Nonparametric tests of density ratio ordering

Brendan K. Beare and Jong-Myun Moon*

Department of Economics, University of California, San Diego

May 1, 2014

Abstract

We study a family of nonparametric tests of density ratio ordering between two continuous probability distributions on the real line. Density ratio ordering is satisfied when the two distributions admit a nonincreasing density ratio. Equivalently, density ratio ordering is satisfied when the ordinal dominance curve associated with the two distributions is concave. To test this property, we consider statistics based on the L^p-distance between an empirical ordinal dominance curve and its least concave majorant. We derive the limit distribution of these statistics when density ratio ordering is satisfied. Further, we establish that, when $1 \leq p \leq 2$, the limit distribution is stochastically largest when the two distributions are equal. When $2 < p \leq \infty$, this is not the case, and in fact the limit distribution diverges to infinity along a suitably chosen sequence of concave ordinal dominance curves. Our results serve to clarify, extend and amend assertions appearing previously in the literature for the cases $p = 1$ and $p = \infty$. We provide numerical evidence confirming their relevance in finite samples.

*We thank Juan Carlos Escanciano, Hiroaki Kaido, Sokbae Lee, Andres Santos, Juwon Seo, Xiaoxia Shi, Joshua Tebbs, Yoon-Jae Whang, the anonymous referees, and seminar participants at Indiana University, Seoul National University, Sungkyunkwan University, the University of Cambridge, the University of Illinois at Urbana-Champaign, the University of Oxford, and Yonsei University for helpful comments.

1 Introduction

Statistical analyses in economics and other fields often involve comparing the shape of two probability distributions on the real line. In particular, there is a large literature on the construction of statistical tests of stochastic dominance between two distributions, and their application to various questions of empirical interest. Maasoumi (2001) provided a review of the state of this literature at the turn of the millennium. Subsequent contributions of importance include Barrett and Donald (2003), Linton et al. (2005, 2010), Schechtman et al. (2008), Delgado and Escanciano (2013) and Davidson and Duclos (2013). Recent work by Bera et al. (2013) on testing the equality of distributions is also relevant. Econometric applications of stochastic dominance testing include studies of income distributions (Anderson, 1996; Davidson and Duclos, 2000, 2013) and efficient portfolio choice (Scaillet and Topaloglou, 2010), among many others.

Stochastic dominance provides one way of ordering two probability distributions on the real line. A stronger notion of stochastic ordering is the *density ratio ordering*, more commonly known as the *likelihood ratio ordering*; we favor the former term. If F and G are two cdfs on $\mathbb{R}$ that admit a nonincreasing density ratio $\mathrm{d}F/\mathrm{d}G$, we say that G *density ratio dominates* F. Density ratio ordering is a stronger property than first-order stochastic dominance: if G density ratio dominates F, then G first-order stochastically dominates F. Further discussion of density ratio ordering may be found in Shaked and Shanthikumar (1994) and Thas (2009).

Density ratio ordering appears frequently as an assumption or implication of various models in economics, finance, and other fields. Roosen and Hennessy (2004) discussed a range of applications, including portfolio choice, crop insurance, mechanism design, and auction theory. Our primary motivation for studying this topic comes from a recent literature in empirical finance on the so-called pricing kernel puzzle. In this literature, the pricing kernel is defined as the ratio of risk neutral and physical densities for the payoff distribution of a market index at a future date. Classical financial theory dictates that this ratio should be nonincreasing; if not, there are perverse implications for the behavior of contingent claims, as discussed by Beare (2011). In the early 2000's, three influential papers by Aït-Sahalia and Lo (2000), Jackwerth (2000) and Rosenberg and Engle (2002) reported nonparametric pricing kernel estimates that exhibited an increasing region in the center of the return distribution, while decreasing elsewhere. This phenomenon has become known as the pricing kernel puzzle, though the statistical significance of departures from monotonicity remains an open question. See Hens and Reichlin (2013) for a recent discussion of the literature in this area.

Compared to the literature on testing stochastic dominance, the literature on testing density ratio ordering is at a much earlier stage of development. Dykstra et al. (1995) and Roosen and Hennessy (2004) proposed tests of density ratio ordering that are applicable to discrete distributions, or to continuous distributions that we first discretize by classifying observations into a finite number of bins. Such discretization has the effect of concentrating the power of the test on a finite dimensional subset of the alternative hypothesis. More recently, Carolan and Tebbs (2005, hereafter CT) proposed a test of density ratio ordering between continuous distributions that avoids unnecessary discretization and is consistent against arbitrary violations of density ratio ordering. In the present paper we build on the contribution of CT by generalizing their approach to a broader class of statistics, providing a rigorous foundation for their results, and clarifying some statements made regarding the asymptotic behavior of their tests.

The testing procedure proposed by CT relies on the insight that, if F and G are continuous cdfs admitting a well-defined density ratio, then G density ratio dominates F if and only if the composition $R = F \circ G^{-1}$ is concave. Here, $G^{-1}(u) = \inf\{y : G(y) \geq u\}$ is the quantile function corresponding to G. The function $R : [0, 1] \to [0, 1]$ is referred to as the *ordinal dominance curve* (odc) for F and G. CT propose to compare $R_{m,n}$, an empirical analogue to R constructed from independent samples of size m and n drawn from F and G, to its *least concave majorant* (lcm) – the smallest concave function lying above $R_{m,n}$. Their test statistic is calculated as the area or maximal distance between $R_{m,n}$ and its lcm, suitably scaled by m and n. CT argue that a conservative test may be obtained by comparing their statistics to the limit distribution they would achieve if in fact $R(u) = u$; this is taken to be the least favorable point in the null.

In this paper we generalize the approach of CT by considering test statistics based on the L^p-distance between $R_{m,n}$ and its lcm. We derive the limit distribution of these test statistics when R is concave and satisfies suitable smoothness conditions. This turns out to be a rather delicate matter. The main technical issue is that the lcm operator fails to satisfy the usual definition of Hadamard differentiability. It does, however, satisfy a weaker property dubbed *Hadamard directional differentiability* by Shapiro (1990, 1991), which suffices for the application of the functional delta method. We obtain the explicit form of the Hadamard directional derivative of the lcm operator. In addition to enabling a rigorous study of the asymptotic behavior of our test statistics, this contribution may be of broader relevance in other contexts. For instance, the tests of stochastic monotonicity and conditional stochastic dominance proposed by Delgado and Escanciano (2012, 2013) are constructed using the lcm operator; our results may conceivably lead to a better understanding of their asymptotic behavior.

Having established the limit distribution of our test statistics at each point in the null hypothesis, we next seek to identify a least favorable point at which to determine critical values. We show that $R(u) = u$ is indeed least favorable when our statistic is constructed using an L^p-distance with $p \in [1, 2]$. This means that the limit distribution of our test statistic when $R(u) = u$ first-order stochastically dominates the limit distribution under any other concave choice of R. Perhaps more surprisingly, we show that this is not the case when $p \in (2, \infty]$. In fact, the limit distribution of our test statistic diverges to infinity along a suitably chosen sequence of concave odcs. This result amends a claim made by CT, who asserted that $R(u) = u$ was least favorable when $p = 1$ and $p = \infty$. The practical implication is that we should require $p \in [1, 2]$ in applications.

The remainder of our paper is organized as follows. In Section 2 we formally describe our statistical framework and the construction of our test. Section 3 contains our derivation of the limit distribution of our test statistics when R is an arbitrary concave odc satisfying suitable smoothness conditions. Our results concerning the differential properties of the lcm operator are given here. In Section 4 we demonstrate that $R(u) = u$ is least favorable when $p \in [1, 2]$, but not when $p \in (2, \infty]$. Section 5 contains a short discussion of local power. Simulation results pertaining to the finite sample relevance of our asymptotic results are provided in Section 6, and we conclude in Section 7. Mathematical proofs are gathered together in the Appendix.

2 Statistical framework and test construction

Let F and G be two continuous cdfs on $\mathbb{R}$ with common support, and let $R = F \circ G^{-1}$. Let Θ denote the collection of nondecreasing, continuously differentiable maps $\theta : [0,1] \to [0,1]$ with $\theta(0) = 0$ and $\theta(1) = 1$. We maintain throughout that our odc $R \in \Theta$, and let R' denote its first derivative. Our null hypothesis is the set $\Theta_0 = \{\theta \in \Theta : \theta \text{ is concave}\}$, and our alternative hypothesis is the set $\Theta_1 = \Theta \setminus \Theta_0$.

Our test is constructed from two samples of real valued random variables $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$. Each sample is iid, with the X_i's having common cdf F and the Y_j's having common cdf G. The two samples are independent of one another. These basic sampling assumptions are taken from CT, and were also maintained by Barrett and Donald (2003) in the context of stochastic dominance testing. In that same context, much weaker sampling assumptions were adopted by Linton et al. (2005, 2010), who allowed for weak dependence within a given sample, and dependence between samples. Such generality appears difficult to accommodate in the present setting. Though it is relatively simple to generalize Lemma 3.1 below to accommodate weak dependence and dependence between samples, and thus proceed to a version of (3.3) involving a more general limit process T_R, further progress in the direction taken by this paper seems difficult. In particular, it is not clear to us how a version of Theorem 4.1, which identifies the least favorable point in the null when $p \in [1,2]$, might be proved under more general dependence conditions.

The construction of our test statistics involves the lcm operator $\mathcal{M}$, which we now formally define. We also take the opportunity to define the restricted lcm operator $\mathcal{M}_{[a,b]}$, which plays no role in the construction of our test statistics, but will prove useful when we describe their limiting behavior in Section 3. In what follows, we let $\ell^\infty([a,b])$ denote the collection of uniformly bounded real valued functions on $[a,b]$ equipped with the uniform metric, and let $\ell^\infty_{\text{co}}([a,b]) = \{f \in \ell^\infty([a,b]) : f \text{ is concave}\}$.

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

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

That is, $\mathcal{M}_{[a,b]} f$ is the pointwise infimum of those functions $g \in \ell^\infty_{\text{co}}([a,b])$ that majorize the restriction of f to $[a, b]$. We write $\mathcal{M}$ as shorthand for $\mathcal{M}_{[0,1]}$, and refer to $\mathcal{M}$ as the lcm operator.

Let $F_m(x) = m^{-1}\sum_{i=1}^m 1\{X_i \leq x\}$, $G_n(y) = n^{-1}\sum_{j=1}^n 1\{Y_j \leq y\}$, and $R_{m,n}(u) = F_m(G_n^{-1}(u))$, where $G_n^{-1}(u) = \inf\{y : G_n(y) \geq u\}$. To test the null hypothesis Θ_0 we consider statistics of the form

$$M_{m,n}^p = c_{m,n}\|\mathcal{M}R_{m,n} - R_{m,n}\|_p. \tag{2.1}$$

Here, $c_{m,n} = (mn/(m+n))^{1/2}$, $\|\cdot\|_p$ is the L^p-norm with respect to Lebesgue measure, and $p \in [1, \infty]$. When $p = 1$ or $p = \infty$, we obtain the two statistics proposed by CT for testing the null of density ratio ordering against its negation; they write $M_{m,n}^{*(12)}$ for $M_{m,n}^1$ and $D_{m,n}^{*(12)}$ for $M_{m,n}^\infty$. More accurately, $M_{m,n}^{*(12)}$ in CT is a computationally convenient approximation to $M_{m,n}^1$ constructed using isotonic regression. The approximation error is asymptotically negligible, as they note on p. 168.

Note that the empirical odc $R_{m,n}$ is invariant to any strictly increasing transformation of the data. That is, if we replace the samples $\{X_i\}_{i=1}^m$ and $\{Y_j\}_{j=1}^n$ with $\{\phi(X_i)\}_{i=1}^m$ and $\{\phi(Y_j)\}_{j=1}^n$, where ϕ is strictly increasing on the support of F and G, then $M_{m,n}^p$ is unaffected, with probability one. It follows that the sampling distribution of $M_{m,n}^p$ is uniquely determined by R. It is for this reason that we define our null and alternative hypotheses as subsets of a space of odcs, and not as subsets of a space of pairs of cdfs.

3 Limit distribution of test statistics

In this section we establish the limit distributions of the test statistics $M_{m,n}^p$ defined in (2.1), with $p \in [1, \infty]$, at each point $R \in \Theta_0$. We first state a well-understood result concerning the weak convergence of the normalized empirical odc $c_{m,n}(R_{m,n} - R)$. In what follows, $\rightsquigarrow$ denotes weak convergence in $\ell^\infty([0, 1])$.

Lemma 3.1. *Suppose $R \in \Theta$. Then as $m \wedge n \to \infty$ with $n/(m+n) \to \lambda \in (0, 1)$, we have $c_{m,n}(R_{m,n} - R) \rightsquigarrow T_R$, where T_R is a random element of $\ell^\infty([0, 1])$ satisfying*

$$T_R(u) = \lambda^{1/2}B_1(R(u)) + (1-\lambda)^{1/2}R'(u)B_2(u), \quad u \in [0, 1],$$

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

Lemma 3.1 is a consequence of Thas (2009, Theorem 7.6), and results cited therein. Hsieh and Turnbull (1996, Theorem 2.2) establish a strong approximation for $c_{m,n}(R_{m,n} - R)$, but weak convergence is sufficient for our purposes.

Suppose $R \in \Theta_0$. Since in this case $\mathcal{M}R = R$, we may write

$$M^p_{m,n} = \|c_{m,n}\left(\mathcal{D}R_{m,n} - \mathcal{D}R\right)\|_p, \tag{3.1}$$

where $\mathcal{D} = \mathcal{M} - \mathcal{I}$, and $\mathcal{I}$ is the identity operator on $\ell^\infty([0,1])$. In view of (3.1) and Lemma 3.1, an obvious approach to establishing the limit distribution of $M^p_{m,n}$ when $R \in \Theta_0$ is to apply the functional delta method to obtain weak convergence of the quantity inside the $\|\cdot\|_p$ in (3.1), and then invoke the continuous mapping theorem. Such an approach requires consideration of the differential properties of the lcm operator. The following definition is adapted from Shapiro (1990, 1991) and Bonnans and Shapiro (2000, Definition 2.45).

Definition 3.1. Let $\mathscr{X}$ and $\mathscr{Y}$ be topological vector spaces over the field of real numbers. A map $\phi : \mathscr{X} \to \mathscr{Y}$ is said to be Hadamard directionally differentiable at $x \in \mathscr{X}$ tangentially to $\mathscr{X}_0 \subseteq \mathscr{X}$ 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 $t_n \in \mathbb{R}_+$ with $z_n \to z \in \mathscr{X}_0$ and $t_n \downarrow 0$. $\phi'_x(z)$ is referred to as the Hadamard directional derivative of ϕ at x in direction z.

Hadamard directional differentiability is a weaker property than Hadamard differentiability because in the latter case we require $\phi'_x(z)$ to be a continuous, linear function of z. Continuity in z follows automatically from Hadamard directional differentiability (Bonnans and Shapiro, 2000, Proposition 2.46), but linearity does not.

In what follows, $C([0,1])$ denotes the set of continuous real valued functions on $[0,1]$.

Lemma 3.2. *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*

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

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* $\mathcal{M}'_R h(u) = h(u)$*.*

Lemma 3.2 establishes the existence and explicit form of the Hadamard directional derivative $\mathcal{M}'_R h$ when $R \in \Theta_0$ and $h \in C([0,1])$. The proof may be found in the Appendix. It is not difficult to see that our directional derivative $\mathcal{M}'_R h$ is not in general a linear function of h. For instance, let $R(u) = u$, $h_1(u) = u^2$ and $h_2(u) = -u^2$. Then $\mathcal{M}'_R h_1(u) = u$, $\mathcal{M}'_R h_2(u) = -u^2$ and $\mathcal{M}'_R(h_1 + h_2)(u) = 0$, implying that $\mathcal{M}'_R(h_1+h_2) \neq \mathcal{M}'_R h_1 + \mathcal{M}'_R h_2$. In fact, $\mathcal{M}'_R h$ is linear in h if and only if R is strictly concave. We will see shortly that our test statistic $M^p_{m,n}$ has a nondegenerate limit distribution only when R is concave but not strictly concave. For this reason, it is critical to establish Hadamard directional differentiability of $\mathcal{M}$ at points $R \in \Theta_0$ that are not strictly concave, rather than merely relying on the Hadamard differentiability of $\mathcal{M}$ at points $R \in \Theta_0$ that are strictly concave.

Standard versions of the functional delta method (see e.g. van der Vaart and Wellner, 1996, Theorem 3.9.4) require the operator under consideration to be Hadamard differentiable. However, Shapiro (1991, Theorem 2.1) established that, with no compensating loss in generality, the functional delta method applies more broadly under Hadamard directional differentiability. Applying Shapiro's result in conjunction with Lemma 3.1 and Lemma 3.2, we deduce that, when $R \in \Theta_0$, as $m \wedge n \to \infty$ with $n/(m+n) \to \lambda \in (0,1)$ we have

$$c_{m,n}\left(\mathcal{D}R_{m,n} - \mathcal{D}R\right) \rightsquigarrow \mathcal{M}'_R T_R - T_R. \tag{3.2}$$

Here we have used the fact that $\mathcal{D}$ is Hadamard directionally differentiable at R tangentially to $C([0,1])$, with $\mathcal{D}'_R = \mathcal{M}'_R - \mathcal{I}$. This is immediate from Lemma 3.2. The random process T_R, defined in Lemma 3.1, lies in $C([0,1])$ with probability one since R is continuously differentiable and the Brownian bridges B_1 and B_2 are continuous with probability one. Thus $\mathcal{M}'_R T_R$ is well-defined.

In view of (3.1), an application of the continuous mapping theorem to (3.2) delivers us

$$M^p_{m,n} \to_d \|\mathcal{M}'_R T_R - T_R\|_p \tag{3.3}$$

as $m \wedge n \to \infty$ with $n/(m+n) \to \lambda \in (0,1)$, whenever $R \in \Theta_0$. Our next result establishes an alternative, more easily interpretable representation for the limit distribution $\|\mathcal{M}'_R T_R - T_R\|_p$. Its statement requires some additional notation. Given an odc $R \in \Theta_0$ that is not strictly concave, there is a unique way to construct a finite or countable union of disjoint closed intervals $[a_k, b_k]$, $k \in K$, such that the restriction of R to each $[a_k, b_k]$ is affine, and such that R is strictly concave over any convex subset of $[0,1] \setminus \cup_{k\in K}[a_k, b_k]$. In fact, recalling Lemma 3.2, these intervals are precisely the intervals $[a_{R,u}, b_{R,u}]$ obtained as we allow u to vary over all points in $(0,1)$ at which R is locally affine. For each $k \in K$, let $d_k = b_k - a_k$, and let $h_k = R(b_k) - R(a_k)$. We suppress the dependence of a_k, b_k, d_k, h_k and K on R in our notation.

Theorem 3.1. *As $m \wedge n \to \infty$ with $n/(m+n) \to \lambda \in (0,1)$, if $R \in \Theta_0$ is not strictly concave then we have*

$$M^p_{m,n} \to_d \left(\sum_{k\in K} \left(\lambda h_k d_k^{2/p} + (1-\lambda) h_k^2 d_k^{(2-p)/p} \right)^{p/2} \|\mathcal{D}B_k\|_p^p \right)^{1/p}$$

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

$$M^\infty_{m,n} \to_d \sup_{k\in K} \left(\lambda h_k + (1-\lambda) h_k^2 d_k^{-1} \right)^{1/2} \|\mathcal{D}B_k\|_\infty,$$

where $\{B_k : k \in K\}$ is a mutually independent collection of standard Brownian bridges on $[0,1]$. If instead $R \in \Theta_0$ is strictly concave, we have $M^p_{m,n} \to_p 0$.

The proof of Theorem 3.1 may be found in the Appendix. We close this section with an enumeration of some of the key ways in which Theorem 3.1 serves to extend and amend certain results of CT.

1. Setting $R(u) = u$ in Theorem 3.1, we find that $M^p_{m,n} \to_d \|\mathcal{D}B\|_p$ for $p \in [1, \infty]$. Theorem 2 of CT asserts this result for the special case $p \in \{1, \infty\}$.

2. Lemma A1 of CT asserts that $M^p_{m,n} \to_p 0$ when R is strictly concave and $p \in \{1, \infty\}$, consistent with the more general statement for $p \in [1, \infty]$ given in Theorem 3.1.

3. Lemma A2 of CT states the limit distribution of $M^1_{m,n}$ when R is concave and piecewise linear. Here we do not allow for piecewise linear R, except in the trivial case where R is the 45° line, because we require that R be continuously differentiable. CT skirt technical issues relating to the smoothness of R in their arguments. If we could drop the requirement that R is continuously differentiable in Theorem 3.1, we would obtain the limit distribution of $M^1_{m,n}$ asserted by CT.

4. Lemma A2 of CT also asserts a limit distribution for $M^\infty_{m,n}$ when R is concave and piecewise linear. This limit distribution is not consistent with what we would obtain in Theorem 3.1 if we could relax the assumption that R is continuously differentiable, and appears to be incorrect. CT do not provide a proof for the claimed limit distribution of $M^1_{m,n}$.

5. Theorem 2 of CT asserts that the limit distribution of $M^\infty_{m,n}$ when R is the 45° line first-order stochastically dominates the limit distributions achieved at other concave odcs. This assertion is contradicted by Theorem 4.2 in Section 4, which follows easily from Theorem 3.1. The source of the inconsistency appears to be the incorrect limit distribution given for $M^\infty_{m,n}$ in Lemma A2 of CT.

4 When is the 45-degree line least favorable?

In Theorem 3.1 we established the limit distribution of $M^p_{m,n}$ at each point $R \in \Theta_0$. To obtain a valid critical value for testing Θ_0 against Θ_1, we would like to be able to say that the limit distribution of $M^p_{m,n}$ is stochastically largest at one particular point in Θ_0, in the sense of being first-order stochastically dominant over the limit distribution of $M^p_{m,n}$ at any other point in Θ_0. In this case, the point in Θ_0 at which the limit distribution of $M^p_{m,n}$ is stochastically largest is said to be least favorable. If we set our critical value c equal to the $1-\alpha$ quantile of the limit distribution of $M^p_{m,n}$ at this least favorable point, and reject Θ_0 when $M^p_{m,n} > c$, then our limiting rejection probability is no greater than α at any point in Θ_0.

When our odc is the 45° line $R(u) = u$, Theorem 3.1 implies that $M^p_{m,n} \to_d \|\mathcal{D}B\|_p$. The following result states that, when $p \in [1, 2]$, this odc is least favorable.

Theorem 4.1. *Suppose $p \in [1, 2]$ and $R \in \Theta_0$. The limit distribution of $M^p_{m,n}$ given in Theorem 3.1 is first-order stochastically dominated by $\|\mathcal{D}B\|_p$, where B is a standard Brownian bridge on $[0, 1]$.*

The proof of Theorem 4.1 may be found in the Appendix. Theorem 2 of CT established this result for $p = 1$. Our proof extends their arguments to allow $p \in [1, 2]$.

Our next result is perhaps more surprising. It implies that, for $M^p_{m,n}$ with $p \in (2, \infty]$, the 45° line $R(u) = u$ is *not* least favorable. In fact, the limit distribution of $M^p_{m,n}$ can be made arbitrarily large with an appropriate choice of $R \in \Theta_0$. This means that there cannot be any point in Θ_0 that is least favorable.

Theorem 4.2. *Suppose $p \in (2, \infty]$. For any $c \in (0, \infty)$ and $\varepsilon \in (0, 1)$, we may choose $R \in \Theta_0$ such that the limit distribution of $M^p_{m,n}$ given in Theorem 3.1 assigns probability of at least $1 - \varepsilon$ to the region (c, ∞).*

Theorem 4.2 contradicts Theorem 2 of CT, which states that $R(u) = u$ is least favorable when $p = \infty$. The proof may be found in the Appendix. The practical implication of Theorem 4.2 is that $M^p_{m,n}$ should not be used for testing Θ_0 against Θ_1 when $p > 2$, because there is no finite critical value that will control asymptotic size everywhere in Θ_0.

The proof of Theorem 4.2 involves showing that the limit distribution of $M^p_{m,n}$ diverges to infinity as we move along a sequence of odcs in Θ_0 which, in a neighbourhood of the origin, are affine and become successively steeper. Along this sequence, the terms $h_1^2 d_1^{(2-p)/p}$ and $h_1^2 d_1^{-1}$ appearing in the limit distributions in Theorem 3.1 diverge to infinity when $p > 2$, delivering the conclusion of Theorem 4.2. Additional regularity conditions may be imposed on Θ so as to rule out such sequences. For instance, we may assume the existence of a $c \in [1, \infty)$ such that $R' \leq c$ for all $R \in \Theta$. Such a condition rules out sequences of odcs along which the limit distribution of $M^p_{m,n}$ diverges, but does not ensure that the 45° line $R(u) = u$ is least favorable when $p > 2$. In simulations reported below (see in particular Figures 4.2(c) and 4.3(c)) we find that odcs exhibiting only a modest degree of steepness at the origin generate limit distributions that are not first-order stochastically dominated by the limit distribution at the 45° line.

At an intuitive level, the anomalous behavior of $M^p_{m,n}$ when $p > 2$ and R is steep near the origin may perhaps be understood in terms of the behavior of $n^{1/2}(G_n^{-1} - G^{-1})$, the empirical quantile process for G. Since $M^p_{m,n}$ is unaffected if we apply a strictly increasing transformation to the data X_i and Y_j, we may assume without loss of generality that F is uniform on $[0, 1]$. In this case, $G = R^{-1}$, and $n^{1/2}(G_n^{-1} - G^{-1}) \rightsquigarrow R'B_2$ in $\ell^\infty([0, 1])$, contributing the second component of the limit process T_R appearing in Lemma 3.1. Weak convergence of $n^{1/2}(G_n^{-1} - G^{-1})$ in $\ell^\infty([0, 1])$ is sensitive to the behavior of the pdf G': we typically require G' to be bounded away from zero on its support, so that the inverse operator is Hadamard differentiable at G (see e.g. van der Vaart and Wellner, 1996, Lemma 3.9.23). But when R is steep near the origin, G' is close to zero near the left endpoint of its support. We therefore cannot expect the weak convergence given in Lemma 3.1 to hold uniformly over the space Θ.

To illustrate the content of Theorem 4.1 and Theorem 4.2, we computed the quartiles of the limit distribution of $M^p_{m,n}$ for $p = 1, 2, \infty$ over a particular family of concave odcs. For $\delta \in [0, 1)$, we let R_δ be the odc given by

$$R_\delta(u) = \begin{cases} ru & \text{if } u \in [0, a] \\ (1 - t(u))^2 ra + 2(1 - t(u))t(u)rm + t(u)^2(b/r + c) & \text{if } u \in (a, b) \\ \frac{1}{r}u + c & \text{if } u \in [b, 1]\,, \end{cases} \tag{4.1}$$

where

$$r = \frac{1+\delta}{1-\delta}, \quad c = \frac{2\delta}{1+\delta}, \quad m = \frac{1-\delta}{2}, \quad a = \frac{4}{5r+4}, \quad b = \frac{(5-4c)r}{5r+4},$$

and

$$t(u) = \frac{1}{a + b - 2m}\left(a - m + \sqrt{m^2 - ab + (a + b - 2m)u}\right), \quad u \in [0, 1].$$

This seemingly mysterious construction is in fact very simple: R_δ is affine to the left of a and to the right of b, and between a and b is given by a quadratic Bézier curve (or conic arc) chosen to make the whole curve continuously differentiable. In Figure 4.1 we graph R_δ for $\delta = 0, 0.2, 0.4, 0.6, 0.8$. When $\delta = 0$, R_δ is the $45°$ line. When $\delta > 0$, R_δ has two affine segments connected by a rightward bend. As δ increases to one, the location of the bend moves toward the upper left corner of the unit square.

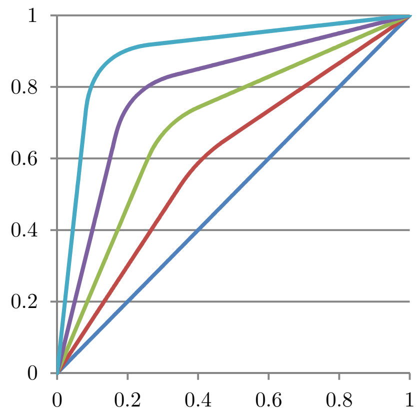
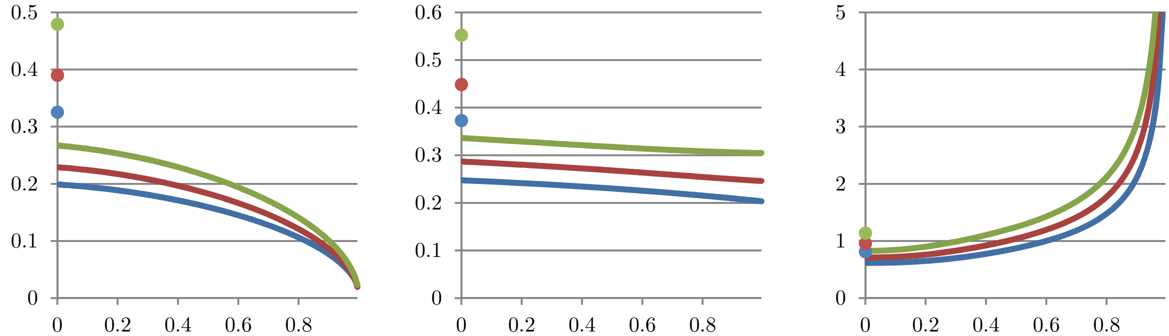


Figure 4.1: Ordinal dominance curves used to compute limiting quartiles in Figure 4.2, as defined in (4.1). We plot the curves corresponding to $\delta = 0, 0.2, 0.4, 0.6, 0.8$. The curves shift upward as δ increases.

The quartiles of the limit distribution of $M^p_{m,n}$ under each odc R_δ, $\delta \in [0, 1)$, are displayed in Figure 4.2. Each panel corresponds to a different value of p. The horizontal axes track the odc parameter δ. In each panel, the three quartile curves were computed over a grid of 200 evenly spaced values of $\delta \in [0, 1)$. At each value of δ, the quartiles were calculated from 10^6 random draws from the limit distribution of $M^p_{m,n}$ given in Theorem 3.1. We used an evenly spaced grid of 10^3 points to obtain a discrete approximation to the Brownian bridges appearing in that limit distribution.

The first thing to notice in Figure 4.2 is that the quartile curves are discontinuous at $\delta = 0$. In each panel, the three quartiles at $\delta = 0$ exceed their limits from the right. The discontinuity occurs because the bend introduced to R_δ when $\delta > 0$ fundamentally alters the limit distribution of $M^p_{m,n}$ given in Theorem 3.1: it is determined by a single Brownian bridge when $\delta = 0$, and by two independent Brownian bridges when $\delta > 0$. At a deeper level, the discontinuity of quartile curves is driven by the fact that the Hadamard directional derivative $\mathcal{M}'_R h$ of the lcm operator $\mathcal{M}$ is, for a fixed direction $h \in C([0, 1])$, not generally a continuous function of $R \in \Theta_0$. This means that different concave odcs that are close in the uniform metric may correspond to radically different limit distributions for $M^p_{m,n}$.

In panels (a) and (b) of Figure 4.2 we see that the limit distributions of $M^1_{m,n}$ and $M^2_{m,n}$

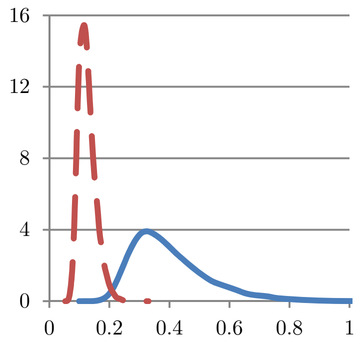


(a) $p = 1$ (b) $p = 2$ (c) $p = \infty$

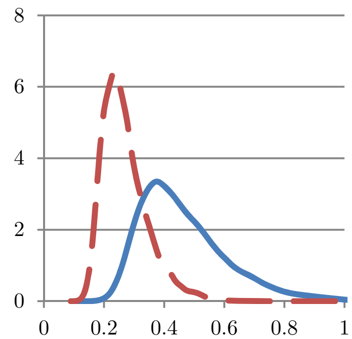
Figure 4.2: Quartiles of the limit distribution of $M^p_{m,n}$, with $p = 1, 2, \infty$. The horizontal axes track the parameter δ indexing the family of concave odcs defined in (4.1). Note that the quartile curves are discontinuous at $\delta = 0$.

do indeed appear to be stochastically largest at $\delta = 0$, as predicted by Theorem 4.1. As δ approaches one, the quartiles of the limit distribution shrink toward zero in panel (a), and gently decrease in panel (b). In panel (c) we see that the quartiles of the limit distribution of $M^\infty_{m,n}$ diverge toward infinity as δ approaches one, consistent with Theorem 4.2.

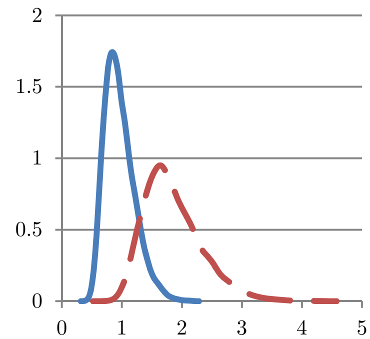
In Figure 4.3 we focus on just two of the odcs plotted in Figure 4.1 – R_0 and $R_{0.8}$ – and plot the entire limit distribution of $M^p_{m,n}$ for $p = 1, 2, \infty$. The limit distributions were computed by applying a kernel smoother to 10^4 draws from the relevant limit distributions, once again approximating the Brownian bridges using an evenly spaced grid of 10^3 points. Consistent with the first-order stochastic dominance predicted by Theorem 4.1, we see in panels (a) and (b) that the limit distribution of $M^p_{m,n}$ at R_0, the 45° line, is indeed mostly to the right of the limit distribution at $R_{0.8}$ if $p = 1$ or $p = 2$. In panel (c), where $p = \infty$, this is no longer true. In fact, not only is it clearly not the case in panel (c) that the limit distribution at R_0 first-order stochastically dominates the limit distribution at $R_{0.8}$, but the reverse may be true: the limit distribution at $R_{0.8}$ appears to first-order stochastically dominate the limit distribution at R_0.



(a) $p = 1$



(b) $p = 2$



(c) $p = \infty$

Figure 4.3: Limit distribution of $M^p_{m,n}$, with $p = 1, 2, \infty$. The solid line plots the limit distribution when R is the 45° line, while the dashed line plots the limit distribution when $R = R_{0.8}$, as defined in (4.1) and plotted in Figure 4.1.

5 Local power

In this section we present two closely related results concerning the local power of the testing procedure discussed above. These results apply for any $p \in [1, \infty]$, though it is clear from the results in the previous section that size control is problematic when $p > 2$. For simplicity, we assume here that the two sample sizes are equal: $m = n$. Proofs may be found in the Appendix.

Consider a sequence of odcs $R^{(1)}, R^{(2)}, \ldots \in \Theta_1$ with common lcm $R \in \Theta_0$. For each n, we observe independent iid samples $(X_1^{(n)}, \ldots, X_n^{(n)})$ and $(Y_1, \ldots, Y_n)$. Each $X_i^{(n)}$ is drawn from the cdf $F^{(n)} = R^{(n)}$, and each Y_j is drawn from the cdf G, which we assume without loss of generality to be uniform on $[0, 1]$. From these samples we compute the empirical cdfs $F_n^{(n)}(x) = n^{-1} \sum_{i=1}^n 1(X_i^{(n)} \le x)$ and $G_n(y) = n^{-1} \sum_{j=1}^n 1(Y_j \le y)$ and the empirical odc $R_n^{(n)}(u) = F_n^{(n)}(G_n^{-1}(u))$. We reject the null hypothesis Θ_0 when our test statistic $c_n \|\mathcal{D} R_n^{(n)}\|_p$, where $c_n = (n/2)^{1/2}$, exceeds some critical value $\tau \in (0, \infty)$.

We are interested in the limiting power of our test against sequences of alternatives $R^{(n)}$ that draw closer to Θ_0, in some sense, as $n \to \infty$. Loosely following the approach of Delgado and Escanciano (2012), we characterize the proximity of $R^{(n)}$ to Θ_0 by the quantity $\|\mathcal{D} R^{(n)}\|_p$, the L^p-distance between $R^{(n)}$ and its lcm. The following result states that, under a uniform boundedness condition on the first derivatives of the odcs $R^{(n)}$, the power of our test converges to one whenever the sequence of odcs approaches their common lcm $R \in \Theta_0$ at a rate slower than $n^{1/2}$.

Theorem 5.1. *Suppose the sequence of odcs* $R^{(1)}, R^{(2)}, \ldots \in \Theta_1$ *has uniformly bounded first derivatives and satisfies* $\lim_{n\to\infty} n^{1/2}\|\mathcal{D}R^{(n)}\|_p = \infty$. *Then*

$$\lim_{n\to\infty} P\left(c_n\|\mathcal{D}R_n^{(n)}\|_p > \tau\right) = 1.$$

Theorem 5.1 implies that our testing procedure is consistent against any fixed alternative $R \in \Theta_1$: merely set $R^{(n)} = R$ for all n.

Our next result implies that, loosely speaking, our testing procedure has high power against certain sequences of odcs in Θ_1 that approach Θ_0 at the rate $n^{1/2}$. Its statement is reminiscent of a result of Delgado and Escanciano (2012, Theorem 2).

Theorem 5.2. *Let* $\beta \in (0,1)$. *There exists* $\eta < \infty$ *such that, if the sequence of odcs* $R^{(1)}, R^{(2)}, \ldots \in \Theta_1$ *has uniformly bounded first derivatives and satisfies* $\liminf_{n\to\infty} n^{1/2}\|\mathcal{D}R^{(n)}\|_p \geq \eta$, *we have*

$$\liminf_{n\to\infty} P\left(c_n\|\mathcal{D}R_n^{(n)}\|_p > \tau\right) \geq \beta.$$

Theorem 5.2 provides an optimistic perspective on the local power of our testing procedure: we have nontrivial power against many sequences in Θ_1 that approach Θ_0 at the rate $n^{1/2}$. Of course, since the limit distribution of our test statistic is zero at all strictly concave members of Θ_0, and since those members constitute boundary points of Θ_0, it is easy to find examples of other sequences in Θ_1 converging to Θ_0 at the rate $n^{1/2}$ for which our limiting power is arbitrarily close to zero.

6 Finite sample simulations

Here we report numerical evidence pertaining to the finite sample relevance of the asymptotic results given in Sections 3, 4 and 5. We calculated the finite sample rejection rates

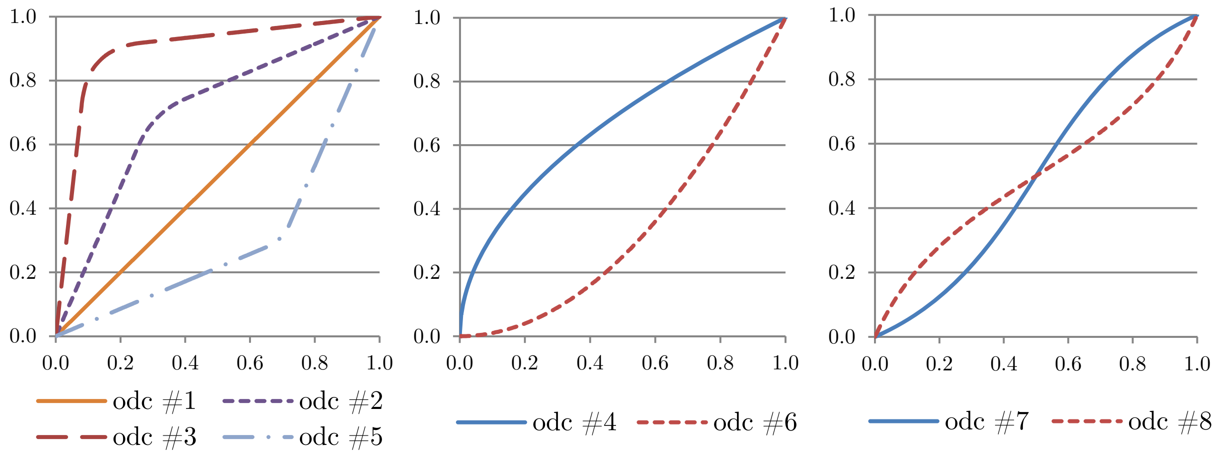


Figure 6.1: Odcs used to produce finite sample rejection rates given in Tables 6.1, 6.2 and 6.3. Odcs #1,2,3,5 are given by (4.1) with $\delta = 0, 0.4, 0.8, -0.4$ respectively. Odc #4 is given by $R(u) = u^{1/2}$. Odc #6 is given by $R(u) = u^2$. Odc #7 is given by $R(u) = 2^{-1} + 2^{-1}\arctan((2u-1)\tan(1))$. Odc #8 is given by $R(u) = 2^{-1} + (2\tan(1))^{-1}\tan(2u-1)$.

of the statistics $M^1_{m,n}$, $M^2_{m,n}$ and $M^\infty_{m,n}$ under eight different choices of R. These eight odcs are plotted in Figure 6.1. The first four odcs are concave, while the second four are nonconcave. We report numerical rejection rates for sample sizes $(m,n) = (20,20)$, $(20,50)$, $(50,50)$, $(200,500)$, $(500,500)$. The nominal size of all tests was 5%. Rejection rates were based on 10^4 experimental replications.

In implementing the tests, we compared each of the statistics $M^p_{m,n}$ to two different critical values (cvs). The first cv is the 0.95 quantile of $\|\mathcal{D}B\|_p$, the limit distribution of $M^p_{m,n}$ when $R(u) = u$. The second cv is the 0.95 quantile of the exact finite sample distribution of $M^p_{m,n}$ when $R(u) = u$; recall from our discussion in Section 2 that the sampling distribution of $M^p_{m,n}$ is uniquely determined by R. Though we lack a closed form expression for the finite sample distribution of $M^p_{m,n}$, it may be obtained by numerical simulation.

The observed rejection rates of $M^p_{m,n}$ when $p = 1$ and $p = 2$ are reported in Table 6.1 and Table 6.2. The results for $p = 1$ and $p = 2$ are mostly similar. As expected, the rejection rates for odc #1, the 45° line, are very close to 0.05 when we use exact cvs. If the exact cvs are correct, our rejection rates should lie within 0.0043 of 0.05 with approximately 95% probability; and indeed, we see that 9 of the 10 rejection frequencies for odc #1

		Asymptotic cvs					Exact cvs				
	m	20	20	50	200	500	20	20	50	200	500
odc #	n	20	50	50	500	500	20	50	50	500	500
1		.030	.032	.032	.046	.044	.053	.051	.046	.052	.049
2		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
3		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
4		.000	.000	.000	.000	.000	.001	.001	.000	.000	.000
5		.629	.808	.961	1.00	1.00	.730	.864	.973	1.00	1.00
6		.465	.612	.849	1.00	1.00	.581	.698	.881	1.00	1.00
7		.046	.049	.077	.590	.870	.084	.078	.111	.612	.883
8		.034	.060	.071	.576	.878	.066	.093	.102	.596	.890

Table 6.1: Finite sample rejection rates for $M^1_{m,n}$. Odcs #1–8 are plotted in Figure 6.1. Nominal size is 0.05. Odcs #1-4 are concave; odcs #5-8 are nonconcave.

		Asymptotic cvs					Exact cvs				
	m	20	20	50	200	500	20	20	50	200	500
odc #	n	20	50	50	500	500	20	50	50	500	500
1		.030	.035	.035	.047	.045	.058	.052	.048	.052	.047
2		.000	.000	.000	.000	.000	.000	.001	.000	.000	.000
3		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
4		.000	.001	.000	.000	.000	.001	.001	.000	.000	.000
5		.649	.824	.968	1.00	1.00	.749	.866	.977	1.00	1.00
6		.451	.600	.836	1.00	1.00	.569	.668	.869	1.00	1.00
7		.064	.080	.129	.751	.946	.109	.114	.165	.765	.949
8		.039	.068	.083	.648	.925	.074	.094	.113	.667	.928

Table 6.2: Finite sample rejection rates for $M^2_{m,n}$. Odcs #1–8 are plotted in Figure 6.1. Nominal size is 0.05. Odcs #1-4 are concave; odcs #5-8 are nonconcave.

using exact cvs fall in this region. With asymptotic cvs the rejection rates are close to 0.05 at the larger sample sizes we consider, but close to 0.03 with smaller samples. The rejection rates for odcs #2–4 are effectively zero using asymptotic or exact cvs. This reflects the fact that our cvs are chosen to control size at odc #1, the least favorable case, and should be conservative at other concave odcs. Turning to odcs #5–8, which are nonconcave, we see that rejection rates increase to one as the sample sizes increase, using

		Asymptotic cvs					Exact cvs				
	m	20	20	50	200	500	20	20	50	200	500
odc #	n	20	50	50	500	500	20	50	50	500	500
1		.013	.027	.025	.045	.042	.054	.051	.048	.050	.047
2		.007	.011	.018	.020	.043	.022	.022	.037	.023	.048
3		.024	.065	.146	.328	.588	.070	.102	.189	.344	.601
4		.000	.001	.000	.000	.000	.002	.002	.001	.000	.000
5		.470	.770	.948	1.00	1.00	.729	.841	.970	1.00	1.00
6		.219	.450	.686	1.00	1.00	.495	.575	.774	1.00	1.00
7		.052	.010	.153	.833	.970	.137	.165	.232	.848	.973
8		.017	.048	.057	.646	.920	.062	.077	.103	.668	.927

Table 6.3: Finite sample rejection rates for $M^{\infty}_{m,n}$. Odcs #1–8 are plotted in Figure 6.1. Nominal size is 0.05. Odcs #1-4 are concave; odcs #5-8 are nonconcave.

either asymptotic or exact cvs. The rejection rates for odcs #7–8 are much lower than the rejection rates for odcs #5–6 at smaller sample sizes. Power is greater using the exact cvs at smaller sample sizes. Comparing the rejection rates for $M^1_{m,n}$ and $M^2_{m,n}$ against the nonconcave odcs, we find that the two tests deliver similar power against odcs #5–6, while $M^2_{m,n}$ outperforms $M^1_{m,n}$ against odcs #7–8.

Rejection rates obtained using the statistic $M^{\infty}_{m,n}$ are reported in Table 6.3. The main feature to observe here is the excessive rejection rate obtained for odc #3, which is concave. Overrejection is not excessive at smaller sample sizes – indeed, with $m = n = 20$ and asymptotic cvs, the test is conservative – but the rejection rate is well in excess of nominal size at larger sample sizes, rising to over 0.5 when $m = n = 500$. This phenomenon illustrates the content of Theorem 4.2, which asserts that when $p > 2$, the limit distribution of $M^p_{m,n}$ diverges to infinity along a suitably chosen sequence of concave odcs. The proof of Theorem 4.2, located in the Appendix, involves showing that the limit distribution of $M^p_{m,n}$ can be made large by choosing a concave odc that is affine and steep near the origin. We see in Figure 6.1 that odc #3 fits this description.

7 Final remarks

In this paper, building on work by Carolan and Tebbs (2005), we studied a family of nonparametric tests of density ratio ordering between two cdfs F and G. We showed that, when $p \in [1, 2]$, pointwise asymptotic size control may be achieved by extracting critical values from the distribution of $M^p_{m,n}$ at $F = G$, the least favorable case. When $p \in (2, \infty]$, this approach breaks down, and the tests have asymptotic size one.

We have assumed throughout that the two samples used to construct the test statistic are iid and independent of one another. This assumption may be implausible in many applications, and its relaxation remains a priority for further research in this area. A second priority is the construction of more powerful tests of density ratio ordering, with critical values chosen to achieve correct asymptotic size at a wider range of points in the null, and not merely at the single point where $F = G$. It is conceivable that some version of the modified bootstrap technique used by Linton et al. (2010) to test for stochastic dominance may be adapted to apply in the present context. We leave the investigation of this possibility to future research.

A Proofs

Proof of Lemma 3.2. Our task is to show that $t_n^{-1}(\mathcal{M}(R + t_n h_n) - R) \to \mathcal{M}'_R h$ for any sequences $t_n \downarrow 0$ and $h_n \to h \in C[0, 1]$. It is known (see e.g. Durot and Tocquet, 2003, Lemma 2.2) that $\sup|\mathcal{M}f - \mathcal{M}g| \leq \sup|f - g|$ for any f, g. Consequently, $t_n^{-1} \sup|\mathcal{M}(R + t_n h_n) - \mathcal{M}(R + t_n h)| \leq \sup|h_n - h|$, and so it suffices for us to show that, for any $h \in C[0, 1]$, $t_n^{-1}(\mathcal{M}(R + t_n h) - R) \to \mathcal{M}'_R h$. We will do this by establishing pointwise monotone convergence $t_n^{-1}(\mathcal{M}(R + t_n h)(u) - R(u)) \downarrow \mathcal{M}'_R h(u)$ at each $u \in [0, 1]$, which implies uniform convergence by Dini's theorem. In what follows, let $\mathcal{M}'_{R,n} h = t_n^{-1}(\mathcal{M}(R + t_n h) - R)$, and let u be a fixed point in $[0, 1]$.

Since R is concave, the supporting hyperplane theorem ensures the existence of an affine function $\xi_u \in C([0, 1])$ such that $\xi_u(u) = R(u)$ and $\xi_u \geq R$. It is known (see e.g. Durot and Tocquet, 2003, Lemma 2.1) that $\mathcal{M}(f + g) = \mathcal{M}f + g$ for any f, g with g affine, and that $\mathcal{M}$ is positive homogeneous of degree one. We therefore have

$$\mathcal{M}'_{R,n}h(u) = t_n^{-1}\mathcal{M}(R + t_n h - \xi_u)(u) = \mathcal{M}\left(h + t_n^{-1}(R - \xi_u)\right)(u). \tag{A.1}$$

Since $\xi_u \geq R$, it is clear from (A.1) that $\mathcal{M}'_{R,n}h(u)$ is decreasing in n, and so it remains only to show the pointwise convergence $\mathcal{M}'_{R,n}h(u) \to \mathcal{M}'_R h(u)$. Let $h_{R,n,u} = h + t_n^{-1}(R - \xi_u)$. We will show that, for any fixed $\delta > 0$,

$$\mathcal{M}h_{R,n,u}(u) = \mathcal{M}_{[(a_{R,u}-\delta)\vee 0,(b_{R,u}+\delta)\wedge 1]}h_{R,n,u}(u) \tag{A.2}$$

for all n sufficiently large. A representation of the lcm in terms of a supremum of secant segments (Carolan, 2002, Lemma 1) allows us to write

$$\mathcal{M}h_{R,n,u}(u) = \sup_{u'\in[0,u]}\sup_{u''\in[u,1]} \frac{(u''-u)h_{R,n,u}(u') + (u-u')h_{R,n,u}(u'')}{u''-u'}, \tag{A.3}$$

with the ratio $0/0$ defined as $h_{R,n,u}(u) = h(u)$ if $u' = u'' = u$. Since ξ_u is affine with $\xi_u(u) = R(u)$, when $u' \neq u''$ the ratio in (A.3) is bounded from above by

$$\mathcal{M}h(u) + t_n^{-1}\left[\frac{(u''-u)R(u') + (u-u')R(u'')}{u''-u'} - R(u)\right]. \tag{A.4}$$

Now, since R is concave and there does not exist a left-neighborhood of $a_{R,u}$ or a right-neighborhood of $b_{R,u}$ in which R is affine, the term in square brackets in (A.4) is negative and bounded away from zero as (u', u'') ranges over the complement of $[(a_{R,u}-\delta)\vee 0, u] \times [u, (b_{R,u}+\delta)\wedge 1]$. By choosing n sufficiently large, we may therefore restrict the suprema in (A.3) to $u' \in [(a_{R,u}-\delta)\vee 0, u]$ and $u'' \in [u, (b_{R,u}+\delta)\wedge 1]$. But the right-hand side of (A.3) is then simply $\mathcal{M}_{[(a_{R,u}-\delta)\vee 0,(b_{R,u}+\delta)\wedge 1]}h_{R,n,u}(u)$ by the representation of Carolan (2002, Lemma 1), and so we have proved that (A.2) holds for all n sufficiently large.

If $a_{R,u} = b_{R,u} = u$, define $\mathcal{M}_{[a_{R,u},b_{R,u}]}h(u) = h(u)$. Then the claimed directional derivative $\mathcal{M}'_R h(u)$ is equal to $\mathcal{M}_{[a_{R,u},b_{R,u}]}h(u)$ regardless of whether R is affine in a neighborhood of u, and in view of (A.1) we now need only show that $\mathcal{M}h_{R,n,u}(u) \to \mathcal{M}_{[a_{R,u},b_{R,u}]}h(u)$.

Since (A.2) holds for all n sufficiently large, we have

$$\begin{aligned} & \limsup_{n\to\infty} \left|\mathcal{M}h_{R,n,u}(u) - \mathcal{M}_{[a_{R,u},b_{R,u}]}h(u)\right| \\ \leq \; & \limsup_{n\to\infty} \left|\mathcal{M}_{[(a_{R,u}-\delta)\vee 0,(b_{R,u}+\delta)\wedge 1]}h_{R,n,u}(u) - \mathcal{M}_{[a_{R,u},b_{R,u}]}h(u)\right|. \end{aligned} \tag{A.5}$$

And since $h_{R,n,u}(u') \leq h(u')$ for all u', with equality when $a_{R,u} \leq u' \leq b_{R_u}$, we have

$$\mathcal{M}_{[a_{R,u},b_{R,u}]}h(u) \leq \mathcal{M}_{[(a_{R,u}-\delta)\vee 0,(b_{R,u}+\delta)\wedge 1]}h_{R,n,u}(u) \leq \mathcal{M}_{[(a_{R,u}-\delta)\vee 0,(b_{R,u}+\delta)\wedge 1]}h(u). \tag{A.6}$$

The lower and upper bound in (A.6) do not depend on n, thereby allowing us to bound the right-hand side of (A.5) by their difference. We thus arrive at the inequality

$$\limsup_{n\to\infty} \left|\mathcal{M}h_{R,n,u}(u) - \mathcal{M}_{[a_{R,u},b_{R,u}]}h(u)\right| \leq \left|\mathcal{M}_{[(a_{R,u}-\delta)\vee 0,(b_{R,u}+\delta)\wedge 1]}h(u) - \mathcal{M}_{[a_{R,u},b_{R,u}]}h(u)\right|. \tag{A.7}$$

The right-hand side of (A.7) depends on δ, which was arbitrary. Letting $\delta \downarrow 0$, the right-hand side of (A.7) vanishes by virtue of the continuity of h, and we are done. □

Proof of Theorem 3.1. When R is strictly concave, Lemma 3.2 implies that $\mathcal{M}'_R T_R = T_R$, and so from (3.3) we have $M^p_{m,n} \to_p 0$, as claimed. Suppose instead that R is concave but not strictly concave. For $u \in [a_k, b_k]$, we know that $R(u) = R(a_k) + h_k d_k^{-1}(u - a_k)$ and $R'(u) = h_k d_k^{-1}$. Therefore, recalling the definition of T_R in Lemma 3.1, we have

$$T_R(u) = \lambda^{1/2}\bar{B}_1\left(R(a_k) + h_k d_k^{-1}(u - a_k)\right) + (1-\lambda)^{1/2}h_k d_k^{-1}\bar{B}_2(u) \quad \forall u \in [a_k, b_k],$$

where $\bar{B}_1$ and $\bar{B}_2$ are independent standard Brownian bridges on $[0,1]$. Let $\bar{W}_1$ and $\bar{W}_2$ be independent Wiener processes on $[0,1]$ such that $\bar{B}_1(u) = \bar{W}_1(u) - u\bar{W}_1(1)$ and $\bar{B}_2(u) = \bar{W}_2(u) - u\bar{W}_2(1)$. We define $T^*_R, T^+_R \in \ell^\infty([0,1])$ as follows. For $u \in [a_k, b_k]$, let

$$\begin{aligned} T^*_R(u) &= \lambda^{1/2}\bar{W}_1\left(R(a_k) + h_k d_k^{-1}(u - a_k)\right) + (1-\lambda)^{1/2}h_k d_k^{-1}\bar{W}_2(u) \\ T^+_R(u) &= \lambda^{1/2}\left(R(a_k) + h_k d_k^{-1}(u - a_k)\right)\bar{W}_1(1) + (1-\lambda)^{1/2}h_k d_k^{-1}u\bar{W}_2(1). \end{aligned}$$

For $u \notin \cup_{k\in K}[a_k, b_k]$, let $T^*_R(u) = T_R(u)$ and $T^+_R(u) = 0$. By construction, $T_R(u) = T^*_R(u) - T^+_R(u)$ for all $u \in [0,1]$. Note that T^+_R is affine over each interval $[a_k, b_k]$. Consequently, for $u \in [a_k, b_k]$ we have $\mathcal{M}_{[a_k,b_k]} T_R(u) = \mathcal{M}_{[a_k,b_k]} T_R^*(u) - T_R^+(u)$, and so

$$\mathcal{M}_{[a_k,b_k]} T_R(u) - T_R(u) = \mathcal{M}_{[a_k,b_k]} T_R^*(u) - T_R^*(u) \quad \forall u \in [a_k, b_k]. \tag{A.8}$$

For $k \in K$ and $u \in [0, 1]$, define

$$\begin{aligned}
\bar{W}_{1,k}(u) &= h_k^{-1/2} \left(\bar{W}_1 \left(R(a_k) + h_k u \right) - \bar{W}_1 \left(R(a_k) \right) \right) \\
\bar{W}_{2,k}(u) &= d_k^{-1/2} \left(\bar{W}_2 \left(a_k + d_k u \right) - \bar{W}_2 \left(a_k \right) \right).
\end{aligned}$$

The self-similarity property of Wiener processes implies that $\bar{W}_{1,k}$ and $\bar{W}_{2,k}$ are themselves Wiener processes. Moreover, the collection of Wiener processes $\{\bar{W}_{j,k} : j \in \{1, 2\}, k \in K\}$ is mutually independent. It follows that the random functions W_k, $k \in K$, defined by

$$W_k(u) = \frac{\lambda^{1/2} h_k^{1/2} \bar{W}_{1,k}(u) + (1-\lambda)^{1/2} h_k d_k^{-1/2} \bar{W}_{2,k}(u)}{\sqrt{\lambda h_k + (1-\lambda) h_k^2 d_k^{-1}}} \quad \forall u \in [0, 1],$$

also form a mutually independent collection of Wiener processes. We may now write

$$T_R^*(u) = \left(\lambda h_k + (1-\lambda) h_k^2 d_k^{-1} \right)^{1/2} W_k \left(d_k^{-1}(u - a_k) \right) + \gamma_{R,k} \quad \forall u \in [a_k, b_k],$$

where $\gamma_{R,k} = \lambda^{1/2} \bar{W}_1(R(a_k)) + (1-\lambda)^{1/2} h_k d_k^{-1} \bar{W}_2(a_k)$. Since $\gamma_{R,k}$ does not depend on u, and $\mathcal{M}_{[a_k,b_k]}$ is positive homogeneous of degree one, we conclude that

$$\mathcal{M}_{[a_k,b_k]} T_R^*(u) - T_R^*(u) = \left(\lambda h_k + (1-\lambda) h_k^2 d_k^{-1} \right)^{1/2} \mathcal{D} W_k \left(d_k^{-1}(u - a_k) \right) \tag{A.9}$$

for all $k \in K$ and $u \in [a_k, b_k]$.

In view of (3.3), it suffices for us to show that $\|\mathcal{M}_R' T_R - T_R\|_p$ has the distribution stated in Theorem 3.1. Suppose first that $p \in [1, \infty)$. From the form of the directional derivative $\mathcal{M}_R'$ given in Lemma 3.2, we see that

$$\|\mathcal{M}_R' T_R - T_R\|_p = \left(\sum_{k \in K} \int_{a_k}^{b_k} \left(\mathcal{M}_{[a_k,b_k]} T_R(u) - T_R(u) \right)^p \mathrm{d}u \right)^{1/p}.$$

(A.8), (A.9), and a simple change of variables may be used to show that

$$\int_{a_k}^{b_k} \left(\mathcal{M}_{[a_k,b_k]}T_R(u) - T_R(u)\right)^p \mathrm{d}u = \left(\lambda h_k d_k^{2/p} + (1-\lambda)h_k^2 d_k^{(2-p)/p}\right)^{p/2} \int_0^1 \mathcal{D}W_k\,(u)^p \,\mathrm{d}u$$

for each $k \in K$. Let $B_k(u) = W_k(u) - uW_k(1)$, a standard Brownian bridge. Since $\mathcal{D}W_k = \mathcal{D}B_k$, this completes our proof for the case $p \in [1,\infty)$. Suppose next that $p = \infty$. Lemma 3.2 implies that

$$\|\mathcal{M}'_R T_R - T_R\|_\infty = \sup_{k\in K} \sup_{u\in[a_k,b_k]} \left(\mathcal{M}_{[a_k,b_k]}T_R(u) - T_R(u)\right).$$

It is immediate from (A.8) and (A.9) that

$$\sup_{u\in[a_k,b_k]} \left(\mathcal{M}_{[a_k,b_k]}T_R(u) - T_R(u)\right) = \left(\lambda h_k + (1-\lambda)h_k^2 d_k^{-1}\right)^{1/2} \|\mathcal{D}W_k\|_\infty.$$

for each $k \in K$. Since $\mathcal{D}W_k = \mathcal{D}B_k$, our proof is complete for the case $p = \infty$ also. □

Proof of Theorem 4.1. If R is strictly concave then $M^p_{mn} \to_p 0$ by Theorem 3.1, and we are done. Suppose R is concave but not strictly concave. For $k \in K$, define

$$l_k = \left(\lambda h_k d_k^{2/p} + (1-\lambda)h_k^2 d_k^{(2-p)/p}\right)^{p/(p+2)}.$$

We need to show that $\|\mathcal{D}B\|_p^p$ first-order stochastically dominates $\sum_{k\in K} l_k^{(p+2)/2}\|\mathcal{D}B_k\|_p^p$. Begin by observing that

$$l_k = h_k^{p/(p+2)} \cdot d_k^{(2-p)/(p+2)} \cdot (\lambda d_k + (1-\lambda)h_k)^{p/(p+2)}.$$

The three exponents on the right-hand side of this equality satisfy

$$\frac{p}{p+2} + \frac{2-p}{p+2} + \frac{p}{p+2} = 1.$$

Therefore, since $p \le 2$, the well-known inequality between weighted geometric and arithmetic means implies that

$$l_k \leq \frac{p}{p+2}h_k + \frac{2-p}{p+2}d_k + \frac{p}{p+2}\left(\lambda d_k + (1-\lambda)h_k\right).$$

Consequently, we have

$$\begin{aligned}\sum_{k\in K} l_k &\leq \sum_{k\in K}\left(\frac{p}{p+2}h_k + \frac{2-p}{p+2}d_k + \frac{p}{p+2}\left(\lambda d_k + (1-\lambda)h_k\right)\right)\\ &= \frac{p(2-\lambda)}{p+2}\sum_{k\in K} h_k + \frac{2-p+\lambda p}{p+2}\sum_{k\in K} d_k\\ &\leq 1, \end{aligned} \tag{A.10}$$

since $\sum_{k\in K} h_k \leq 1$ and $\sum_{k\in K} d_k \leq 1$. The inequality (A.10) ensures the existence of a collection of closed intervals $[a_k^*, b_k^*] \subseteq [0,1]$, $k \in K$, such that the intersection of any two intervals contains at most one point, and such that $b_k^* - a_k^* = l_k$ for each $k \in K$. Let W be a Wiener process with $B(u) = W(u) - uW(1)$. Recalling that $\mathcal{D}B = \mathcal{D}W$, we have

$$\|\mathcal{D}B\|_p^p = \int_0^1 (\mathcal{M}W(u) - W(u))^p \,\mathrm{d}u \geq \sum_{k\in K}\int_{a_k^*}^{b_k^*}\left(\mathcal{M}_{[a_k^*,b_k^*]}W(u) - W(u)\right)^p \mathrm{d}u.$$

Since W is self-similar with independent increments, we may define a mutually independent collection of Wiener processes $\{W_k : k \in K\}$ such that

$$\int_{a_k^*}^{b_k^*}\left(\mathcal{M}_{[a_k^*,b_k^*]}W(u) - W(u)\right)^p \mathrm{d}u = l_k^{(p+2)/2}\int_0^1 (\mathcal{M}W_k(u) - W_k(u))^p \,\mathrm{d}u = l_k^{(p+2)/2}\|\mathcal{D}W_k\|_p^p$$

for each $k \in K$. Consequently, we have $\|\mathcal{D}B\|_p^p \geq \sum_{k\in K} l_k^{(p+2)/2}\|\mathcal{D}W_k\|_p^p$. Letting $B_k(u) = W_k(u) - uW_k(1)$, we obtain the desired first-order stochastic dominance. □

Proof of Theorem 4.2. Suppose $p \in (2,\infty)$. In view of Theorem 3.1, it is clear that the limit distribution of $M_{m,n}^p$ may be made to assign arbitrarily small probability to $[0,c]$ if we can choose $R \in \Theta_0$ to make $h_1^2 d_1^{(2-p)/p}$ sufficiently large. But since $p > 2$, this can always be achieved by making R linear and sufficiently steep near the origin. For instance, if we set $R = R_\delta$, the odc defined in (4.1), then $d_1 = (4-4\delta)/(9+\delta)$ and $h_1 = (4+4\delta)/(9+\delta)$. As $\delta \to 1$, we have $d_1 \to 0$ and $h_1 \to 0.8$, and hence $h_1^2 d_1^{(2-p)/p} \to \infty$ when $p \in (2,\infty)$.

We can therefore make $h_1^2 d_1^{(2-p)/p}$ arbitrarily large by choosing δ sufficiently close to one. Similarly, we can make the limit distribution of $M_{m,n}^{\infty}$ assign arbitrarily small probability to $[0, c]$ by choosing $R \in \Theta_0$ to make $h_1^2 d_1^{-1}$ sufficiently large, and this too is always possible by choosing δ sufficiently close to one in (4.1). □

Proof of Theorem 5.1. Begin by writing

$$c_n \|\mathcal{D}R_n^{(n)}\|_p \geq c_n \|\mathcal{D}R^{(n)}\|_p - c_n \| \left(\mathcal{M}R_n^{(n)} - \mathcal{M}R^{(n)}\right) - \left(R_n^{(n)} - R^{(n)}\right) \|_p. \tag{A.11}$$

Since the first term on the right-hand side of (A.11) diverges to infinity, it suffices for us to show that the second term is $O_p(1)$. By Durot and Tocquet (2003, Lemma 2.2) we have $\|\mathcal{M}R_n^{(n)} - \mathcal{M}R^{(n)}\|_\infty \leq \|R_n^{(n)} - R^{(n)}\|_\infty$, so if we can show that $c_n \|R_n^{(n)} - R^{(n)}\|_\infty = O_p(1)$ we are done.

Let $F = R$, the common lcm of the $R^{(n)}$'s, and let $X_1, \ldots, X_n$ be an iid sample drawn from F, independent of $Y_1, \ldots, Y_n$. Let F_n denote the empirical cdf of that sample. Note that $X_i \overset{d}{=} F^{-1}(F^{(n)}(X_i^{(n)}))$ for each i, where $\overset{d}{=}$ signifies equality of distribution. Consequently, as random elements of $\ell^\infty([0,1])$, we have $F_n^{(n)}(\cdot) \overset{d}{=} F_n(F^{-1}(F^{(n)}(\cdot)))$. We may therefore write

$$\begin{aligned} c_n \left(R_n^{(n)}(\cdot) - R^{(n)}(\cdot)\right) &\overset{d}{=} c_n \left(F_n(F^{-1}(F^{(n)}(G_n^{-1}(\cdot)))) - F^{(n)}(G^{-1}(\cdot))\right) \\ &= c_n \left(F_n(F^{-1}(F^{(n)}(G_n^{-1}(\cdot)))) - F^{(n)}(G_n^{-1}(\cdot))\right) \\ &\quad + c_n \left(F^{(n)}(G_n^{-1}(\cdot)) - F^{(n)}(G^{-1}(\cdot))\right) \end{aligned} \tag{A.12}$$

The first term on the right-hand side of (A.12) is uniformly $O_p(1)$ since $n^{1/2}(F_n(F^{-1}(\cdot)) - (\cdot)) \rightsquigarrow B$ in $\ell^\infty([0,1])$. The second term on the right-hand side of (A.12) is uniformly $O_p(1)$ since the first derivatives of the $F^{(n)}$'s are uniformly bounded and $n^{1/2}(G_n^{-1} - G^{-1}) \rightsquigarrow B$ in $\ell^\infty([0,1])$. Thus $c_n(R_n^{(n)}(\cdot) - R^{(n)}(\cdot))$ is uniformly $O_p(1)$, and the proof is complete. □

Proof of Theorem 5.2. The argument here is nearly the same as that used in the proof of Theorem 5.1. Let Z_n denote the second term on the right-hand side of (A.11), so that $c_n \|\mathcal{D}R_n^{(n)}\|_p \geq c_n \|\mathcal{D}R^{(n)}\|_p - Z_n$. Then $P(c_n \|\mathcal{D}R_n^{(n)}\|_p > \tau) \geq P(Z_n < c_n \|\mathcal{D}R^{(n)}\|_p - \tau)$. Since $Z_n = O_p(1)$ and $\liminf_{n\to\infty} n^{1/2} \|\mathcal{D}R^{(n)}\|_p \geq \eta$, for large enough η we will have $P(Z_n < c_n \|\mathcal{D}R^{(n)}\|_p - \tau) \geq \beta$ for all n sufficiently large. □

References

Aït-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics* **94** 9–51.

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

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.

Bera, A. K., Ghosh, A. and Xiao, Z. (2013). A smooth test for the equality of distributions. *Econometric Theory* **29** 419–446.

Bonnans, J. F. and Shapiro, A. (2000). *Perturbation Analysis of Optimization Problems.* Springer-Verlag, New York.

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.

Davidson, R. and Duclos, J. -Y. (2013). Testing for restricted stochastic dominance. *Econometric Reviews* **32** 84–125.

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.

Durot, C. and Tocquet, A. -S. (2003). On the distance between the empirical process and its concave majorant in a monotone regression framework. *Annales de l'Institut Henri Poincaré* (B) **39** 217–240.

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.

Hens, T. and Reichlin, C. (2013). Three solutions to the pricing kernel puzzle. *Review of Finance* **17** 1065–1098.

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

Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* **13** 433–451.

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.

Rosenberg, J. V. and Engle, R. F. (2002). Empirical pricing kernels. *Journal of Financial Economics* **64** 341–372.

Shaked, M. and Shanthikumar, J. (1994). *Stochastic Orders and their Applications.* Academic Press, 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.

Scaillet, O. and Topaloglou, N. (2010). Testing for stochastic dominance efficiency. *Journal of Business and Economic Statistics* **28** 169–180.

Schechtman, E., Shelef, A., Shlomo, Y. and Zitikis, R. (2008). Testing hypotheses about absolute concentration curves and marginal conditional stochastic dominance. *Econometric Theory* **24** 1044–1062.

Thas, O. (2009). *Comparing Distributions.* Springer-Verlag, New York.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.