqrt{2\mathbf{s}^T \boldsymbol{\Sigma} \mathbf{s}}}\right),$$

where Φ is the DF of a standard normal RV. It follows that $\mathbb{P}(\mathbf{s}^T \mathbf{X} \leq \mathbf{s}^T \mathbf{Y})$ is maximized when the ratio $\mathbf{s}^T(\boldsymbol{\nu} - \boldsymbol{\mu})/\sqrt{\mathbf{s}^T \boldsymbol{\Sigma} \mathbf{s}}$ is maximized. From the Cauchy–Schwarz inequality we have

$$(\text{A.10}) \qquad \frac{\mathbf{s}^T(\boldsymbol{\nu} - \boldsymbol{\mu})}{\sqrt{\mathbf{s}^T \boldsymbol{\Sigma} \mathbf{s}}} \leq \sqrt{(\boldsymbol{\nu} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\nu} - \boldsymbol{\mu})}$$

for all $\mathbf{s}$. It is now easily verified that $\mathbf{s} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\nu} - \boldsymbol{\mu})$ maximizes the left-hand side of (A.10). □

PROOF OF PROPOSITION 3.2. Let Q_q, $q = 1, \ldots, 4$, be the four quadrants. It is clear that maximizing (3.1) is equivalent to maximizing

$$(\text{A.11}) \qquad \Psi'_{n,m}(\mathbf{s}) = \sum_{\mathbf{Z}_{ij} \in Q_2} \mathbb{I}_{(\mathbf{s}^T \mathbf{Z}_{ij} \geq 0)} + \sum_{\mathbf{Z}_{ij} \in Q_4} \mathbb{I}_{(\mathbf{s}^T \mathbf{Z}_{ij} \geq 0)}.$$

It is also clear that for any $\mathbf{s}$ the indicators $\mathbb{I}_{(\mathbf{s}^T \mathbf{Z}_{ij} \geq 0)}$ are independent of the length of $\mathbf{Z}_{ij}$ which we therefore take to have length unity. Observe that the value of (A.11) is constant in the intervals $(\theta_{[i]}, \theta_{[i+1]})$ where $\theta_{[i]}$ are defined in Algorithm 3.1. At each point $\theta_{[i]}$, $i = 0, \ldots, M + 1$, the value of (A.11) may increase or

decrease. It follows that for all $\mathbf{s} \in \mathcal{S}_+^{p-1}$ $\Psi'_{n,m}(\mathbf{s}) \in \{\Psi'_{n,m}(\mathbf{s}_{[0]}), \ldots, \Psi'_{n,m}(\mathbf{s}_{[M+1]})\}$ where $\mathbf{s}_{[i]}$ are defined in Algorithm 3.1. Therefore the maximum value of (3.1) is an element of the above list. Now suppose that $\mathbf{s}_{[i]}$ is a global maximizer of (A.11). Clearly either $\Psi'_{n,m}(\mathbf{s}_{[i]}) = \Psi'_{n,m}(\mathbf{s}_{[i-1]})$ or $\Psi'_{n,m}(\mathbf{s}_{[i]}) = \Psi'_{n,m}(\mathbf{s}_{[i+1]})$ must hold, in which case any value in $[\theta_{[i-1]}, \theta_{[i]}]$ or $[\theta_{[i]}, \theta_{[i+1]}]$ is a global maximizer. This concludes the proof. $\square$

PROOF OF THEOREM 3.1. Using Hajek's projection and for any $\mathbf{s}$, we may write

$$(A.12) \qquad \Psi_{n,m}(\mathbf{s}) = \Psi(\mathbf{s}) + n^{-1} \sum_{i=1}^{n} \psi_1(\mathbf{X}_i, \mathbf{s}) + m^{-1} \sum_{j=1}^{m} \psi_2(\mathbf{Y}_j, \mathbf{s}) + R_{n,m}(\mathbf{s}),$$

where

$$\psi_1(\mathbf{X}_i, \mathbf{s}) = \bar{G}(\mathbf{s}^T \mathbf{X}_i) - \Psi(\mathbf{s}),$$

$$\psi_2(\mathbf{Y}_j, \mathbf{s}) = F(\mathbf{s}^T \mathbf{Y}_j) - \Psi(\mathbf{s})$$

and $R_{n,m}(\mathbf{s})$ is a remainder term. Here $\bar{G}(\mathbf{s}^T \mathbf{x}) = \mathbb{P}(\mathbf{s}^T \mathbf{Y} \geq \mathbf{s}^T \mathbf{x})$, $F(\mathbf{s}^T \mathbf{y}) = \mathbb{P}(\mathbf{s}^T \mathbf{X} \leq \mathbf{s}^T \mathbf{y})$ and $\Psi(\mathbf{s}) = \mathbb{E}(\bar{G}(\mathbf{s}^T \mathbf{X}_i)) = \mathbb{E}(F(\mathbf{s}^T \mathbf{Y}_j))$. Clearly $\mathbb{E}[\psi_1(\mathbf{X}_i, \mathbf{s})] = \mathbb{E}[\psi_2(\mathbf{Y}_j, \mathbf{s})] = 0$ for all i and j, so by the strong law of large numbers $n^{-1} \sum_{i=1}^{n} \psi_1(\mathbf{X}_i, \mathbf{s})$ and $m^{-1} \sum_{j=1}^{m} \psi_2(\mathbf{Y}_j, \mathbf{s})$ both converge to zero with probability one. Now,

$$\sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} |\Psi_{n,m}(\mathbf{s}) - \Psi(\mathbf{s})| \leq \sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \left| n^{-1} \sum_{i=1}^{n} \psi_1(\mathbf{X}_i, \mathbf{s}) \right| + \sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \left| m^{-1} \sum_{j=1}^{m} \psi_2(\mathbf{Y}_j, \mathbf{s}) \right|$$

$$+ \sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} |R_{n,m}(\mathbf{s})|.$$

The set $\mathcal{S}_+^{p-1}$ is compact, and the function $\psi_1(\mathbf{x}, \mathbf{s})$ is continuous in $\mathbf{s} \in \mathcal{S}_+^{p-1}$ for all values of $\mathbf{x}$ and bounded [in fact $|\psi_1(\mathbf{x}, \mathbf{s})| \leq 2$]. Thus the conditions in Theorem 3.1 in DasGupta (2008) are satisfied, and it follows that $\sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} |n^{-1} \sum_{i=1}^{n} \psi_1(\mathbf{X}_i, \mathbf{s})| \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$. Similarly $\sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} |m^{-1} \times \sum_{i=1}^{m} \psi_2(\mathbf{Y}_i, \mathbf{s})| \overset{\text{a.s.}}{\to} 0$ as $m \to \infty$. Since $\Psi_{n,m}(\mathbf{s})$ is bounded all its moments exist; therefore from Theorem 5.3.3 in Serfling (1980) we have that with probability one $R_{n,m}(\mathbf{s}) = o(1/N)$. Moreover it is clear that the latter holds uniformly for all $\mathbf{s}$. Thus,

$$\sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} |\Psi_{n,m}(\mathbf{s}) - \Psi(\mathbf{s})| \overset{\text{a.s.}}{\to} 0 \qquad \text{as } n, m \to \infty.$$

By assumption $\Psi(\mathbf{s}_{\max}) > \Psi(\mathbf{s})$ for all $\mathbf{s} \in \mathcal{S}_+^{p-1} \setminus \mathbf{s}_{\max}$ so we can apply Theorem 2.12 in Kosorok (2008) to conclude that

$$\hat{\mathbf{s}}_{\max} \overset{\text{a.s.}}{\to} \mathbf{s}_{\max};$$

that is, $\hat{\mathbf{s}}_{\max}$ is strongly consistent. This completes the first part of the proof.

Since the densities of $\mathbf{X}$ and $\mathbf{Y}$ are differentiable, it follows that $\Psi(\mathbf{s})$ is continuous and twice differentiable. In particular at $\mathbf{s}_{\max} \in \mathcal{S}_+^{p-1}$, the matrix $-\nabla^2 \Psi(\mathbf{s}_{\max})$ exists and is positive definite. A Taylor expansion implies that

$$\sup_{\|\mathbf{s}-\mathbf{s}_{\max}\| < \delta} \Psi(\mathbf{s}) - \Psi(\mathbf{s}_{\max}) \le -C\delta^2.$$

It is also obvious that

$$n^{-1} \sum_{i=1}^n \psi_1(\mathbf{X}_i, \mathbf{s}_{\max}) = O_p(1/\sqrt{N})$$

and

$$m^{-1} \sum_{j=1}^m \psi_2(\mathbf{Y}_j, \mathbf{s}_{\max}) = O_p(1/\sqrt{N}).$$

Finally as noted above $|R_{n,m}(\mathbf{s})| = O(1/N)$ for all $\mathbf{s}$ as $n, m \to \infty$. Therefore by Theorem 1 in Sherman (1993) we have that

(A.13) $$\hat{\mathbf{s}}_{\max} = \mathbf{s}_{\max} + O_p(N^{-1/2});$$

that is, $\hat{\mathbf{s}}_{\max}$ converges to $\mathbf{s}_{\max}$ at a $N^{1/2}$ rate. This completes the second part of the proof.

The functions $\Psi(\mathbf{s})$, $\psi_1(\mathbf{X}_i, \mathbf{s})$ and $\psi_2(\mathbf{Y}_j, \mathbf{s})$ on the right-hand side of (A.12) all admit a quadratic expansion. A bit of algebra shows that for $\mathbf{s}$ in an $O_p(N^{-1/2})$ neighborhood of $\mathbf{s}_{\max}$, we have

(A.14)
$$\Psi_{n,m}(\mathbf{s}) = \Psi(\mathbf{s}_{\max}) + (\mathbf{s} - \mathbf{s}_{\max})^T \frac{\mathbf{M}_{n,m}}{N^{1/2}}$$
$$+ \frac{1}{2}(\mathbf{s} - \mathbf{s}_{\max})^T \mathbf{V}(\mathbf{s} - \mathbf{s}_{\max}) + o_p(1/N),$$

where

$$\mathbf{M}_{n,m} = \frac{1}{\sqrt{\lambda_{n,m}}} \frac{\sum_{i=1}^n \nabla \psi_1(\mathbf{X}_i, \mathbf{s}_{\max})}{n^{1/2}} + \frac{1}{\sqrt{1 - \lambda_{n,m}}} \frac{\sum_{j=1}^m \nabla \psi_2(\mathbf{Y}_j, \mathbf{s}_{\max})}{m^{1/2}}$$

$\lambda_{n,m} = n/N$, for $j = 1, 2$ the function $\nabla \psi_j(\cdot, \mathbf{s}_{\max})^T$ is the gradient of $\psi_j(\cdot, \mathbf{s})$ evaluated at $\mathbf{s}_{\max}$, and the matrix $\mathbf{V}$ is given by

$$\mathbf{V} = \mathbb{E}(\nabla^2 \psi_1(\mathbf{X}, \mathbf{s}_{\max})) + \mathbb{E}(\nabla^2 \psi_2(\mathbf{Y}, \mathbf{s}_{\max})).$$

Note that the $o_p(1/N)$ term in (A.14) absorbs $R_{n,m}(\mathbf{s})$ in (A.12) as well as the higher-order terms in the quadratic expansions of $\Psi(\mathbf{s})$, $n^{-1} \sum_{i=1}^n \psi_1(\mathbf{X}_i, \mathbf{s})$ and $m^{-1} \sum_{j=1}^m \psi_2(\mathbf{Y}_j, \mathbf{s})$. Now by the CLT and Slutzky's theorem, we have that

$$\mathbf{M}_{n,m} \Rightarrow N(0, \boldsymbol{\Delta}),$$

where

$$\mathbf{\Delta} = \frac{1}{\lambda} \mathbb{E}\big(\nabla \psi_1(\mathbf{X}, \mathbf{s}_{\max}) \nabla \psi_1(\mathbf{X}, \mathbf{s}_{\max})^T\big)$$

$$+ \frac{1}{1 - \lambda} \mathbb{E}\big(\nabla \psi_2(\mathbf{Y}, \mathbf{s}_{\max}) \nabla \psi_2(\mathbf{Y}, \mathbf{s}_{\max})^T\big).$$

Finally it follows by Theorem 2 in Sherman (1993) that

$$N^{1/2}(\hat{\mathbf{s}}_{\max} - \mathbf{s}_{\max}) \Rightarrow N(0, \mathbf{\Sigma}),$$

where $\mathbf{\Sigma} = \mathbf{V}^{-1} \mathbf{\Delta} V^{-1}$, completing the proof. $\square$

PROOF OF THEOREM 3.2. Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are discrete RVs with finite support. Let $p_a = \mathbb{P}(\mathbf{X} = \mathbf{x}_a) > 0$ and $q_b = \mathbb{P}(\mathbf{Y} = \mathbf{y}_b) > 0$ where $a = 1, \dots, A$ and $b = 1, \dots, B$; A and B are finite. Define the set $\mathbf{S}_{ab} = \{\mathbf{s} \in \mathcal{S}_+^{p-1} : \mathbf{s}^T \mathbf{x}_a \leq \mathbf{s}^T \mathbf{y}_b\}$. A simple argument shows that

$$\Psi(\mathbf{s}) = \mathbb{P}\big(\mathbf{s}^T \mathbf{X} \leq \mathbf{s}^T \mathbf{Y}\big) = \sum_{a=1}^A \sum_{b=1}^B \mathbb{I}_{(\mathbf{s} \in \mathbf{S}_{ab})} p_a q_b = \sum_{k=1}^K \alpha_k \mathbb{I}_{(\mathbf{s} \in \mathbf{S}_k)},$$

where $K \leq 2^{AB}$ is finite, the sets $\mathbf{S}_k$ are distinct, $\bigcup_{k=1}^K \mathbf{S}_k = \mathcal{S}_+^{p-1}$ and $\alpha_k = \sum_{(a,b) \in J_k} p_a q_b$ with $J_k = \{(a,b) : \mathbf{S}_k \cap \mathbf{S}_{ab} \neq \varnothing\}$. Thus $\Psi(\mathbf{s})$ is a simple function on $\mathcal{S}_+^{p-1}$, and $\mathbf{S}_{\max}$ is the set associated with the largest α_k. We will assume, without any loss of generality, that $\alpha_1 > \alpha_k$ for all $k \geq 2$. Now note that

$$\Psi_{n,m}(\mathbf{s}) = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \mathbb{I}_{(\mathbf{s}^T \mathbf{X}_i \leq \mathbf{s}^T \mathbf{Y}_j)} = \sum_{a=1}^A \sum_{b=1}^B \frac{n_a m_b}{nm} \mathbb{I}_{(\mathbf{s} \in \mathbf{S}_{ab})},$$

where $n_a = \sum_{i=1}^n \mathbb{I}_{(\mathbf{X}_i = \mathbf{x}_a)}, m_b = \sum_{j=1}^m \mathbb{I}_{(\mathbf{Y}_i = \mathbf{y}_b)}$ where $a = 1, \dots, A$ and $b = 1, \dots, B$. Clearly $\Psi_{n,m}(\mathbf{s})$ is also a simple function. Moreover for large enough n and m we will have $n_a > 0$ and $m_b > 0$ for all $a = 1, \dots, A$ and $b = 1, \dots, B$, and consequently $\Psi_{n,m}(\mathbf{s})$ is defined over the same sets as $\Psi(\mathbf{s})$, that is,

$$\Psi_{n,m}(\mathbf{s}) = \sum_{k=1}^K \hat{\alpha}_k \mathbb{I}_{(\mathbf{s} \in \mathbf{S}_k)},$$

where $\hat{\alpha}_k = \sum_{(a,b) \in J_k} \hat{p}_a \hat{q}_b$ with $\hat{p}_a = n_a/n$ and $\hat{q}_b = m_b/m$. Furthermore the maximizer of $\Psi_{n,m}(\mathbf{s})$ is any $\mathbf{s} \in \mathbf{S}_k$ provided that $\mathbf{S}_k$ is associated with the largest $\hat{\alpha}_k$. Hence,

$$\mathbb{P}(\hat{\mathbf{s}}_{\max} \notin \mathbf{S}_{\max}) = \mathbb{P}\Big(\arg \max_{1 \leq k \leq K} \hat{\alpha}_k \neq 1\Big) = \mathbb{P}\left(\bigcup_{k=2}^K \{\hat{\alpha}_1 \leq \hat{\alpha}_k\}\right)$$

(A.15)

$$\leq \sum_{k=2}^K \mathbb{P}(\hat{\alpha}_1 \leq \hat{\alpha}_k) \leq (K - 1) \max_{2 \leq k \leq K} \mathbb{P}(\hat{\alpha}_1 \leq \hat{\alpha}_k).$$

A bit of rearranging shows that

$$\mathbb{P}(\hat{\alpha}_1 \leq \hat{\alpha}_k) = \mathbb{P}\left(\sum_{(a,b)\in J_1} \hat{p}_a \hat{q}_b - \sum_{(a,b)\in J_k} \hat{p}_a \hat{q}_b \leq 0 \right)$$

$$= \mathbb{P}\left(n^{-1} m^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} Z_{ij}^{(k)} \leq 0 \right),$$

where

$$Z_{ij}^{(k)} = \sum_{(a,b)\in J_1 \setminus J_k} \mathbb{I}_{(\mathbf{X}_i=\mathbf{x}_a)} \mathbb{I}_{(\mathbf{Y}_j=\mathbf{y}_b)} - \sum_{(a,b)\in J_k \setminus J_1} \mathbb{I}_{(\mathbf{X}_i=\mathbf{x}_a)} \mathbb{I}_{(\mathbf{Y}_j=\mathbf{y}_b)}.$$

Note that $Z_{ij}^{(k)}$ may be viewed as a kernel of a two sample U-statistic. Moreover

$$-|J_k \setminus J_1| \leq Z_{ij}^{(k)} \leq |J_k \setminus J_1|$$

is bounded (here $|\cdot|$ denotes set cardinality) and $\mathbb{E}(Z_{ij}^{(k)}) = \mu_k = \mathbb{E}(\hat{\alpha}_1 - \hat{\alpha}_k) > 0$ by assumption. Applying Theorem 2 and the derivations in Section 5b in Hoeffding (1963) we have that

$$(A.16) \quad \mathbb{P}\left(n^{-1} m^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} Z_{ij}^{(k)} \leq 0 \right) \leq \exp\left(-\frac{N \min(\lambda_{n,m}, 1-\lambda_{n,m}) \mu_k^2}{2(|J_k \setminus J_1| + |J_1 \setminus J_k|)^2} \right),$$

where $\lambda_{n,m} = n/N \to \lambda \in (0,1)$ as $n, m \to \infty$. Finally from (A.15) and (A.16) we have that

$$\mathbb{P}(\hat{\mathbf{s}}_{\max} \notin \mathbf{S}_{\max}) \leq C_1 \exp(-C_2 N),$$

where $C_1 = K - 1$ and $C_2 = \min(\lambda, 1-\lambda) \min 2\mu_k^2 (|J_k \setminus J_1| + |J_1 \setminus J_K|)^{-2}$ completing the proof. $\quad\square$

PROOF OF THEOREM 3.3.   Choose $\varepsilon > 0$. We have already seen that under the stated conditions, $\Psi(\mathbf{s})$ is continuous, and therefore for each $\mathbf{s}$ the set $B_{\mathbf{s},\varepsilon} = \{\mathbf{s}' : |\Psi(\mathbf{s}') - \Psi(\mathbf{s})| < \varepsilon\}$ is open. The collection $\{B_{\mathbf{s},\varepsilon} : \mathbf{s} \in \mathcal{S}_+^{p-1}\}$ is an open cover for $\mathcal{S}_+^{p-1}$. Since $\mathcal{S}_+^{p-1}$ is compact there exists a finite subcover $B_{\mathbf{s}_1,\varepsilon}, \ldots, B_{\mathbf{s}_K,\varepsilon}$ for $\mathcal{S}_+^{p-1}$ where $K < \infty$. Hence each $\mathbf{s}$ belongs to some $B_{\mathbf{s}_i,\varepsilon}$ and therefore

$$\sup_{\mathbf{s}\in\mathcal{S}_+^{p-1}} |\Psi_N(\mathbf{s}) - \Psi(\mathbf{s})|$$

$$\leq \max_{1\leq i\leq K} \sup_{\mathbf{s}\in B_{\mathbf{s}_i,\varepsilon}} |\Psi_N(\mathbf{s}) - \Psi_N(\mathbf{s}_i)|$$

$$+ \max_{1\leq i\leq K} |\Psi_N(\mathbf{s}_i) - \Psi(\mathbf{s}_i)| + \max_{1\leq i\leq K} \sup_{\mathbf{s}\in B_{\mathbf{s}_i,\varepsilon}} |\Psi(\mathbf{s}_i) - \Psi(\mathbf{s})|.$$

By construction $\sup_{\mathbf{s} \in B_{\mathbf{s}_i, \varepsilon}} |\Psi(\mathbf{s}_i) - \Psi(\mathbf{s})| < \varepsilon$ for all i. By the law of large numbers $\Psi_N(\mathbf{s}_i) \overset{\text{a.s.}}{\to} \Psi(\mathbf{s}_i)$ as $N \to \infty$ for each $i = 1, \ldots, K$. Since K is finite,

$$\max_{1 \le i \le K} |\Psi_N(\mathbf{s}_i) - \Psi(\mathbf{s}_i)| \le \sum_{i=1}^{K} |\Psi_N(\mathbf{s}_i) - \Psi(\mathbf{s}_i)| \overset{\text{a.s.}}{\to} 0.$$

Now for any $\mathbf{s}' \in B_{\mathbf{s}_i, \varepsilon}$, $\Psi_N(\mathbf{s}') \overset{\text{a.s.}}{\to} \Psi(\mathbf{s}')$, $\Psi_N(\mathbf{s}_i) \overset{\text{a.s.}}{\to} \Psi(\mathbf{s}_i)$ and $|\Psi(\mathbf{s}') - \Psi(\mathbf{s}_i)| < \varepsilon$. This implies that we can choose N large enough so $|\Psi_N(\mathbf{s}') - \Psi_N(\mathbf{s}_i)| < 2\varepsilon$. Moreover this bound holds for all $\mathbf{s}' \in B_{\mathbf{s}_i, \varepsilon}$ and i so

$$\limsup |\Psi_N(\mathbf{s}) - \Psi(\mathbf{s})| \le 3\varepsilon$$

on $\mathcal{S}_+^{p-1}$. Since ε is arbitrary we conclude that $\sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} |\Psi_N(\mathbf{s}) - \Psi(\mathbf{s})| \overset{\text{a.s.}}{\to} 0$ as $N \to \infty$. By assumption $\Psi(\mathbf{s}_{\max}) > \Psi(\mathbf{s})$ for all $\mathbf{s} \in \mathcal{S}_+^{p-1} \setminus \mathbf{s}_{\max}$, so we can apply Theorem 2.12 in Kosorok (2008) to conclude that

$$\hat{\mathbf{s}}_{\max} \overset{\text{a.s.}}{\to} \mathbf{s}_{\max},$$

that is, $\hat{\mathbf{s}}_{\max}$ is strongly consistent. This completes the first part of the proof.

We have already seen that

$$\sup_{\|\mathbf{s} - \mathbf{s}_{\max}\| < \delta} \Psi(\mathbf{s}) - \Psi(\mathbf{s}_{\max}) \le -C\delta^2$$

holds. We now need to bound $\mathbb{E}^* \sup_{\|\mathbf{s} - \mathbf{s}_{\max}\| < \delta} N^{1/2} |(\mathbb{P}_N(\bar{\Psi}(\mathbf{s}) - \bar{\Psi}(\mathbf{s}_{\max})))|$, where $\mathbb{E}^*$ denotes the outer expectation and $\bar{\Psi}(\mathbf{s}) = \mathbb{I}_{(\mathbf{s}^T \mathbf{Z} \ge 0)} - \Psi(\mathbf{s})$. We first note that the bracketing entropy of the upper half-planes is of the order δ/ε^2. The envelope function of the class $\mathbb{I}_{(\mathbf{s}^T \mathbf{z} \ge 0)} - \mathbb{I}_{(\mathbf{s}_{\max}^T \mathbf{z} \ge 0)}$ where $\|\mathbf{s} - \mathbf{s}_{\max}\| < \delta$ is bounded by $\mathbb{I}_{(\mathbf{s}^T \mathbf{z} \ge 0 > \mathbf{s}_{\max}^T \mathbf{z})} + \mathbb{I}_{(\mathbf{s}_{\max}^T \mathbf{z} \ge 0 > \mathbf{s}^T \mathbf{z})}$ whose squared L_2 norm is

$$(A.17) \qquad \mathbb{P}(\mathbf{s}^T \mathbf{Z} \ge 0 > \mathbf{s}_{\max}^T \mathbf{Z}) + \mathbb{P}(\mathbf{s}_{\max}^T \mathbf{Z} \ge 0 > \mathbf{s}^T \mathbf{Z}).$$

Note that we may replace the RV $\mathbf{Z}$ in (A.17) with the RV $\mathbf{Z}' = \mathbf{Z}/\|\mathbf{Z}\|$ whose mass is concentrated on the unit sphere. The condition that $\|\mathbf{s} - \mathbf{s}_{\max}\| < \delta$ implies that the angle between $\mathbf{s}$ and $\mathbf{s}_{\max}$ is of the order $O(\delta)$, and therefore $\mathbb{P}(\mathbf{s}^T \mathbf{Z}' \ge 0 > \mathbf{s}_{\max}^T \mathbf{Z}')$ is computed as surface integral on a spherical wedge with maximum width δ. It follows that (A.17) is bounded by $2A_{p-1}\delta\|h'\|_\infty$ where A_{p-1} is the area of $\mathcal{S}_+^{p-1}$, and $\|h'\|_\infty$ is the supremum of the density of $\mathbf{Z}'$. Clearly $\|h'\|_\infty < \infty$ since the density of $\mathbf{Z}$ is bounded by assumption. Thus by Corollary 19.35 in van der Vaart (2000) we have

$$\mathbb{E}^* \sup_{\|\mathbf{s} - \mathbf{s}_{\max}\| < \delta} N^{1/2} |\mathbb{P}_N(\bar{\Psi}(\mathbf{s}) - \bar{\Psi}(\mathbf{s}_{\max}))| \le C\delta^{1/2}.$$

It now follows that

$$(A.18) \qquad \Psi_N(\hat{\mathbf{s}}_{\max}) \ge \sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \Psi_N(\mathbf{s}) - o_p(N^{-2/3}),$$

which implies by Theorem 5.52 in van der Vaart (2000) and Theorem 14.4 of Kosorok (2008) that

$$\hat{\mathbf{s}}_{\max} = \mathbf{s}_{\max} + O_p(N^{-1/3});$$

that is, $\hat{\mathbf{s}}_{\max}$ converges to $\mathbf{s}_{\max}$ at a cube root rate. This completes the second part of the proof.

The limit distribution is derived by verifying the conditions in Theorem 1.1 of Kim and Pollard (1990), denoted henceforth by KP. First note that (A.18) is condition (i) in KP. Since $\hat{\mathbf{s}}_{\max}$ is consistent, condition (ii) also holds, and condition (iii) holds by assumption. The differentiability of the density of $\mathbf{Z}$ implies that $\Psi(\mathbf{s})$ is twice differentiable. The uniqueness of the maximizer implies that $-\nabla^2 \Psi(\mathbf{s}_{\max})$ is positive definite, and hence condition (iv) holds; see also Example 6.4 in KP for related calculations. Condition (v) in KP is equivalent to the existence of the limit $H(\mathbf{u}, \mathbf{v}) = \lim_{\alpha \to \infty} \alpha \mathbb{E}(\bar{\Psi}(\mathbf{s}_{\max} + \mathbf{u}/\alpha)\bar{\Psi}(\mathbf{s}_{\max} + \mathbf{v}/\alpha))$ which can be rewritten as

$$\lim_{\beta \to 0} \frac{1}{\beta} \big[\mathbb{P}((\mathbf{s}_{\max} + \beta\mathbf{u})^T \mathbf{Z} \geq 0, (\mathbf{s}_{\max} + \beta\mathbf{v})^T \mathbf{Z} \geq 0)$$
$$- \mathbb{P}((\mathbf{s}_{\max} + \beta\mathbf{u})^T \mathbf{Z} \geq 0)\mathbb{P}((\mathbf{s}_{\max} + \beta\mathbf{v})^T \mathbf{Z} \geq 0) \big].$$

With some algebra we find that this limit exists and equals

$$\int_{\mathbf{z} \in \mathcal{S}_+^{p-1}} \delta(\mathbf{s}_{\max}^T \mathbf{z})(\mathbf{z}^T(\mathbf{u} + \mathbf{v}))h(\mathbf{z})\, d\mathbf{z}$$
$$- \int_{\mathbf{z} \in \mathcal{S}_+^{p-1}} \delta(\mathbf{s}_{\max}^T \mathbf{z})(\mathbf{z}^T \mathbf{u})h(\mathbf{z})\, d\mathbf{z} \int_{\mathbf{z} \in \mathcal{S}_+^{p-1}} \delta(\mathbf{s}_{\max}^T \mathbf{z})(\mathbf{z}^T \mathbf{v})h(\mathbf{z})\, d\mathbf{z},$$

where $\delta(\mathbf{s}_{\max}^T \mathbf{z})$ is the usual Dirac function; hence integration is with respect to the surface measure on $\{\mathbf{s}_{\max}^T \mathbf{z} = 0\}$. It follows that condition also (v) holds. Conditions (vi) and (vii) were verified in the second part of the proof. Thus we may apply Theorem 1.1 in KP to get

$$N^{1/3}(\hat{\mathbf{s}}_{\max} - \mathbf{s}_{\max}) \Rightarrow \arg\max\{-Q(\mathbf{s}) + \mathbb{W}(\mathbf{s}) : \mathbf{s} \in \mathcal{S}_+^{p-1}\},$$

where by KP $Q(\mathbf{s}) = \mathbf{s}^T \nabla^2 \Psi(\mathbf{s}_{\max})\mathbf{s}$ and $\mathbb{W}(\mathbf{s})$ is a zero mean Gaussian process with covariance function $H(\mathbf{u}, \mathbf{v})$. This completes the proof. $\square$

PROOF OF PROPOSITION 3.3.   Note that

$$\Psi(\mathbf{s}) = \mathbb{P}(\mathbf{s}^T \mathbf{X} \leq \mathbf{s}^T \mathbf{Y}) = \mathbb{P}(\mathbf{s}^T \mathbf{Z} \geq 0) = \mathbb{P}\left(\frac{\mathbf{s}^T \mathbf{Z} - \mathbf{s}^T \boldsymbol{\delta}}{\sqrt{\mathbf{s}^T \boldsymbol{\Sigma} \mathbf{s}}} \geq -\frac{\mathbf{s}^T \boldsymbol{\delta}}{\sqrt{\mathbf{s}^T \boldsymbol{\Sigma} \mathbf{s}}} \right)$$
$$= 1 - F\left(-\frac{\mathbf{s}^T \boldsymbol{\delta}}{\sqrt{\mathbf{s}^T \boldsymbol{\Sigma} \mathbf{s}}} \right).$$

Now, by assumption the DF F is independent of $\mathbf{s}$. Therefore $\Psi(\mathbf{s})$ is uniquely maximized on $\mathcal{S}_+^{p-1}$ if and only if the function

$$\varkappa(\mathbf{s}) = \frac{\mathbf{s}^T \boldsymbol{\delta}}{\sqrt{\mathbf{s}^T \boldsymbol{\Sigma} \mathbf{s}}}$$

is uniquely maximized on $\mathcal{S}_+^{p-1}$. If $\boldsymbol{\Sigma} = \boldsymbol{I}$, then $\varkappa(\mathbf{s}) = \mathbf{s}^T \boldsymbol{\delta}$, and we wish to maximize a linear function on $\mathcal{S}_+^{p-1}$. It is easily verified (by using ideas from linear programming) that the maximizer is unique if $\boldsymbol{\delta} \geq \mathbf{0}$ which is true by assumption. Incidentally, it is easy to show directly that $\varkappa(\mathbf{s})$ is maximized at $\mathbf{s}^*/\|\mathbf{s}^*\|$ where

$$\mathbf{s}^* = (\delta_1 \mathbb{I}_{(\delta_1 \geq 0)}, \dots, \delta_p \mathbb{I}_{(\delta_p \geq 0)}).$$

Now let $\boldsymbol{\Sigma} \neq \boldsymbol{I}$ and assume that a unique maximizer does not exist; that is, suppose that $\varkappa(\mathbf{s})$ is maximized by both $\mathbf{s}_1$ and $\mathbf{s}_2$. It is clear that $\varkappa(\lambda_1 \mathbf{s}_1) = \varkappa(\lambda_2 \mathbf{s}_2)$ for all $\lambda_1, \lambda_2 > 0$; that is, the value of $\varkappa(\cdot)$ is constant along rays through the origin. The rays passing through $\mathbf{s}_1$ and $\mathbf{s}_2$, respectively, intersect the ellipsoid $\mathbf{s}^T \boldsymbol{\Sigma} \mathbf{s} = 1$ at the points $\mathbf{p}_1$ and $\mathbf{p}_2$. It follows that $\varkappa(\mathbf{p}_1) = \varkappa(\mathbf{p}_2)$, moreover $\mathbf{p}_1$ and $\mathbf{p}_2$ maximize $\varkappa(\cdot)$ on the ellipsoid. Now since $\mathbf{p}_1^T \boldsymbol{\Sigma} \mathbf{p}_1 = 1 = \mathbf{p}_2^T \boldsymbol{\Sigma} \mathbf{p}_2$ we must have $\mathbf{p}_1^T \boldsymbol{\delta} = \mathbf{p}_2^T \boldsymbol{\delta}$. Recall that a linear function on ellipsoid is uniquely maximized (just like on a sphere; see the comment above). Therefore we must have $\mathbf{p}_1 = \mathbf{p}_2$ which implies that $\mathbf{s}_1 = \mathbf{s}_2$ as required. $\square$

PROOF OF THEOREM 3.4. If $\mathbf{X} =_{\mathrm{st}} \mathbf{Y}$, then for all $\mathbf{s}$ we have $\mathbf{s}^T \mathbf{X} =_{\mathrm{st}} \mathbf{s}^T \mathbf{Y}$. By assumption both $\mathbf{s}^T \mathbf{X}$ and $\mathbf{s}^T \mathbf{Y}$ are continuous RVs, so $\mathbb{P}(\mathbf{s}^T \mathbf{X} \leq \mathbf{s}^T \mathbf{Y}) = 1/2$. Suppose now that both $\mathbf{X} \preceq_{\mathrm{st}} \mathbf{Y}$ and $\mathbb{P}(\mathbf{s}^T \mathbf{X} \leq \mathbf{s}^T \mathbf{Y}) > 1/2$ for some $\mathbf{s} \in \mathcal{S}_+^{p-1}$, hold. Then we must have $\mathbf{X} \prec_{l\text{-st}} \mathbf{Y}$. Since $\mathbf{X} \preceq_{\mathrm{st}} \mathbf{Y}$ we have $X_j \preceq_{\mathrm{st}} Y_j$ for $1 \leq j \leq p$. One of these inequalities must be strict; otherwise $\mathbf{X} =_{\mathrm{st}} \mathbf{Y}$ contradicts the fact that $\mathbf{X} \prec_{l\text{-st}} \mathbf{Y}$. Now use Theorem 1 in Davidov and Peddada (2011) to complete the proof. $\square$

PROOF OF THEOREM 3.5. The functions ψ_1 and ψ_2 defined in the proof of Theorem 3.1 are Donsker; cf. Example 19.7 in van der Vaart (2000). Hence by the theory of empirical processes applied to (A.12), we find that

$$(\text{A.19}) \qquad N^{1/2}\big(\Psi_{n,m}(\mathbf{s}) - \Psi(\mathbf{s})\big) \Rightarrow \mathbb{G}(\mathbf{s}),$$

where $\mathbb{G}(\mathbf{s})$ is a zero mean Gaussian process, and convergence holds for all $\mathbf{s} \in \mathcal{S}_+^{p-1}$. We also note that (A.19) is a two-sample U-processes. A central limit theorem for such processes is described by Neumeyer (2004). Hence by the continuos mapping theorem, and under H_0, we have $N^{1/2}(\Psi_{n,m}(\hat{\mathbf{s}}_{\max}) - 1/2)) \Rightarrow \sup_{\mathbf{s} \in \mathcal{S}_+^{p-1}} \mathbb{G}(\mathbf{s})$ where the covariance function of $\mathbb{G}(\mathbf{s})$, denoted by $C(\mathbf{u}, \mathbf{v})$, is given

by

$$(A.20) \quad \frac{1}{\lambda}\mathbb{P}\big(\mathbf{u}^T\mathbf{X}_1 \le \mathbf{u}^T\mathbf{X}_2, \mathbf{v}^T\mathbf{X}_1 \le \mathbf{v}^T\mathbf{X}_3\big)$$

$$+ \frac{1}{1-\lambda}\mathbb{P}\big(\mathbf{u}^T\mathbf{X}_1 \le \mathbf{u}^T\mathbf{X}_2, \mathbf{v}^T\mathbf{X}_3 \le \mathbf{v}^T\mathbf{X}_2\big) - \frac{1}{4\lambda(1-\lambda)},$$

where $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ are i.i.d. from the common DF. $\quad\square$

PROOF OF THEOREM 3.6. Suppose that $\mathbf{X} \prec_{l\text{-st}} \mathbf{Y}$. Then for some $\mathbf{s}_* \in \mathcal{S}_+^{p-1}$ we have $\mathbf{s}_*^T\mathbf{X} \prec_{\text{st}} \mathbf{s}_*^T\mathbf{Y}$ which implies that $\mathbb{P}(\mathbf{s}_*^T\mathbf{X} \le \mathbf{s}_*^T\mathbf{Y}) > 1/2$. By definition $\mathbb{P}(\mathbf{s}_{\max}^T\mathbf{X} \le \mathbf{s}_{\max}^T\mathbf{Y}) \ge \mathbb{P}(\mathbf{s}_*^T\mathbf{X} \le \mathbf{s}_*^T\mathbf{Y})$ so $\Psi(\mathbf{s}_{\max}) > 1/2$. It follows from the proof of Theorem 3.1 that $\Psi_{n,m}(\hat{\mathbf{s}}_{\max}) \to \Psi(\mathbf{s}_{\max})$ with probability one. Thus,

$$S_{n,m} = N^{1/2}\big(\Psi_{n,m}(\hat{\mathbf{s}}_{\max}) - 1/2\big) \xrightarrow{\text{a.s.}} \infty \qquad \text{as } n, m \to \infty.$$

Therefore by Slutzky's theorem,

$$\mathbb{P}(S_{n,m} > q_{n,m,1-\alpha}; H_1) \to 1 \qquad \text{as } n, m \to \infty,$$

where $q_{n,m,1-\alpha}$ is the critical value for an α level test based on samples of size n and m and $q_{n,m,1-\alpha} \to q_{1-\alpha}$. Hence the test based on $S_{n,m}$ is consistent. Consistency for $I_{n,m}$ and $I_{n,m}^+$ is established in a similar manner.

Now assume that $\mathbf{X} \preceq_{l\text{-st}} \mathbf{Y} \preceq_{l\text{-st}} \mathbf{Z}$ so that $\mathbf{s}^T\mathbf{Y} \preceq_{\text{st}} \mathbf{s}^T\mathbf{Z}$ for all $\mathbf{s} \in \mathcal{S}_+^{p-1}$. Fix $\mathbf{x}_i$, $i = 1, \ldots, n$, and choose $\mathbf{s} \in \mathcal{S}_+^{p-1}$. Without any loss of generality assume that $\mathbf{s}^T\mathbf{x}_1 \le \mathbf{s}^T\mathbf{x}_2 \le \cdots \le \mathbf{s}^T\mathbf{x}_n$. Define $U_j = \sum_{i=1}^n \mathbb{I}_{(\mathbf{s}^T\mathbf{x}_i \le \mathbf{s}^T\mathbf{Y}_j)}$ and $V_j = \sum_{i=1}^n \mathbb{I}_{(\mathbf{s}^T\mathbf{x}_i \le \mathbf{s}^T\mathbf{Z}_j)}$. Clearly U_j and V_j take values in $J = \{0, \ldots, n\}$. Now, for $k \in J$ we have

$$\mathbb{P}(U_j \ge k) = \mathbb{P}\big(\mathbf{s}^T\mathbf{Y}_j \ge \mathbf{s}^T\mathbf{x}_k\big) \le \mathbb{P}\big(\mathbf{s}^T\mathbf{Z}_j \ge \mathbf{s}^T\mathbf{x}_k\big) = \mathbb{P}(V_j \ge k),$$

where we use the fact that $\mathbf{s}^T\mathbf{Y} \preceq_{\text{st}} \mathbf{s}^T\mathbf{Z}$. It follows that $U_j \preceq_{\text{st}} V_j$ for $j = 1, \ldots, m$. Moreover $\{U_j\}$ and $\{V_j\}$ are all independent and it follows from Theorem 1.A.3 in Shaked and Shanthikumar (2007) that $\sum_{j=1}^m U_j \preceq_{\text{st}} \sum_{j=1}^m V_j$. Thus $\sum_{i=1}^n \sum_{j=1}^m \mathbb{I}_{(\mathbf{s}^T\mathbf{x}_i \le \mathbf{s}^T\mathbf{Y}_j)} \preceq_{\text{st}} \sum_{i=1}^n \sum_{j=1}^m \mathbb{I}_{(\mathbf{s}^T\mathbf{x}_i \le \mathbf{s}^T\mathbf{Z}_j)}$. The latter holds for every value of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and therefore it holds unconditionally as well, that is,

$$\sum_{i=1}^n \sum_{j=1}^m \mathbb{I}_{(\mathbf{s}^T\mathbf{X}_i \le \mathbf{s}^T\mathbf{Y}_j)} \preceq_{\text{st}} \sum_{i=1}^n \sum_{j=1}^m \mathbb{I}_{(\mathbf{s}^T\mathbf{X}_i \le \mathbf{s}^T\mathbf{Z}_j)}.$$

It follows that $\Psi_{n,m}^{\mathbf{X},\mathbf{Y}}(\mathbf{s}) \preceq_{\text{st}} \Psi_{n,m}^{\mathbf{X},\mathbf{Z}}(\mathbf{s})$ for all $\mathbf{s} \in \mathcal{S}_+^{p-1}$ where $\Psi_{n,m}^{\mathbf{X},\mathbf{Y}}(\mathbf{s})$ and $\Psi_{n,m}^{\mathbf{X},\mathbf{Z}}(\mathbf{s})$ are defined in (3.1) and the superscripts emphasize the different arguments used to evaluate them. Thus

$$\Psi_{n,m}^{\mathbf{X},\mathbf{Y}}(\hat{\mathbf{s}}_{\max}^{\mathbf{X},\mathbf{Y}}) \preceq_{\text{st}} \Psi_{n,m}^{\mathbf{X},\mathbf{Z}}(\hat{\mathbf{s}}_{\max}^{\mathbf{X},\mathbf{Y}}) \preceq_{\text{st}} \Psi_{n,m}^{\mathbf{X},\mathbf{Z}}(\hat{\mathbf{s}}_{\max}^{\mathbf{X},\mathbf{Z}})$$

and as a consequence $\mathbb{P}(S_{n,m}^{X,Y} > q_{n,m,1-\alpha}) \le \mathbb{P}(S_{n,m}^{X,Z} > q_{n,m,1-\alpha})$ as required. The monotonicity of the power function of $I_{n,m}$ and $I_{n,m}^+$ follows immediately from the fact that $\Psi_{n,m}^{\mathbf{X},\mathbf{Y}}(\mathbf{s}) \preceq_{\text{st}} \Psi_{n,m}^{\mathbf{X},\mathbf{Z}}(\mathbf{s})$ for all $\mathbf{s} \in \mathcal{S}_+^{p-1}$. $\quad\square$

REFERENCES

ABREVAYA, J. and HUANG, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73** 1175–1204. MR2149245

ARCONES, M. A., KVAM, P. H. and SAMANIEGO, F. J. (2002). Nonparametric estimation of a distribution subject to a stochastic precedence constraint. *J. Amer. Statist. Assoc.* **97** 170–182. MR1947278

AUDET, C., BÉCHARD, V. and LE DIGABEL, S. (2008). Nonsmooth optimization through mesh adaptive direct search and variable neighborhood search. *J. Global Optim.* **41** 299–318. MR2398937

BEKELE, B. N. and THALL, P. F. (2004). Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. *J. Amer. Statist. Assoc.* **99** 26–35. MR2061885

BICKEL, P. J. and SAKOV, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statist. Sinica* **18** 967–985. MR2440400

DASGUPTA, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer, New York. MR2664452

DAVEY, B. A. and PRIESTLEY, H. A. (2002). *Introduction to Lattices and Order*, 2nd ed. Cambridge Univ. Press, New York. MR1902334

DAVIDOV, O. (2012). Ordered inference, rank statistics and combining p-values: A new perspective. *Stat. Methodol.* **9** 456–465. MR2871445

DAVIDOV, O. and HERMAN, A. (2011). Multivariate stochastic orders induced by case-control sampling. *Methodol. Comput. Appl. Probab.* **13** 139–154. MR2755136

DAVIDOV, O. and HERMAN, A. (2012). Ordinal dominance curve based inference for stochastically ordered distributions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **74** 825–847.

DAVIDOV, O. and PEDDADA, S. (2011). Order-restricted inference for multivariate binary data with application to toxicology. *J. Amer. Statist. Assoc.* **106** 1394–1404. MR2896844

DELAGDO, M., RODRIGUEZ-POO, J. M. and WOLF, M. (2001). Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator. *Econom. Lett.* **73** 241–250.

DING, Y. and ZHANG, X. (2004). Some stochastic orders of Kotz-type distributions. *Statist. Probab. Lett.* **69** 389–396. MR2091758

FANG, K. T., KOTS, S. and NG, K. W. (1989). *Symmetric Multivariate and Related Distributions*. Chapman & Hall, London.

FISHER, N. I. and HALL, P. (1989). Bootstrap confidence regions for directional data. *J. Amer. Statist. Assoc.* **84** 996–1002. MR1134489

HÁJEK, J., ŠIDÁK, Z. and SEN, P. K. (1999). *Theory of Rank Tests*, 2nd ed. Academic Press, San Diego, CA. MR1680991

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

HU, J., HOMEM-DE-MELLO, T. and MEHROTRA, S. (2011). Concepts and applications of stochastically weighted stochastic dominance. Unpublished manuscript. Available at http://www.optimization-online.org/DB_FILE/2011/04/2981.pdf.

IVANOVA, A. and MURPHY, M. (2009). An adaptive first in man dose-escalation study of NGX267: Statistical, clinical, and operational considerations. *J. Biopharm. Statist.* **19** 247–255. MR2518371

JOE, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London. MR1462613

JOHNSON, R. and WICHERN, D. (1998). *Applied Multivariate Statistical Analysis*. Prentice Hall, New York.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist*. **18** 191–219. MR1041391

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York. MR2724368

LEE, S. M. S. (1999). On a class of *m* out of *n* bootstrap confidence intervals. *J. R. Stat. Soc. Ser. B Stat. Methodol*. **61** 901–911. MR1722246

LUCAS, L. A. and WRIGHT, F. T. (1991). Testing for and against a stochastic ordering between multivariate multinomial populations. *J. Multivariate Anal*. **38** 167–186. MR1131713

MOSER, V. (2000). Observational batteries in neurotoxicity testing. *International Journal of Toxicology* **19** 407–411.

NEUMEYER, N. (2004). A central limit theorem for two-sample U-processes. *Statist. Probab. Lett*. **67** 73–85. MR2039935

NTP (2003). NTP toxicology and carcinogenesiss studies of citral (microencapsulated) (CAS No. 5392-40-5) in F344/N rats and B6C3F1 mice (feed studies).

PEDDADA, S. D. (1985). A short note on Pitman's measure of nearness. *Amer. Statist*. **39** 298–299. MR0817793

PEDDADA, S. D. and CHANG, T. (1996). Bootstrap confidence region estimation of the motion of rigid bodies. *J. Amer. Statist. Assoc*. **91** 231–241. MR1394077

PITMAN, E. J. G. (1937). The closest estimates of statistical parameters. *Proceedings of the Cambridge Philosophical Society* **33** 212–222.

PRICE, C. J., REALE, M. and ROBERTSON, B. L. (2008). A direct search method for smooth and non-smooth unconstrained optimization problems. *ANZIAM J*. **48** 927–948.

ROY, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist*. **24** 220–238. MR0057519

SAMPSON, A. R. and WHITAKER, L. R. (1989). Estimation of multivariate distributions under stochastic ordering. *J. Amer. Statist. Assoc*. **84** 541–548. MR1010343

SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist*. **38** 1953–1977. MR2676880

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

SHAKED, M. and SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. Springer, New York. MR2265633

SHERMAN, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61** 123–137. MR1201705

VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge.

DEPARTMENT OF STATISTICS
UNIVERSITY OF HAIFA
MOUNT CARMEL, HAIFA 31905
ISRAEL
E-MAIL: davidov@stat.haifa.ac.il

BIOSTATISTICS BRANCH
NATIONAL INSTITUTE
  OF ENVIRONMENTAL HEALTH SCIENCES
111 T.W. ALEXANDER DRIVE
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709
USA
E-MAIL: peddada@niehs.nih.gov