An improved bootstrap test of density ratio ordering

beare, brendan and shi, xiaoxia

5 August 2015

An improved bootstrap test of density ratio ordering

Brendan K. Beare[1] and Xiaoxia Shi[2]

[1]Department of Economics, University of California, San Diego
[2]Department of Economics, University of Wisconsin, Madison

August 4, 2015

Abstract

Two probability distributions with common support are said to exhibit density ratio ordering when they admit a nonincreasing density ratio. Existing statistical tests of the null hypothesis of density ratio ordering are known to be conservative, with null limiting rejection rates below the nominal significance level whenever the two distributions are unequal. We show how a bootstrap procedure can be used to shrink the critical values used in existing procedures such that the limiting rejection rate is increased to the nominal significance level on the boundary of the null. This improves power against nearby alternatives. Our procedure is based on preliminary estimation of a contact set, the form of which is obtained from a novel representation of the Hadamard directional derivative of the least concave majorant operator. Numerical simulations indicate that improvements to power can be very large in moderately sized samples.

We thank Zhonglin Li and Juwon Seo for research assistance, and Andres Santos and seminar participants at the University of Texas at Austin, Hong Kong University of Science and Technology, University of Tokyo, University of Sydney, Pennsylvania State University, University College London, University of Copenhagen, Aarhus University, and Hitotsubashi University for helpful comments.

1 Introduction

Let F and G be cumulative distribution functions (cdfs) on the real line $\mathbb{R}$, with common support. When F and G admit a nonincreasing density ratio $\mathrm{d}F/\mathrm{d}G$, we say that there is *density ratio ordering* between F and G. Density ratio ordering implies, but is not implied by, first order stochastic dominance. While first order stochastic dominance provides a suitable ordering between distributions in many applications, there are times when economic or financial models indicate that density ratio ordering is the more appropriate property to consider. For instance, Beare (2011) shows that, in a simple one period pricing model, a failure of density ratio ordering between the risk neutral and physical payoff distributions associated with a market portfolio has perverse implications for the behavior of contingent claims. See also Beare and Schmidt (2015) for related empirical analysis. Other contexts in which density ratio ordering plays a key role, including mechanism design and auction theory, are discussed by Roosen and Hennessy (2004).

Statistical methods for testing the null hypothesis of stochastic dominance between two cdfs are already well established; see e.g. Anderson (1996), Davidson and Duclos (2000), Barrett and Donald (2003), Linton et al. (2005, 2010), Donald and Hsu (2015) and the survey article by Maasoumi (2001). Less work has been done on testing the null hypothesis of density ratio ordering. Dykstra et al. (1995) and Roosen and Hennessy (2004) dealt with the case where F and G are discrete distributions. The more delicate case where F and G are continuous distributions was studied by Carolan and Tebbs (2005) and Beare and Moon (2015). These authors exploit the fact that, in the continuous case, density ratio ordering is equivalent to the concavity of the *ordinal dominance curve* (odc): the composition of F with G^{-1}, the quantile function for G. They consider a statistic constructed from the difference between an empirical estimate of the odc and its *least concave majorant* (lcm). It is compared to a critical value that delivers a limiting rejection rate equal to nominal size when $F = G$, and below nominal size when $F \neq G$ but density ratio ordering is satisfied.

The contribution of this paper is a modification to the density ratio ordering test of Carolan and Tebbs (2005) and Beare and Moon (2015) that improves power. We retain the test statistic used by those authors, but compare it to a data dependent critical value computed using the bootstrap. This has the effect of raising the limiting rejection rate of the test to the nominal significance level on the boundary of the null; more precisely, at those points in the null where the limit distribution of the test statistic is nondegenerate. Consequently, power is improved at nearby points in the alternative. Our bootstrap procedure requires preliminary estimation of a *contact set*, and has a similar flavor to the bootstrap procedures used by Linton et al. (2010) and Donald and Hsu (2015) to improve the power of the test of stochastic dominance proposed by Barrett and Donald (2003).

The main technical hurdles we face when studying the asymptotic properties of our procedure relate to the differential properties of the lcm operator. Beare and Moon (2015) showed that this operator fails to be Hadamard differentiable at all points in the null, but instead satisfies a weaker smoothness condition dubbed *Hadamard directional differentiability* by Shapiro (1990). Hadamard directional differentiability suffices for the application of the functional delta method, which is the key device used by Beare and Moon (2015) to determine the asymptotic behavior of their test statistic. However, as shown by Dümbgen (1993) and discussed further in a recent working paper by Fang and Santos (2014), standard bootstrap inference can be problematic when working with operators that are Hadamard directionally differentiable but not Hadamard differentiable. We propose a modified bootstrap procedure with good asymptotic and finite sample properties. Our primary technical innovation is a new representation of the Hadamard directional derivative of the lcm operator that expresses the derivative at each point in the null in terms of an estimable subset of the unit cube: our contact set.

The remainder of our paper is structured as follows. In Section 2 we introduce our sampling framework and test statistic, including a discussion of the directional differentiability of the lcm operator, and an explanation of how this property can be used to derive relevant asymptotic results under the null. In Section 3 we present our main results, including our new representation of the directional derivative of the lcm operator. We explain how this representation can be used to develop a modified bootstrap procedure based on preliminary estimation of the contact set, and establish conditions under which this procedure raises the limiting rejection rate of our test to the nominal significance level on the boundary of the null. Section 4 provides a discussion of some practical issues that arise in the implementation of our procedure, including the numerical computation of suprema and integrals, and the selection of a tuning parameter used in the contact set estimation. Section 5 reports numerical evidence on the finite sample performance of our procedure, and final remarks are given in Section 6. Mathematical proofs of the results stated in Section 3 are collected in the Appendix.

2 Test statistic

Here we introduce the test of density ratio ordering studied by Carolan and Tebbs (2005) and Beare and Moon (2015), including details sufficient to provide a basis for our discussion of bootstrap critical values in Section 3. In Section 2.1 we define the null and alternative hypotheses, state the sampling framework, and explain the construction of the test statistic. In Section 2.2 we review results given by Beare and Moon (2015) on the differential properties of the lcm operator, including a discussion of the distinction between Hadamard differentiability and Hadamard directional differentiability. These results are used in Section 2.3 to give a brief derivation of the limit distribution of our test statistic under the null hypothesis, again following Beare and Moon.

2.1 Statistical framework

Our data consist of two independent and identically distributed samples of real valued random variables $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$, mutually independent of one another. We let F denote the common cdf of the X_i's and G denote the common cdf of the Y_j's, and assume that F and G are continuous and strictly increasing on their common support. Our goal is to test the hypothesis that the odc $R = F \circ G^{-1}$ is concave, where $G^{-1}(u) = \inf\{y : G(y) \geq u\}$ is the quantile function corresponding to G. Let Θ denote the collection of strictly increasing, continuously differentiable maps $\theta : [0, 1] \to [0, 1]$ with $\theta(0) = 0$ and $\theta(1) = 1$, and let $\Theta_0 = \{\theta \in \Theta : \theta \text{ is concave}\}$. We maintain throughout that $R \in \Theta$, and write R' for its first derivative. We seek to test the null hypothesis $H_0 : R \in \Theta_0$ against the alternative hypothesis $H_1 : R \in \Theta \setminus \Theta_0$.

Let $\ell^\infty([a, b])$ denote the collection of uniformly bounded real valued functions on $[a, b]$ equipped with the uniform norm. The following definition is taken from Beare and Moon (2015, Def. 2.1).

Definition 2.1. Given a closed interval $[a,b] \subseteq [0,1]$, the lcm over $[a,b]$ is the operator $\mathcal{M}_{[a,b]} : \ell^\infty([0,1]) \to \ell^\infty([a,b])$ that maps each $f \in \ell^\infty([0,1])$ to the function

$$\mathcal{M}_{[a,b]}f(u) = \inf\{g(u) : g \in \ell^\infty([a,b]), \, g \text{ is concave, and } f \leq g \text{ on } [a,b]\}, \quad u \in [a,b].$$

We write $\mathcal{M}$ as shorthand for $\mathcal{M}_{[0,1]}$, and refer to $\mathcal{M}$ as the lcm operator.

Following Carolan and Tebbs (2005), we take as our estimator of R the empirical odc $R_{m,n} = F_m \circ G_n^{-1}$, where

$$F_m(\cdot) = \frac{1}{m}\sum_{i=1}^{m} 1(X_i \leq \cdot), \quad G_n(\cdot) = \frac{1}{n}\sum_{j=1}^{n} 1(Y_j \leq \cdot)$$

are the empirical cdfs of (X_i) and (Y_j) respectively. Our test statistic is

$$M_{m,n} = c_{m,n} \left\| \mathcal{M}R_{m,n} - R_{m,n} \right\|_p,$$

where $c_{m,n} = (mn/(m+n))^{1/2}$, $\| \cdot \|_p$ is the L^p-norm with respect to Lebesgue measure on $[0,1]$, and $p \in [1,\infty]$. This statistic was proposed by Carolan and Tebbs (2005) for $p = 1$ and $p = \infty$, while Beare and Moon (2015) considered the more general family of statistics indexed by $p \in [1,\infty]$

The empirical odc $R_{m,n}$ is unaffected with probability one if we replace our observations X_i and Y_j with $\psi(X_i)$ and $\psi(Y_j)$ for any real valued ψ strictly increasing on the common support of F and G. Taking $\psi = G$ normalizes the cdf of the $\psi(X_i)$'s to be R and the cdf of the $\psi(Y_j)$'s to be uniform on $[0,1]$, and so we see that the distribution of $M_{m,n}$ is uniquely determined by R. Consequently, it makes sense to talk about the distribution of $M_{m,n}$ at different points in Θ; different pairs of cdfs (F,G) give rise to the same distribution for $M_{m,n}$ whenever they correspond to the same odc $R \in \Theta$.

In the asymptotic theory to be developed shortly, we will let the two sample sizes m and n tend to infinity simultaneously, with $n/(m+n) \to \lambda \in (0,1)$. Formally, we can think of m as being implicitly a function of n, with $m(n) \to \infty$ and $n/(m(n)+n) \to \lambda \in (0,1)$ as $n \to \infty$. We might therefore consider indexing all sample statistics only by n, and never by m or m,n. However, for concreteness, we continue to index sample statistics by

m and/or n where appropriate, consistent with Carolan and Tebbs (2005) and Beare and Moon (2015).

2.2 Differential properties of the lcm operator

The arguments used by Beare and Moon (2015) to determine the null limiting behavior of $M_{m,n}$ rely critically on an understanding of the differential properties of the operator $\mathcal{M}$. The following definition is adapted from Dümbgen (1993).

Definition 2.2. Let $\mathscr{X}$ and $\mathscr{Y}$ be real Banach spaces. A map $\phi : \mathscr{X} \to \mathscr{Y}$ is said to be Hadamard directionally differentiable at $x \in \mathscr{X}$ tangentially to a linear space $\mathscr{X}_0 \subseteq \mathscr{X}$ if there exists a map $\phi'_x : \mathscr{X}_0 \to \mathscr{Y}$ such that

$$\phi'_x(z) = \lim_{n\to\infty} \frac{\phi(x + t_n z_n) - \phi(x)}{t_n}$$

for any sequences $z_n \in \mathscr{X}$ and $t_n \in (0, 1)$ with $z_n \to z \in \mathscr{X}_0$ and $t_n \downarrow 0$. We refer to $\phi'_x(z)$ as the Hadamard directional derivative of ϕ at x in direction z. If ϕ'_x is linear then we say that ϕ is Hadamard differentiable at x tangentially to $\mathscr{X}_0$, and we refer to $\phi'_x(z)$ as the Hadamard derivative of ϕ at x in direction z.

A Hadamard directional derivative is automatically continuous and positive homogeneous of degree one, but may be nonlinear. Linearity turns out to be unimportant for applications of the functional delta method (Shapiro, 1991), but is vitally important for establishing bootstrap consistency (Dümbgen, 1993; Fang and Santos, 2014). A closely related version of differentiability called quasi-Hadamard differentiability has been studied by Beutner et al. (2012) and Volgushev and Shao (2014). Beutner and Zähle (2010, 2012) also study a version of differentiability that they call quasi-Hadamard differentiability, but in their case the derivative is automatically linear because they consider general sequences t_n converging to zero, and not merely those converging to zero from above.

It turns out that, at points $R \in \Theta_0$, the lcm operator $\mathcal{M}$ is Hadamard directionally differentiable but not in general Hadamard differentiable. The following result, in which $C([0, 1])$ denotes the space of continuous real valued functions on $[0, 1]$ equipped with the uniform norm, was proved by Beare and Moon (2015, Lem. 3.2).

Lemma 2.1. *If $R \in \Theta_0$ then $\mathcal{M}$ is Hadamard directionally differentiable at R tangentially to $C([0,1])$. Given $h \in C([0,1])$, if R is affine in a neighborhood of $u \in (0,1)$, then we have $\mathcal{M}'_R h(u) = \mathcal{M}_{[a_{R,u}, b_{R,u}]} h(u)$, where*

$$\begin{aligned} a_{R,u} &= \sup\{u' \in (0,u] : R \text{ is not affine in a neighborhood of } u'\}, \\ b_{R,u} &= \inf\{u' \in [u,1) : R \text{ is not affine in a neighborhood of } u'\}, \end{aligned}$$

and we define $\inf \varnothing = 1$ *and* $\sup \varnothing = 0$. *If R is not affine in a neighborhood of $u \in (0,1)$, or if $u \in \{0,1\}$, then $\mathcal{M}'_R h(u) = h(u)$.*

We illustrate the content of Lemma 2.1 with an example in Figure 2.1. In panel (a) we display the odc R at which we wish to differentiate $\mathcal{M}$. It is affine over the intervals $[0,a]$ and $[b,1]$, and strictly concave over the interval $[a,b]$. We also display the direction h in which we wish to differentiate, a sinusoid. In panel (b) we display $\mathcal{M}'_R h$, the Hadamard directional derivative of $\mathcal{M}$ at R in direction h. It has three distinct parts. Over the intervals $[0,a]$ and $[b,1]$, where R is affine, the directional derivative is given by the restricted lcms $\mathcal{M}_{[0,a]}h$ and $\mathcal{M}_{[b,1]}h$ respectively. Over the interval $[a,b]$, where R is strictly concave, the directional derivative is h. In panel (c) we display $\mathcal{M}'_R(-h)$, the Hadamard directional derivative of $\mathcal{M}$ at R in direction $-h$. Comparing $\mathcal{M}'_R h$ and $\mathcal{M}'_R(-h)$ in panels (b) and (c), we observe that $\mathcal{M}'_R h \neq -\mathcal{M}'_R(-h)$, implying that $\mathcal{M}'_R$ cannot be linear. Consequently, $\mathcal{M}$ is not Hadamard differentiable at R tangentially to $C([0,1])$ in the example depicted. In fact, as noted by Beare and Moon (2015), $\mathcal{M}$ is Hadamard differentiable at $R \in \Theta_0$ tangentially to $C([0,1])$ if and only if R is strictly concave.

2.3 Limit distribution under concavity

Let $\mathcal{A} : \ell^\infty([0,1]) \to \mathbb{R}$ be the operator

$$\mathcal{A}f = \|\mathcal{M}f - f\|_p, \quad f \in \ell^\infty([0,1]).$$

When R is concave our test statistic $M_{m,n}$ may be written as

$$M_{m,n} = c_{m,n}\left(\mathcal{A}R_{m,n} - \mathcal{A}R\right).$$

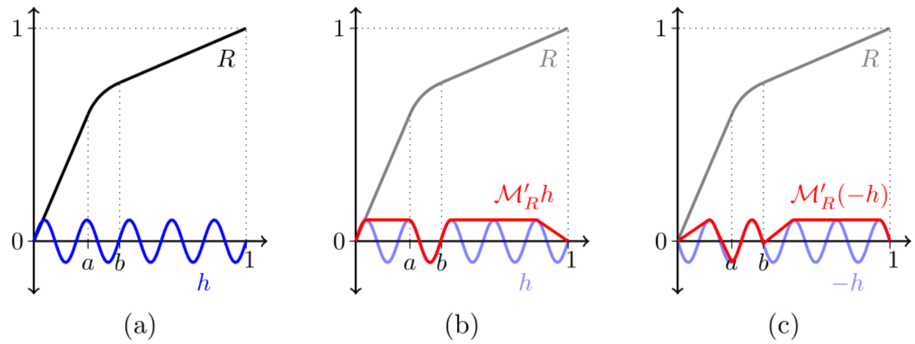


Figure 2.1: Panel (b) displays the Hadamard directional derivative $\mathcal{M}'_R h$ for the particular choice of R and h shown in panel (a). Panel (c) displays $\mathcal{M}'_R(-h)$.

Two ingredients suffice for us to establish the limit distribution of $M_{m,n}$ at each $R \in \Theta_0$. First, we require weak convergence of the empirical odc process $c_{m,n}(R_{m,n} - R)$ to a suitable limit, and second, we require the operator $\mathcal{A}$ to satisfy a smoothness condition sufficient for the application of the functional delta method. The former ingredient has been available at least since Hsieh and Turnbull (1996, Thm. 2.2); the following statement is taken from Beare and Moon (2015, Lem. 3.1), with $\rightsquigarrow$ denoting weak convergence in a metric space in the sense of Hoffmann-Jørgensen.

Lemma 2.2. *Suppose $R \in \Theta$. Then as $m \wedge n \to \infty$ with $n/(m+n) \to \lambda \in (0,1)$, we have $c_{m,n}(R_{m,n} - R) \rightsquigarrow T$, where T has the form*

$$T(u) = \lambda^{1/2} B_1(R(u)) + (1-\lambda)^{1/2} R'(u) B_2(u), \quad u \in [0,1],$$

and B_1 and B_2 are independent standard Brownian bridges on $[0,1]$.

It remains to establish a smoothness condition on $\mathcal{A}$ sufficient for the application of the functional delta method. With Lemma 2.1 in hand, a routine application of the chain rule for Hadamard directionally differentiable operators (Shapiro, 1990, Prop. 3.6) establishes that $\mathcal{A}$ is Hadamard directionally differentiable at $R \in \Theta_0$ tangentially to $C([0,1])$, with directional derivative

$$\mathcal{A}'_R h = \left\| \mathcal{M}'_R h - h \right\|_p, \quad h \in C([0,1]).$$

Though textbook treatments of the functional delta method typically impose Hadamard differentiability upon the operator in question, it is sufficient to impose the weaker requirement of Hadamard directional differentiability. This was proved by Shapiro (1991, Thm. 2); for a more recent statement, see Fang and Santos (2014, Thm. 2.1). We thus arrive at the following result.

Theorem 2.1. *Suppose $R \in \Theta$. Then as $m \wedge n \to \infty$ with $n/(m+n) \to \lambda \in (0,1)$, we have $M_{m,n} \to_d \mathcal{A}'_R T$.*

From Lemma 2.2 we see that the law of T is uniquely determined by R, and hence the law of $\mathcal{A}'_R T$ is also uniquely determined by R. Beare and Moon (2015, Thm. 4.1) proved that, for $p \in [1,2]$, $\mathcal{A}'_R T$ is stochastically dominated by $\mathcal{A}'_I T = \|\mathcal{M}B - B\|_p$, where I is the identity map on $[0,1]$ and B is a Brownian bridge. We therefore refer to $R = I$ as the *least favorable case* (lfc) and may construct a conservative test of concavity by using as a critical value the relevant quantile of the law of $\mathcal{A}'_I T$. If we reject the null hypothesis of concavity when M_n exceeds this critical value, then the limiting rejection rate of our test is α at the lfc $R = I$, and is no greater than α at all other $R \in \Theta_0$. The idea of using a fixed critical value to control size at the lfc is due to Carolan and Tebbs (2005), and requires us to choose $p \in [1,2]$, as $R = I$ is no longer least favorable when $p \in (2,\infty]$ (Beare and Moon, 2015, Thm. 4.2).

The disadvantage of using a fixed critical value to set the limiting rejection rate equal to α at the lfc $R = I$ is that the limiting rejection rate may be well below α at other $R \in \Theta_0$. Indeed, since $\mathcal{A}'_R T = 0$ when R is strictly concave, the limiting rejection rate at all strictly concave $R \in \Theta_0$ is zero. Numerical results reported by Beare and Moon (2015) also indicate that, with $\alpha = 0.05$ and in sample sizes as large as 500, the rejection rate is effectively zero at some members of Θ_0 that are not strictly concave, and are in fact affine over wide portions of their domain. This is problematic because any concave member of Θ may be approximated arbitrarily well in the uniform metric by a nonconcave member of Θ, suggesting that power against relevant nonconcave alternatives may be close to zero.

3 Bootstrap critical values

Our main results are in this section. In Section 3.1 we give a novel representation of the Hadamard directional derivative of the lcm operator and explain how it can be used to express the null limit distribution of $M_{m,n}$ in terms of a contact set and the weak limit of the empirical odc process. In Section 3.2 we discuss the estimation of this contact set. In Section 3.3 we show how the estimated contact set can be used to bootstrap critical values in a way that yields a limiting rejection rate equal to the nominal significance level at all points in the null where R is not strictly concave. Proofs of all results are collected in the Appendix.

3.1 An alternative representation of $\mathcal{M}'_R$

Begin by defining the set

$$A = \{(u, v, w) \in [0, 1]^3 : v \leq u \leq w\}.$$

Let $\mathcal{S} : \ell^\infty([0, 1]) \to \ell^\infty(A)$ be the operator

$$\mathcal{S}f(u, v, w) = \frac{(w - u)f(v) + (u - v)f(w)}{w - v}, \quad f \in \ell^\infty([0, 1]), \quad (u, v, w) \in A,$$

where for $v = w$ we define $\mathcal{S}f(u, v, w) = f(u)$. We may view $\mathcal{S}f(u, v, w)$ as the approximation to $f(u)$ obtained by linearly interpolating between the values taken by f at v and w. We note the following property of $\mathcal{S}$ for later use.

Lemma 3.1. *$\mathcal{S}$ is a linear isometry.*

With the operator $\mathcal{S}$ and odc R we define the set

$$B = \{(u, v, w) \in A : \mathcal{S}R(u, v, w) = R(u)\},$$

and the family of cross-sections

$$B(u) = \{(v, w) \in [0, 1]^2 : (u, v, w) \in B\}, \quad u \in [0, 1].$$

The set B always contains the main diagonal $u = v = w$ of the unit cube, and that the cross-section $B(u)$ always includes the point (u, u).

Our alternative representation of the Hadamard directional derivative of the lcm operator—compare to Lemma 2.1 above—is as follows.

Lemma 3.2. *The Hadamard directional derivative of $\mathcal{M}$ at $R \in \Theta_0$ in direction $h \in C([0,1])$ satisfies*

$$\mathcal{M}'_R h(u) = \sup_{(v,w)\in B(u)} \mathcal{S}h(u,v,w), \quad u \in [0,1].$$

In view of Theorem 2.1 and Lemma 3.2, when $R \in \Theta_0$ the weak limit $\mathcal{A}'_R T$ of our test statistic $M_{m,n}$ satisfies

$$\mathcal{A}'_R T = \left\| \sup_{(v,w)\in B(\cdot)} \tilde{\mathcal{S}}T(\cdot, v, w) \right\|_p,$$

where $\tilde{\mathcal{S}} : \ell^\infty([0,1]) \to \ell^\infty(A)$ is the operator

$$\tilde{\mathcal{S}}f(u,v,w) = \mathcal{S}f(u,v,w) - f(u), \quad f \in \ell^\infty([0,1]), \quad (u,v,w) \in A.$$

The weak limit $\mathcal{A}'_R T$ is uniquely determined by the law of T and the set B. In this sense, B plays a similar role to the so-called contact set used by Linton et al. (2010) to characterize the null limit distribution of their statistic for testing stochastic dominance. We shall borrow their terminology and refer to B as our contact set. Contact sets also play a key role in the analyses of Lee and Whang (2009), Anderson et al. (2012) and Lee et al. (2014), although in these papers there arise significant additional technical complications owing to the lack of a weak convergence result analogous to Lemma 2.2.

3.2 Contact set estimation

To implement our bootstrap procedure we require a preliminary estimate of the unknown contact set B. We now present three candidate estimators of B, denoted $B_{m,n}$, $B'_{m,n}$ and $B''_{m,n}$. By construction, $B_{m,n} \subseteq B'_{m,n} \subseteq B''_{m,n}$. Under the null hypothesis, the three estimators closely approximate B with probability approaching one; see Lemma 3.3 below. Under the alternative hypothesis, there can be large differences between the three estimators that persist asymptotically. We will see later that a smaller estimated contact set delivers a smaller critical value, improving the probability of rejecting the null hypothesis. Our preferred contact set estimator is therefore $B_{m,n}$, but we also discuss $B'_{m,n}$ and $B''_{m,n}$ for expository purposes.

Our three contact set estimators make use of a tuning parameter $\delta_{m,n} \in (0, \infty)$. This tuning parameter is required to converge to zero as the sample sizes m and n increase, but not too quickly.

Assumption 3.1. As $m \wedge n \to \infty$ with $n/(m+n) \to \lambda \in (0,1)$, we have $\delta_{m,n} \to 0$ and $c_{m,n}\delta_{m,n} \to \infty$.

The results given in this section are valid for any choice of $\delta_{m,n}$ that satisfies Assumption 3.1 with probability one. In Section 4.2 we suggest an approach to choosing $\delta_{m,n}$ in practice.

The largest of our contact set estimators, $B''_{m,n}$, is also the most obvious: we simply set

$$B''_{m,n} = \left\{(u,v,w) \in A : |\tilde{\mathcal{S}}R_{m,n}(u,v,w)| \leq \delta_{m,n}\right\}.$$

The estimated set $B''_{m,n}$ contains those triples $(u,v,w) \in A$ for which $\mathcal{S}R_{m,n}(u,v,w)$ is close to $R_{m,n}(u)$, with closeness defined in terms of the tuning parameter value $\delta_{m,n}$.

In large samples, $B''_{m,n}$ can be expected to provide a good approximation to B regardless of whether the null hypothesis is true. For our purposes, a better estimator of B is one that provides a good approximation to B when the null hypothesis is satisfied, but is as small as possible otherwise. Consider the possible contact sets B that may obtain when the null hypothesis is satisfied. When R is concave, if B contains some triple $(u,v,w) \in A$, then it must be the case that $R(t) = \mathcal{M}R(t)$ for $t \in \{u,v,w\}$. Our second contact set estimator is constructed to exclude members of $B''_{m,n}$ that appear very likely to violate this condition:

$$B'_{m,n} = B''_{m,n} \cap \{t \in [0,1] : \mathcal{M}R_{m,n}(t) \leq R_{m,n}(t) + \delta_{m,n}\}^3.$$

Yet more can be said about the form of B when the null hypothesis is satisfied. When R

is concave, if B contains some triple $(u,v,w) \in A$, then it must also contain each triple $(u,v',w') \in A$ for which $v' \in [v,u]$ and $w' \in [u,w]$. This motivates our smallest and preferred contact set estimator $B_{m,n}$, defined as

$$B_{m,n} = \left\{(u,v,w) \in B'_{m,n} : (u,v',w') \in B'_{m,n} \text{ for all } (v',w') \in [v,u] \times [u,w]\right\}.$$

Our next result states that, with high probability, $B_{m,n}$, $B'_{m,n}$ and $B''_{m,n}$ each provide a good outer-approximation to our contact set B when the null hypothesis is satisfied.

Lemma 3.3. *Suppose $R \in \Theta_0$ and Assumption 3.1 is satisfied. Then as $m \wedge n \to \infty$ with $n/(m+n) \to \lambda \in (0,1)$, we have $P(B \subseteq B_{m,n} \subseteq B'_{m,n} \subseteq B''_{m,n} \subseteq B^{\epsilon}) \to 1$ for any $\epsilon > 0$, where*

$$B^{\epsilon} = \left\{a \in A : \inf_{b \in B} \|b - a\| \leq \epsilon\right\},$$

the ϵ-enlargement of B.

3.3 Bootstrap procedure

In short, our bootstrap approximation to the weak limit $\mathcal{A}'_R T$ of $M_{m,n}$ works by simulating the distribution of $M^*_{m,n} = \hat{\mathcal{A}}'_{m,n} T^*_{m,n}$ conditional on our data, where $T^*_{m,n}$ is a bootstrap version of T, and $\hat{\mathcal{A}}'_{m,n} : \ell^{\infty}([0,1]) \to \mathbb{R}$ is the data dependent operator

$$\hat{\mathcal{A}}'_{m,n} f = \left\| \sup_{(v,w) \in B_{m,n}(\cdot)} \tilde{\mathcal{S}} f(\cdot, v, w) \right\|_p, \quad f \in \ell^{\infty}([0,1]).$$

The estimated operator $\hat{\mathcal{A}}'_{m,n}$ is determined by the estimated contact set $B_{m,n}$; note that $B_{m,n}(u)$ is a cross-section of $B_{m,n}$, defined in the same way as $B(u)$. Our approach places us in the general framework used by Fang and Santos (2014) to explore the use of bootstrap inference when standard differentiability conditions are violated.

To obtain $T^*_{m,n}$, we first construct bootstrap versions of F_m and G_n by setting

$$F^*_m(\cdot) = \frac{1}{m} \sum_{i=1}^{m} V^*_{i,m} 1(X_i \leq \cdot), \quad G^*_n(\cdot) = \frac{1}{n} \sum_{j=1}^{n} W^*_{j,n} 1(Y_j \leq \cdot),$$

where the weights $V_m^* = (V_{1,m}^*, \ldots, V_{m,m}^*)$ and $W_n^* = (W_{1,n}^*, \ldots, W_{n,n}^*)$ are drawn independently of the data and of one another from the multinominal distribution with probabilities spread evenly over the categories $1, \ldots, m$ and $1, \ldots, n$ respectively. From F_m^* and G_n^* we construct $R_{m,n}^* = F_m^* \circ G_n^{*-1}$, our bootstrap version of $R_{m,n}$. We then set $T_{m,n}^* = c_{m,n}(R_{m,n}^* - R_{m,n})$.

The following result establishes that the law of $T_{m,n}^*$ conditional on our data provides an accurate approximation to the law of T with high probability. Weak convergence conditional on the data in probability is meant in the sense of Kosorok (2008, pp. 19-20); see also Volgushev and Shao (2014, p. 411).

Lemma 3.4. *Suppose $R \in \Theta$. Then as $m \wedge n \to \infty$ with $n/(m+n) \to \lambda \in (0,1)$, we have $T_{m,n}^* \rightsquigarrow T$ conditional on the data in probability.*

The law of $T_{m,n}^*$ conditional on the data can be simulated: we simply compute large numbers of realizations of $T_{m,n}^*$ corresponding to repeated draws of the multinomial weights V_m^* and W_n^*. In order to obtain suitable critical values for our test statistic, we seek to approximate the law of its weak limit $\mathcal{A}_R' T$ when $R \in \Theta_0$. If $\mathcal{A}$ were Hadamard differentiable at $R \in \Theta_0$ tangentially to $C([0,1])$, we could deduce from the functional delta method for the bootstrap that $c_{m,n}\left(\mathcal{A}R_{m,n}^* - \mathcal{A}R_{m,n}\right) \rightsquigarrow \mathcal{A}_R' T$ conditional on the data in probability, which would justify the use of the law of $c_{m,n}(\mathcal{A}R_{m,n}^* - \mathcal{A}R_{m,n})$ conditional on the data as an approximation to the law of $\mathcal{A}_R' T$. Unfortunately we cannot apply the delta method for the bootstrap in this fashion unless R is strictly concave, because it is only at the strictly concave members of Θ_0 that $\mathcal{A}$ is Hadamard differentiable. Though $\mathcal{A}$ is Hadamard directionally differentiable at all $R \in \Theta_0$, it was shown by Dümbgen (1993) that directional differentiability does not suffice for the application of the functional delta method for the bootstrap, and that the naïve bootstrap typically fails when working with operators that are not fully Hadamard differentiable.

In view of the failure of the naïve bootstrap we take an alternative route, and approximate the law of $\mathcal{A}_R' T$ using the law of $M_{m,n}^* = \hat{\mathcal{A}}_{m,n}' T_{m,n}^*$ conditional on the data. For a test with nominal size $\alpha \in (0,1)$ we take as our critical value

$$\mu_{m,n}(\alpha) = \inf\{x : P(M_{m,n}^* \leq x \mid X_1, \ldots, X_m, Y_1, \ldots, Y_n) \geq 1 - \alpha\},$$

the $(1-\alpha)$-quantile of the distribution of $M^*_{m,n}$ conditional on the data.

Theorem 3.1. *Suppose $R \in \Theta_0$ and Assumption 1 is satisfied. Then as $m \wedge n \to \infty$ with $n/(m+n) \to \lambda \in (0,1)$, we have $M^*_{m,n} \rightsquigarrow \mathcal{A}'_R T$ conditional on the data in probability. If in addition R is not strictly concave, we have $P(M_{m,n} > \mu_{m,n}(\alpha)) \to \alpha$.*

Theorem 3.1 establishes that our bootstrap procedure delivers a test with limiting rejection rate equal to nominal size whenever R is concave but not strictly concave. These R are precisely those points in Θ_0 at which the limit distribution of $M_{m,n}$ is nondegenerate, and form what Linton et al. (2010) refer to as the *boundary of the null*. Of course, this notion of boundary differs from the usual topological one; in the uniform topology, every member of Θ_0 is the limit of a sequence in Θ_1, and so Θ_0 is its own boundary.

A shortcoming of Theorem 3.1 is that it says nothing about the limiting rejection rate of our test when R is strictly concave. In this case, both $M_{m,n}$ and $\mu_{m,n}(\alpha)$ converge in probability to zero, and we cannot say much of substance about their relative magnitudes without investigating their higher order asymptotic behavior, which seems difficult. In a related context, Andrews and Shi (2013, p. 625) have proposed a technical remedy to this problem: instead of using $\mu_{m,n}(\alpha)$ as our critical value, we can use $\mu_{m,n}(\alpha) + \epsilon$ or $\mu_{m,n}(\alpha) \vee \epsilon$, where $\epsilon > 0$ is some small fixed value. The presence of ϵ prevents our critical value from converging in probability to zero alongside $M_{m,n}$ when R is strictly concave, ensuring a limiting rejection rate of zero. For further discussion, see Fang and Santos (2014, Rem. 3.12) and Donald and Hsu (2015, p. 13). We have found in numerical simulations with $p = 1$ and $p = 2$ that in practice it is unnecessary to modify the critical value in this fashion. Our test appears to be very conservative at strictly concave choices of R, and also at many concave choices of R that are not strictly concave.

We have not discussed power properties of our test. In fact, it is simple to show that, under mild regularity conditions, our test has power approaching one against any sequence of nonconcave odcs that approach the null at a rate slower than $n^{-1/2}$, and nonvanishing power against some sequences of nonconcave odcs that approach the null at the rate $n^{-1/2}$. We omit the formal statement and proof of these claims, which can be given in virtually identical fashion to those of Theorems 5.1 and 5.2 of Beare and Moon (2015).

4 Practical implementation

Here we provide some pragmatic guidelines for implementing our testing procedure. In Section 4.1 we provide a step-by-step guide to the computation of our test statistic and bootstrap critical value, avoiding abstract operations such as suprema over infinite sets and integration, and instead using only operations that are easily implementable using standard numerical software packages. A method for choosing the tuning parameter $\delta_{m,n}$ is suggested in Section 4.2.

4.1 Numerical computation

What follows is a step-by-step recipe for computing our test statistic and critical value. All steps provide an exact calculation, with the exception of step 3(v), which uses a summation to numerically approximate an integral. The approximation error should be negligible unless n is very small.

1. Compute the test statistic.

   (i) Order the two samples as $X_{(1)} \leq \cdots \leq X_{(m)}$ and $Y_{(1)} \leq \cdots \leq Y_{(n)}$.

   (ii) Set $R_{m,n}(0) = 0$ and for $i = 1, \ldots, n$ compute

   $$R_{m,n}\left(\frac{i}{n}\right) = \frac{1}{m} \max\{j = 1, \ldots, m : X_{(j)} \leq Y_{(i)}\},$$

   with the maximum over the empty set defined to be zero.

   (iii) For $j = 0, \ldots, n-1$ and $i = j+1, \ldots, n$ and $k = i, \ldots, n$ compute

   $$\mathcal{S}R_{m,n}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) = \frac{(k-i)R_{m,n}(j/n) + (i-j)R_{m,n}(k/n)}{k-j},$$

   and for $i = 0, \ldots, n$ set $\mathcal{S}R_{m,n}(i/n, i/n, i/n) = R_{m,n}(i/n)$.

   (iv) Set $\mathcal{M}R_{m,n}(1) = R_{m,n}(1)$ and for $i = 1, \ldots, n$ compute

   $$\mathcal{M}R_{m,n}\left(\frac{i-1}{n}\right) = \max_{j=1,\ldots,i} \max_{k=i,\ldots,n} \mathcal{S}R_{m,n}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right).$$

(v) Compute $M_{m,n}$. For $p=1$ we have

$$M_{m,n} = \frac{c_{m,n}}{n} \sum_{i=1}^{n} \left[\frac{1}{2} \mathcal{M}R_{m,n}\left(\frac{i-1}{n}\right) + \frac{1}{2} \mathcal{M}R_{m,n}\left(\frac{i}{n}\right) - R_{m,n}\left(\frac{i}{n}\right) \right].$$

For $p=2$ we have

$$M_{m,n} = \frac{c_{m,n}}{n^{1/2}} \left(\sum_{i=1}^{n} \left\{ \frac{1}{3} \left[\mathcal{M}R_{m,n}\left(\frac{i}{n}\right) - \mathcal{M}R_{m,n}\left(\frac{i-1}{n}\right) \right]^2 \right.\right.$$
$$\left.\left. + \left[\mathcal{M}R_{m,n}\left(\frac{i-1}{n}\right) - R_{m,n}\left(\frac{i}{n}\right) \right] \left[\mathcal{M}R_{m,n}\left(\frac{i}{n}\right) - R_{m,n}\left(\frac{i}{n}\right) \right] \right\} \right)^{1/2}.$$

2. Determine which of the relevant points in the unit cube belong to the estimated contact set.

(i) For $i = 0, \ldots, n$ and $j = 0, \ldots, i$ and $k = i, \ldots, n$ set $b'_{i,j,k} = 1$ if both

$$\mathcal{M}R_{m,n}\left(\frac{l}{n}\right) \leq R_{m,n}\left(\frac{l}{n}\right) + \delta_{m,n} \quad \text{for } l = i, j, k$$

and

$$\left| \mathcal{S}R_{m,n}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) - R_{m,n}\left(\frac{i}{n}\right) \right| \leq \delta_{m,n}$$

are satisfied, and set $b'_{i,j,k} = 0$ otherwise.

(ii) For $i = 0, \ldots, n$ and $j = 0, \ldots, i$ and $k = i, \ldots, n$, set $b_{i,j,k} = 1$ if $b'_{i,j',k'} = 1$ for all $j' = j, \ldots, i$ and all $k' = i, \ldots, k$, and set $b_{i,j,k} = 0$ otherwise.

3. Generate the bootstrap critical value.

(i) Generate bootstrap samples $X_1^*, \ldots, X_m^*$ and $Y_1^*, \ldots, Y_n^*$ by drawing with replacement from the original samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$.

(ii) For $i = 0, \ldots, n$ and $j = 0, \ldots, i$ and $k = i, \ldots, n$ compute $R_{m,n}^*(i/n)$ and $\mathcal{S}R_{m,n}^*(i/n, j/n, k/n)$ by following the procedure in steps 1(i)-1(iii).

(iii) For $i = 0, \ldots, n$ and $j = 0, \ldots, i$ and $k = i, \ldots, n$ compute

$$\tilde{S}T^*_{m,n}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) = c_{m,n}\left[\mathcal{S}R^*_{m,n}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) - R^*_{m,n}\left(\frac{i}{n}\right) - \mathcal{S}R_{m,n}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) + R_{m,n}\left(\frac{i}{n}\right)\right].$$

(iv) For $i = 0, \ldots, n$ compute

$$H^*_{m,n}\left(\frac{i}{n}\right) = \max_{j=0,\ldots,i}\max_{k=i,\ldots,n} b_{i,j,k}\tilde{S}T^*_{m,n}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right).$$

(v) Exact computation of $M^*_{m,n}$ is complicated. We suggest using the numerical approximation

$$M^*_{m,n} \approx \left[\frac{1}{n}\sum_{i=1}^{n} H^*_{m,n}\left(\frac{i}{n}\right)^p\right]^{1/p}.$$

(vi) Repeat steps 3(i)-3(v) N times, for some large N, to obtain a large number of realizations of $M^*_{m,n}$. Our bootstrap critical value $\mu_{m,n}(\alpha)$ is set equal to the $[\alpha N]$-th largest of these realizations. We reject the null if $M_{m,n} > \mu_{m,n}(\alpha)$. As a p-value we may take the smallest q such that $M_{m,n} > \mu_{m,n}(q)$.

4.2 Tuning parameter selection

Under Assumption 3.1 we are free to choose any tuning parameter $\delta_{m,n}$ that satisfies $\delta_{m,n} \to 0$ and $c_{m,n}\delta_{m,n} \to \infty$ as our sample sizes m and n increase. That is all well and good for the purposes of asymptotic thought experiments, but not a lot of help when it comes to choosing $\delta_{m,n}$ in practice. Some degree of ad hocery is difficult to avoid.

The following procedure for choosing $\delta_{m,n}$ has worked well for us in numerical simulations when $p = 1$ and $p = 2$. For a grid of candidate tuning parameters, use Monte Carlo simulation to compute the rejection rate of the test when $R = I$, the least favorable case for $p = 1$ and $p = 2$. Then, choose the smallest tuning parameter that yields a rejection rate acceptably close to the nominal size α. We have found in numerical simulations that the rejection rate of our test is below α at $R = I$ when $\delta_{m,n}$ is chosen very large, and rises above α at $R = I$ when $\delta_{m,n}$ becomes sufficiently small, so this should typically be possible. The selected tuning parameter will control the finite sample rejection rate at $R = I$ by construction, and we have found in numerical simulations that it delivers a finite sample rejection rate below nominal size at other points in the null.

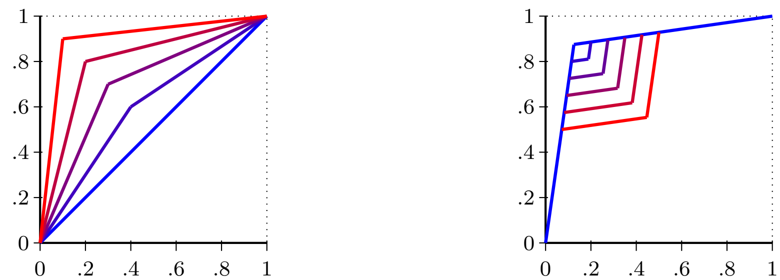
5 Finite sample performance

To investigate the finite sample performance of our proposed testing procedure we used Monte Carlo simulation to compute rejection rates at a range of ordinal dominance curves satisfying the null or alternative hypothesis. Here we report results obtained for equally sized samples with $m = n = 200$. Results for other sample sizes we investigated were qualitatively similar. For each ordinal dominance curve considered, we used 10000 Monte Carlo replications to compute rejection rates. We used the method of Giacomini et al. (2013) to reduce computation time, so bootstrap critical values were based on 10000 bootstrap samples drawn over the full set of Monte Carlo replications. Rejection rates were computed using $p = 1$ and $p = 2$. A tuning parameter value of $\delta_{m,n} = 0.08$ was used; at this value, preliminary simulations of the kind described in Section 4.2 indicated that the rejection rates at $R = I$ were close to but below 0.05.

The ordinal dominance curves used in our simulations were drawn from two parametric families. To investigate the behavior of our test when R is concave, we considered the parametrization

$$R_\gamma^0(u) = \begin{cases} \frac{1+\gamma}{1-\gamma}u & \text{if} \quad 0 \leq u \leq \frac{1-\gamma}{2} \\ \frac{1-\gamma}{1+\gamma}u + \frac{2\gamma}{1+\gamma} & \text{if} \quad \frac{1-\gamma}{2} \leq u \leq 1, \end{cases}$$

with $\gamma \in [0, 1)$. In panel (a) of Figure 5.1 we graph R_γ^0 for several values of γ. At $\gamma = 0$ the graph of R_γ^0 is the $45°$ line, while for $\gamma > 0$ the graph is piecewise affine with a single kink located at a point that moves toward the upper-left corner of the unit square as $\gamma \to 1$. This is the same family of curves considered in numerical simulations reported by Beare and Moon (2015, Figure 1), except that we have not bothered to smooth away the single kink appearing when $\gamma > 0$. This means that our kinked ordinal dominance curves violate the continuous differentiability condition imposed on members of Θ; however,



(a) R^0_γ for $\gamma \in \{0, 0.2, 0.4, 0.6, 0.8\}$. (b) R^1_γ for $\gamma \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

Figure 5.1: Ordinal dominance curves used to evaluate finite sample size and power.

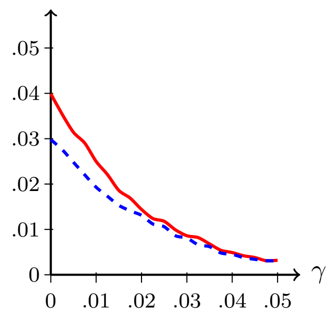
we have found that applying a small degree of smoothing to R^0_γ to restore continuous differentiability makes essentially no difference to the rejection rates computed.

To investigate the power of our test, we considered the parametrization

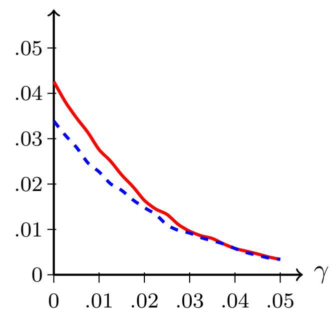
$$R^1_\gamma(u) = \begin{cases} 7u & \text{if} \quad 0 \leq u \leq \frac{7-3\gamma}{56} \\ \frac{1}{7}u + \frac{42-18\gamma}{49} & \text{if} \quad \frac{7-3\gamma}{56} \leq u \leq \frac{7+18\gamma}{56} \\ 7u - \frac{18\gamma}{7} & \text{if} \quad \frac{7+18\gamma}{56} \leq u \leq \frac{1+3\gamma}{8} \\ \frac{1}{7}u + \frac{6}{7} & \text{if} \quad \frac{1+3\gamma}{8} \leq u \leq 1, \end{cases}$$

with $\gamma \in [0, 7/3]$. In panel (b) of Figure 5.1 we graph R^1_γ for several values of γ. When $\gamma = 0$ we see that R^1_γ is a piecewise affine concave function with a single kink, and in fact we have $R^1_0 = R^0_{0.75}$. When $\gamma > 0$, R^1_γ is a piecewise affine nonconcave function with three kinks. As γ increases, R^1_γ moves further away from the concave function R^1_0; intuitively, we can think of R^1_γ as moving deeper into the alternative region as γ increases. Strictly speaking R^1_γ does not belong to Θ due to the violation of continuous differentiability, but as with R^0_γ this is a purely technical issue that can be overcome by applying a negligible degree of smoothing at kink points.

Figure 5.2 displays the rejection rates we computed for the concave ordinal dominance curves R^0_γ. We report rejection rates using a fixed critical value as in Carolan and Tebbs (2005) and Beare and Moon (2015) and using the bootstrap critical values proposed here. Nominal size was 0.05. In two panels corresponding to $p = 1$ and $p = 2$ we plot the



(a) $p = 1$



(b) $p = 2$

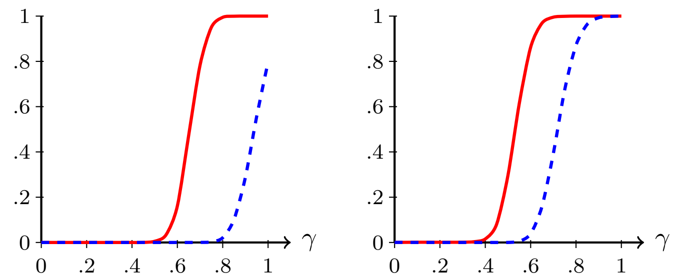
Figure 5.2: Null rejection rates with fixed critical values (dashed) and bootstrap critical values (solid).

rejection rates against the parameter γ.

The results for $p = 1$ and $p = 2$ are similar. In both cases the rejection rates using the fixed and bootstrap critical values are a little below the nominal size at $\gamma = 0$, the least favorable case. They drop very rapidly to zero as γ increases, becoming indistinguishable from zero at around $\gamma = 0.05$, and staying at that level as γ rises to one; we do not bother to plot the rejection rates for $\gamma > 0.05$. This is puzzling, because our theoretical results indicate that the limiting rejection rate using the bootstrap critical value should be 0.05 at all $\gamma \in [0, 1)$. We will say more about this shortly.

Figure 5.3 displays power curves for the family of ordinal dominance curves R^1_γ. The results for $p = 1$ and $p = 2$ are similar: power curves for both tests rise from zero to one as γ increases, with the test using bootstrap critical values easily outperforming the test using fixed critical values. With $p = 1$ and $\gamma = 0.8$, or with $p = 2$ and $\gamma = 0.6$, the improvement in power brought about by our bootstrap procedure is close to one. Comparing the power curves for $p = 1$ and $p = 2$, we see better performance with $p = 2$.

Why are the null rejection rates for the bootstrap test plotted in Figure 5.2 not approximately flat at 0.05, as suggested by Theorem 3.1? The most obvious answer would be that our sample sizes of $m = n = 200$ are too small, but in fact we have found in unreported simulations that the problem persists with much larger sample sizes. It is possible that the



(a) $p = 1$ (b) $p = 2$

Figure 5.3: Power curves with fixed critical values (dashed) and bootstrap critical values (solid).

limited finite sample relevance of Theorem 3.1 is a reflection of the fact that it establishes pointwise size control only at those points in the null where the ordinal dominance curve is not strictly concave. If the limiting rejection rates at nearby strictly concave ordinal dominance curves are zero or close to zero, that may go some way toward explaining the extremely low finite sample null rejection rates we observe when $p = 1$ and $p = 2$.

6 Final remarks

We have been concerned in this paper with the problem of testing whether a ratio of pdfs is nonincreasing. We proposed a bootstrap procedure based on preliminary estimation of a contact set that can deliver substantially greater power than existing tests based on fixed critical values. Numerical simulations indicate that our procedure remains conservative when $p = 1$ or $p = 2$.

It may be possible to adapt the methods developed here to more general hypothesis testing problems that can be formulated in terms of the concavity of some estimable function R, not necessarily an odc. If we have an estimator R_n of R such that $n^{1/2}(R_n - R)$ converges weakly to a continuous limit then, following the approach taken in this paper, it should be possible to use the functional delta method to determine the limit distribution of a test statistic $M_n := n^{1/2}\|\mathcal{M}R_n - R_n\|_p$, and to use Lemma 3.2 to motivate a bootstrap procedure based on preliminary estimation of a suitable contact set. A recent working paper by Seo (2014) takes this approach to construct a more powerful bootstrap version of a test of stochastic monotonicity proposed by Delgado and Escanciano (2012). There is an additional level of dimensionality to her problem, so the relevant contact set turns out to be a subset of the four dimensional unit hypercube. Similar improvements can presumably be made to a test of conditional stochastic dominance also proposed by Delgado and Escanciano (2013). More broadly, our results may be relevant in any situation where the lcm operator is used to construct a statistical test of concavity.

A Proofs

Here we provide proofs of all results stated in Section 3.

Proof of Lemma 3.1. Linearity is obvious, so we have $\sup|\mathcal{S}f_1 - \mathcal{S}f_2| = \sup|\mathcal{S}(f_1 - f_2)|$ for $f_1, f_2 \in \ell^\infty([0,1])$. Let $g = f_1 - f_2$. Since $\mathcal{S}g(u,v,w)$ is a convex combination of $g(v)$ and $g(w)$, it is bounded in absolute value by $\max\{|g(v)|, |g(w)|\} \leq \sup|g|$. And since $\mathcal{S}g(u,u,u) = g(u)$, we have $g(u) \leq \sup|\mathcal{S}g|$. Consequently, $\sup|\mathcal{S}g| = \sup|g|$, and our claim is proved. □

Proof of Lemma 3.2. Suppose first that R is affine in a neighborhood of u. In this case Lemma 2.1 implies that $\mathcal{M}'_R h(u) = \mathcal{M}_{[a_{R,u}, b_{R,u}]} h(u)$. Applying a result of Carolan (2002, Lemma 1) expressing the lcm as a supremum of secant segments, we may write

$$\mathcal{M}_{[a_{R,u}, b_{R,u}]} h(u) = \sup_{a_{R,u} \leq v \leq u} \sup_{u \leq w \leq b_{R,u}} \mathcal{S}h(u,v,w).$$

Since R is concave, the rectangle $[a_{R,u}, u] \times [u, b_{R,u}]$ is precisely the cross-section $B(u)$, and our claim is proved. Next suppose that R is not affine in a neighborhood of u. Since R is concave, for all $(v,w) \in B(u)$ we must have either $v = u$ or $w = u$, or both, and so $\sup_{(v,w)\in B(u)} \mathcal{S}h(u,v,w) = h(u)$. But Lemma 2.1 implies that $\mathcal{M}'_R h(u) = h(u)$, and so our claim is proved in this case also. □

Proof of Lemma 3.3. Since $B_{m,n} \subseteq B'_{m,n} \subseteq B''_{m,n}$ by construction, it suffices to show that

$P(B''_{m,n} \subseteq B^\epsilon) \to 1$ and that $P(B \subseteq B_{m,n}) \to 1$. We first show that $P(B''_{m,n} \subseteq B^\epsilon) \to 1$. Since $\tilde{\mathcal{S}}R$ is continuous and is equal to zero precisely on the contact set B, we have $\inf_{a \in A \setminus B^\epsilon} |\tilde{\mathcal{S}}R(a)| > 0$. Lemma 2.2 and the continuity of $\tilde{\mathcal{S}}$ imply the weak convergence $\tilde{\mathcal{S}}R_{m,n} \rightsquigarrow \tilde{\mathcal{S}}R$, so we also have

$$\sup_{a \in B''_{m,n}} |\tilde{\mathcal{S}}R(a)| = \sup_{a \in B''_{m,n}} |\tilde{\mathcal{S}}R_{m,n}(a)| + o_p(1) \leq \delta_{m,n} + o_p(1) = o_p(1),$$

the last equality following from Assumption 3.1. It follows that

$$P\left\{ \sup_{a \in B''_{m,n}} |\tilde{\mathcal{S}}R(a)| < \inf_{a \in A \setminus B^\epsilon} |\tilde{\mathcal{S}}R(a)| \right\} \to 1.$$

Consequently, $P(B''_{m,n} \cap (A \setminus B^\epsilon) = \varnothing) \to 1$, and so $P(B''_{m,n} \subseteq B^\epsilon) \to 1$.

We next show that $P(B \subseteq B_{m,n}) \to 1$. Using the linearity of $\tilde{\mathcal{S}}$ and the fact that $\tilde{\mathcal{S}}R(a) = 0$ for all $a \in B$, we find that

$$\sup_{a \in B} |\tilde{\mathcal{S}}R_{m,n}(a)| = c_{m,n}^{-1} \sup_{a \in B} |\tilde{\mathcal{S}}\left(c_{m,n}(R_{m,n} - R)\right)(a)|.$$

Therefore, since $\tilde{\mathcal{S}}(c_{m,n}(R_{m,n} - R)) \rightsquigarrow \tilde{\mathcal{S}}T$ by Lemma 2.2 and the continuous mapping theorem, we conclude in view of Assumption 3.1 that $\sup_{a \in B} |\tilde{\mathcal{S}}R_{m,n}(a)| = o_p(\delta_{m,n})$. This shows that $P(B \subseteq B''_{m,n}) \to 1$. Further, since R is concave, we may use the triangle inequality to write

$$\sup_{u \in [0,1]} |\mathcal{M}R_{m,n}(u) - R_{m,n}(u)| \leq \sup_{u \in [0,1]} |\mathcal{M}R_{m,n}(u) - \mathcal{M}R(u)| + \sup_{u \in [0,1]} |R_{m,n}(u) - R(u)|.$$

Both terms on the right-hand side of this inequality are $o_p(\delta_{m,n})$ under Assumption 3.1, and so $P(\mathcal{M}R_{m,n}(u) \leq R_{m,n}(u) + \delta_{m,n}) \to 1$ for every $u \in [0,1]$. Combined with the fact that $P(B \subseteq B''_{m,n}) \to 1$, this shows that $P(B \subseteq B'_{m,n}) \to 1$. Finally, we observe that when R is concave the cross-sections $B(u)$ are closed intervals, and so for each triple $(u, v, w) \in B$ we also have $(u, v', w') \in B$ for all pairs $(v', w') \in [v, u] \times [u, w]$. Since $P(B \subseteq B'_{m,n}) \to 1$, this shows that $P(B \subseteq B_{m,n}) \to 1$. $\square$

Proof of Lemma 3.4. This follows from Lemma 2.2 by applying the functional delta method for the bootstrap (see e.g. Kosorok, 2008, Theorem 2.9), provided that the mapping from pairs of distributions to the corresponding ordinal dominance curve satisfies a suitable Hadamard differentiability condition. This may be verified using well-known results on the Hadamard differentiability of the inverse and composition operators (see e.g. Kosorok, 2008, Lemmas 12.2 & 12.8(ii)) and the chain rule (see e.g. Kosorok, 2008, Lemma 6.19) provided that the density of G is bounded away from zero on its support. But in fact we may assume without loss of generality that G is the uniform distribution on $[0,1]$, since $R_{m,n}$ and $R^*_{m,n}$ are unaffected with probability one if we replace our observations X_i and Y_j with $G(X_i)$ and $G(Y_j)$. □

Proof of Theorem 3.1. By Lemma 3.3 there exists a sequence $\epsilon_n \downarrow 0$ such that

$$P(B \subseteq B_{m,n} \subseteq B'_{m,n} \subseteq B''_{m,n} \subseteq B^{\epsilon_n}) \to 1.$$

Let $g_n : \ell^\infty([0,1]) \to \mathbb{R}$ be the map $g_n(f) = \| \sup_{(v,w)\in B^{\epsilon_n}(\cdot)} \tilde{\mathcal{S}} f(\cdot, v, w)\|_p$, and let $g = \mathcal{A}'_R$, so that in view of Lemma 3.2 we have

$$P(g(T^*_{m,n}) \leq M^*_{m,n} \leq g_n(T^*_{m,n})) \to 1. \tag{A.1}$$

We will show that, for any sequence f_n in $\ell^\infty([0,1])$ with $f_n \to f_\infty \in C([0,1])$, we have

$$g_n(f_n) \to g(f_\infty). \tag{A.2}$$

The convergence (A.2) is the result of the following argument:

$$\begin{aligned}
|g_n(f_n) - g(f_\infty)| &\leq |g_n(f_n) - g_n(f_\infty)| + |g_n(f_\infty) - g(f_\infty)| \\
&\leq \sup_{a\in B^{\epsilon_n}} |\tilde{S} f_n(a) - \tilde{S} f_\infty(a)| + |g_n(f_\infty) - g(f_\infty)| \\
&\leq 2\|f_n - f_\infty\|_\infty + |g_n(f_\infty) - g(f_\infty)| \\
&\leq 2\|f_n - f_\infty\|_\infty + \sup_{(a_n,a'_n)\in B\times B^{\epsilon_n}:\|a_n - a'_n\|\leq \epsilon_n} |\tilde{\mathcal{S}} f_\infty(a_n) - \tilde{\mathcal{S}} f_\infty(a'_n)| \to 0.
\end{aligned}$$

Here, the first and second inequalities follow from the triangle inequality, the third inequality holds by Lemma 3.1, the fourth inequality holds by the definition of g_n and g, and the convergence to zero holds because $f_n \to f_\infty$, $\epsilon_n \downarrow 0$ and $\tilde{S} f_\infty$ is uniformly continuous. Lemma 3.4 together with (A.2) allows us to apply the extended continuous mapping theorem (see e.g. Dümbgen, 1993, p. 136) to obtain $g_n(T^*_{m,n}) \rightsquigarrow g(T)$ and $g(T^*_{m,n}) \rightsquigarrow g(T)$ conditional on the data in probability. In view of (A.1) and the definition of g, we conclude that $M^*_{m,n} \rightsquigarrow \mathcal{A}'_R T$ conditional on the data in probability.

It is clear from Theorem 3.1 of Beare and Moon (2015) that when R is not strictly concave the distribution function of $\mathcal{A}'_R T$ is continuous everywhere and strictly increasing on $[0,\infty)$. Continuity everywhere combined with the weak convergence $M^*_{m,n} \rightsquigarrow \mathcal{A}'_R T$ conditional on the data in probability implies (Kosorok, 2008, Lemma 10.11(i)) that

$$\sup_{x\in\mathbb{R}} \left| P(M^*_{m,n} \leq x \mid X_1,\ldots,X_m,Y_1,\ldots,Y_n) - P(\mathcal{A}'_R T \leq x) \right| = o_p(1). \tag{A.3}$$

Let $\mu(\alpha) = \inf\{x : P(\mathcal{A}'_R T \leq x) \geq 1-\alpha\}$, the $(1-\alpha)$-quantile of $\mathcal{A}'_R T$. Since the distribution function of $\mathcal{A}'_R T$ is strictly increasing at $\mu(\alpha)$, the continuous mapping theorem applied to (A.3) yields $\mu_{m,n}(\alpha) = \mu(\alpha) + o_p(1)$. It now follows from the weak convergence $M_{m,n} \rightsquigarrow \mathcal{A}'_R T$ ensured by Theorem 2.1, and the continuity of the distribution function of $\mathcal{A}'_R T$ at $\mu(\alpha)$, that $P(M_{m,n} > \mu_{m,n}(\alpha)) \to \alpha$ as claimed. □

References

Anderson, G. (1996). Nonparametric tests of stochastic dominance in income distributions. *Econometrica* **64** 1183–1193.

Anderson, G., Linton, O. and Whang, Y. -J. (2012). Nonparametric estimation and inference about the overlap of two distributions. *Journal of Econometrics* **171** 1–23.

Andrews, D. W. K. and Shi, X. (2013). Inference based on conditional moment inequalities. *Econometrica* **81** 609–666.

Barrett, G. F. and Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica* **71** 71–104.

Beare, B. K. (2011). Measure preserving derivatives and the pricing kernel puzzle. *Journal of Mathematical Economics* **47** 689–697.

Beare, B. K. and Moon, J. -M. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory* **31** 471–492.

Beare, B. K. and Schmidt, L. D. W. (2015). An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics*, in press.

Beutner, E., Wu, W. B. and Zähle, H. (2012). Asymptotics for statistical functionals of long-memory sequences. *Stochastic Processes and their Applications* **122**, 910–929.

Beutner, E. and Zähle, H. (2010). A modified functional delta method and its application to the estimation of risk functionals. *Journal of Multivariate Analysis* **101**, 2452–2463.

Beutner, E. and Zähle, H. (2012). Deriving the asymptotic distribution of u- and v-statistics of dependent data using weighted empirical processes. *Bernoulli* **18**, 803–822.

Carolan, C. A. (2002). The least concave majorant of the empirical distribution function. *Canadian Journal of Statistics* **30** 317–328.

Carolan, C. A. and Tebbs, J. M. (2005). Nonparametric tests for and against likelihood ratio ordering in the two sample problem. *Biometrika* **92** 159–171.

Davidson, R. and Duclos, J. -Y. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica* **68** 1435–1464.

Delgado, M. A. and Escanciano, J. C. (2012). Distribution-free tests of stochastic monotonicity. *Journal of Econometrics* **170** 68–75.

Delgado, M. A. and Escanciano, J. C. (2013). Conditional stochastic dominance testing. *Journal of Business and Economic Statistics* **31** 16–28.

Donald, S. G. and Hsu, Y. -C. (2015). Improving the power of tests of stochastic dominance. *Econometric Reviews*, in press.

Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields* **95** 125–140.

Dykstra, R., Kochar, S. and Robertson, T. (1995). Inference for likelihood ratio ordering in the two-sample problem. *Journal of the American Statistical Association* **90** 1034–1040.

Fang, Z. and Santos, A. (2014). Inference on directionally differentiable functions. Preprint, UC San Diego.

Giacomini, R., Politis, D. N. and White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory* **29** 567–589.

Hsieh, F. and Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics* **24** 25–40.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics, New York.

Lee, S. and Whang, Y. -J. (2009). Nonparametric tests of conditional treatment effects. CeMMAP Working Paper No. 36/09, Institute for Fiscal Studies.

Lee, S., Song, K. and Whang, Y. -J. (2014). Testing for a general class of functional inequalities. CeMMAP Working Paper No. 09/14, Institute for Fiscal Studies.

Linton, O., Maasoumi, E. and Whang, Y. -J. (2005). Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies* **72** 735–765.

Linton, O., Song, K. and Whang, Y. -J. (2010). An improved bootstrap test of stochastic dominance. *Journal of Econometrics* **154** 186–202.

Maasoumi, E. (2001). Parametric and nonparametric tests of limited domain and ordered hypotheses in economics. In Baltagi, B. (Ed.), *A Companion to Econometric Theory*. Blackwell Pub., Oxford.

Roosen, J. and Hennessy, D. A. (2004). Testing for the monotone likelihood ratio assumption. *Journal of Business and Economic Statistics* **28** 358–366.

Seo, J. (2014). Tests of stochastic monotonicity with improved size and power properties. Preprint, UC San Diego.

Shapiro, A. (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications* **66** 477–487.

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

Volgushev, S. and Shao, X. (2014). A general approach to the joint asymptotic analysis of statistics from subsamples. *Electronic Journal of Statistics* **8**, 390–431.