Tests for the Second Order Stochastic Dominance Based on L-Statistics

José R. Berrendero and Javier Cárcamo
Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049 Madrid

Abstract

We use some characterizations of convex and concave-type orders to define discrepancy measures useful in two testing problems involving stochastic dominance assumptions. The results are connected with the mean value of the order statistics and have a clear economic interpretation in terms of the expected cumulative resources of the poorest (or richest) in random samples. Our approach mainly consists in comparing the estimated means in ordered samples of the involved populations. The test statistics we derive are functions of L-statistics and are generated through estimators of the mean order statistics. We illustrate some properties of the procedures with simulation studies and an empirical example.

Keywords: Convex order; second order stochastic dominance; Lorenz order; order statistics; hypothesis testing.

1 INTRODUCTION

Stochastic orders have shown to be useful notions in several areas of economics, such as inequality analysis, risks analysis or portfolio insurance. Since the beginning of the 1970s, stochastic dominance rules have been an essential tool in the comparison and analysis of poverty and income inequality. More recently, stochastic orders have also played an important role in the development of the theory of decision under risk and in actuarial sciences where they have been used to compare and measure different risks. Therefore, it is of major interest to acquire a deep understanding of the meaning and implications of the stochastic dominance assumptions. The construction of suitable empirical tools to make inferences about such assumptions is also clearly worthwhile.

The influential papers by Atkinson (1970) and Shorrocks (1983) are examples of theoretical works that provided a far-reaching insight into the importance of the stochastic dominance rules. In them, it was shown that the so called *Lorenz dominance* can be interpreted in terms of social welfare for increasing concave, but otherwise arbitrary, income-utility functions. The book by Lambert (1993) also supplies a nice and general exposition on this subject and other topics related to the theory of income distributions. On the other hand, the books by Goovaerts *et al.* (1990), Kaas *et al.* (1994) and Denuit *et al.* (2005) provide different applications of stochastic orders to actuarial sciences and risk analysis.

From the empirical point of view, there are many papers in the econometric literature that propose different kinds of statistical tests for hypotheses involving different stochastic orders. In them, income distributions (or financial risks) are compared according to different criteria. For instance, Anderson (1996) uses Pearson's goodness of fit type tests whereas the approach in Barrett and Donald (2003) and in Denuit *et al.* (2007) is inspired in the Kolmogorov-Smirnov statistic. See also Kaur *et al.* (1994), McFadden (1989) and Davidson and Duclos (2000) for other related procedures.

In this paper we analyze convex and concave-type orderings between two integrable random variables X and Y. We recall that X is said to be *less or equal to Y in the convex order*, and we write $X \leq_{\mathrm{cx}} Y$, if $\mathrm{E}(f(X)) \leq \mathrm{E}(f(Y))$, for every convex function f for which the previous expectations are well defined. The increasing convex order, to be denoted $\leq_{\mathrm{icx}}$, is defined analogously, but imposing on the convex functions to be also non decreasing. By replacing "convex" by "concave" in the definitions above, we obtain the *concave order* ($\leq_{\mathrm{cv}}$) and the *increasing concave order* ($\leq_{\mathrm{icv}}$).

The increasing convex and concave orders can be defined equivalently by:

$$X \leq_{\rm icx} Y \Longleftrightarrow \int_t^{\infty} \Pr(X > u)\,{\rm d}u \leq \int_t^{\infty} \Pr(Y > u)\,{\rm d}u, \quad t \in \mathbb{R},$$
$$X \leq_{\rm icv} Y \Longleftrightarrow \int_{-\infty}^{t} \Pr(X \leq u)\,{\rm d}u \geq \int_{-\infty}^{t} \Pr(Y \leq u)\,{\rm d}u, \quad t \in \mathbb{R}.$$

In this way, it is apparent that the increasing convex order compares the right tail of the distributions, while the increasing concave one focusses on the lowest part of the distributions. This difference leads to different uses of these two orders.

For instance, in actuarial sciences the risk associated to large-loss events is extremely important. Since convex functions take larger values when its argument is sufficiently large, if $X \leq_{\rm icx} Y$ holds, then Y is more likely to take "extreme values" than X (see Corollary 1 (a) below for the precise statement of this fact). Therefore, the risk associated with X is preferible to the one with Y. Actually, the partial order relations $\leq_{\rm icx}$ and $\leq_{\rm cx}$ are extensively used in the theory of decision under risk, where they are called *stop-loss order* and *stop-loss order with equal means*.

On the other hand, when comparing income distributions, it is sensible to analyze carefully the lowest part of the distributions, that is, the stratus in the populations with less resources. This is the reason why the increasing concave order is mainly considered in the literature of social inequality and welfare measurement under the name of the *second order stochastic dominance*. In this case, if $X \leq_{\rm icv} Y$, the distribution of the wealth in Y is considered to be more even than in X. If the populations to be compared have different positive expectation, it is very usual to normalize them by dividing by their respective means and then check if they are comparable with respect to the concave order. This approach leads to the *Lorenz order*, which is an essential tool in economics.

The convex and the concave orders are dual, and every property satisfied by one of them can be translated to the other one due to the relationships between the convex and the concave functions. Actually, it is easy to see that $X \leq_{\rm cx} Y$ if and only if $Y \leq_{\rm cv} X$, and $X \leq_{\rm icx} Y$ if and only if $-Y \leq_{\rm icv} -X$.

In this work, we use some characterizations of the (increasing) convex and concave orders related to the mean order statistics to define discrepancy measures useful in testing problems involving stochastic dominance assumptions. The characterizations have a clear economic interpretation in terms of the expected cumulative resources of the poorest (or richest) in a randomly selected sample of individuals from the population. The considered discrepancy measures yield in turn a testing approach quite different from those quoted above. The estimators which appear are functions of L-statistics and thus, in some situations, an asymptotic theory can be developed. However, in other cases computational techniques such as bootstrap are also required.

The paper is structured as follows: Section 2 includes some theoretical results linking the orderings with the mean order statistics. In Sections 3 and 4 we address two different testing problems: a) testing whether two distributions are ordered with respect to the increasing convex or concave order versus the alternative that they are not, and b) testing whether two distributions are equal against the alternative that one strictly dominates the other. Sections 5 and 6 include various simulation studies. A real data example is considered in Section 7. Section 8 summarizes the main conclusions of the paper. Finally, the proofs are collected in the appendix.

2 CONVEX-TYPE ORDERS AND MEAN ORDER STATISTICS

Throughout the paper, X and Y are integrable random variables with distribution functions F and G, respectively, and we denote by F^{-1} and G^{-1} their quantile functions, i.e., $F^{-1}(t) := \inf\{x : F(x) \geq t\}$, $0 < t < 1$. For a real function ω on $[0, 1]$, we define

$$\Delta_\omega(X, Y) := \int_0^1 \left(G^{-1}(t) - F^{-1}(t)\right) \omega(t)\, \mathrm{d}t, \tag{1}$$

whenever the above integral exists. The following theorem shows that $X \leq_{\mathrm{cx}} Y$ is characterized by the fact that (1) is nonnegative for non decreasing ω. Moreover, under a strict domination, $\Delta_\omega(X, Y)$ is necessarily positive for increasing weight functions. In the following, "$=_{\mathrm{st}}$" indicates the equality in distribution.

Theorem 1. *Let $\mathcal{I}$ denote the class of non decreasing real functions on $[0, 1]$ and $\mathcal{I}_0$ the subset of functions $\omega \in \mathcal{I}$ with the property $\omega(0) \geq 0$. Also, $\mathcal{I}^*$ stands for the subclass of strictly increasing functions of $\mathcal{I}$ and $\mathcal{I}_0^* := \mathcal{I}_0 \cap \mathcal{I}^*$. We have:*

(a) *$X \leq_{\mathrm{cx}} Y$ if and only if $\Delta_\omega(X, Y) \geq 0$, for all $\omega \in \mathcal{I}$. The equivalence remains true if "$\leq_{\mathrm{cx}}$" and "$\mathcal{I}$" are replaced by "$\leq_{\mathrm{icx}}$" and "$\mathcal{I}_0$", respectively.*

(b) *If $X \leq_{\mathrm{cx}} Y$ and $\Delta_\omega(X, Y) = 0$ for some $\omega \in \mathcal{I}^*$, then $X =_{\mathrm{st}} Y$. The result still holds if "$\leq_{\mathrm{cx}}$" and "$\mathcal{I}^*$" are replaced by "$\leq_{\mathrm{icx}}$" and "$\mathcal{I}_0^*$", respectively.*

The distance (1) is closely related to the expected value of the order statistics. For $k \geq 1$, let $(X_1, \ldots, X_k)$ and $(Y_1, \ldots, Y_k)$ be random samples from X and Y, respectively. $X_{i:k}$ and $Y_{i:k}$ denote the associated i-th order statistics, $i = 1, \ldots, k$.

For $k \geq 1$ and $1 \leq m \leq k$, let us denote by $S_{m:k}(X)$ (resp. $s_{m:k}(X)$) the expectation of the sum of the m greatest (resp. smallest) order statistics, i.e.,

$$S_{m:k}(X) := \sum_{i=k+1-m}^{k} \mathrm{E}X_{i:k}, \qquad s_{m:k}(X) := \sum_{i=1}^{m} \mathrm{E}X_{i:k}. \tag{2}$$

If X measures the income level of the individuals in a population, the function $S_{m:k}(X)$ (resp. $s_{m:k}(X)$) is nothing but the expected cumulative income of the m richest (resp. poorest) individuals out of a random sample of size k from the population. On the other hand, if X is a risk, $S_{m:k}(X)$ (resp. $s_{m:k}(X)$) measures the expected loss of the m largest (resp. lowest) loss events out of k as X.

Corollary 1. (a) *When $X \leq_{\text{icx}} Y$ we have, $S_{m:k}(X) \leq S_{m:k}(Y)$, for all $k \geq 1$ and $1 \leq m \leq k$. If additionally $S_{m:k}(X) = S_{m:k}(Y)$ for some $k \geq 2$ and $1 \leq m < k$, then $X =_{\text{st}} Y$.*

(b) *When $X \leq_{\text{icv}} Y$, we have $s_{m:k}(X) \leq s_{m:k}(Y)$, for all $k \geq 1$ and $1 \leq m \leq k$. If additionally $s_{m:k}(X) = s_{m:k}(Y)$ for some $k \geq 2$ and $1 \leq m < k$, then $X =_{\text{st}} Y$.*

The first statements in (a) and (b) above are also consequences of Corollary 2.1 in de la Cal and Cárcamo (2006). However, the second parts are new and required to derive the tests in Section 4.

By using Theorem 1 similar results can be obtained for the expected difference between the resources of the richest and the poorest. For instance, $X \leq_{\text{icv}} Y$ also implies that $\mathrm{E}(Y_{k:k} - Y_{1:k}) \leq \mathrm{E}(X_{k:k} - X_{1:k})$ for all $k \geq 2$. That is, the expected gap between the resources of the richest and the poorest individual (out of k) is lower for Y. Moreover, if this expected gap is equal for some $k \geq 2$, we necessarily have $X =_{\text{st}} Y$. Some results in the direction of Theorem 1 can also be found in Sordo and Ramos (2007).

3 TESTING STOCHASTIC DOMINANCE AGAINST NO DOMINANCE

When it is assumed that two populations are ordered, it is important to ensure that assumption is consistent with the data at hand. Accordingly, we consider the problem of testing the null hypothesis $\mathrm{H}_0 : \mathrm{X} \leq_{\text{icx}} \mathrm{Y}$ against the alternative

$\mathrm{H}_1 : \mathrm{X} \nleq_{\mathrm{icx}} \mathrm{Y}$ using two independent random samples $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ from X and Y, respectively. A test for the second order stochastic dominance $\mathrm{H}_0 : \mathrm{X} \leq_{\mathrm{icv}} \mathrm{Y}$ versus $\mathrm{H}_1 : \mathrm{X} \nleq_{\mathrm{icv}} \mathrm{Y}$ can be accomplished from the test considered in this section by just changing the sign of the data and exchanging the roles played by X and Y. For relevant papers dealing with the same testing problem we refer to Barrett and Donald (2003) and the references therein.

Our approach consists in comparing estimates of the mean order statistics of the two variables. Following the same lines as in the proof of Lemma 4.4 in de la Cal and Cárcamo (2006), it is easy to give a characterization of the alternative hypothesis using the functions $S_{m:k}(X)$ defined in (2):

$$\mathrm{H}_1 \text{ is true} \Longleftrightarrow \max_{1 \leq m < k} \big(S_{m:k}(X) - S_{m:k}(Y) \big) > 0, \quad \text{for all } k \geq k_0, \tag{3}$$

where k_0 depends on the distribution of the variables X and Y.

As a consequence, a sensible procedure to solve the testing problem is to estimate the above quantity and reject H_0 whenever the estimate is large enough. Suppose that H_1 holds and that the value of k is chosen so that $k \to \infty$ when $n_1, n_2 \to \infty$. Due to (3), we will eventually find a value of k for which there is at least a significant positive difference $\widehat{S}_{m:k}(X) - \widehat{S}_{m:k}(Y)$ (for some m), where $\widehat{S}_{m:k}(X)$ and $\widehat{S}_{m:k}(Y)$ are estimators of the quantities $S_{m:k}(X)$ and $S_{m:k}(Y)$, respectively. Hence, this procedure is expected to be asymptotically consistent. Indeed, the precise consistency result is established below.

We adopt a plug-in approach to estimate $S_{m:k}(X)$ and $S_{m:k}(Y)$ and we replace F and G with the empirical distributions F_{n_1} and G_{n_2}. That is,

$$\widehat{S}_{m:k}(X) := \sum_{j=k+1-m}^{k} \mathrm{E}_{F_{n_1}}(X_{j:k}) \quad \text{and} \quad \widehat{S}_{m:k}(Y) := \sum_{j=k+1-m}^{k} \mathrm{E}_{G_{n_2}}(Y_{j:k}).$$

In the end, our estimators are L-statistics since it can be readily shown that

$$\mathrm{E}_{F_n}(X_{j:k}) = \sum_{i=1}^{n} \left[j \binom{k}{j} \int_{(i-1)/n}^{i/n} t^{j-1}(1-t)^{k-j}\, \mathrm{d}t \right] X_{i:n}. \tag{4}$$

Once we know how to estimate $S_{m:k}(X)$ and $S_{m:k}(Y)$, we define the statistics

$$\hat{\Lambda}_{k,n_1,n_2} := \frac{1}{k} \max_{1 \leq m < k} \big(\widehat{S}_{m:k}(X) - \widehat{S}_{m:k}(Y) \big). \tag{5}$$

Our proposal is to reject H_0 when an appropriately normalized version of $\hat{\Lambda}_{k,n_1,n_2}$ is large enough, that is, we use the following critical region:

$$\left\{(1/n_1 + 1/n_2)^{-1/2}\hat{\Lambda}_{k,n_1,n_2} > c\right\}, \tag{6}$$

where $k \to \infty$ if $n_1, n_2 \to \infty$, and the critical value $c > 0$ is chosen so that the test has a preselected significance level, α, in the limiting case under H_0, i.e., in the case $X =_{\mathrm{st}} Y$. The following result collects some properties of the test defined by the previous rejection region. The limits in the theorem are taken as n_1 and n_2 go to infinity in a way such that $n_1/(n_1 + n_2) \to \lambda \in (0, 1)$. We note that the value of k implicitly depends on n_1 and n_2 since we have to estimate the expectations of the order statistics $\mathrm{E}X_{i:k}$ and $\mathrm{E}Y_{i:k}$ with samples of sizes n_1 and n_2, respectively. However, for the following asymptotic result it is only required that $k \to \infty$, without any additional restriction.

Theorem 2. *Let X be a random variable such that $\int_0^\infty \sqrt{\Pr(|X| > x)}\,\mathrm{d}x < \infty$.*

(a) *Under* H_1*,* $\lim \Pr(\text{reject } \mathrm{H}_0) = 1$*.*

(b) *Under* H_0*, if* $X \neq_{\mathrm{st}} Y$*, then* $\lim \Pr(\text{reject } \mathrm{H}_0) = 0$*.*

The condition $\mathrm{E}|X|^{2+\delta} < \infty$, for some $\delta > 0$, implies $\int_0^\infty \sqrt{\Pr(|X| > x)}\,\mathrm{d}x < \infty$. Therefore, the procedure described above yields a consistent test under a condition only slightly stronger than the existence of the second moment. If one is only willing to assume $\mathrm{E}|X|^\gamma < \infty$ for some $1 < \gamma < 2$, then the rate of convergence is slower and the consistency requires to modify slightly the test (see Remark 1 in the Appendix). In such a case, the critical region (6) has to be replaced by

$$\left\{(1/n_1 + 1/n_2)^{1/\gamma - 1}\hat{\Lambda}_{k,n_1,n_2} > c\right\}. \tag{7}$$

The previous theorem does not require the variables to have finite support, neither the continuity of the distribution functions. However, it does not make any statement over the asymptotic size of the test, that is, it is not shown that $\limsup_{n_1,n_2\to\infty} \sup \Pr(\text{reject } \mathrm{H}_0) = \alpha$, where the supremum is taken over all X and Y under the null.

Theorem 2 assumes that we know the distribution of the test statistics when $X =_{\mathrm{st}} Y$ so that we can determine the threshold value c beyond which we reject for each significance level α. Unfortunately, it seems extremely difficult to derive such a distribution, which in general depends on the underlying unknown distribution of the data. To apply the test in practice we propose to rely on a bootstrap approximation to simulate p-values, according to the following scheme:

1. Compute $\hat{\Lambda}_{m,n_1,n_2}$ from the original samples $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$.

2. Consider the pooled data set $Z_1, \ldots, Z_n$, with $n = n_1 + n_2$, and resample with replacement to obtain $Z_1^*, \ldots, Z_n^*$. Divide this bootstrap sample into two parts $X_1^*, \ldots, X_{n_1}^*$ and $Y_1^*, \ldots, Y_{n_2}^*$. Use these two parts to compute a bootstrap version of the test statistics $\hat{\Lambda}^*_{m,n_1,n_2}$.

3. Repeat step 2 a large number B of times, yielding B bootstrap test statistics $\hat{\Lambda}^{*(b)}_{m,n_1,n_2}$, $b = 1, \ldots, B$.

4. The p-value of the proposed test is given by $p := \text{Card}\{\hat{\Lambda}^{*(b)}_{m,n_1,n_2} > \hat{\Lambda}_{m,n_1,n_2}\}/B$. We reject at a given level α whenever $p < \alpha$.

The null hypothesis of the test is composite. By resampling from the pooled sample we approximate the distribution of our test statistics when $X =_{\text{st}} Y$, which represents the least favorable case for H_0. If the probability of rejection in this case is approximately α, then α is expected to be an upper bound for the probability of rejection under other less advantageous situations. This idea has been confirmed in the simulation study carried out in Section 5. Similar bootstrap approximations have been successfully applied in an alike context in Abadie (2002) and Barrett and Donald (2003).

4 TESTING STOCHASTIC EQUALITY AGAINST STRICT DOMINANCE

In this section we want to test if two variables are equally distributed against the alternative that they are strictly ordered. Hence, we provide a test for the null hypothesis $\text{H}_0 : X =_{\text{st}} Y$ against the alternative $\text{H}_1 : X \leq_{\text{icx}} Y$ and $X \neq_{\text{st}} Y$ (or its dual, $\text{H}_1 : X \leq_{\text{icv}} Y$ and $X \neq_{\text{st}} Y$).

This kind of unidirectional tests may appear naturally in economics where a certain change in the scenario, for example a change in tax policy or a technology shock, is expected to produce a decrease or increase in the variability of the variables such as income or stock return, and therefore is not unreasonable to assume that the distributions are ordered. The goal is to determine if they are different.

In other occasions we can apply a two step procedure. First, we use the test described in Section 3 to show that it is reasonable to assume that the variables are ordered and afterward apply this test. One possible solution to control the significance level of the final test is to use the *Bonferroni method*. For example, to test the two hypotheses on the same data at 0.05 significance level, instead of using a p-value threshold of 0.05, one would use 0.025.

It is important to remark that these tests use the additional information that the variables are ordered, and thus the corresponding power is by far much higher than the power of the usual tests of equality of distribution, as the Kolmogorov-Smirnov test, for instance. Some illustrations regarding this question are included in Subsection 6.3 (see also Figure 2).

Let us consider first the test for the increasing convex order. Theorem 1 and Corollary 1 provide a large number of potential discrepancy measures on which our test statistic could be based on. Among them, we have opted for a relatively simple one, namely, the difference between the expected value of the maximum of $k \geq 2$ observations. By selecting $m = 1$ in Corollary 1 (a), we obtain:

If $X \leq_{\mathrm{icx}} Y$*, then* $\mathrm{E}X_{k:k} \leq \mathrm{E}Y_{k:k}$*. Moreover, if additionally* $\mathrm{E}X_{k:k} = \mathrm{E}Y_{k:k}$ *for some* $k \geq 2$*, then* $X =_{\mathrm{st}} Y$*.*

Therefore, we consider the discrepancy $\Delta_k(X, Y) := \mathrm{E}Y_{k:k} - \mathrm{E}X_{k:k}$, $k \geq 2$. $\Delta_k(X, Y) = 0$ under H_0, while $\Delta_k(X, Y) > 0$ under H_1. A natural idea is to estimate $\Delta_k(X, Y)$ and reject H_0 whenever the estimate is large enough. As before, to estimate $\Delta_k(X, Y)$ we replace F and G with the empirical distributions, F_{n_1} and G_{n_2}. The resulting estimators are obtained by setting $j = k$ in (4).

$$\hat{\Delta}_{k,n_1,n_2} := \sum_{i=1}^{n_2} \omega_{i,k,n_2} Y_{i:n_2} - \sum_{i=1}^{n_1} \omega_{i,k,n_1} X_{i:n_1}, \tag{8}$$

where for $k \geq 2$, $n \geq 1$ and $1 \leq i \leq n$, the weights are given by

$$\omega_{i,k,n} := \left(\frac{i}{n}\right)^k - \left(\frac{i-1}{n}\right)^k. \tag{9}$$

Theorem 3. *Assume that* H_0 *holds and that the common distribution* F *satisfies*

$$\int_{-\infty}^{\infty} F(x)^{2k-2} x^2 \mathrm{d}F(x) < \infty. \tag{10}$$

Assume also that n_1/n *tends to* $\lambda \in (0, 1)$*, as* $n \to \infty$*, where* $n := n_1 + n_2$*. Let* U *be a random variable uniformly distributed on* $(0, 1)$*. Define the random variable*

$$W := kU^{k-1}F^{-1}(U) + k(k-1)\int_U^1 t^{k-2}F^{-1}(t)\,\mathrm{d}t \tag{11}$$

and let σ_W^2 *be the variance of* W*. Then, as* $n_1 \to \infty$*,*

$$(1/n_1 + 1/n_2)^{-1/2}\,\hat{\Delta}_{k,n_1,n_2} \longrightarrow_d \mathrm{N}(0, \sigma_W^2), \tag{12}$$

where the symbol $\longrightarrow_d$ *stands for the convergence in distribution, and* $\mathrm{N}(0, \sigma_W^2)$ *is a normal random variable with mean* 0 *and variance* σ_W^2*.*

Hence, to test H_0 against H_1, we can use the simple critical region

$$\left\{(1/n_1 + 1/n_2)^{-1/2}\hat{\Delta}_{k,n_1,n_2}/\hat{\sigma}_W > z\right\}, \tag{13}$$

where $\hat{\sigma}_W$ is any consistent estimate of σ_W and z is a quantile of the standard normal distribution that depends on a suitable significance level.

To estimate σ_W we proceed as follows: given U with uniform distribution on (0,1), we obtain a pseudo-value of W, $\hat{W}$, by replacing in (11) the unknown true distribution F by the pooled empirical distribution F_n, $n = n_1 + n_2$, computed from the observations of the two samples (since under H_0 both samples come from the same distribution). Let $Z_{1:n} \leq \cdots \leq Z_{n:n}$ be the order statistics of the pooled sample. After some computations, $\hat{W}$ can be expressed as:

$$\hat{W} = k\left(\frac{\lceil nU\rceil}{n}\right)^{k-1} Z_{\lceil nU\rceil:n} + k\sum_{i=\lceil nU\rceil+1}^{n} \omega_{i,k-1,n} Z_{i:n},$$

where $\lceil\cdot\rceil$ is the ceiling function and $\omega_{i,k-1,n}$ is defined as in (9). Finally, we generate a large number of pseudo-values and compute its standard deviation. According to our computations, 5000 pseudo-values are enough to obtain a precise estimate $\hat{\sigma}_W$. The consistency of this estimator is analyzed in Proposition 1 of the Appendix.

If we are interested in the alternative hypothesis $H_1 : X \leq_{\text{icv}} Y$ and $X \neq_{\text{st}} Y$, some modifications of the procedure described above are needed. In this case, the appropriate discrepancy measures are $\Gamma_k := \mathrm{E}Y_{1:k} - \mathrm{E}X_{1:k}$, $k \geq 2$. This quantity can be estimated by setting $j = 1$ in (4). That is,

$$\hat{\Gamma}_{k,n_1,n_2} := \sum_{i=1}^{n_2} \gamma_{i,k,n_2} Y_{i:n_2} - \sum_{i=1}^{n_1} \gamma_{i,k,n_1} X_{i:n_1},$$

where now the weights are given by

$$\gamma_{i,k,n} := \left(1 - \frac{i-1}{n}\right)^k - \left(1 - \frac{i}{n}\right)^k. \tag{14}$$

Comparing (9) with (14) we see that when we are interested in the second order dominance, we use an L-statistics that places more weight on the lowest order statistics whereas for the increasing convex order, the highest order statistics receive more weight.

The proof of the following theorem is analogous to that of Theorem 3 so that it is omitted.

Theorem 4. *Assume that* H_0 *holds and that the common distribution* F *satisfies*

$$\int_{-\infty}^{\infty} (1 - F(x))^{2k-2} x^2 \, \mathrm{d}F(x) < \infty. \tag{15}$$

Assume also that n_1/n *tends to* $\lambda \in (0,1)$, *as* $n \to \infty$, *where* $n := n_1 + n_2$. *Let* U *be a random variable uniformly distributed on* $(0,1)$. *Define*

$$V := -k(1-U)^{k-1} F^{-1}(U) + k(k-1) \int_U^1 (1-t)^{k-2} F^{-1}(t) \, \mathrm{d}t$$

and let σ_V^2 *be the variance of* V. *Then, as* $n_1 \to \infty$,

$$(1/n_1 + 1/n_2)^{-1/2} \, \hat{\Gamma}_{k,n_1,n_2} \longrightarrow_d \mathrm{N}(0, \sigma_V^2).$$

The procedure to estimate the asymptotic standard deviation σ_V is also analogous to that proposed for σ_W. In this case the pseudo-values are

$$\hat{V} = -k \left(1 - \frac{\lceil nU \rceil}{n}\right)^{k-1} Z_{\lceil nU \rceil : n} + k \sum_{\lceil nU \rceil + 1}^{n} \gamma_{i,k-1,n} Z_{i:n},$$

where $\gamma_{i,k-1,n}$ is defined in (14). For an appropriate normal quantile z, H_0 is rejected in the critical region $\left\{(1/n_1 + 1/n_2)^{-1/2} \hat{\Gamma}_{k,n_1,n_2}/\hat{\sigma}_V > z\right\}$.

An appealing aspect of Theorem 4 is that it guarantees the asymptotic normality of the test statistic under remarkably mild conditions. For instance, when the variables are nonnegative (which is the case of most interesting economic variables) the condition $\mathrm{E}X < \infty$ implies $x[1 - F(x)] \to 0$ as $x \to \infty$, which in turn implies (15) for $k \geq 2$. Therefore, in this important case, the finiteness of the expectation is all what is needed to ensure the asymptotic normality whereas the asymptotic behavior of most related test statistics in the literature involves the existence of the second moment. See for instance, Aly (1990, Theorem 2.1), Marzec and Marzec (1991, Theorem 2.1) and Belzunce *et al.* (2005, Theorem 2.1).

5 MONTE CARLO RESULTS: STOCHASTIC DOMINANCE AGAINST NO DOMINANCE

To investigate the properties of the test of Section 3 for small samples, we have carried out a simulation study inspired by that of Barrett and Donald (2003). We consider the test for the second order stochastic dominance, that is, $\mathrm{H}_0 : X \leq_{\mathrm{icv}} Y$ versus $\mathrm{H}_1 : X \nleq_{\mathrm{icv}} Y$ across five different models (M1–M5). The models are related to log-normal distributions, which are frequently found in welfare analysis. We consider three mutually independent standard normal variables Z, Z' and Z''. In all models $X = \exp(0.85 + 0.6Z)$ is fixed. In the first three models, $Y = \exp(\mu + \sigma Z')$ is also log-normal. The models only differ in the values of the parameters μ and σ:

- M1: $\mu = 0.85$ and $\sigma = 0.6$. Hence, H_0 is true and $X =_{\mathrm{st}} Y$.
- M2: $\mu = 0.6$ and $\sigma = 0.8$. In this case, H_0 is false.
- M3: $\mu = 1.2$ and $\sigma = 0.2$. In this model, H_0 is true but $X \neq_{\mathrm{st}} Y$.

In the last two models, Y is a mixture of two log-normal distributions:

$$Y = 1_{\{U \geq 0.1\}} \exp(\mu_1 + \sigma_1 Z') + 1_{\{U < 0.1\}} \exp(\mu_2 + \sigma_2 Z''),$$

where 1_A stands for the indicator function of the set A, U is a uniform [0,1] random variable also independent of the normal variables Z, Z' and Z''.

- M4: $\mu_1 = 0.8$, $\sigma_1 = 0.5$, $\mu_2 = 0.9$ and $\sigma_2 = 0.9$. In this case, H_0 is false.
- M5: $\mu_1 = 0.85$, $\sigma_1 = 0.4$, $\mu_2 = 0.4$ and $\sigma_2 = 0.9$. Here H_0 is again false.

We have simulated samples with sizes $n_1 = n_2 = 50$ and $n_1 = n_2 = 100$, and then we have applied the test based on (6) with $k = \lceil n_1/10 \rceil$ (denoted by CX10), with $k = \lceil n_1/20 \rceil$ (denoted by CX20), and a test proposed by Barrett and Donald (2003) for the same testing problem (denoted by BD). We have used the bootstrap scheme described in Section 3 with $B = 1000$ to approximate the p-value. Accordingly, we have selected the Barrett-Donald test using the same bootstrap scheme, namely, the one called KSB2 in that paper.

For each model, we have performed 1000 Monte Carlo replications of the experiment and recorded the rejection rates at the significance levels $\alpha = 0.01$ and $\alpha = 0.05$ for both tests. The results can be found in Table 1.

Both tests behave similarly under H_0: under M1, we obtain rejection rates not far from the nominal significance levels. This suggests that the bootstrap approximation works well for the three tests. Under M3, H_0 is true but $X \neq_{\mathrm{st}} Y$ so that we expect a rejection rate below the nominal significance level. Observe that the tests never reject H_0 in this case. Regarding models for which H_1 is true, neither of the tests is uniformly better than the others: CX10 and CX20 are more powerful than BD under M2, but BD is more powerful than CX10 and CX20 under the mixtured models M4 and M5.

α	n	Test	M1	M2	M3	M4	M5
0.01	50	BD	0.013	0.085	0.000	0.029	0.068
0.01	50	CX10	0.010	0.198	0.000	0.026	0.030
0.01	50	CX20	0.010	0.242	0.000	0.021	0.023
0.01	100	BD	0.004	0.127	0.000	0.053	0.136
0.01	100	CX10	0.004	0.299	0.000	0.044	0.070
0.01	100	CX20	0.007	0.421	0.000	0.043	0.038
0.05	50	BD	0.055	0.248	0.000	0.140	0.237
0.05	50	CX10	0.047	0.420	0.000	0.095	0.116
0.05	50	CX20	0.051	0.496	0.000	0.078	0.087
0.05	100	BD	0.043	0.391	0.000	0.185	0.368
0.05	100	CX10	0.039	0.609	0.000	0.168	0.217
0.05	100	CX20	0.040	0.705	0.000	0.144	0.138

Table 1: Rejection rates for BD, CX10 and CX20 tests with bootstrap p-value under models M1–M5.

6 MONTE CARLO RESULTS: STOCHASTIC EQUALITY AGAINST STRICT DOMINANCE

We have carried out a simulation study to assess the performance of the tests proposed in Section 4 for finite sample sizes. Also we want to illustrate some ideas about the choice of the parameter k.

6.1 General description and results for fixed k

We have considered a situation in which X has a Weibull distribution with shape parameter 10 and scale parameter $1/\Gamma(1 + 1/10)$. On the other hand, Y has a Weibull distribution with shape parameter θ and scale parameter $1/\Gamma(1 + 1/\theta)$, for $\theta = 6, 8, 10$. The scale parameters have been chosen so that $\mathrm{E}X = \mathrm{E}Y = 1$, which is the most unfavorable situation to detect deviations from H_0. The null hypothesis corresponds to $\theta = 10$. All the considered pairs of variables are ordered since they have the same mean and the difference between their density functions has two crossing points (see Shaked and Shanthikumar (2006, Theorem 3.A.44, p. 133)). For all the described combinations, we simulate couples of independent samples with sizes $n_1 = n_2 = 50, 100$ and apply the test based on the critical region (13) for $k = 2, 4, 6, 10$ at the significance level $\alpha = 0.05$. After replicating this experiment 1000 times, we registered the proportion of times for which H_0 was rejected, that is, the empirical power of the tests.

In Figure 1 the resulting empirical power curves ($n_1 = n_2 = 50$ and $n_1 = n_2 = 100$) are represented. The horizontal dotted line corresponds to the nominal significance level of the test. From the results of the experiment it is apparent that the largest values of k perform clearly better than the smallest ones. However, as k increases the improvement seems to be less significant. These results point out that the power of the test may strongly depend on the value of k. In the following subsection we address the problem of choosing an appropriate value of this parameter.

We also performed similar simulation studies with both gamma and Student t random variables. The results were remarkably similar and have been omitted for the sake of briefness.

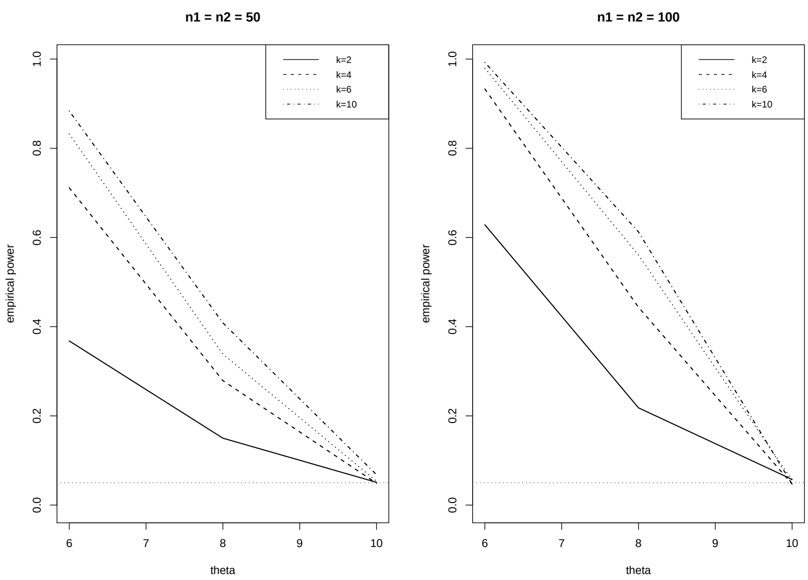

Figure 1: Empirical power curves for the test (13) under different Weibull alternatives, sample sizes 50 and 100, and several values of k.

6.2 Data-driven selection of k

There are two factors that should be taken into account in the selection of k. The first one is the ability of the discrepancy measure $\Delta_k(X, Y)$ to detect deviations from the null hypothesis. From this point of view we should choose the value k that maximizes $\Delta_k(X,Y)$. In some important situations it can be shown that $\Delta_k(X,Y)$ increases with k and then we should choose k as large as possible in these cases. However, since X and Y are unknown, in practice we use the estimate $\hat{\Delta}_{k,n_1,n_2}$ instead of $\Delta_k(X,Y)$. Therefore, another important factor to be considered is the variability of $\hat{\Delta}_{k,n_1,n_2}$. It is intuitively clear that as k increases (for fixed n_1 and n_2), it is more difficult to estimate $\Delta_k(X,Y)$, so that an increase in the variance of $\hat{\Delta}_{k,n_1,n_2}$ should be expected. As a consequence, a large value of k might not be a good choice regarding this second aspect. The results displayed in Figure 1 collect the overall effect of the two factors under the Weibull model.

A simple measure to quantify which of the two factors is more influential is the inverse of the coefficient of variation. If X is a random variable with finite second moment, we denote by $\mathrm{CV}^{-1} := \mathrm{E}X/\sigma_X$ the inverse of the coefficient of variation (σ_X being the standard deviation of X). Let us denote by CV_k^{-1} the inverse of the coefficient of variation of the test statistics $\hat{\Delta}_{k,n_1,n_2}$ given in (8). Overall, a high value of CV_k^{-1} could generate a test with a good power. A reasonable data-driven choice of k would be then the value of k that provides the highest estimated value of CV_k^{-1}. In practice, a standard bootstrap procedure can be used to estimate CV_k^{-1}.

A hindrance of this approach is that the asymptotic result for fixed k derived in Section 4 is no longer applicable. The procedure described above automatically selects a value of k "against the null" so that the critical value prescribed by equation (13) is too liberal. Again, bootstrap techniques may help to approximate the appropriate critical value for a given significance level. The method is fairly similar to the one described in Section 3 and the details are omitted.

Since we use bootstrap both for estimating the coefficient of variation and for approximating the critical level, our procedure is computationally expensive. Fortunately, using a small number of bootstrap samples yields acceptable results. In Table 2, we report the empirical significance level and power of the test, with k automatically selected as described above using only 10 bootstrap samples to estimate the coefficient of variation and 200 bootstrap samples to approximate the critical value. The experiment has been replicated 1000 times and $\alpha = 0.05$ is the nominal significance level. We see that the data-driven selection of k yields good power and at the same time allows us to control the significance level.

Weibull	$\theta = 6$	$\theta = 8$	$\theta = 10$
$n_1 = n_2 = 50$	0.845	0.309	0.067
$n_1 = n_2 = 100$	0.988	0.542	0.052

Table 2: Rejection rates for the test of equality against strict dominance when k is automatically selected. The nominal significance level is $\alpha = 0.05$. The last column corresponds to the null hypothesis.

6.3 Comparison with the Kolmogorov-Smirnov test

As it was mentioned at the beginning of the Section 4, the tests generated by this approach take into account the important information of the ordering between the two variables. Therefore, the power of these tests is expected to be higher than the power of the usual omnibus tests for equality of the distributions in the literature (which work against all and not just ordered alternatives). To illustrate this point, we have compared the empirical power of our test (both with fixed $k = 10$ and data-driven selected k) with that of the Kolmogorov-Smirnov test, under the Weibull model described in Subsection 6.1 with sample sizes $n_1 = n_2 = 50, 100$. The results are summarized in Figure 2. We see that the tests proposed in this section have (uniformly) by far a much higher power than the Kolmogorov-Smirnov test. Note also that the automatic procedure to select k yields similar results to the case $k = 10$ (which is the best one across all the considered values, see Figure 1).

7 AN EMPIRICAL EXAMPLE

To illustrate the tests of Sections 3 and 4, we discuss a data set previously considered in Barrett and Donald (2003). The data set is drawn from the *Canadian Family Expenditure Survey* from the years 1978 and 1986. We are interested in the comparison of the income distributions in these years.

We normalize incomes dividing the data in each sample by its average, and analyze whether the resulting distributions are ordered with respect to the concave order. This is equivalent to comparing the distributions according to the Lorenz order.

Some descriptive graphics of the normalized incomes can be found in Figure 3. In the panels on the left we have plotted the empirical distribution functions for the pre-tax and post-tax normalized income data. The corresponding kernel density

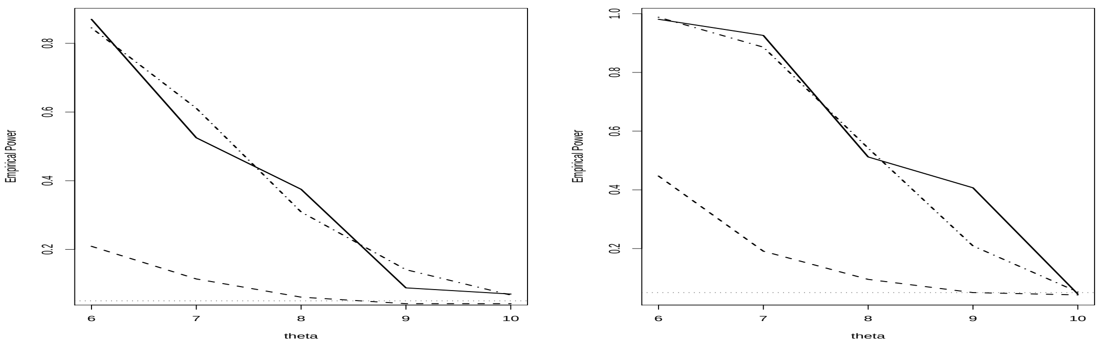

Figure 2: Empirical power curves for the test (13) with fixed $k = 10$ (solid line), automatic selection of k (dotted-dashed line), and the Kolmogorov-Smirnov test (dashed line) under the Weibull model. The sample sizes are $n_1 = n_2 = 50$ (left) and 100 (right). Horizontal dotted line corresponds to significance level $\alpha = 0.05$.

estimates are plotted in the right panels. Notice that the empirical distribution functions for 1978 and 1986 are rather similar. From the estimated densities we notice that qualitative aspects of both distributions (positive skewness, slight bimodality) are also comparable.

Figure 4 displays the difference between the integrated empirical quantile functions of the normalized samples. This difference is exactly the empirical counterpart of the function given in Lemma 1 in the appendix. This function is for the most part positive and this fact suggests that both distributions could be ordered according to the Lorenz order. A more formal evaluation of this ordering property can be achieved by applying the test developed in Section 3. The null hypothesis is that the income distribution of 1978 dominates in the Lorenz order that of 1986. To be more precise, if X_{1978} and X_{1986} denote the variables in the years 1978 and 1986, we test $\mathrm{H}_0 : X_{1986}/\mathrm{E}X_{1986} \leq_{\mathrm{cv}} X_{1978}/\mathrm{E}X_{1978}$ against H_1 : not H_0 (or equivalently, $\mathrm{H}_0 : X_{1978}/\mathrm{E}X_{1978} \leq_{\mathrm{cx}} X_{1986}/\mathrm{E}X_{1986}$ against the same alternative). Therefore, we have applied the test based on the critical region (6), where the critical value c has been approximated using $B = 500$ bootstrap samples. Also, several values of k have been considered. The corresponding p-values are displayed in Table 3.

Since the p-values are quite large, the conclusion is that the null hypothesis cannot be rejected. This also means that the negative parts of the functions

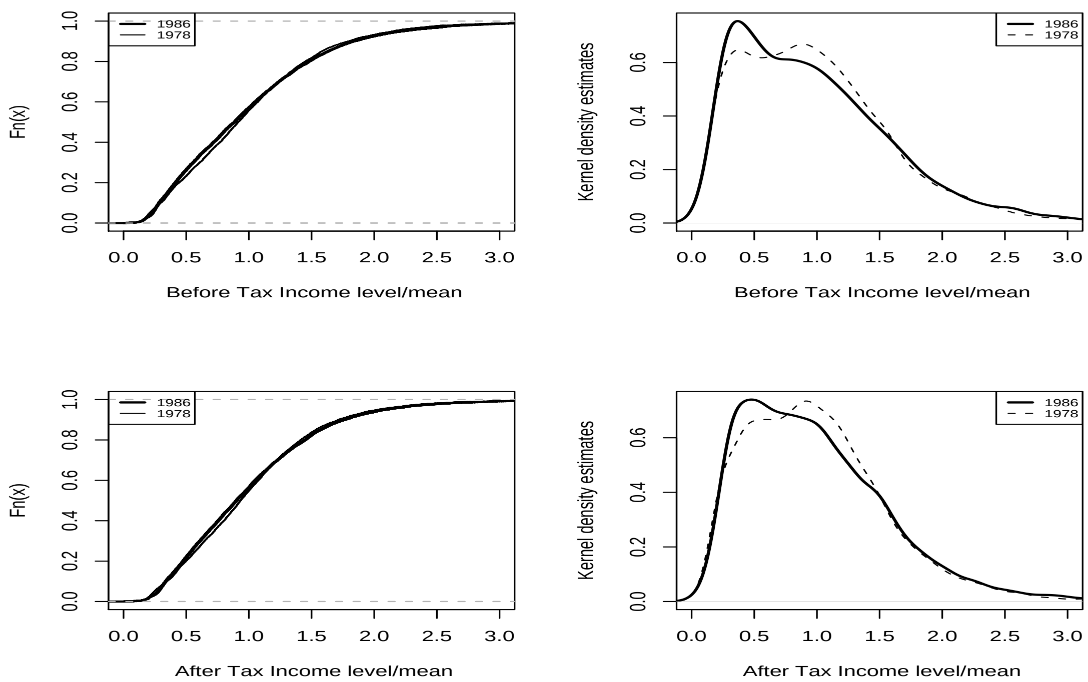

Figure 3: Empirical distribution functions and kernel density estimates for pre-tax and post-tax income data in 1978 and 1986.

depicted in Figure 4 are not significant. The p-values are almost the same for the considered values of k. Moreover, p-values are also similar when considering the after tax or before tax incomes.

Since the assumption that both distributions are ordered is acceptable, a natural question is if the distributions are equal (according to this order), or if one *strictly* dominates the other. We note that the equality for the Lorenz order is the equality in distribution up to dilations of the variables. To answer this question we use the test introduced in Section 4. We have used bootstrap estimates (based on $B = 500$ resamples) of the inverse of the coefficient of variation of the discrepancies, as described in Subsection 6.2, to explore which values of k are more suitable. A graphical representation of these estimates can be found in Figure 5. It turns out that the maximum is attained at $\hat{k} = 14$ (before tax) and $\hat{k} = 11$ (after tax). Accordingly, we carry out the tests corresponding to these values, where the p-values are approximated using again bootstrap. The resulting p-values are 0.006 (before tax) and 0.002 (after tax). Therefore, our conclusion is that the income

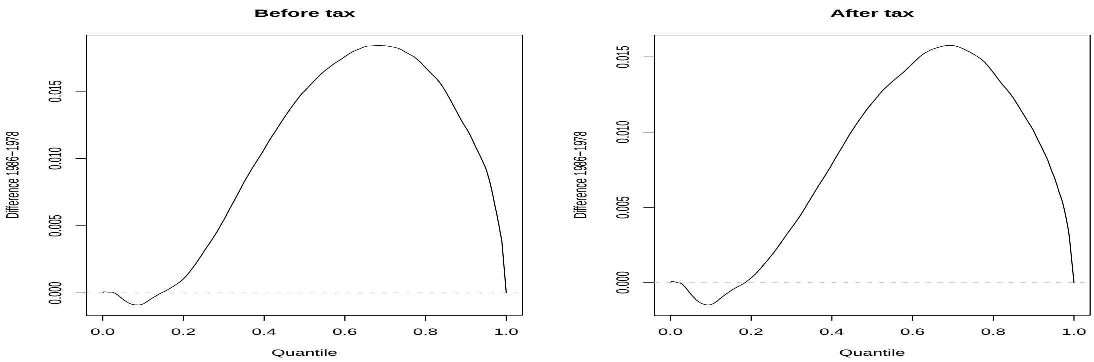

Figure 4: Difference between the integrated quantile functions (1986 minus 1978).

k	$\lceil \min\{n_1, n_2\}/100 \rceil$	$\lceil \min\{n_1, n_2\}/500 \rceil$	$\lceil \min\{n_1, n_2\}/1000 \rceil$
Before tax	0.666	0.628	0.636
After tax	0.590	0.592	0.598

Table 3: p-values for testing the null hypothesis that the income distribution of 1978 dominates in the Lorenz order that of 1986 versus the alternative that both distributions are not ordered.

distribution in 1978 strictly dominates that of 1986, according to Lorenz order. In this sense, we conclude that the income distribution in 1978 was more even than in 1986. The conclusion does not depend upon the mean income level since we have used normalized incomes and neither upon the consideration of incomes before tax or after tax. Moreover, we can assert that $X_{1986} \neq_{\text{st}} aX_{1978}$, for all $a > 0$, that is, the situation in 1986 was not a dilation of that in 1978.

8 CONCLUSIONS

We propose a new approach to solve two different testing problems related to the second order stochastic dominance. First, we discuss a test for stochastic dominance versus no dominance. The technique consist in comparing the estimated expected cumulative resources of the m-poorest in random samples of size k of the populations. We derive the asymptotic consistency of the method and approximate the p-values via bootstrap. The simulation studies show that the methodology works well. However, in this work it is not proved the asymptotic consistency of

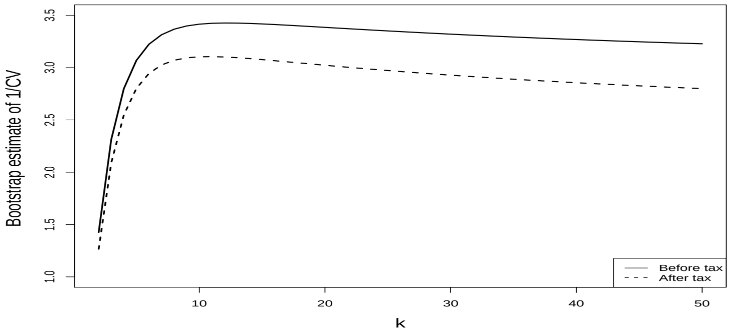

Figure 5: Bootstrap estimates of the inverse of the coefficient of variation of $\hat{\Delta}_{k,n_1,n_2}$ as a function of k.

the considered bootstrap scheme. The selection of the tuning parameter k affects the power of the resulting test and further research will be needed to understand better the influence of this parameter. Also, we consider a test of stochastic equality against strict domination. In this case, we compare the estimated expected maxima or minima of random samples of size k of the populations. The estimators of the discrepancies are L-statistics and we show that their distributions are asymptotically normal (for all k). Again, the choice of k has an impact on the power of the tests. For this reason, we derive a data-driven selection of k to obtain a value of this parameter generating a powerful test. One important advantage of this last test is that it is much more powerful than the usual tests of equality in distribution since we include the additional information that the variables are stochastically ordered.

APPENDIX: PROOFS

Without loss of generality we assume the functions $\omega \in \mathcal{I}$ are right continuous. For $\omega \in \mathcal{I}$, μ_ω is the Lebesgue-Stieltjes measure defined by $\mu_\omega((a,b]) = \omega(b) - \omega(a)$, $(a,b] \subset [0,1]$.

The following lemma is a consequence of the results in Rüschendorf (1981), where 1_A stands for the indicator function of the set A.

Lemma 1. *$X \leq_{\mathrm{cx}} Y$ if and only if $\Delta_{1_{(t,1]}}(X,Y) \geq 0$, for $0 \leq t < 1$, with equality for $t = 0$. The equivalence remains true if "$\leq_{\mathrm{cx}}$" is replaced by "$\leq_{\mathrm{icx}}$" and the restriction for $t = 0$ is dropped.*

Lemma 2. *Let f be a continuous and nonnegative function on $[0,1]$. If $\int_{[0,1]} f \, \mathrm{d}\mu_\omega = 0$, for some $\omega \in \mathcal{I}^*$, then $f \equiv 0$ on $[0,1]$.*

Proof. If for some $t_0 \in [0,1]$, $f(t_0) > 0$, by the continuity of f there would exist a nonemply interval $(a,b] \subset [0,1]$ such that $f > 0$ on $(a,b]$. Since $\omega \in \mathcal{I}^*$, we have $\mu_\omega((a,b]) > 0$. Therefore, $f > 0$ on an interval with positive measure, which contradicts the assumption on the value of the integral. □

Proof of Theorem 1. We restrict to the case $X \leq_{\mathrm{icx}} Y$. The proof for the convex order is analogous taking into account $X \leq_{\mathrm{cx}} Y$ if and only if $X \leq_{\mathrm{icx}} Y$ and $\mathrm{E}X = \mathrm{E}Y$. From Lemma 1, $\Delta_\omega(X,Y) \geq 0$ for all $\omega \in \mathcal{I}_0$ implies $X \leq_{\mathrm{icx}} Y$. For the reverse implication, let us assume that $X \leq_{\mathrm{icx}} Y$. For $\omega \in \mathcal{I}_0$, by Fubini's theorem, we have

$$\begin{aligned} \Delta_\omega(X,Y) &= \omega(0)(\mathrm{E}Y - \mathrm{E}X) + \int_0^1 \left(G^{-1}(t) - F^{-1}(t)\right) \left(\int_{[0,t]} \mathrm{d}\mu_\omega(s) \right) \mathrm{d}t \\ &= \omega(0)(\mathrm{E}Y - \mathrm{E}X) + \int_{[0,1]} \Delta_{1_{(s,1]}}(X,Y) \, \mathrm{d}\mu_\omega(s). \end{aligned} \tag{16}$$

By Lemma 1, $\Delta_{1_{(\cdot,1]}}(X,Y)$ is nonnegative on $[0,1]$ and since $\mathrm{E}X \leq \mathrm{E}Y$ provided $X \leq_{\mathrm{icx}} Y$, we obtain that $\Delta_\omega(X,Y) \geq 0$ and this completes the proof of part (a).

To show (b), let us assume that $X \leq_{\mathrm{icx}} Y$ and there exists a function $\omega \in \mathcal{I}_0^*$ such that $\Delta_\omega(X,Y) = 0$. By (16), we have $\int_{[0,1]} \Delta_{1_{(s,1]}}(X,Y) \, \mathrm{d}\mu_\omega(s) = 0$. By Lemma 1, the function $\Delta_{1_{(\cdot,1]}}(X,Y)$ is nonnegative on $[0,1]$ and it is trivially continuous. We apply Lemma 2 to conclude $\Delta_{1_{(\cdot,1]}}(X,Y) \equiv 0$ on $[0,1]$. This implies $F^{-1} = G^{-1}$ a.e. on $[0,1]$ and thus $X =_{\mathrm{st}} Y$. □

Proof of Corollary 1. Fix $k \geq 1$ and $1 \leq i \leq k$. Following the same lines as in de la Cal and Cárcamo (2006, Lemma 4.3), we obtain

$$S_{m:k}(Y) - S_{m:k}(X) = \Delta_\omega(X,Y) \quad \text{with} \quad \omega(t) = k \ \Pr(\beta_{k-m:k-1} \leq t), \tag{17}$$

where $\beta_{k-m:k-1}$ ($m < k$) is a Beta$(k-m, m)$ random variable and $\beta_{0,k-1}$ is degenerate at 0. This implies that $\omega \in \mathcal{I}_0$ and for $k \geq 2$ and $1 \leq m < k$, $\omega \in \mathcal{I}_0^*$. Hence, Corollary 1 (a) follows from Theorem 1. Part (b) is analogous. □

Proof of Theorem 2. We only give a proof for part (a) since (b) is analogous. Given a random sample $X_1, \ldots, X_{n_1}$ (resp. $Y_1, \ldots, Y_{n_1}$) of X (resp. Y), we denote by F_{n_1} and $F_{n_1}^{-1}$ (resp. G_{n_2} and $G_{n_2}^{-1}$) the empirical distribution and quantile functions of the sample. By (17) and (1), for all m and k, we obtain

$$\begin{aligned}\left|\frac{\hat{S}_{m:k}(X)-\hat{S}_{m:k}(Y)}{k}\right| &= \left|\int_0^1 \left(F_{n_1}^{-1}(t)-G_{n_2}^{-1}(t)\right)\Pr(\beta_{k-m:k-1}\le t)\,\mathrm{d}t\right| \\ &\le \int_0^1 \left|F_{n_1}^{-1}(t)-F^{-1}(t)\right|\,\mathrm{d}t+\int_0^1 \left|F^{-1}(t)-G_{n_2}^{-1}(t)\right|\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} |F_{n_1}(t)-F(t)|\,\mathrm{d}t+\int_{-\infty}^{\infty} |F(t)-G_{n_2}(t)|\,\mathrm{d}t. \end{aligned} \tag{18}$$

Therefore, the statistics $\hat{\Lambda}_{k,n_1,n_2}$ given in (5) satisfies

$$(1/n_1+1/n_2)^{-1/2}|\hat{\Lambda}_{k,n_1,n_2}| \le \sqrt{n_1}\int_{-\infty}^{\infty}|F_{n_1}(t)-F(t)|\,\mathrm{d}t+\sqrt{n_2}\int_{-\infty}^{\infty}|F(t)-G_{n_2}(t)|\,\mathrm{d}t.$$

Now, if $X =_{\mathrm{st}} Y$ and X fulfills $\int_0^\infty \sqrt{\Pr(|X|>x)}\,\mathrm{d}x<\infty$, Theorem 2.1 (b) in del Barrio *et al.* (1999) ensures that for $i=1,2$ the sequences $\left\{\sqrt{n_i}\int_{-\infty}^{\infty}|F(t)-F_{n_i}(t)|\,\mathrm{d}t\right\}_{n_i\ge 1}$ are bounded in probability. This directly implies that $\left\{(1/n_1+1/n_2)^{-1/2}|\hat{\Lambda}_{k,n_1,n_2}|\right\}_{n_1,n_2\ge 1}$ is bounded in probability. Hence the value c that appears in (6) is bounded.

On the other hand, if H_1 is true, a similar argument as in the proof of Lemma 4.4 in de la Cal and Cárcamo (2006) shows that there exist an $\epsilon_0>0$ and a positive integer k_0 such that for all $k\ge k_0$

$$\frac{1}{k}\max_{1\le m<k}\{S_{m:k}(X)-S_{m:k}(Y)\}>\epsilon_0. \tag{19}$$

A similar reasoning as in (18), the integrability of X and Glivenko-Cantelli yield

$$\left|\frac{\hat{S}_{m:k}(X)-S_{m:k}(X)}{k}\right| \le \int_{-\infty}^{\infty}|F(t)-F_{n_1}(t)|\,\mathrm{d}t\to 0,\quad \text{a.s. as } n_1\to\infty, \tag{20}$$

uniformly in m and k. Now, we have that

$$\frac{\hat{S}_{m:k}(X)-\hat{S}_{m:k}(Y)}{k}=\frac{\hat{S}_{m:k}(X)-S_{m:k}(X)}{k}+\frac{S_{m:k}(X)-S_{m:k}(Y)}{k}+\frac{S_{m:k}(Y)-\hat{S}_{m:k}(Y)}{k}. \tag{21}$$

For n_1 and n_2 large enough, (20) ensures that for all m and k

$$\left|\frac{\hat{S}_{m:k}(X)-S_{m:k}(X)}{k}\right|,\ \left|\frac{S_{m:k}(Y)-\hat{S}_{m:k}(Y)}{k}\right|<\epsilon_0/4. \tag{22}$$

Finally, equations (19), (21) and (22) imply that for $k \geq k_0$, $\hat{\Lambda}_{k,n_1,n_2} > \epsilon_0/2$ (almost surely) for n_1 and n_2 large enough. This implies that, under H_1, when n_1 and $n_2 \to \infty$ and $k \geq k_0$, $(1/n_1 + 1/n_2)^{-1/2}\hat{\Lambda}_{k,n_1,n_2} \to \infty$, a.s., and since the rejection region in determined by a bounded quantity c, we conclude that (a) holds and the proof is complete. □

Remark 1. Under the condition $\mathrm{E}|X|^\gamma < \infty$ for some $1 < \gamma < 2$ instead of $\int_0^\infty \sqrt{\Pr(|X| > x)}\,\mathrm{d}x < \infty$, the same proof but using Theorem 2.2 in del Barrio *et al.* (1999) yields the consistency of the test given by the critical region (7).

Proof of Theorem 3. First, we shall show that the L-statistic $\sum_{i=1}^{n_1} \omega_{i,k,n_1} X_{i:n_1}$, where the weights ω_{i,k,n_1} are defined as in (9), is asymptotically equivalent to $(k/n_1)\sum_{i=1}^{n_1}(i/n_1)^{k-1}X_{i:n_1}$. Notice that for all k, n_1 and $1 \leq i \leq n_1$, there exists $\theta_i \in ((i-1)/n_1, i/n_1)$ such that $\omega_{i,k,n_1} = k\int_{(i-1)/n_1}^{i/n_1} t^{k-1}\,\mathrm{d}t = (k/n_1)\theta_i^{k-1}$. Then,

$$\left|\sum_{i=1}^{n_1}\omega_{i,k,n_1}X_{i:n_1} - \frac{k}{n_1}\sum_{i=1}^{n_1}\left(\frac{i}{n_1}\right)^{k-1}X_{i:n_1}\right|$$
$$\leq \frac{k}{n_1}\sum_{i=1}^{n_1}\left|\left(\frac{i}{n_1}\right)^{k-1} - \theta_i^{k-1}\right| \cdot |X_{i:n_1}| \leq \frac{k(k-1)}{n_1}\frac{\sum_{i=1}^{n_1}|X_{i:n_1}|}{n_1}.$$

Since X has finite expectation, the last quantity is $O_p(1/n_1)$.

Next, we apply Li *el al.* (2001, Theorem 2.1) with $H(u) = u$ and $J(u) = ku^{k-1}$ to obtain

$$\sqrt{n_1}\left(\frac{k}{n_1}\sum_{i=1}^{n_1}\left(\frac{i}{n_1}\right)^{k-1}X_{i:n_1} - \mu\right) \longrightarrow_d \mathrm{N}(0, \sigma_W^2), \quad n_1 \to \infty,$$

where $\mu := \mathrm{E}X_{k:k}$ and

$$\begin{aligned}
\sigma_W^2 &:= \mathrm{Var}\left[J(U)F^{-1}(U) + \mu + \int_0^1 (1_{\{U\leq t\}} - t)J'(t)F^{-1}(t)\,\mathrm{d}t\right] \\
&= \mathrm{Var}\left[kU^{k-1}F^{-1}(U) + k(k-1)\int_0^1 (1_{\{U\leq t\}} - t)t^{k-2}F^{-1}(t)\,\mathrm{d}t\right] \\
&= \mathrm{Var}(W),
\end{aligned}$$

with W defined in (11). For the last equality, take into account that

$$\int_0^1 (1_{\{U\leq t\}} - t)t^{k-2}F^{-1}(t)\,\mathrm{d}t = \int_U^1 t^{k-2}F^{-1}(t)\,\mathrm{d}t - \int_0^1 t^{k-1}F^{-1}(t)\,\mathrm{d}t,$$

and the second term is not random. Condition (10) is needed to ensure that conditions (i)-(iii) in Li *el al.* (2001, Theorem 2.1) hold. By the asymptotic equivalence established above and the assumption $n_1/n \to \lambda$ we also have

$$\sqrt{n}\left(\sum_{i=1}^{n_1} \omega_{i,k,n_1} X_{i:n_1} - \mu\right) \longrightarrow_d \mathrm{N}(0, \sigma_W^2/\lambda), \quad n \to \infty.$$

Following the same lines, we can also show

$$\sqrt{n}\left(\sum_{i=1}^{n_2} \omega_{i,k,n_2} Y_{i:n_2} - \mu\right) \longrightarrow_d \mathrm{N}(0, \sigma_W^2/(1-\lambda)), \quad n \to \infty.$$

Finally, since both samples are independent we deduce

$$\sqrt{n}\hat{\Delta}_{k,n_1,n_2} \longrightarrow_d \mathrm{N}(0, \sigma_W^2/[\lambda(1-\lambda)]), \quad n \to \infty,$$

which in turn implies (12). □

Next proposition shows the consistency of the estimators of the asymptotic standard deviations σ_W and σ_V described in Section 4. Although we ask for a finite second moment, we believe the conditions (10) and (15) are enough.

Proposition 1. *If the variable X has finite second moment, the estimators $\hat{\sigma}_W$ and $\hat{\sigma}_V$ described in Section 4 are consistent.*

Proof. We only give a proof for $\hat{\sigma}_W$ since the one for $\hat{\sigma}_V$ is analogous. If $\mathrm{E}X^2 < \infty$, it is easy to check that $\mathrm{E}W^2 < \infty$, where W is defined in (11). Therefore, recalling that $\hat{W}$ is the empirical counterpart of W, we only need to show that $\mathrm{E}(\hat{W} - W)^2 \to 0$, as the sample size $n \to \infty$. Some computations show

$$\mathrm{E}(\hat{W} - W)^2 \le k^2 \|F_n^{-1} - F^{-1}\|_2^2 + k^2(k^2-1)\|F_n^{-1} - F^{-1}\|_1,$$

where F_n^{-1} is the empirical quantile function and $\|\cdot\|_1$, $\|\cdot\|_2$ stand for the L^1 and L^2 norms, respectively. The existence of the second moment of X guarantees that $\|F_n^{-1} - F^{-1}\|_2$, $\|F_n^{-1} - F^{-1}\|_1 \to 0$, as $n \to \infty$ and the proof is complete. □

ACKNOWLEDGMENTS

The authors are thankful to the referees and the associated editor of the journal for their valuable comments and suggestions which have led to an improved version of the paper. Furthermore, they pointed to us the importance of the testing problem discussed in Section 3.

This research was supported by the Spanish MEC, grants MTM2007-66632 and MTM2008-06281-C02-02, and Comunidad de Madrid, grant S-0505/ESP/0158.
The e-mail addresses of the authors are:

– joser.berrendero@uam.es (José R. Berrendero)

– javier.carcamo@uam.es (Javier Cárcamo, corresponding author)

REFERENCES

Abadie, A. (2002), "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292.

Aly, E-E. A.A. (1990), "A Simple Test for Dispersive Ordering," *Statistics & Probability Letters*, 9, 323–325.

Anderson G. (1996), "Nonparametric Tests of Stochastic Dominance in Income Distributions," *Econometrica*, 64, 1183–1193.

Atkinson, A. B. (1970), "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 244–263.

Barrett, G. F., and Donald, S. G. (2003), "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 71–104.

del Barrio, E., Giné, E., and Matrán, C. (1999), "Central Limit Theorems for the Wassertein Distance between the Empirical and the True Distribution," *The Annals of Probability*, 27, 1009–1071.

Belzunce, F., Pinar, J. F., and Ruiz, J. M. (2005), "On Testing the Dilation Order and HNBUE Alternatives," *Annals of the Institute of Statistical Mathematics*, 57, 803–815.

de la Cal, J., and Cárcamo, J. (2006), "Stochastic Orders and Majorization of Mean Order Statistics," *Journal of Applied Probability*, 43, 704–712.

Davidson, R., and Duclos, J. Y. (2000), "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica*, 68, 1435–1464.

Denuit, M., Dhaene, J., Goovaerts, M. and Kass, R. (2005), *Actuarial Theory for Dependent Risks*, Wiley, New York.

Denuit, M., Goderniaux, A. C., and Scaillet, O. (2007), "A Kolmogorov-Smirnov-Type Test for Shortfall Dominance against Parametric Alternatives," *Technometrics*, 49, 88–99.

Goovaerts, M. J., Kaas, R., and van Heerwaarden, A. E. (1990), *Effective Actuarial Methods*, North-Holland, Amsterdam.

Kaur, A., Prakasa Rao, B. L. S., and Singh, H. (1994), "Testing for Second-Order Stochastic Dominance of Two Distributions," *Econometric Theory*, 10, 849–866.

Kaas, R., van Heerwaarden, A. E., and Goovaerts, M. J. (1994), *Ordering of Actuarial Risks*, Cairo Education Series: Amsterdam.

Lambert, P. J. (1993), *The Distribution and Redistribution of Income, a Mathematical Analysis,* Manchester University Press, Manchester, 2nd edition.

Li, D., Rao, M. B., and Tomkins, R. J. (2001), "The Law of the Iterated Logarithm and Central Limit Theorem for L-Statistics," *Journal of Multivariate Analysis*, 78, 191–217.

McFadden, D. (1989), "Testing for Stochastic Dominance," in *Studies in the Economics of Uncertainty: In Honor of Josef Hadar*, ed. Fomby, T. B., and Seo., T. K., New York, Berlin, London, and Tokyo: Springer.

Marzec, L., and Marzec, P. (1991), "On Testing Equality in Dispersion of Two Probability Distributions," *Biometrika*, 78, 923–925.

Rüschendorf, L. (1981), "Ordering of Distributions and Rearrangement of Functions," *The Annals of Probability*, 9, 276–283.

Sordo, M. A., and Ramos, H. M. (2007), "Characterization of stochastic orders by L-functionals," *Statistical Papers*, 48, 249–263.

Shaked, M., and Shanthikumar, J. G. (2006), *Stochastic Orders*, Springer Series in Statistics.

Shorrocks, A. F. (1983), "Ranking Income Distributions," *Economica*, 50, 3–17.