Submitted to the Annals of Statistics
arXiv: math.PR/0000000

NONPARAMETRIC GOODNESS-OF-FIT TESTS FOR UNIFORM STOCHASTIC ORDERING

By Chuan-Fa Tang, Dewei Wang and Joshua M. Tebbs

Department of Statistics, University of South Carolina

We propose L^p distance-based goodness-of-fit (GOF) tests for uniform stochastic ordering with two continuous distributions F and G, both of which are unknown. Our tests are motivated by the fact that when F and G are uniformly stochastically ordered, the ordinal dominance curve $R = FG^{-1}$ is star-shaped. We derive asymptotic distributions and prove that our testing procedure has a unique least favorable configuration of F and G for $p \in [1, \infty]$. We use simulation to assess finite-sample performance and demonstrate that a modified, one-sample version of our procedure (e.g., with G known) is more powerful than the one-sample GOF test suggested by Arcones and Samaniego (2000, *Annals of Statistics*). We also discuss sample size determination. We illustrate our methods using data from a pharmacology study evaluating the effects of administering caffeine to prematurely born infants.

1. Introduction. Suppose X and Y are continuous random variables with distribution functions F and G, respectively. In many applications, it is of interest to compare F and G. The ordinal dominance curve (ODC), which plots $(G(t), F(t))$ for $-\infty \leq t \leq \infty$, is a useful graphical tool that facilitates such a comparison (Bamber, 1975; Hsieh and Turnbull, 1996; Carolan and Tebbs, 2005; Davidov and Herman, 2012). The ODC can also be defined as $R = FG^{-1}$, where $G^{-1}(u) = \inf\{t : G(t) \geq u\}$ is the quantile function of G. When $F = G$, the ODC follows the main diagonal of the unit square, the so-called equal distribution line.

We consider order-restricted comparisons of F and G. Define $\overline{F} = 1 - F$ and $\overline{G} = 1 - G$. These are the survivor functions if X and Y are lifetime random variables, although herein we do not require X and Y to be nonnegative. Denote the corresponding densities by f and g, respectively. If $\overline{F} \leq \overline{G}$, then X and Y are stochastically ordered; this is written as $F \leq_{\mathrm{S}} G$ and means informally that X "tends to be smaller" than Y. Two stronger

*This work was supported by Grant R01 AI121351 from the National Institutes of Health.

MSC 2010 subject classifications: Primary 62G10; secondary 62F30.

Keywords and phrases: Brownian bridge, Hazard rate ordering, Least favorable distribution, Order-restricted inference, Ordinal dominance curve, Star-shaped ordering.

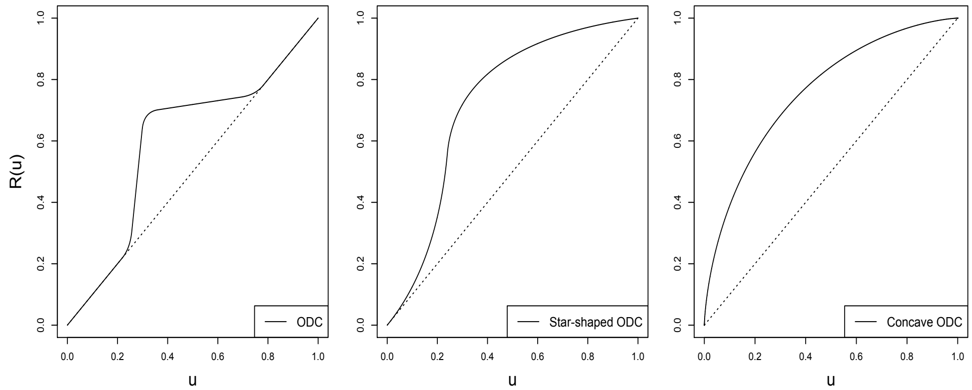

FIG 1. *Ordinal dominance curves. Left:* $F \leq_{\mathrm{S}} G$. *Middle:* $F \leq_{\mathrm{US}} G$. *Right:* $F \leq_{\mathrm{LR}} G$. *In each subfigure, the equal distribution line is shown dotted.*

orders are the uniform stochastic order and the likelihood ratio order. When $\overline{F}/\overline{G}$ is nonincreasing, X and Y satisfy a uniform stochastic order, written $F \leq_{\mathrm{US}} G$. When f/g is nonincreasing, X and Y satisfy a likelihood ratio order, written $F \leq_{\mathrm{LR}} G$. It is easy to show these orderings follow the nested structure: $F \leq_{\mathrm{LR}} G \Longrightarrow F \leq_{\mathrm{US}} G \Longrightarrow F \leq_{\mathrm{S}} G$. A comprehensive account of these and other orderings is given in Shaked and Shanthikumar (2007).

Different stochastic orderings give rise to different functional forms of the ODC. The weakest ordering $F \leq_{\mathrm{S}} G$ holds if and only if R is at least as large as the equal distribution line; i.e., $R(u) \geq u$, for $0 \leq u \leq 1$. The strongest ordering $F \leq_{\mathrm{LR}} G$ holds if and only if R is concave. The intermediate ordering $F \leq_{\mathrm{US}} G$ holds if and only if R is *star-shaped* (Lehmann and Rojo, 1992). One way to characterize a star-shaped ODC is that the slope of the secant line from the point $(1,1)$ to $(u, R(u))$; i.e., $r(u) = \{1-R(u)\}/(1-u)$, is nonincreasing in u. Figure 1 gives examples of ODCs that correspond to stochastic, uniform stochastic, and likelihood ratio orderings. This figure demonstrates the utility of the ODC in characterizing how two distributions are ordered and how the structure $F \leq_{\mathrm{LR}} G \Longrightarrow F \leq_{\mathrm{US}} G \Longrightarrow F \leq_{\mathrm{S}} G$ manifests itself graphically in the ODC.

This article is motivated by a pharmacology study evaluating the effects of administering caffeine to prematurely born infants in Columbia, South Carolina; see Section 5. Among 404 infants in the study, $m = 127$ were administered caffeine and $n = 277$ were not. Each infant was then followed until he or she was discharged from the hospital. All infants were eventually discharged and were alive at the time of discharge; i.e., no discharge times were censored. One of the goals of the study was to understand how the

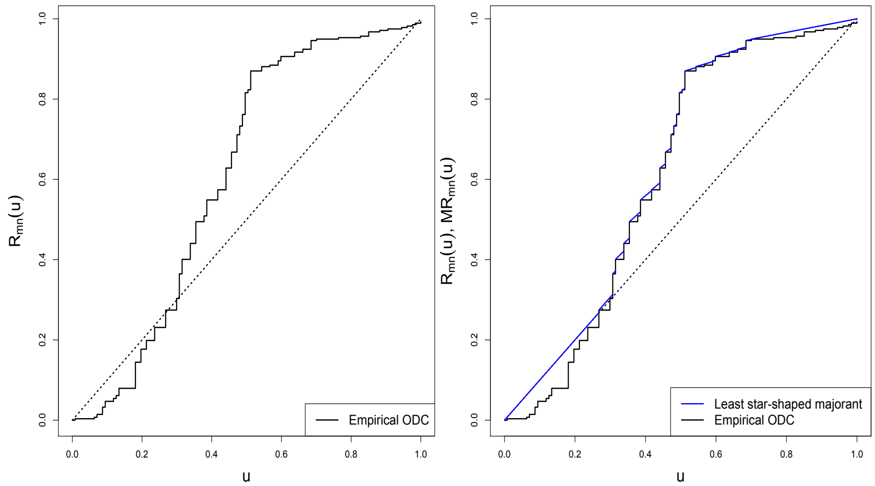

FIG 2. *Premature infant data. Left: The sample ODC* $R_{mn}(u) = F_m\{G_n^{-1}(u)\}$ *for the time to discharge (*F *= caffeine;* G *= no caffeine). Right: The least star-shaped majorant* $\mathcal{M}R_{mn}$ *is shown in blue. In each subfigure, the equal distribution line is shown dotted.*

distributions of discharge times F (caffeine) and G (no caffeine) compared for the two groups. In Figure 2 (left), we display the sample ODC for the data, which is defined as $R_{mn}(u) = F_m\{G_n^{-1}(u)\}$, for $0 \le u \le 1$, where F_m and G_n are the empirical distribution functions and $G_n^{-1}(u) = \inf\{t : G_n(t) \ge u\}$ is the empirical quantile function. The sample ODC and its large-sample properties were described in Hsieh and Turnbull (1996).

On the basis of Figure 2, which stochastic ordering, if any, characterizes the true relationship between the discharge time distributions? There is a substantive literature on nonparametric tests for stochastic orderings with two or more distributions; see Davidov and Herman (2012), El Barmi and McKeague (2016), and the references therein. In the two-sample case, most of this literature describes tests where the equal distribution assumption $F = G$ is treated as the null hypothesis and the ordering (i.e., $F \le_{\mathrm{S}} G$, $F \le_{\mathrm{US}} G$, or $F \le_{\mathrm{LR}} G$) is placed in the alternative. A potential drawback with this type of test is that it is constructed assuming a specific order-restricted class of alternatives; if the assumed class is incorrect, the test may lead to misleading or vacuous conclusions. For example, applying tests of this type to the premature infant data, we obtain the following results:

- testing $F = G$ versus $F \le_{\mathrm{S}} G$: p-value < 0.00002 (Davidov and Herman, 2012)
- testing $F = G$ versus $F \le_{\mathrm{US}} G$: p-value < 0.00001 (Arcones and

Samaniego, 2000)
- testing $F = G$ versus $F \leq_{\mathrm{LR}} G$: p-value < 0.00001 (Carolan and Tebbs, 2005).

Each test clearly dictates that the infant data are not consistent with $F = G$. However, we are no closer to identifying which specific ordering (if any) holds in this setting.

In this light, we consider goodness-of-fit (GOF) testing procedures instead. By "goodness-of-fit," we mean the procedure places the ordering in the null hypothesis and attempts to detect departures from the ordering. By comparison, the literature on nonparametric GOF tests with two distributions is more sparse, perhaps because this type of testing problem is more difficult. The primary reason for the added difficulty is that the ordering can hold under different configurations of F and G. Therefore, one must determine the least favorable configuration of the two distributions before the test can be performed; i.e., so that the probability of type I error can be controlled. Carolan and Tebbs (2005) proposed nonparametric GOF tests for likelihood ratio ordering with two continuous distributions by using the least concave majorant of the sample ODC. This work was generalized and improved upon by Beare and Moon (2015) in the econometrics literature, who considered likelihood ratio ordering and its applications in finance.

GOF tests for uniform stochastic ordering have been proposed but only in limited settings. Dardanoni and Forcina (1998) considered likelihood-based tests against uniform stochastic ordering in a two-way contingency table. Park et al. (1998) used a nonparametric maximum likelihood approach to formulate GOF tests with two or more continuous distributions, but only after data from these distributions have been assigned to disjoint intervals in the form of counts. This essentially discretizes the problem and results in testing against uniform stochastic ordering among several multinomial distributions. Furthermore, this formulation gives rise to non-unique least favorable configurations that depend on how the intervals are selected, the number of distributions, and even the significance level used. Finally, in the two-population setting, Arcones and Samaniego (2000) suggested a GOF test for uniform stochastic ordering based on the family of order-restricted estimators in Mukerjee (1996). However, these authors assume that one of the population distributions is known (e.g., G is known) and do not determine the least favorable configuration for their procedure. Instead, the authors use critical values from an upper bound asymptotic distribution which leads to a conservative test.

In this article, we propose a family of GOF tests for uniform stochastic ordering with two continuous distributions F and G; that is, we are interested in testing $H_0 : F \leq_{\text{US}} G$ versus $H_1 : F \nleq_{\text{US}} G$, where both distributions are unknown. Motivated by the ODC approaches taken in Carolan and Tebbs (2005) and Beare and Moon (2015), we construct test statistics for H_0 versus H_1 based on the L^p difference between the sample ODC and its *least star-shaped majorant* (defined in Section 2). We then derive asymptotic distributions and prove that our testing procedure has a unique least favorable configuration for $p \in [1, \infty]$. Interestingly, this theoretical result is different from the finding in Beare and Moon (2015), who showed that when using L^p distance-based GOF tests for likelihood ratio ordering, the least favorable configuration exists only when $p \in [1, 2]$. Furthermore, unlike Park et al. (1998), our approach does not require one to discretize the support of the distributions which can only lead to a loss in power. Finally, we show that the one-sample version of our test (e.g., with G known) is not as conservative as the test proposed by Arcones and Samaniego (2000) and is generally better equipped to detect departures from H_0.

Formulating L^p distance-based GOF tests for uniform stochastic ordering in the two-sample problem is technically challenging. It is not possible to simply modify the proofs in Carolan and Tebbs (2005) and Beare and Moon (2015) under likelihood ratio ordering; see Section 3. At the same time, establishing that such an ordering exists has great practical implications. For example, if X and Y are lifetime random variables (and are absolutely continuous), then $F \leq_{\text{US}} G$ is equivalent to the corresponding hazard rates being ordered. This is an important characterization in reliability and survival analysis applications. Our interest in uniform stochastic ordering is motivated by our collaboration with researchers in the premature infant study discussed earlier. Letting X and Y denote the times to discharge for the caffeine and no-caffeine groups, respectively, uniform stochastic ordering holds if and only if $\text{pr}(X > t | X > t_0) \leq \text{pr}(Y > t | Y > t_0)$, for all t, t_0 satisfying $t > t_0 \geq 0$. In other words, no matter how much time $t_0 \geq 0$ has subsequently passed, administering caffeine is consistent with shorter discharge times. Note that, in this context, stochastic ordering requires that the relationship above hold only initially (i.e., when $t_0 = 0$). Uniform stochastic ordering guarantees this type of dominance will hold for all $t_0 \geq 0$.

2. Testing procedure. Suppose that $X_1, X_2, ..., X_m$ are independent and identically distributed (iid) from F and that $Y_1, Y_2, ..., Y_n$ are iid from G. We assume the two samples are independent and that both F and G are unknown. Let $R = FG^{-1}$ denote the corresponding ODC. For our asymptotic results in Section 3 to hold, as in Hsieh and Turnbull (1996), we assume F and G have continuous densities f and g and that the first derivative of

R is bounded over $[0,1]$. Throughout this article, we denote the parameter space of R by Θ, the collection of nondecreasing, continuously differentiable functions from $[0,1]$ to $[0,1]$. Under our assumptions, the hypotheses $H_0 : F \leq_{\mathrm{US}} G$ and $H_1 : F \not\leq_{\mathrm{US}} G$ can be expressed equivalently as

$$H_0 : R \in \Theta_0 = \{\theta \in \Theta : \theta \text{ is star-shaped}\} \quad \text{and} \quad H_1 : R \in \Theta_1 = \Theta \setminus \Theta_0.$$

Recall that $\theta \in \Theta$ is star-shaped if and only if $\{1-\theta(u)\}/(1-u)$ is nonincreasing in u.

Let $R_{mn} = R_{mn}(u) = F_m\{G_n^{-1}(u)\}$ denote the sample ODC, defined in Section 1. Informally, our testing procedure is based on measuring the distance between R_{mn} and an estimate of R subject to the constraint that $F \leq_{\mathrm{US}} G$. Towards defining this restricted estimator, let $l([0,1])$ denote the collection of bounded functions on $[0,1]$. For any $h \in l([0,1])$, its least star-shaped majorant is defined as

$$\mathcal{M}h = \inf\{h^* \in l([0,1]) : h \leq h^* \text{ and } h^* \text{ is star-shaped}\};$$

i.e., $\mathcal{M}h$ is the smallest star-shaped function in $l([0,1])$ that is at least as large as h. Throughout our work, we call $\mathcal{M} : l([0,1]) \mapsto l([0,1])$ the least star-shaped majorant operator. Just as R_{mn} is an estimator of R under no restriction (Hsieh and Turnbull, 1996), the least star-shaped majorant $\mathcal{M}R_{mn}$ is an estimator of R under $H_0 : F \leq_{\mathrm{US}} G$. Using Lemma 1 in the supplementary article (Tang et al., 2016), we show that this restricted estimator can be calculated as

$$\mathcal{M}R_{mn}(u) = 1 - \min_{\substack{v \in \mathcal{V}\cup\{0\} \\ v \leq u}} \left\{\frac{1-R_{mn}(v)}{1-v}\right\}(1-u),$$

for $0 \leq u < 1$, where $\mathcal{V}$ is the set of discontinuous (jump) points of R_{mn} and $\mathcal{M}R_{mn}(1) = 1$. Figure 2 (right) shows the least star-shaped majorant of the sample ODC for the premature infant data described in Section 1.

Our testing procedure utilizes the sample ODC R_{mn} and its least star-shaped majorant $\mathcal{M}R_{mn}$. Specifically, we propose the family of test statistics

$$M_{mn}^p = c_{mn}\|\mathcal{M}R_{mn} - R_{mn}\|_p,$$

where $c_{mn} = \{mn/(m+n)\}^{1/2}$ is a normalizing constant and $\|\cdot\|_p$ is the L^p norm with respect to Lebesgue measure. We allow for $p \in [1,\infty]$; i.e., $\|h\|_p = (\int_{[0,1]} |h(u)|^p du)^{1/p}$ when $p < \infty$ and $\|h\|_\infty = \sup_{u\in[0,1]} |h(u)|$. For example, when $p = 1$, $\|\mathcal{M}R_{mn} - R_{mn}\|_1$ equals the area between the two estimators; when $p = \infty$, $\|\mathcal{M}R_{mn} - R_{mn}\|_\infty$ equals the largest vertical distance between the estimators. For any $p \in [1,\infty]$, clearly large values of M_{mn}^p are evidence against H_0.

3. Theoretical results. In this section, we first describe the asymptotic distribution of M^p_{mn} for any star-shaped ODC; i.e., for any $R \in \Theta_0$. We then demonstrate that, for any $p \in [1, \infty]$, all null distributions are dominated stochastically by the asymptotic distribution of M^p_{mn} under $R(u) = u$, that is, when $F = G$. From this least favorable distribution, we can find the critical value $c_{\alpha,p}$ that satisfies $\lim_{m,n\to\infty} \mathrm{pr}(M^p_{mn} \geq c_{\alpha,p}) = \alpha$ when $F = G$ and $\lim_{m,n\to\infty} \mathrm{pr}(M^p_{mn} \geq c_{\alpha,p}) \leq \alpha$ when $H_0 : F \leq_{\mathrm{US}} G$ is true. In other words, rejecting H_0 when $M^p_{mn} \geq c_{\alpha,p}$ is an asymptotic size α decision rule. Finally, we examine relevant asymptotic distributions when $R \in \Theta_1$ and then characterize large-sample power properties. We also discuss sample size calculations to detect departures from H_0. All theorems are proved in Section 7. Additional technical details are provided in the supplementary article (Tang et al., 2016).

3.1. *Asymptotic results under H_0.* Let $\mathcal{I}$ denote the identity operator on $l([0, 1])$ and define $\mathcal{D} = \mathcal{M} - \mathcal{I}$. When H_0 is true; i.e., when $R \in \Theta_0$, note that $\mathcal{M}R = R$ and

$$M^p_{mn} = c_{mn}\|\mathcal{M}R_{mn} - R_{mn}\|_p = c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p.$$

At first glance, establishing the limiting distribution of M^p_{mn} under H_0 might seem to be straightforward, that is, one could simply start with the asymptotic distribution of $c_{mn}(R_{mn} - R)$ described in Hsieh and Turnbull (1996) and apply the functional delta method (see, e.g., Section 3.9 in van der Vaart and Wellner, 1996) and continuous mapping theorem. This was the approach taken by Beare and Moon (2015) with their L^p distance-based GOF test statistics under likelihood ratio ordering. In our setting, this direct approach is not possible because whereas the least concave majorant operator in Beare and Moon (2015) is Hadamard directionally differentiable (Shapiro, 1990, 1991), the least star-shaped majorant operator $\mathcal{M}$ (and hence $\mathcal{D}$) is not always so; see Lemma 5 in the supplementary article (Tang et al., 2016). Fortunately, this does not create insurmountable problems because weak convergence of $c_{mn}(\mathcal{D}R_{mn} - \mathcal{D}R)$ is not a necessary prerequisite to derive the asymptotic distribution of $c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p$.

Before we state the asymptotic distribution of M^p_{mn} for any $R \in \Theta_0$, we need to describe R precisely because these distributions depend completely on the shape of R. Recall that when $R \in \Theta_0$, the slope function $r(u) = \{1 - R(u)\}/(1 - u)$ is nonincreasing in u. When $r(u)$ is strictly decreasing over $[0, 1]$, we say that R is *strictly* star-shaped. When $R \in \Theta_0$ is not strictly star-shaped, then, analogously to Beare and Moon (2015), there exists a unique collection (finite or countable) of closed, pairwise disjoint intervals of the form $[a_k, b_k]$, $0 \leq a_k < b_k \leq 1$, where

- the slope $r(u)$ is constant over each interval (i.e., R is affine over each interval)
- no two intervals possess the same value of $r(u)$.

In this case, we say that $R \in \Theta_0$ is *non-strictly* star-shaped. The reason we bifurcate Θ_0 using "strictly" and "non-strictly" descriptors is that the nondegenerate part of the asymptotic distribution of M_{mn}^p depends only on those regions where R is non-strictly star-shaped. If R is strictly star-shaped over $[0,1]$, the distribution of M_{mn}^p collapses to zero in the limit.

To make our description of the asymptotic distributions precise, we therefore introduce the following notation. For $0 \le a < b \le 1$, define

$$\mathcal{M}_{[a,b]}^{(1,0)} h = \inf\{h^* \in l([0,1]) : h \le h^* \text{ and } h^* \text{ is star-shaped over } [a,b] \text{ with kernel } (1,0)\}.$$

A general definition of what it means for a function h^* to be star-shaped with kernel (c,d) is given directly before Lemma 1 in the supplementary article (Tang et al., 2016). For any $h \in l([0,1])$, the function $\mathcal{M}_{[a,b]}^{(1,0)} h$ has two defining characteristics. First, $\mathcal{M}_{[a,b]}^{(1,0)} h(u) = h(u)$ whenever $u \notin [a,b]$. Second, over $[a,b]$, $\mathcal{M}_{[a,b]}^{(1,0)} h$ is the smallest function (at least as large as h) that is star-shaped with kernel $(1,0)$; i.e., the slope function $-\mathcal{M}_{[a,b]}^{(1,0)} h(u)/(1-u)$ over $[a,b]$ is nonincreasing in u. The importance of the functional operator $\mathcal{M}_{[a,b]}^{(1,0)} : l([0,1]) \mapsto l([0,1])$ becomes clear as we state our first main result.

Theorem 1 *Suppose $R \in \Theta_0$ and let $\mathcal{B}$ denote a standard Brownian bridge. The asymptotic results below hold when $\min\{m,n\} \to \infty$ and $n/(m+n) \to \lambda \in (0,1)$.*

(a) *If R is strictly star-shaped over $[0,1]$, then $M_{mn}^p \xrightarrow{d} 0$ for all $p \in [1,\infty]$.*

(b) *If R is non-strictly star-shaped, then for $p \in [1,\infty)$,*

$$M_{mn}^p \xrightarrow{d} \left\{ \sum_k \left[\lambda R'(a_k) + (1-\lambda)\{R'(a_k)\}^2\right]^{p/2} \int_{a_k}^{b_k} \left\{ \mathcal{D}_{[a_k,b_k]}^{(1,0)} \mathcal{B}(u) \right\}^p du \right\}^{1/p};$$

when $p = \infty$,

$$M_{mn}^p \xrightarrow{d} \sup_k \left\{ \left[\lambda R'(a_k) + (1-\lambda)\{R'(a_k)\}^2\right]^{1/2} \sup_{u \in [a_k,b_k]} \left\{ \mathcal{D}_{[a_k,b_k]}^{(1,0)} \mathcal{B}(u) \right\} \right\}.$$

In both asymptotic distributions, R' is the derivative of R and $\mathcal{D}_{[a_k,b_k]}^{(1,0)} = \mathcal{M}_{[a_k,b_k]}^{(1,0)} - \mathcal{I}$.

From Theorem 1, one can see that when $F \leq_{\mathrm{US}} G$, the only randomness in the asymptotic distribution of M_{mn}^p arises from the non-strictly star-shaped regions $[a_k, b_k]$ and is described probabilistically by the $\mathcal{D}_{[a_k,b_k]}^{(1,0)}\mathcal{B}$ processes. Furthermore, when $F = G$, the asymptotic distribution of M_{mn}^p simplifies to $\|\mathcal{D}_{[0,1]}^{(1,0)}\mathcal{B}\|_p$ for all $p \in [1, \infty]$. When $p = 1$, for example, this quantity describes the distribution of the area between the least star-shaped majorant of a standard Brownian bridge $\mathcal{B}$ and $\mathcal{B}$ itself. When $p = \infty$, $\|\mathcal{D}_{[0,1]}^{(1,0)}\mathcal{B}\|_\infty$ describes the distribution of the sup-norm distance between these two processes. Readers familiar with the GOF tests for likelihood ratio ordering in Carolan and Tebbs (2005) and Beare and Moon (2015) will no doubt recognize the homology between our Theorem 1 and the corresponding results in these articles. However, as noted earlier, GOF tests for uniform stochastic ordering present their own set of mathematical challenges and different conclusions are reached about the existence of a least favorable configuration.

Theorem 2 *Suppose $R \in \Theta_0$. For any $p \in [1, \infty]$, the asymptotic distribution of M_{mn}^p is ordinary stochastically smaller than $\|\mathcal{D}_{[0,1]}^{(1,0)}\mathcal{B}\|_p$; i.e.,*

$$\lim_{\substack{m,n\to\infty \\ n/(m+n)\to\lambda}} \mathrm{pr}_{R\in\Theta_0}(M_{mn}^p \geq t) \leq \mathrm{pr}(\|\mathcal{D}_{[0,1]}^{(1,0)}\mathcal{B}\|_p \geq t),$$

for all $t \in \mathbb{R}$, where λ is defined in Theorem 1.

Theorem 2 establishes that when using M_{mn}^p to test $H_0 : F \leq_{\mathrm{US}} G$ versus $H_1 : F \nleq_{\mathrm{US}} G$, the equal distribution line $R(u) = u$ represents the least favorable configuration of F and G for all $p \in [1, \infty]$. Proving this result involves showing that each of the $\mathcal{D}_{[a_k,b_k]}^{(1,0)}\mathcal{B}$ processes in Theorem 1 are mutually independent, a somewhat startling discovery because each process shares the same Brownian bridge $\mathcal{B}$ and each operator $\mathcal{D}_{[a_k,b_k]}^{(1,0)}$ shares the same kernel point $(1, 0)$. The practical utility of Theorem 2 is that, for any $p \in [1, \infty]$, we can determine the critical value that maximizes the probability of type I error over all configurations of F and G in Θ_0. This result is different than the conclusion reached in Beare and Moon (2015), who showed that when testing against likelihood ratio ordering using L^p distance-based statistics involving the least concave majorant of R_{mn}, $R(u) = u$ is the least favorable configuration when $p \in [1, 2]$ and for $p > 2$ the least favorable configuration does not exist. Careful inspection of Theorem 1 and some intuition sheds insight on why this is true. When R is star-shaped, but not strictly star-shaped, each of the derivatives $R'(a_k)$ in Theorem 1 satisfies $R'(a_k) \leq 1$. However, when $F \leq_{\mathrm{LR}} G$, there is no guarantee these derivatives are uniformly bounded for all concave R and hence anomalous limiting behavior can result when p is too large.

For given values of the significance level α and $p \in [1, \infty]$, denote the $1-\alpha$ quantile of $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$ by $c_{\alpha,p}$; i.e., $c_{\alpha,p}$ solves $\alpha = \mathrm{pr}(\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p \geq c_{\alpha,p})$. To approximate the distribution of $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$, we generated 100,000 Brownian bridge paths on a grid of 100,000 equally spaced points in $[0,1]$, and, for each $p \in \{1, 2, 3, 5, \infty\}$, we calculated $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$ for each path. For each p, these 100,000 values were used to approximate the density function of $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p$ and quantiles $c_{\alpha,p}$, for $\alpha = 0.01$, 0.05, and 0.10. These functions and the selected quantiles $c_{\alpha,p}$ are provided in the supplementary article (Tang et al., 2016).

3.2. *Asymptotic results under H_1.* The difference between the asymptotic distribution of M^p_{mn} under $H_0 : R \in \Theta_0$ and that under $H_1 : R \in \Theta_1$ arises from the non-star-shaped regions of R. To characterize a non-star-shaped ODC $R \in \Theta_1$, start with $\mathcal{M}R$, which is star-shaped, and note that (as in Section 3.1) one can partition the unit interval $[0,1]$ as $[0,1] = S \cup (\cup_k S_k)$, where $\mathcal{M}R$ is strictly star-shaped over S and non-strictly star-shaped over pairwise disjoint intervals of the form $S_k = [a_k, b_k]$, $0 \leq a_k < b_k \leq 1$, for $k = 1, 2, \ldots$. One can further partition each S_k as $S_k = S_{k1} \cup S_{k2}$, where $S_{k1} = \{u \in S_k : \mathcal{M}R(u) = R(u)\}$ and $S_{k2} = \{u \in S_k : \mathcal{M}R(u) > R(u)\}$. Each S_{k1} must contain a_k so it is never empty, and the non-star-shaped regions of R can be written as $\cup_k S_{k2}$. In other words, $R \in \Theta_0$ when $\cup_k S_{k2}$ is empty and $R \in \Theta_1$ otherwise.

In general, these types of regions contribute differently to the limiting distribution of M^p_{mn}. Over the strictly star-shaped region S, $\mathcal{M}R(u) = R(u)$ for all u and the L^p norm of $c_{mn}\{\mathcal{D}R_{mn}(u) - \mathcal{D}R(u)\}$ converges in distribution to 0, as in Section 3.1. To clearly describe the contribution over the S_k regions, we introduce new notation. For any $h \in l([0,1])$, define the functional operator $\mathcal{L}_{S_k} : l([0,1]) \mapsto l([0,1])$ according to

$$\mathcal{L}_{S_k} h(u) = -\inf_{\substack{v \in S_{k1} \\ v \leq u}} \left\{ \frac{-h(v)}{1-v} \right\} (1-u) I_{S_k}(u) + h(u) I_{S_k^c}(u), \quad \text{for } u \in [0,1),$$

where $I_A(\cdot)$ is the indicator function over the set A and A^c denotes the complement of A. When $u = 1$, $\mathcal{L}_{S_k} h(u) = \max\{h(1), 0\}$ or $h(1)$ depending on whether the singleton $\{1\} \in S_{k1}$ or not; see Appendix C in the supplementary article (Tang et al., 2016). Using this new operator, we now characterize asymptotic distributions for any ODC $R \in \Theta$ with those in $\Theta_1 = \Theta \setminus \Theta_0$ of particular interest. A discussion on the large-sample power properties of our testing procedure follows.

Theorem 3 *Suppose $R \in \Theta$. Using the notation described in this subsection,*

$$c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p \xrightarrow{d} \left\{ \sum_k \int_{u \in S_k} \left| \mathcal{L}_{S_k} T_R^\lambda(u) - T_R^\lambda(u) \right|^p du \right\}^{1/p}$$

for $p \in [1, \infty)$; when $p = \infty$,

$$c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p \xrightarrow{d} \sup_k \sup_{u \in S_k} \left| \mathcal{L}_{S_k} T_R^\lambda(u) - T_R^\lambda(u) \right|.$$

Both results hold as $\min\{m, n\} \to \infty$ and $n/(m+n) \to \lambda \in (0,1)$. In both cases, $T_R^\lambda(u) = \lambda^{1/2}\mathcal{B}_1(R(u)) + (1-\lambda)^{1/2} R'(u)\mathcal{B}_2(u)$, $0 \le u \le 1$, where $\mathcal{B}_1$ and $\mathcal{B}_2$ denote two independent standard Brownian bridges.

Four remarks are in order. First, the process $T_R^\lambda = \{T_R^\lambda(u), 0 \le u \le 1\}$ in Theorem 3 is well known; as noted earlier, it represents the asymptotic distribution of $c_{mn}(R_{mn} - R)$ for any $R \in \Theta$; see, e.g., Theorem 2.2 in Hsieh and Turnbull (1996). Second, the asymptotic distributions identified in Theorem 3 apply for any $R \in \Theta$, but we show in the supplementary article (Tang et al., 2016) that they quickly reduce to those in Theorem 1 when $R \in \Theta_0$. Third, our L^p tests are consistent for $p \in [1, \infty]$. To see why, consider the sup-norm ($p = \infty$) case in Theorem 3 and note that, by the triangle inequality,

$$\begin{aligned} \mathrm{pr}_{R \in \Theta_1}(M_{mn}^\infty \ge c_{\alpha,\infty}) &= \mathrm{pr}_{R \in \Theta_1}\big(c_{mn}\,\|\mathcal{D}R_{mn}\|_\infty \ge c_{\alpha,\infty}\big) \\ &\ge \ \mathrm{pr}_{R \in \Theta_1}\big(c_{mn}\,\|\mathcal{D}R_{mn} - \mathcal{D}R\|_\infty \le c_{mn}\,\|\mathcal{D}R\|_\infty - c_{\alpha,\infty}\big) \end{aligned}$$

which can be approximated by

$$\mathrm{pr}_{R \in \Theta_1}\Big(\sup_k \sup_{u \in S_k} \left| \mathcal{L}_{S_k} T_R^\lambda(u) - T_R^\lambda(u) \right| \le c_{mn}\,\|\mathcal{D}R\|_\infty - c_{\alpha,\infty}\Big).$$

It is easy to show that $\sup_k \sup_{u \in S_k} |\mathcal{L}_{S_k} T_R^\lambda(u) - T_R^\lambda(u)|$ is bounded and that, for any $R \in \Theta_1$, $c_{mn}\|\mathcal{D}R\|_\infty \to \infty$, as $\min\{m,n\} \to \infty$, which establishes our claim. The finite p argument is analogous. Fourth, approximate lower bounds on the power, like the one above in the sup-norm case, can be used for sample size calculations. For an ODC $R \in \Theta_1$ deemed to be clinically relevant, one can determine numerically the smallest m and n that solve $\mathrm{pr}_{R \in \Theta_1}(\sup_k \sup_{u \in S_k} |\mathcal{L}_{S_k} T_R^\lambda(u) - T_R^\lambda(u)| \le c_{mn}\|\mathcal{D}R\|_\infty - c_{\alpha,\infty}) = 1 - \beta$, where $\beta \in (0,1)$. The resulting solution will be inexorably conservative but still potentially useful for planning purposes. We illustrate this approach with examples in Section 4.

We conclude this section with a brief discussion on local power. This discussion is ultimately not dissimilar from the local power discussion in Beare and Moon (2015) under likelihood ratio ordering. However, our interest in local power arises because we want to compare the one-sample version of our testing procedure to the GOF test suggested by Arcones and Samaniego (2000). This one-sample comparison is given in Section 4.3. The two-sample discussion is given now. Let $\{R^{(r)}, r = 1, 2, ..., \}$ denote a sequence of ODCs in Θ_1. For each $r \geq 1$, denote the corresponding distributions by $F^{(r)}$ and $G^{(r)}$ from which we have independent random samples $X_1^{(r)}, X_2^{(r)}, ..., X_m^{(r)}$ and $Y_1^{(r)}, Y_2^{(r)}, ..., Y_n^{(r)}$, respectively. We examine local power properties by letting $R^{(r)}$ approach Θ_0 in the sense that $\|\mathcal{D}R^{(r)}\|_p = \|\mathcal{M}R^{(r)} - R^{(r)}\|_p \to 0$ as $r \to \infty$ at different rates. Using the notation in this paragraph, our last theorem summarizes the salient results.

Theorem 4 *Suppose the first derivative of $R^{(r)} \in \Theta_1$ is uniformly bounded over $[0, 1]$ for all r. Suppose $p \in [1, \infty]$. All limits stated below assume that $\max\{m, n\} = O(r)$ and $n/(m + n) \to \lambda \in (0, 1)$, as $r \to \infty$.*
(a) *If $\lim c_{mn}\|\mathcal{D}R^{(r)}\|_p = \infty$, then $\lim \mathrm{pr}_{R^{(r)} \in \Theta_1}(M_{mn}^p > c_{\alpha,p}) = 1$.*
(b) *For any $\beta \in (0, 1)$, there exists $\eta_p(\beta) > 0$ such that*

$$\liminf \mathrm{pr}_{R^{(r)} \in \Theta_1}(M_{mn}^p > c_{\alpha,p}) \geq 1 - \beta$$

whenever $\liminf c_{mn}\|\mathcal{D}R^{(r)}\|_p \geq \eta_p(\beta)$.

Part (a) of Theorem 4 indicates that when $\|\mathcal{D}R^{(r)}\|_p$ converges to 0 at a rate slower than c_{mn}^{-1}, $c_{mn}\|\mathcal{D}R^{(r)}\|_p$ diverges and the power of our test converges to 1. Part (b) guarantees that when $c_{mn}\|\mathcal{D}R^{(r)}\|_p$ remains bounded away from zero, the power of our test is still nontrivial; i.e., it does not converge to 0. This occurs when the "amount of information" c_{mn} increases and the "departure" $\|\mathcal{D}R^{(r)}\|_p$ decreases, and both do so at the same rate.

4. Simulation evidence.

We use simulation to assess the finite-sample performance of our tests. In Section 4.1, we consider fixed ODCs under both $H_0 : F \leq_{\mathrm{US}} G$ and $H_1 : F \nleq_{\mathrm{US}} G$ to estimate type I error probability and power, respectively, and we illustrate the sample size calculations described in Section 3.2. In Section 4.2, we modify our testing procedure to allow for one of the population distributions to be known and compare this modified test to the one-sample GOF test in Arcones and Samaniego (2000). Local power results are provided in Section 4.3.

4.1. *Fixed ODC comparisons.*

We consider four ODCs satisfying $R \in \Theta_0$ (R_1, R_2, R_3, and R_4) and four ODCs satisfying $R \in \Theta_1$ (R_5, R_6, R_7, and R_8).

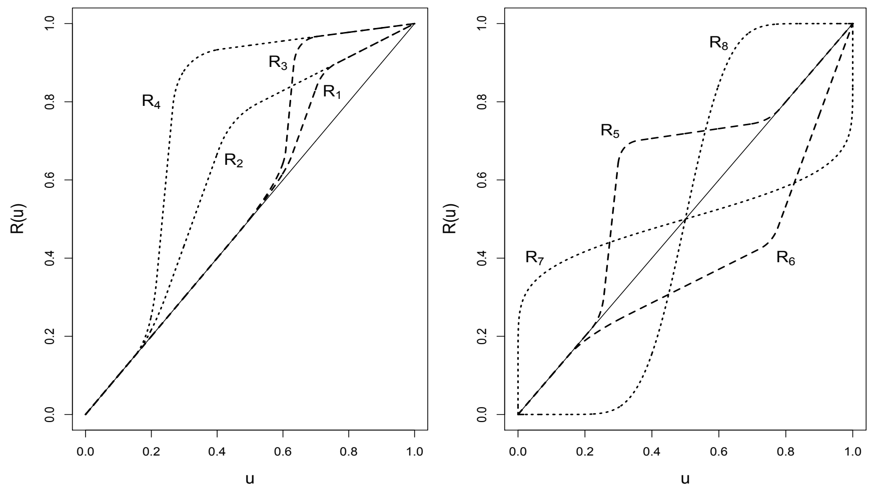

FIG 3. *Left: Star-shaped ODCs; i.e., $R_i \in \Theta_0$. Right: Non-star-shaped ODCs; i.e., $R_i \in \Theta_1$. A description of each curve is given in the supplementary article* (Tang et al., 2016).

The H_0 ODCs (Figure 3, left) are each members of a family of star-shaped ODCs that we describe in the supplementary article (Tang et al., 2016). The H_1 ODCs (Figure 3, right) are not star-shaped and are also described in Tang et al. (2016). We also consider $R_0 = R_0(u) = u$, for $u \in [0, 1]$, to examine finite-sample performance under the least favorable configuration $F = G$. All of our results are based on $10,000$ Monte Carlo data sets using independent samples from F and G with sample sizes m and n, respectively. To generate the samples, we let $F(u) = R_i(u)$ and $G(u) = u$, for $u \in [0, 1]$. We then sample $X_1, X_2, ..., X_m$ from F using the inverse cumulative distribution function technique and independently sample $Y_1, Y_2, ..., Y_n$ from a uniform$(0, 1)$ distribution. This provides independent samples for each ODC R under consideration.

Table S.2 in the supplementary article (Tang et al., 2016) gives Monte Carlo estimates of the probability of rejecting $H_0 : F \leq_{\mathrm{US}} G$ for different sample sizes, values of $p \in \{1, 2, \infty\}$, and $\alpha = 0.05$. We experimented with other values of p (i.e., $p = 3$ and $p = 5$) but obtained results similar to those when $p = 2$. Of initial interest is the finite-sample performance when $F = G$. With 10,000 simulated data sets, the margin of error associated with the size estimates under $F = G$, assuming a 99 percent confidence level, is approximately 0.006. Therefore, one notes that our tests with $p = 1$ and $p = 2$ are slightly anticonservative with small samples and otherwise operate closely to the nominal level. Furthermore, examining the rejection rates for the other star-shaped ODCs (R_1, R_2, R_3, and R_4) supports Theorem 2 which, for $p \in [1, \infty]$, guarantees the probability of type I error will be at its maximum under $F = G$. Likewise, powers for the non-star-shaped ODCs (R_5, R_6, R_7, and R_8) all approach unity as m and n become large. This reinforces our consistency claim.

We also use the non-star-shaped ODCs in Figure 3 to illustrate sample size determination. For $p \in [1, \infty]$ and for a given $R \in \Theta_1$, denote by $d_{R,\beta,p}$ the $1 - \beta$ quantile of the asymptotic distributions in Theorem 3. Using our lower bound on the asymptotic power from Section 3.2 and taking $m = n$ (for simplicity), we obtain a closed-form expression for the minimum sample size necessary to detect the departure $\|\mathcal{D}R\|_p = \|\mathcal{M}R - R\|_p$ with probability $1 - \beta$ when using an asymptotic size α test; i.e.,

$$m = 2\left(\frac{d_{R,\beta,p} + c_{\alpha,p}}{\|\mathcal{D}R\|_p}\right)^2, \quad \text{for } p \in [1, \infty].$$

With $\alpha = 0.05$ and $1 - \beta = 0.8$, the supplementary article (Tang et al., 2016) tables these solutions for each non-star-shaped ODC in Figure 3 and for each $p \in \{1, 2, \infty\}$. For example, for the R_5 ODC, which corresponds to F and G being stochastically ordered (but not uniformly stochastically ordered), the minimum sample size solutions for $p \in \{1, 2, \infty\}$, respectively, are $m = 634$, $m = 461$, and $m = 582$. Such sample sizes might seem dispiritingly large; however, it is not surprising these solutions are conservative. We describe in Section 6 alternative approaches that should reduce this conservatism.

4.2. *Comparison with Arcones and Samaniego* (2000). We now turn our attention to the special case of testing $H_0 : F \leq_{\mathrm{US}} G$ versus $H_1 : F \not\leq_{\mathrm{US}} G$ where G is known. Arcones and Samaniego (2000), who focused largely on optimal estimation of F (with $F \leq_{\mathrm{US}} G$ and G known), also suggested a conservative large-sample procedure to test against H_0. Their proposed test statistic, which we denote by D_m, can be expressed as a function of the one-sample ODC $R_m = F_m G^{-1}$; specifically,

$$D_m = m^{1/2} \sup_{0 \leq v \leq u \leq 1} [(1 - v)\{1 - R_m(u)\} - (1 - u)\{1 - R_m(v)\}].$$

However, instead of deriving a least favorable (asymptotic) distribution for inference, the authors proved that the asymptotic distribution of D_m is bounded above by $2 \sup_{u \in [0,1]} |\mathcal{B}(u)|$, where $\mathcal{B}$ is a standard Brownian bridge, and selected their critical value $c^{\mathrm{AS}}_{\alpha/2}$ to satisfy $\alpha = \mathrm{pr}(\sup_{u \in [0,1]} |\mathcal{B}(u)| \geq c^{\mathrm{AS}}_{\alpha/2})$. On the other hand, one-sample versions of our GOF procedure are available and use the test statistics

$$M^p_m = m^{1/2}\left[\int_{[0,1]} \{\mathcal{D}R_m(u)\}^p du\right]^{1/p} \quad \text{and} \quad M^\infty_m = m^{1/2} \sup_{u \in [0,1]} \{\mathcal{D}R_m(u)\},$$

where $\mathcal{D}$ is the operator defined in Section 3.1 and $R_m(u) = F_m\{G^{-1}(u)\}$. The limiting distributions in Theorem 1 also apply here as $m \to \infty$; in addition, it is straightforward to modify the proof of Theorem 2 to conclude that $F = G$ admits the least favorable configuration for $p \in [1, \infty]$ in the known G case.

For different sample sizes m (now corresponding to F only), Table S.3 in the supplementary article (Tang et al., 2016) gives small-sample rejection rates of our one-sample tests and the test from Arcones and Samaniego (2000), both performed using $\alpha = 0.05$. We used techniques similar to those described in Section 3.1 to approximate the critical value $c^{\mathrm{AS}}_{\alpha/2} = c^{\mathrm{AS}}_{0.025} = 1.359$ and performed all simulations in the same way as before except G is now known. Clearly, there is a price to be paid for using the test based on the D_m statistic when $F = G$; type I error probability estimates remain significantly below the nominal level for all $m \leq 200$. On the other hand, our $p = 1$ and $p = 2$ tests are only minimally conservative when $m \leq 75$, and our sup-norm ($p = \infty$) test performs nominally even when $m = 20$. In addition, the sup-norm test can be markedly more powerful at detecting non-star-shaped alternatives with small to moderately sized samples.

4.3. *Local power analysis.* A consequence of Theorem 3 is that, for any fixed $R \in \Theta_1$, our L^p GOF tests are consistent for all $p \in [1, \infty]$. To glean additional insight on which values of p might be preferred in practice, we investigate the power associated with local alternatives. Starting in the lower left corner, Figure 4 depicts a sequence of ODCs in Θ_1 that approach Θ_0 (moving from lower left to upper right). Each ODC shown in Figure 4 belongs to a family of ODCs described in the supplementary article (Tang et al., 2016); the defining feature of this family is that it is indexed by a single parameter $\delta \in [0, 0.5]$. The $\delta = 0$ member, say $R_{(0)}$, is the initial ODC in the lower left corner of Figure 4; the $\delta = 0.5$ member $R_{(0.5)}$, shown in the upper right, is the limiting ODC in Θ_0. ODCs $R_{(\delta)}$ with intermediate values of $\delta \in (0, 0.5)$ are also identified in Figure 4.

In our testing problem, a local power analysis involves examining a sequence of ODCs $\{R^{(r)}, r = 1, 2, ..., \}$ in Θ_1 that converges to Θ_0 at different rates. We do so here by using the family of ODCs just described. Specifically, we consider the rates $\zeta_r \in \{\log r, r^{2/5}, r^{1/2}\}$. For each ζ_r, we first choose a sequence of constants $\delta^{(r)}$ such that $\lim_{r\to\infty} \zeta_r |\delta^{(r)} - 0.5| = c_{\zeta_r} > 0$ and then select members from our ODC family identified by $R^{(r)} = R_{(\delta^{(r)})}$, for $r = 1, 2, ...$ The resulting sequence $R^{(r)}$ satisfies $\|\mathcal{D}R^{(r)}\|_p = \|\mathcal{M}R^{(r)} - R^{(r)}\|_p \to 0$ and $\zeta_r \|\mathcal{D}R^{(r)}\|_p \to c^*_{\zeta_r,p} > 0$, both as $r \to \infty$. This investigation allows us to learn more about the practical aspects of Theorem 4 (i.e., with

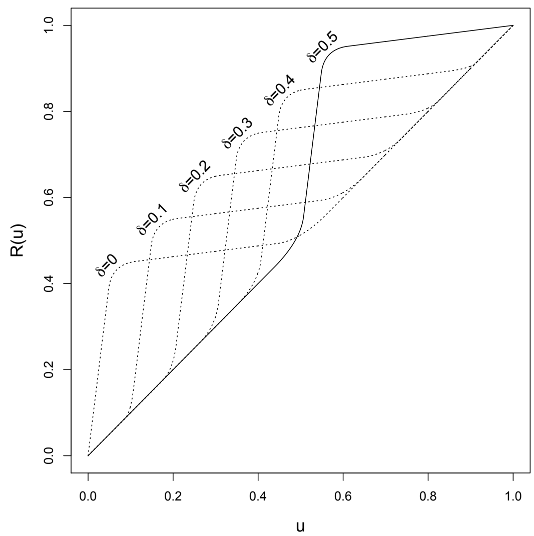

FIG 4. *Local power family of ODCs indexed by* $\delta \in [0, 0.5]$. *The* $\delta = 0$ *member* $R_{(0)}$ *is the initial ODC in* Θ_1*; the* $\delta = 0.5$ *member* $R_{(0.5)}$ *is the limiting ODC in* Θ_0. *This family is described in the supplementary article* (Tang et al., 2016).

both F and G unknown). We also use these ODC sequences, one for each rate ζ_r, to compare the one-sample versions of our tests with the test in Arcones and Samaniego (2000).

For each $r \in \{50, 100, 500, 1000, 5000, 10000\}$, we simulated 10,000 independent random samples, $X_1^{(r)}$, $X_2^{(r)}$, ..., $X_m^{(r)}$ from $F^{(r)}$ and $Y_1^{(r)}$, $Y_2^{(r)}$, ..., $Y_n^{(r)}$ from $G^{(r)}$, where $F^{(r)}(u) = R^{(r)}(u)$ and $G^{(r)}(u) = u$, $0 \leq u \leq 1$, and $m = n = r$. Figure 5 (top row) shows the estimated powers of our $\alpha = 0.05$ tests associated with each rate: $\zeta_r = \log r$ (left), $\zeta_r = r^{2/5}$ (middle), and $\zeta_r = r^{1/2}$ (right). Note that with $m = n = r$, considering the slower rates $\zeta_r = \log r$ and $\zeta_r = r^{2/5}$ allows us to assess part (a) of Theorem 4, while the fastest rate $\zeta_r = r^{1/2}$ allows us to assess part (b). Both parts are supported by our empirical results in Figure 5. For the slower rates, the powers approach unity as expected; however, we find that there is no decisively preferred value of p among $p \in \{1, 2, \infty\}$. On the other hand, when $\zeta_r = r^{1/2}$, the $p = 1$ powers hover only slightly above 0.3 for all r, while the $p = 2$ and $p = \infty$ powers still approach unity.

Switching to the one-sample problem, we find quite different results. For each rate ζ_r, Figure 5 (bottom row) displays the estimated powers of our one-sample $\alpha = 0.05$ tests which use M_m^1, M_m^2, and M_m^∞. Powers were estimated in the same way as for the two-sample case except now we treat $G^{(r)}(u) = u$

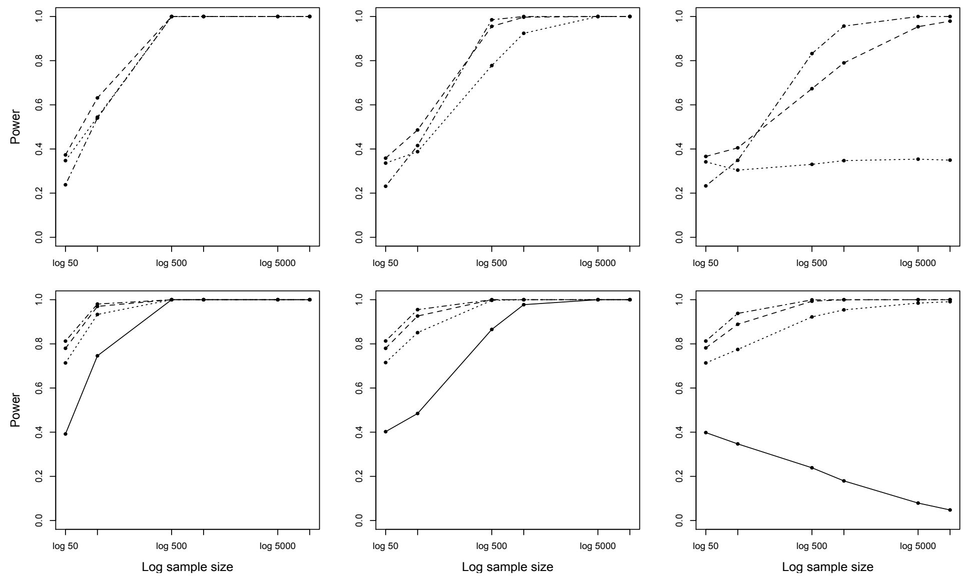

FIG 5. *Local power results with $\alpha = 0.05$. Left: $\zeta_r = \log r$. Middle: $\zeta_r = r^{2/5}$. Right: $\zeta_r = r^{1/2}$. Top: Two-sample case. Bottom: One-sample case. Our L^p results are shown dotted for $p = 1$, dashed for $p = 2$, and dot-dashed for $p = \infty$.* Arcones and Samaniego (2000) *results (one-sample case only) are shown using a solid line.*

as known and take $m = r$. In this setting, the sup-norm test consistently provides the largest power, followed by the $p = 2$ test and the $p = 1$ test. In addition, all three distance-based tests outperform the corresponding $\alpha = 0.05$ Arcones and Samaniego (2000) test in terms of local power, especially at the fastest rate $\zeta_r = r^{1/2}$ where $\mathrm{pr}_{R^{(r)} \in \Theta_1}(D_m > c^{\mathrm{AS}}_{0.025})$ appears to decrease towards zero.

5. Premature infant data.

Caffeine is commonly used to treat newborn infants for apnea of prematurity (Schmidt et al., 2006) and to prevent the onset of respiratory distress syndrome, bronchopulmonary dysplasia, and extubation failure (Cox et al., 2015). Known as "the silver bullet" in the treatment of prematurely born infants at risk for these and other acute conditions (Aranda et al., 2010), caffeine is widely regarded within the neonatal care community to be safe and cost effective. It has also been approved by the United States Food and Drug Administration for use with preterm infants due to its history of providing beneficial outcomes with no long-term adverse side effects (Dobson and Hunt, 2013).

We now analyze the data from the study described in Section 1; for complete details, see Cox et al. (2015). Because assessing the use of caffeine with premature infants was a central focus of this study, we consider only those infants who were classified as "premature;" i.e., newborns whose gestational age was at or below 37 weeks. With F and G denoting the discharge time distributions for the caffeine and no-caffeine groups, respectively, recall that Figure 2 displays the sample ODC R_{mn} and its least star-shaped majorant $\mathcal{M}R_{mn}$, calculated from samples of size $m = 127$ from F and $n = 277$ from G. As noted in Section 1, we performed the test in Davidov and Herman (2012) with these data and concluded that $F \leq_{\mathrm{S}} G$ was strongly supported over $F = G$. We also performed the GOF tests in Beare and Moon (2015) and concluded that $F \leq_{\mathrm{LR}} G$ would be rejected at $\alpha = 0.05$; the L^1 and L^2 statistics based on the least concave majorant of R_{mn} are 0.717 and 0.999, respectively, which are larger than the corresponding 0.95 quantiles 0.664 and 0.753 identified by their least favorable distributions.

We therefore assess whether or not the data in Figure 2 are consistent with uniform stochastic ordering. Testing $H_0 : F \leq_{\mathrm{US}} G$ versus $H_1 : F \nleq_{\mathrm{US}} G$ based on the least star-shaped majorant of R_{mn}, our GOF test statistics are $M^1_{mn} = 0.170$, $M^2_{mn} = 0.263$, and $M^\infty_{mn} = 0.949$, each of which is well below the $\alpha = 0.10$ critical values identified in the supplementary article (0.496, 0.586, and 1.219, respectively), that is, H_0 cannot be discounted at any reasonable level of significance. Therefore, not only does caffeine therapy provide point-of-care health benefits and improved long-term outcomes for prematurely born infants, our analysis suggests that treating these infants with caffeine may also lead to hospital discharge times that are uniformly stochastically smaller than those for infants not treated with caffeine.

6. Concluding remarks.

When two distributions F and G satisfy uniform stochastic ordering, F and G when conditioned on the interval $[t_0, \infty)$, for any $t_0 \in \mathbb{R}$, also satisfy uniform stochastic ordering. This desirable property could be exploited to increase the power of our tests under H_1 and simultaneously reduce the sample sizes necessary to detect departures from H_0. To see how, suppose that uniform stochastic ordering is suspected to be violated when $t > t_0$, either from historical information or from observing data in related applications. In this situation, one could apply our tests after conditioning to determine if R is non-star-shaped over the smaller region $[G^{-1}(t_0), 1]$ and calculate sample sizes to detect departures over it instead of over $[0, 1]$. A similar approach was suggested by Carolan and Tebbs (2005) for detecting departures from likelihood ratio ordering. In the same spirit, Beare and Moon (2015) suggest that bootstrapping samples over departure regions could help to increase the power of GOF tests for likelihood ratio ordering. This strategy may also be fruitful in our setting, allowing one to reduce the conservatism arising from relying on the least favorable distribution over the entire unit interval.

We believe that our GOF tests could be generalized to allow for different types of censored data, but the theory underpinning these extensions would not be trivial. For example, with random right-censored data, there would be nothing to prevent one from simply replacing the empirical survival functions $\overline{F}_m$ and $\overline{G}_n$ with Kaplan-Meier estimators of $\overline{F}$ and $\overline{G}$ and then calculating R_{mn} and $\mathcal{M}R_{mn}$ using these estimates. However, asymptotic distributions of the corresponding test statistics may depend heavily on the latent censoring distributions, and there is no guarantee that the least favorable configuration of F and G will exist. Future work could investigate censored-data extensions of majorant-based inference−not only with uniform stochastic ordering, but with other orderings as well.

Finally, estimating distributions under a uniform stochastic ordering assumption has received considerable attention for two populations; see, e.g., Rojo and Samaniego (1993), Mukerjee (1996), and Arcones and Samaniego (2000). We view the one- and two-sample tests proposed herein as helpful inference procedures to determine if the uniform stochastic ordering assumption is plausible and hence restricted estimation methods for F and G are warranted. An anonymous referee has suggested that developing pointwise confidence intervals for $R(u)$ under a uniform stochastic ordering constraint may be a worthwhile next step. We agree and comment on this further after Lemma 4 in the supplementary article (Tang et al., 2016). Another interesting avenue for future research would be to generalize our majorant-based tests to more than two populations. Estimation techniques in this setting are available in Dykstra et al. (1991) and El Barmi and Mukerjee (2016).

7. Proofs. In this section, we provide the proofs of Theorems 1-4. Lemmas cited in this section are stated and proved in the supplementary article (Tang et al., 2016), henceforth referred to as "the supplementary article."

Proof of Theorem 1. We start with the asymptotic distribution of R_{mn}, suitably centered and scaled. Applying Theorem 2.2 in Hsieh and Turnbull (1996), it follows that $c_{mn}(R_{mn}-R)$ converges weakly to T_R^λ as $\min\{m,n\} \to \infty$ and $n/(m+n) \to \lambda \in (0,1)$, where T_R^λ satisfies $T_R^\lambda(u) = \lambda^{1/2}\mathcal{B}_1(R(u)) + (1-\lambda)^{1/2}R'(u)\mathcal{B}_2(u)$, for $0 \le u \le 1$, and $\mathcal{B}_1$ and $\mathcal{B}_2$ are independent standard Brownian bridges. When $R \in \Theta_0$, $\mathcal{D}R = 0$ and $M_{mn}^p = c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p$.

Define the functional operator $d\mathcal{D}_R : l([0,1]) \mapsto l([0,1])$ by

$$d\mathcal{D}_R h(u) = \begin{cases} \max\{h(1), 0\} - h(1), & \text{if } u = 1 \\ \mathcal{M}^{(1,0)}_{[a_k,b_k]} h(u) - h(u), & \text{if } \exists k \text{ such that } a_k \le u \le b_k \\ 0, & \text{otherwise,} \end{cases}$$

for $h \in l([0,1])$. Denote by $C([0,1])$ the collection of all real continuous functions with domain $[0,1]$. If $\mathcal{D}$ is Hadamard directionally differentiable tangentially to $C([0,1])$ at R, then $d\mathcal{D}_R$ is the Hadamard directional derivative of $\mathcal{D}$. Applying the functional delta method and continuous mapping theorem yields $M^p_{mn} \xrightarrow{d} \|d\mathcal{D}_R T^\lambda_R\|_p$ for $p \in [1,\infty]$. Those situations in which $\mathcal{D}$ is Hadamard directionally differentiable are described in Lemma 5 in the supplementary article.

When $\mathcal{D}$ is not Hadamard directionally differentiable, the functional delta method and continuous mapping theorem cannot be applied. However, by using Lemma 6 in the supplementary article, we are able to prove that $M^p_{mn} \xrightarrow{d} \|d\mathcal{D}_R T^\lambda_R\|_p$ anyway. For convenience, let $Z_{mn} = c_{mn}(R_{mn} - R)$ and $Z = T^\lambda_R$. From Theorem 12.2 in Billingsley (1999) and Skorohod's representation theorem (see, e.g., Theorem 6.7 in Billingsley, 1999), there exist random elements Z'_{mn} and Z' defined on a common probability space with $Z'_{mn} \overset{L}{=} Z_{mn}$ and $Z' \overset{L}{=} Z$ such that $\|Z'_{mn} - Z'\|_\infty \to 0$ almost surely. The notation "$\overset{L}{=}$" denotes that two processes are equivalent in distribution. Define $R'_{mn} = c^{-1}_{mn} Z'_{mn} + R$. From Lemma 6 in the supplementary article, because c^{-1}_{mn} decreases to 0 and $\|Z'_{mn} - Z'\|_\infty \to 0$ almost surely, then for all $p \in [1,\infty]$ we have

$$\lim_{\substack{m,n\to\infty \\ n/(m+n)\to\lambda}} c_{mn} \|\mathcal{D}R'_{mn} - \mathcal{D}R\|_p = \|d\mathcal{D}_R Z'\|_p$$

almost surely. Because $c_{mn}\|\mathcal{D}R'_{mn} - \mathcal{D}R\|_p \overset{d}{=} c_{mn}\|\mathcal{D}R_{mn} - \mathcal{D}R\|_p$ and also $\|d\mathcal{D}_R Z'\|_p \overset{d}{=} \|d\mathcal{D}_R T^\lambda_R\|_p$, where the notation "$\overset{d}{=}$" means equal in distribution, we have

$$\lim_{\substack{m,n\to\infty \\ n/(m+n)\to\lambda}} c_{mn} \|\mathcal{D}R_{mn} - \mathcal{D}R\|_p \overset{d}{=} \|d\mathcal{D}_R T^\lambda_R\|_p.$$

This shows that $M^p_{mn} \xrightarrow{d} \|d\mathcal{D}_R T^\lambda_R\|_p$ for all $p \in [1,\infty]$.

When R is strictly star-shaped over $[0,1]$, it is easy to see that $\|d\mathcal{D}_R T^\lambda_R\|_p = 0$ which quickly establishes part (a). The remainder of the proof focuses on establishing part (b). When R is non-strictly star-shaped,

$$\|d\mathcal{D}_R T_R^\lambda\|_p = \left[\sum_k \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]} T_R^\lambda(u)\}^p du\right]^{1/p}$$

for $p \in [1,\infty)$ and $\|d\mathcal{D}_R T_R^\lambda\|_p = \sup_k\{\sup_{u\in[a_k,b_k]} \mathcal{D}^{(1,0)}_{[a_k,b_k]} T_R^\lambda(u)\}$ for $p = \infty$. Using Lemma 1 in the supplementary article, we write $\mathcal{D}^{(1,0)}_{[a_k,b_k]} T_R^\lambda(u) = \sup_{v\in[a_k,u]} Q_k(u,v)$ for $u \in [a_k,b_k]$, where

$$Q_k(u,v) = \left(\frac{1-u}{1-v}\right) T_R^\lambda(v) - T_R^\lambda(u), \quad \text{for } v \in [a_k,u].$$

In Lemma 8 in the supplementary article, we show that the processes

$$\{Q_k(u,v),\ a_k \le v \le u < b_k\}$$

are mutually independent across k. Therefore, $\{\mathcal{D}^{(1,0)}_{[a_k,b_k]} T_R^\lambda(u), u \in [a_k,b_k]\}$ are also mutually independent. To prove further results, we note that over each non-strictly star-shaped region $[a_k,b_k]$, we can write $R(u)$ as a linear function; i.e., $R(u) = 1 - R'(a_k)(1-u)$. Thus, from Lemma 2 in the supplementary article, we have

$$\mathcal{D}^{(1,0)}_{[a_k,b_k]} T_R^\lambda(u) = \mathcal{D}^{(1,0)}_{[a_k,b_k]}\{W_R^\lambda(u) - l_{R,k}^\lambda(1)\},$$

for all k, where $W_R^\lambda(u) = \lambda^{1/2}\mathcal{W}_1(R(u)) + (1-\lambda)^{1/2}R'(u)\mathcal{W}_2(u)$,

$$l_{R,k}^\lambda(u) = \lambda^{1/2}\{1 - R'(a_k)(1-u)\mathcal{W}_1(1) + (1-\lambda)^{1/2}R'(a_k)u\mathcal{W}_2(1)\},$$

and $\mathcal{W}_1$ and $\mathcal{W}_2$ are independent standard Wiener processes; i.e., $\mathcal{W}_i$, for $i = 1,2$, satisfies $\mathcal{B}_i(u) = \mathcal{W}_i(u) - u\mathcal{W}_i(1)$, $0 \le u \le 1$, for $i = 1,2$. Based on the properties of a standard Wiener process, it follows that for $u \in [a_k,b_k]$,

$$\begin{aligned}\mathcal{W}_i(R(u)) - \mathcal{W}_i(1) &= \mathcal{W}_i(1 - R'(a_k)(1-u)) - \mathcal{W}_i(1)\\ &\overset{L}{=} R'(a_k)^{1/2}\{\mathcal{W}_i(u) - \mathcal{W}_1(1)\},\end{aligned}$$

for $i = 1,2$. Furthermore, for $u \in [a_k,b_k]$, we have $R'(u) = R'(a_k)$ and

$$\begin{aligned}W_R^\lambda(u) - l_{R,k}^\lambda(1) &\overset{L}{=} \lambda^{1/2}R'(a_k)^{1/2}\{\mathcal{W}_1(u) - \mathcal{W}_1(1)\}\\ &\qquad + (1-\lambda)^{1/2}R'(a_k)\{\mathcal{W}_2(u) - \mathcal{W}_2(1)\}\\ &\overset{L}{=} \{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{1/2}\{\mathcal{W}(u) - \mathcal{W}(1)\},\end{aligned}$$

where $\mathcal{W}$ is a standard Wiener process. The last equivalence (in distribution) follows because both right-hand side processes above are Gaussian, they have the same mean $E\{W_R^\lambda(u) - l_{R,k}^\lambda(1)\} = 0$, for $u \in [a_k, b_k]$, and they have the same covariance $\operatorname{cov}\{W_R^\lambda(u_1) - l_{R,k}^\lambda(1), W_R^\lambda(u_2) - l_{R,k}^\lambda(1)\} = \{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}\min\{1-u_1, 1-u_2\}$, for $u_1, u_2 \in [a_k, b_k]$. Using Lemma 2 in the supplementary article again, we have

$$\begin{aligned}\mathcal{D}_{[a_k,b_k]}^{(1,0)}&\{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{1/2}\{\mathcal{W}(u) - \mathcal{W}(1)\}\\ &= \{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{1/2}\mathcal{D}_{[a_k,b_k]}^{(1,0)}\mathcal{B}(u),\end{aligned}$$

where $\mathcal{B}$ is a standard Brownian bridge formed by $\mathcal{W}$; i.e., $\mathcal{B}(u) = \mathcal{W}(u) - u\mathcal{W}(1)$, for $u \in [0,1]$. We can therefore write

$$\begin{aligned}\int_{a_k}^{b_k}&\{\mathcal{D}_{[a_k,b_k]}^{(1,0)}T_R^\lambda(u)\}^p du\\ &\stackrel{d}{=} \{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{p/2}\int_{a_k}^{b_k}\{\mathcal{D}_{[a_k,b_k]}^{(1,0)}\mathcal{B}(u)\}^p du,\end{aligned}$$

for $p \in [1,\infty)$, and

$$\sup_{u\in[a_k,b_k]}\mathcal{D}_{[a_k,b_k]}^{(1,0)}T_R^\lambda(u) \stackrel{d}{=} \{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{1/2}\sup_{u\in[a_k,b_k]}\mathcal{D}_{[a_k,b_k]}^{(1,0)}\mathcal{B}(u),$$

for $p = \infty$. For $p \in [1,\infty)$, we have shown that $\int_{a_k}^{b_k}\{\mathcal{D}_{[a_k,b_k]}^{(1,0)}T_R^\lambda(u)\}^p du$ are mutually independent. One can show that $\int_{a_k}^{b_k}\{\mathcal{D}_{[a_k,b_k]}^{(1,0)}\mathcal{B}(u)\}^p du$ are also mutually independent by replacing $T_R^\lambda(\cdot)$ with $\mathcal{B}(\cdot)$ in the definition of $Q_k(u,v)$ and repeating the same argument. Therefore, we have

$$\begin{aligned}\sum_k&\int_{a_k}^{b_k}\{\mathcal{D}_{[a_k,b_k]}^{(1,0)}T_R^\lambda(u)\}^p du\\ &\stackrel{d}{=} \sum_k\{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{p/2}\int_{a_k}^{b_k}\{\mathcal{D}_{[a_k,b_k]}^{(1,0)}\mathcal{B}(u)\}^p du,\end{aligned}$$

which completes the proof for $p \in [1,\infty)$. Completing the proof for the $p = \infty$ case is analogous. □

Proof of Theorem 2. When $F = G$, the ODC is $R_0 = R_0(u) = u$, $0 \le u \le 1$, and $T_{R_0}^\lambda \stackrel{d}{=} \mathcal{B}$. Because R_0 is non-strictly star-shaped over $[0,1]$, Theorem 1 yields $M_{mn}^p \xrightarrow{d} \|d\mathcal{D}_{R_0}T_{R_0}^\lambda\|_p \stackrel{d}{=} \|\mathcal{D}_{[0,1]}^{(1,0)}\mathcal{B}\|_p$ when $F = G$ for

$p \in [1, \infty]$. It therefore suffices to show $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p \geq_{\mathrm{S}} \|d\mathcal{D}_R T^\lambda_R\|_p$ for $p \in [1, \infty]$ and for any other $R \in \Theta_0$. If $R \in \Theta_0$ is strictly star-shaped, then from Theorem 1, $\|d\mathcal{D}_R T^\lambda_R\|_p = 0$ for $p \in [1, \infty]$ and hence $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p \geq_{\mathrm{S}} \|d\mathcal{D}_R T^\lambda_R\|_p$. If $R \in \Theta_0$ is non-strictly star-shaped, then for $p \in [1, \infty)$,

$$\begin{aligned}\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p = \left[\int_0^1 \{\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}(u)\}^p \, du\right]^{1/p} &\geq \left[\sum_k \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}(u)\}^p \, du\right]^{1/p} \\ &\geq \left[\sum_k \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]}\mathcal{B}(u)\}^p \, du\right]^{1/p}. \qquad (7.1)\end{aligned}$$

The first and second inequalities above hold because $\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}(u) \geq 0$ and also $\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}(u) \geq \mathcal{D}^{(1,0)}_{[a_k,b_k]}\mathcal{B}(u) \geq 0$, for all $u \in [0, 1]$. Because $\lambda \in (0, 1)$ and $R'(a_k) \leq 1$ for all k, $\lambda R'(a_k) + (1 - \lambda)R'(a_k)^2 \leq 1$ and the rightmost side of (7.1) is greater than or equal to

$$\left[\sum_k \{\lambda R'(a_k) + (1-\lambda)R'(a_k)^2\}^{p/2} \int_{a_k}^{b_k} \{\mathcal{D}^{(1,0)}_{[a_k,b_k]}\mathcal{B}(u)\}^p du\right]^{1/p} \stackrel{d}{=} \|d\mathcal{D}_R T^\lambda_R\|_p.$$

Therefore, for $R \in \Theta_0$ non-strictly star-shaped, we have $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_p \geq_{\mathrm{S}} \|d\mathcal{D}_R T^\lambda_R\|_p$ for $p \in [1, \infty)$. Showing $\|\mathcal{D}^{(1,0)}_{[0,1]}\mathcal{B}\|_\infty \geq_{\mathrm{S}} \|d\mathcal{D}_R T^\lambda_R\|_\infty$ for $R \in \Theta_0$ non-strictly star-shaped is analogous. □

Proof of Theorem 3. When $R \in \Theta_1$, we redefine the functional operator $d\mathcal{D}_R : l([0, 1]) \mapsto l([0, 1])$ in Theorem 1 by

$$d\mathcal{D}_R h(u) = \begin{cases} -h(1), & \text{if } u = 1,\ R(u) < 1 \\ \max\{h(1), 0\} - h(1), & \text{if } u = 1,\ R(u) = 1 \\ \mathcal{L}_{S_k} h(u) - h(u), & \text{if } \exists k \text{ such that } u \in S_k \setminus \{1\} \\ 0, & \text{otherwise.} \end{cases}$$

The proof proceeds in the same manner as in Theorem 1. If $\mathcal{D}$ is not Hadamard directionally differentiable, one can use Skorohod's representation theorem and part (b) of Lemma 7 in the supplementary article to obtain the result. □

Proof of Theorem 4. For convenience, all limits stated in this proof assume that $\max\{m, n\} = O(r)$ and $n/(m + n) \to \lambda \in (0, 1)$, as $r \to \infty$. We have independent random samples $X^{(r)}_1, X^{(r)}_2, ..., X^{(r)}_m$ and $Y^{(r)}_1, Y^{(r)}_2, ..., Y^{(r)}_n$ from $F^{(r)}$ and $G^{(r)}$, respectively. The sample ODC is $R^{(r)}_{mn} = F^{(r)}_m(G^{(r)}_n)^{-1}$, where $F_m^{(r)}$ and $(G_n^{(r)})^{-1}$ are the empirical distribution and empirical quantile functions, respectively. Our test statistic is $M_{mn}^p = c_{mn}\|\mathcal{D}R_{mn}^{(r)}\|_p$. By the triangle inequality,

$$\begin{aligned}&\mathrm{pr}_{R^{(r)}\in\Theta_1}(M_{mn}^p \geq c_{\alpha,p})\\ &\qquad\geq \mathrm{pr}_{R^{(r)}\in\Theta_1}(c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p < c_{mn}\|\mathcal{D}R^{(r)}\|_p - c_{\alpha,p})\end{aligned}$$

for all $p \in [1,\infty]$. Therefore, to prove part (a), it suffices to show that $c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p = O_P(1)$.

From Lemma 3 in the supplementary article, it follows that $\|\mathcal{M}R_{mn}^{(r)} - \mathcal{M}R^{(r)}\|_\infty \leq \|R_{mn}^{(r)} - R^{(r)}\|_\infty$, which implies $\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_\infty \leq 2\|R_{mn}^{(r)} - R^{(r)}\|_\infty$. Because L^p norms are dominated by the sup-norm, it therefore suffices to show $c_{mn}\|R_{mn}^{(r)} - R^{(r)}\|_\infty$ is $O_P(1)$. To accomplish this, we decompose $c_{mn}(R_{mn}^{(r)} - R^{(r)})$ into two parts:

$$\begin{aligned}c_{mn}(R_{mn}^{(r)} - R^{(r)}) &= c_{mn}\{F_m^{(r)}(G_n^{(r)})^{-1} - F^{(r)}(G_n^{(r)})^{-1}\}\\ &\quad + c_{mn}\{F^{(r)}(G_n^{(r)})^{-1} - F^{(r)}(G^{(r)})^{-1}\}.\end{aligned} \tag{7.2}$$

Define the two independent empirical processes

$$U_m(u) = \frac{1}{m}\sum_{i=1}^m I\{F^{(r)}(X_i^{(r)}) \leq u\}$$

and $V_n(u) = n^{-1}\sum_{i=1}^n I\{G^{(r)}(Y_i^{(r)}) \leq u\}$, for $0 \leq u \leq 1$. This allows us to rewrite $F_m^{(r)}$ as $U_m F^{(r)}$ and $F^{(r)}(G_n^{(r)})^{-1}$ as $R^{(r)}V_n^{(r)}$. Consequently, the two terms on the right-hand side of Equation (7.2) can be written as

$$\begin{aligned}&c_{mn}\{F_m^{(r)}(G_n^{(r)})^{-1} - F^{(r)}(G_n^{(r)})^{-1}\}\\ &\qquad = c_{mn}[U_m\{F^{(r)}(G_n^{(r)})^{-1}\} - U\{F^{(r)}(G_n^{(r)})^{-1}\}]\end{aligned} \tag{7.3}$$

and

$$c_{mn}\{F^{(r)}(G_n^{(r)})^{-1} - F^{(r)}(G^{(r)})^{-1}\} = c_{mn}(R^{(r)}V_n - R^{(r)}V), \tag{7.4}$$

where $U(\cdot)$ and $V(\cdot)$ both represent the cumulative distribution function of a uniform distribution on $[0,1]$. These expressions allow us to unify all random samples (from different distributions) to be uniformly distributed.

We are now ready to show that the sup-norms of the right-hand sides of Equations (7.3) and (7.4) are uniformly bounded in probability. We begin with the uniform processes. From Theorem 3 in Komlós et al. (1975), there exist versions of independent standard Brownian bridges $\mathcal{B}_1^{(m)}$ and $\mathcal{B}_2^{(n)}$ such that, almost surely,

$$\|\sqrt{m}(U_m - U) - \mathcal{B}_1^{(m)}\|_\infty = o(m^{-1/2}(\log m)^2) \tag{7.5}$$

$$\|\sqrt{n}(V_n - V) - \mathcal{B}_2^{(n)}\|_\infty = o(n^{-1/2}(\log n)^2). \tag{7.6}$$

Because $\lim c_{mn}/(\lambda^{1/2}\sqrt{m}) = 1$, we have $\|c_{mn}(U_m - U) - \lambda^{1/2}\mathcal{B}_1^{(m)}\|_\infty = o(m^{-1/2}(\log m)^2)$ from Equation (7.5). Consequently, the sup-norm of the right-hand side of Equation (7.3) is less than or equal to

$$\|c_{mn}[U_m\{F^{(r)}(G_n^{(r)})^{-1}\} - U\{F^{(r)}(G_n^{(r)})^{-1}\}] - \lambda^{1/2}\mathcal{B}_1^{(m)}\{F^{(r)}(G_n^{(r)})^{-1}\}\|_\infty + \|\lambda^{1/2}\mathcal{B}_1^{(m)}\{F^{(r)}(G_n^{(r)})^{-1}\}\|_\infty$$

which is less than or equal to

$$\|c_{mn}(U_m - U) - \lambda^{1/2}\mathcal{B}_1^{(m)}\|_\infty + \|\lambda^{1/2}\mathcal{B}_1^{(m)}\|_\infty = o(m^{-1/2}(\log m)^2) + O_P(1).$$

The $O_P(1)$ term arises because $\mathcal{B}_1^{(m)}$ is bounded with probability 1. Likewise, the $o(m^{-1/2}(\log m)^2)$ term comes from Equation (7.5). Therefore, we have shown that the sup-norm of the right-hand side of Equation (7.3), that is, $\|c_{mn}\{F_m^{(r)}(G_n^{(r)})^{-1} - F^{(r)}(G_n^{(r)})^{-1}\}\|_\infty = O_P(1)$.

For the right-hand side of Equation (7.4), we use the mean value theorem to write

$$R^{(r)}V_n(u) - R^{(r)}V(u) = \dot{R}^{(r)}(\tau_u)\{V_n(u) - V(u)\},$$

where $\dot{R}^{(r)}$ denotes the derivative of $R^{(r)}$ and where τ_u is between $V_n(u)$ and $V(u)$. Therefore,

$$\sup_{u\in[0,1]} \left|\sqrt{n}\{R^{(r)}V_n(u) - R^{(r)}V(u)\} - \dot{R}^{(r)}(\tau_u)\mathcal{B}_2^{(n)}(u)\right| = \sup_{u\in[0,1]} \left|\dot{R}^{(r)}(\tau_u)[\sqrt{n}\{V_n(u) - V(u)\} - \mathcal{B}_2^{(n)}(u)]\right|$$

which is less than or equal to

$$\|\dot{R}^{(r)}\|_\infty \|\sqrt{n}(V_n - V) - \mathcal{B}_2^{(n)}\|_\infty = O(1)o(n^{-1/2}(\log n)^2) = o(n^{-1/2}(\log n)^2).$$

The $O(1)$ term above comes from the assumption that the derivative of $R^{(r)}$ is uniformly bounded for all r over $[0,1]$. Likewise, the $o(n^{-1/2}(\log n)^2)$ term comes from Equation (7.6). Therefore, because $\lim c_{mn}/\{(1-\lambda)^{1/2}\sqrt{n}\} = 1$

and because $\mathcal{B}_2^{(n)}$ is bounded with probability 1, we have shown the sup-norm of the right-hand side of Equation (7.4), that is, $c_{mn}\|R^{(r)}V_n^{(r)} - R^{(r)}V\|_\infty = O_P(1)$. Finally, from Equation (7.2), we have $c_{mn}\|R_{mn}^{(r)} - R^{(r)}\|_\infty = O_P(1) + O_P(1) = O_P(1)$, which establishes part (a).

To prove part (b), let $q_{\beta,p}^{(r)}$ denote the $1-\beta$ quantile of the finite-sample distribution of $c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p$; i.e., $q_{\beta,p}^{(r)}$ solves $\text{pr}_{R^{(r)}\in\Theta_1}(c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p \le q_{\beta,p}^{(r)}) = 1-\beta$. We have already shown $c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p = O_P(1)$, so $\sup_r q_{\beta,p}^{(r)} \equiv q_{\beta,p} < \infty$. Therefore,

$$\liminf \text{pr}_{R^{(r)}\in\Theta_1}(c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p < q_{\beta,p}) \ge 1-\beta.$$

Set $\eta_p(\beta) = q_{\beta,p} + c_{\alpha,p}$. Whenever $\liminf c_{mn}\|\mathcal{D}R\|_p \ge \eta_p(\beta)$, it follows from the triangle inequality that $\liminf \text{pr}_{R^{(r)}\in\Theta_1}(M_{mn}^p \ge c_{\alpha,p})$ is greater than or equal to

$$\begin{aligned}\liminf \text{pr}_{R^{(r)}\in\Theta_1}&(c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p < c_{mn}\|\mathcal{D}R^{(r)}\|_p - c_{\alpha,p}) \\ &\ge \liminf \text{pr}_{R^{(r)}\in\Theta_1}(c_{mn}\|\mathcal{D}R_{mn}^{(r)} - \mathcal{D}R^{(r)}\|_p < q_{\beta,p}) \ge 1-\beta.\end{aligned}$$

This completes the proof of part (b). □

Acknowledgements. We thank the Associate Editor and two anonymous referees for their helpful comments on an earlier version of this article. We also thank Dr. Christina Cox for her collaboration on the premature infant study.

SUPPLEMENTARY MATERIAL

Supplement to "Nonparametric goodness-of-fit tests for uniform stochastic ordering" (DOI: COMPLETED BY THE TYPESETTER; .pdf). In the supplementary article (Tang et al., 2016), we state and prove lemmas that are cited in this manuscript. These lemmas describe theoretical properties of the least star-shaped majorant operator, including Hadamard directional differentiability. We also provide the estimated densities of $\|\mathcal{D}_{[0,1]}^{(1,0)}\mathcal{B}\|_p$ and critical values $c_{\alpha,p}$ for our tests. Finally, we describe the families of ODCs used in Section 4 and give finite-sample simulation results and sample size calculations.

References.

Aranda, J., Beharry, K., Valencia, G., Natarajan, G., and Davis, J. (2010). Caffeine impact on neonatal morbidities. *Journal of Maternal-Fetal and Neonatal Medicine* **23** 20–23.

ARCONES, M. and SAMANIEGO, F. (2000). On the asymptotic distribution theory of a class of consistent estimators of a distribution satisfying a uniform stochastic ordering constraint. *Annals of Statistics* **28** 116–150. MR1762906

BAMBER, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* **12** 387–415. MR0384214

BEARE, B. and MOON, J. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory* **31** 471–492. MR3348455

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*. Wiley, New York.

CAROLAN, C. and TEBBS, J. (2005). Nonparametric tests for and against likelihood ratio ordering in the two-sample problem. *Biometrika* **92** 159–171. MR2158617

COX, C., HASHEM, N., TEBBS, J., BOOKSTAVER, B., and ISKERSKY, V. (2015). Evaluation of caffeine and the development of necrotizing enterocolitis. *Journal of Neonatal-Perinatal Medicine* **8** 339–347.

DARDANONI, V. and FORCINA, A. (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *Journal of the American Statistical Association* **93** 1112–1123. MR1649205

DAVIDOV, O. and HERMAN, A. (2012). Ordinal dominance curve based inference for stochastically ordered distributions. *Journal of the Royal Statistical Society, Series B* **74** 825–847. MR2988908

DOBSON, N. and HUNT, C. (2013). Caffeine use in neonates: Indications, pharmacokinetics, clinical effects, outcomes. *NeoReviews* **14** 540–550.

DYKSTRA, R., KOCHAR, S., and ROBERTSON, T. (1991). Statistical inference for uniform stochastic ordering in several populations. *Annals of Statistics* **19** 870–888. MR1105849

EL BARMI, H. and MCKEAGUE, I. (2016). Testing for uniform stochastic ordering via empirical likelihood. *Annals of the Institute of Statistical Mathematics* **68** 955–976. MR3538581

EL BARMI, H. and MUKERJEE, H. (2016). Consistent estimation of survival functions under uniform stochastic ordering: The k-sample case. *Journal of Multivariate Analysis* **144** 99–109. MR3434943

HSIEH, F. and TURNBULL, B. (1996). Nonparametric estimation of the receiver operating characteristic curve. *Annals of Statistics* **24** 25–40. MR2686607

KOMLÓS, J., MAJOR, P., and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **32** 111–131.

LEHMANN, E. and ROJO, J. (1992). Invariant directional orderings. *Annals of Statistics* **20** 2100–2110. MR1193328

MUKERJEE, H. (1996). Estimation of survival functions under uniform stochastic ordering. *Journal of the American Statistical Association* **91** 1684–1689. MR1439110

PARK, C., LEE, C., and ROBERTSON, T. (1998). Goodness-of-fit test for uniform stochastic ordering among several populations. *Canadian Journal of Statistics* **26** 69–81.

ROJO, J. and SAMANIEGO, F. (1993). On estimating a survival curve subject to a uniform stochastic ordering constraint. *Journal of the American Statistical Association* **88** 566–572. MR1224382

SCHMIDT, B., ROBERTS, R., DAVIS, P., DOYLE, L., BARRINGTON, K., OHLSSON, A., SOLIMANO, A., and TIN, W. (2006). Caffeine therapy for apnea of prematurity. *New England Journal of Medicine* **354** 2112–2121.

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

SHAPIRO, A. (1990). On concepts of directional differentiability. *Journal of Optimization*

Theory and Applications **66** 477–487. MR1080259

SHAPIRO, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research* **30** 169–186. MR1118896

TANG, C., WANG, D., and TEBBS, J. (2016). Supplement to "Nonparametric goodness-of-fit tests for uniform stochastic ordering."

VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, SC 29208, USA
E-MAIL: tebbs@stat.sc.edu
URL: http://people.stat.sc.edu/tebbs