Received: 13 October 2023 | Revised: 27 May 2024 | Accepted: 15 October 2024

DOI: 10.1111/sjos.12761

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

A new class of nonparametric tests for second-order stochastic dominance based on the Lorenz P–P plot

Tommaso Lando | **Sirio Legramanti**

Department of Economics, University of Bergamo, Bergamo, Italy

Correspondence
Tommaso Lando, Department of Economics, University of Bergamo, Bergamo, Italy.
Email: tommaso.lando@unibg.it

Funding information
Italian funds ex MURST 60%

Abstract

Given samples from two non-negative random variables, we propose a family of tests for the null hypothesis that one random variable stochastically dominates the other at the second order. Test statistics are obtained as functionals of the difference between the identity and the Lorenz P–P plot, defined as the composition between the inverse unscaled Lorenz curve of one distribution and the unscaled Lorenz curve of the other. We determine upper bounds for such test statistics under the null hypothesis and derive their limit distribution, to be approximated via bootstrap procedures. We then establish the asymptotic validity of the tests under relatively mild conditions and investigate finite-sample properties through simulations. The results show that our testing approach can be a valid alternative to classic methods based on the difference in the integrals of the cumulative distribution functions, which require bounded support and struggle to detect departures from the null in some cases. The same approach can be extended to a family of fractional-degree stochastic orders, including the first order as a limiting case.

KEYWORDS

bootstrap, Lorenz curve, stochastic order, test consistency

1 INTRODUCTION

The theory of stochastic orders deals with the problem of comparing pairs of random variables, or the corresponding distributions, with respect to concepts such as size, variability (or riskiness), shape, aging, or combinations of these aspects. The main notion in this context is generally referred to as the *usual stochastic order* or *first-order stochastic dominance* (FSD), which expresses the concept of one random variable being *stochastically larger* than the other (Shaked & Shantikumar, 2007). For this reason, FSD has important applications in all those fields in which "more" is preferable to "less", clearly including economics. However, FSD is a restrictive criterion, and there are many real-world applications in which it is not satisfied. This has pushed economic theorists to develop finer concepts, which formed the theory of *stochastic dominance* (SD), taking into account variability and shape, in addition to size (Fishburn, 1980; Hadar & Russell, 1969; Hanoch & Levy, 1969; Muliere & Scarsini, 1989; Wang & Young, 1998; Whitmore & Findlay, 1978). In this regard, the most commonly used SD relation is the *second-order SD* (SSD), expressing a preference for the random variable which is stochastically larger or at least less risky, therefore combining size and dispersion into a single preorder. This has applications in economics, finance, operations research, reliability, and many other fields in which decision makers typically prefer larger or at least less uncertain outcomes.

Given a pair of samples from two unknown random variables of interest, statistical methods may be employed to establish whether such variables are stochastically ordered. In particular, we focus on a major problem in nonparametric statistics, that is testing the null hypothesis of dominance versus the alternative of nondominance. About SSD, several procedures are available in the literature, some of which are described in the book of Whang (2019). We will now recall a few of these approaches. Davidson and Duclos (2000) proposed a test for SSD based on the difference between the integrals of the cumulative distribution functions (CDF). The problem with this test is that dominance is evaluated only on a fixed grid, which may lead to inconsistency. Barrett and Donald (2003) employed a similar approach, combined with bootstrap methods, to formulate a class of tests that are consistent under the assumption that the distributions under analysis are supported on a compact interval. Donald and Hsu (2016) leveraged a less conservative approach to determine critical values compared to Barrett and Donald (2003), avoiding the use of the least favorable configuration. We refer the reader to Linton et al. (2005, 2010) for other relevant approaches. Note that all the aforementioned papers deal more generally with finite-order SD, and then obtain SSD as a special case. Alternatively, other works focused on tests for the so-called *Lorenz dominance*, which is a scale-free version of SSD that applies to non-negative random variables. For example, Barrett et al. (2014) proposed a class of consistent tests for the Lorenz dominance that rely on the distance between empirical Lorenz curves. In this case, supports may be unbounded. Critical values are determined by approximating the limit distribution of a stochastic upper bound of the test statistic, similar to Barrett and Donald (2003). Sun and Beare (2021) used a different and less conservative bootstrap approach to improve the power of such tests, and established asymptotic properties under less restrictive distributional assumptions.

The main idea of this article follows from noticing that some stochastic orders, including FSD, can be expressed and tested via the classic P–P plot, also referred to as the *ordinal dominance curve* (Beare & Clarke, 2022; Beare & Moon, 2015; Davidov & Herman, 2012; Hsieh & Turnbull, 1996; Schmid & Trede, 1996; Tang et al., 2017). Following a similar approach, we propose a new class of nonparametric tests for SSD between non-negative random variables, in which the test statistic is based on what we refer to as the *Lorenz P–P plot* (LPP), a kind of P–P plot based on *unscaled*

Lorenz curves. More precisely, the LPP is obtained as the functional composition of the inverse unscaled Lorenz curve of one distribution and the unscaled Lorenz curve of the other. The key property of the LPP is that, under SSD, it stays above the identity function on the unit interval. Therefore, the LPP stands out as a promising tool for detecting deviations from the null hypothesis of SSD. Namely, any functional that quantifies the positive part of the difference between the identity and the LPP can be used to construct a test statistic. This gives rise to a whole class of tests, depending on the choice of the functional. The p-values of such tests can then be computed via bootstrap procedures. In particular, we use a similar idea as in Barrett et al. (2014) to asymptotically bound the size of the test, and establish its consistency via the functional delta method. Note that the consistency of our family of tests is established without requiring a bounded support, which represents an advantage compared to classic methods based on integrals of CDFs. Moreover, our simulation studies show that our tests are often more reliable than the established KSB3 test by Barrett and Donald (2003), which may have problems detecting violations of the null hypothesis in some cases.

The LPP may also be used to define families of fractional-degree orders “between” FSD and SSD (or beyond SSD) via a simple transformation. In this regard, we propose a method to define a continuum of SD relations, called *transformed SD*, in the spirit of Müller et al. (2017), Lando and Bertoli-Barsotti (2020), and Huang et al. (2020). Interestingly, our tests can be easily adapted to this more general family of orders by simply transforming the samples through the same transformation used in the definition of transformed SD. In particular, FSD can be obtained as a limiting case, in which the empirical LPP of the transformed sample tends to the classic empirical P–P plot. This opens up the possibility of applying our class of tests to a wide family of stochastic orders.

The article is organized as follows. Section 2 introduces the LPP and describes the idea behind the proposed family of tests. In Section 3, we propose an estimator of the LPP and study its properties. The empirical process associated with the LPP is investigated in Section 4, where we establish a weak convergence result that can be used to derive asymptotic properties of the tests. In Section 5, we establish bounded size under the null hypothesis and consistency under the alternative one, for both independent and paired samples. The extension to a family of fractional-degree orders is discussed in Section 6. In Section 7, we illustrate the finite-sample properties of the tests through simulation studies, focusing on tests arising from sup-norm and integral-based functionals. Finally, Section 8 contains our concluding remarks. All tables and proofs are reported in the Appendix.

2 | PRELIMINARIES

Throughout this article, H denotes a general CDF supported on the non-negative half line, with finite mean μ_H. In particular, we consider a pair of non-negative random variables X and Y with CDFs F and G, respectively, and finite expectations. When F and G are absolutely continuous, we will denote their densities with f and g, respectively. Given that stochastic orders depend only on distribution functions, for any order relation $\succ$ we may write $X \succ Y$ or $F \succ G$ interchangeably.

Let $L^p(0,1)$, for $p \geq 1$, be the class of real-valued functions on the unit interval equipped with the L^p norm $\|.\|_p$, that is, for $v \in L^p(0,1)$, $\|v\|_p = (\int_0^1 |v(t)|^p dt)^{1/p}$, and $L^\infty(0,1)$ be the class of bounded real-valued functions equipped with the uniform norm $\|.\|_\infty$ ($\|v\|_\infty = \sup_{t\in[0,1]} |v(t)|$). Moreover, let $C[0,1]$ be the space of continuous real-valued functions on [0,1] also equipped with the uniform norm. Henceforth, “increasing” means “nondecreasing” and “decreasing” means “nonincreasing”. Given a function r, we denote with $r_+ = \max(0,r)$ its positive part. If r is increasing, $r^{-1}(y) = \inf\{x : r(x) \geq y\}$ denotes its left-continuous generalized inverse. Finally, $\rightsquigarrow$ denotes weak convergence, while $\to_p$ denotes convergence in probability.

2.1 | Stochastic dominance

We say that X is larger than Y with respect to FSD, denoted as $X \geq_1 Y$, if $F(x) \leq G(x), \forall x$. Equivalently, $X \geq_1 Y$ if and only if $\mathbb{E}u(X) \geq \mathbb{E}u(Y)$ for any increasing function u. Within an economic framework, coherently with the expected-utility approach, one may assume that X and Y represent monetary lotteries and u is a utility function. Under this perspective, FSD represents all nonsatiable decision makers, that is, all those with an increasing utility, and therefore can be seen as one of the strongest ordering principles. On the other hand, FSD has a limited range of applicability since, in real-world applications, CDFs often cross and hence distributions cannot be ordered using this criterion.

For this reason, weaker ordering relations have been introduced, among which the most important is the SSD. We say that X is larger than Y with respect to SSD, denoted as $X \geq_2 Y$, if $\int_{-\infty}^{x} F(t)dt \leq \int_{-\infty}^{x} G(t)dt, \forall x$. Equivalently, $X \geq_2 Y$ if and only if $\mathbb{E}u(X) \geq \mathbb{E}u(Y)$ for any increasing and concave function u. In economics, SSD generally represents all nonsatiable and risk-averse decision makers, expressing a preference for the random variable with larger values or smaller dispersion. For example, $X \geq_2 Y$ entails that $\mathbb{E}X \geq \mathbb{E}Y$ and, in case of equality, $Var(X) \leq Var(Y)$ and $\gamma(X) \leq \gamma(Y)$, where γ denotes the Gini coefficient. The above definitions may be generalized to kth order SD, denoted as $X \geq_k Y$, $k = 1, 2, 3, \ldots$, and represented using the following integral inequality $F^{[k]}(x) \leq G^{[k]}(x), \forall x$, where $H^{[1]} = H$ and $H^{[k]}(x) = \int_{-\infty}^{x} H^{[k-1]}(t)dt$, for $k \in \{2, 3, \ldots\}$.

Besides the classic definitions of SD discussed above, different notions—often including FSD and SSD as special or limiting cases—have been studied in the literature. Notable examples are the *inverse* SD (Muliere & Scarsini, 1989), which is based on recursive integration of the quantile function instead of the CDF, and coincides with classic SD at degrees 1 and 2, and also some fractional-degree SD relations that interpolate FSD and SSD (see, e.g., Müller et al., 2017), as discussed in more detail in Section 6.

2.2 | The Lorenz P–P plot

The goal of this article is to test the null hypothesis $\mathcal{H}_0 : X \geq_2 Y$ versus the alternative $\mathcal{H}_1 : X \ngeq_2 Y$. This requires estimating some kind of distance between the situation of dominance and the situation of nondominance. The classic solution (Barrett & Donald, 2003; Davidson & Duclos, 2000) is to construct test statistics based on an empirical version of the difference $\int_{-\infty}^{x} F(t)dt - \int_{-\infty}^{x} G(t)dt$, which is expected to be large, at least at some point, if $\mathcal{H}_0$ is false. However, the main issue with the usual definition of SSD, based on these integrated CDFs, is that such integrals are unbounded in $[0, \infty)$, so there are no uniformly consistent estimators for them unless both distributions have bounded support. Not by chance, Barrett and Donald (2003) require that F and G have common bounded support $[0, a]$, with a finite, to derive consistent tests of stochastic dominance of order k, including SSD. To avoid this limitation, we rely on an alternative but equivalent definition of SSD in terms of the unscaled Lorenz curve, which is always bounded. In particular, we observe that some stochastic orders may be alternatively expressed in terms of a Q–Q plot (Lando et al., 2023) or a P–P plot (Lehmann & Rojo, 1992). Similarly, SSD can be characterized using the modified P–P plot described below.

Let H^{-1} be the (left-continuous) quantile function of the CDF H. The unscaled Lorenz curve of H is defined as $L_H(p) = \int_0^p H^{-1}(t)dt, p \in [0, 1]$. The symbol L_H is often used for the scaled version of the Lorenz curve, that is L_H/μ_H, while we use L_H to denote the *unscaled* Lorenz curve, for the sake of simplicity. Note that $L_H : [0, 1] \to [0, \mu_H]$ is increasing, convex and continuous in the unit interval. Then, L_H^{-1} is increasing and concave in $[0, \mu_H]$. However, for technical reasons, we let $L_H^{-1}(y) = 1$ for $y > \mu_H$, so that $L_H^{-1} : [0, \infty) \to [0, 1]$.

Now, given the pair of CDFs F and G, consider the increasing continuous function

$$Z(p) = L_G^{-1} \circ L_F(p), \qquad p \in [0, 1],$$

which takes values in $[0, 1 \wedge L_G^{-1}(\mu_F)]$, where $x \wedge y$ denotes the minimum between two real numbers x and y. Letting $\nu = 1 \wedge L_F^{-1}(\mu_G)$, note that, if $\mu_G < \mu_F$, then $\nu < 1$ and we set $Z(p) = 1$ for $p \in (\nu, 1]$. Given some point $y = L_F(p)$, for $p \in [0, 1]$, the graph of Z is a P–P plot with coordinates $(L_F^{-1}(y), L_G^{-1}(y))$, which will be referred to as the Lorenz P–P plot (LPP). Within an economic framework, $Z(p)$ returns the probability given by G to the average level of income corresponding to $L_F(p)$. In particular, if such a level cannot be reached under G, we have $Z(p) = 1$. The LPP is scale-free, like the classic P–P plot; in particular, if X and Y are both multiplied by the same positive scale factor, then Z remains unchanged.

To see how Z can be leveraged to characterize SSD, first recall that $X \geq_2 Y$ if and only if $L_F(p) \geq L_G(p),\ \forall p \in [0, 1]$; see, for example, Shaked and Shantikumar (2007, ch. 4). Such a relation can be equivalently expressed in terms of Z:

$$X \geq_2 Y \iff Z(p) \geq p, \quad \forall p \in [0, 1]. \tag{1}$$

It is generally complicated to obtain an explicit expression of Z for parametric probabilistic models. Explicit calculations for the case of a Weibull versus a unit exponential distribution are provided in Example 1 below, while a graphical illustration is given in Figure 1. Differently, and more importantly for our testing purposes, the LPP can be computed quite easily in the empirical case, as discussed in Section 3.

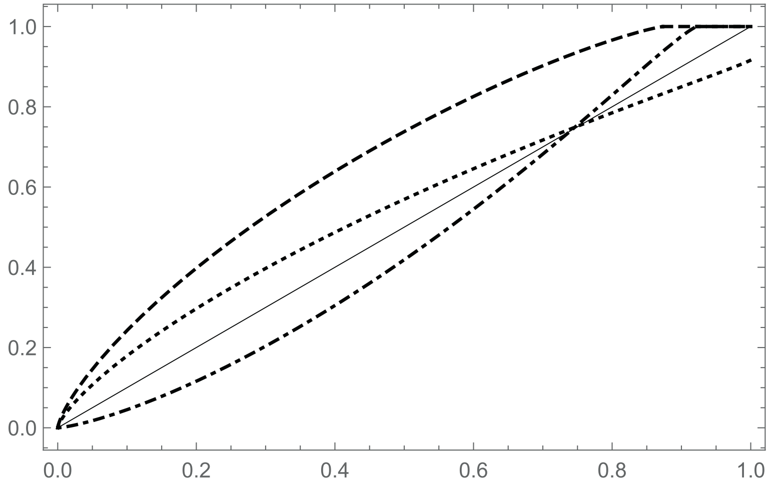

FIGURE 1 The LPP in Example 1 for the case $a = 2, b = 1.5$ (dashed), in which $F \geq_2 G$, and for the cases $a = 2, b = 0.8$ (dotted) and $a = 0.6, b = 1.2$ (dot-dashed), in which $F \ngeq_2 G$.

Example 1. Consider the Weibull distribution $F(x) = 1 - \exp(-(x/b)^a)$, with $a, b > 0$, and the unit exponential distribution, $G(x) = 1 - \exp(-x)$, both supported on $x \geq 0$. In this case, the LPP has the following expression (see the Appendix for details):

$$Z(p) = 1 \wedge \mathcal{R}\left[1 - \exp\left(1 + W_{-1}\left(\frac{b\big(\Gamma(1+1/a) - \Gamma\big(1+1/a, -\log(1-p)\big)\big) - 1}{e}\right)\right)\right],$$

where $\mathcal{R}$ indicates the real part of a complex number, $\Gamma(\cdot, x)$ is the incomplete gamma function and W_{-1} is the Lambert function (Corless et al., 1996). Using the properties of SSD and the crossing conditions described in Shaked and Shantikumar (2007), it is easy to verify that $F \geq_2 G$ if and only if $a \geq 1$ and $\mu_F = b\,\Gamma(1+1/a) \geq \mu_G = 1$. Figure 1 shows the behavior of Z when $F \geq_2 G$ and $F \ngeq_2 G$.

2.3 | Detecting deviations from SSD

Denote the identity function by I. The representation of SSD in (1) can be leveraged to construct a test. In fact, $\mathcal{H}_0 : X \geq_2 Y$ is false if and only if $I - Z$ is strictly positive at some point in the unit interval. Accordingly, departures from SSD can be detected by quantifying the positive part of the difference between I and Z. This may be represented by some functional $\mathcal{T}$ applied to the difference $I - Z \in C[0,1]$. In particular, we propose a family of test statistics obtained as empirical versions of the functionals

$$\mathcal{T}_p(I - Z) = \|(I - Z)_+\|_p,$$

for $p \geq 1$, including $p = \infty$. It can be shown that such functionals satisfy the following properties.

Proposition 1. *For every $v_1, v_2 \in C[0,1]$ and for every $p \geq 1$,*

1. *If $v_1(x) = 0, \forall x \in [0,1]$, then $\mathcal{T}_p(v_1) = 0$;*
2. *if $v_1(x) \leq 0, \forall x \in [0,1]$, then $\mathcal{T}_p(v_2) \leq \mathcal{T}_p(v_2 - v_1), \forall v_2$;*
3. *if $v_1(x) > 0$ for some $x \in [0,1]$, then $\mathcal{T}_p(v_1) > 0$;*
4. *$|\mathcal{T}_p(v_1) - \mathcal{T}_p(v_2)| \leq \|v_1 - v_2\|_\infty$;*
5. *$c\mathcal{T}_p(v_1) = \mathcal{T}_p(cv_1)$, for any positive constant $c > 0$;*
6. *$\mathcal{T}_p$ is convex;*
7. *for any $p_2 \geq p_1 \geq 1$, $\mathcal{T}_{p_2}(v_1) \geq \mathcal{T}_{p_1}(v_1)$.*

Henceforth, we will denote simply by $\mathcal{T}$ any general functional satisfying the above properties 1–6. These properties determine a family of functionals which may be used to obtain consistent tests. In particular, Properties 2 and 3 completely characterize SSD, in that $\mathcal{T}(I - Z) = 0$ if and only if $X \geq_2 Y$, while $\mathcal{T}(I - Z) > 0$ if and only if $X \ngeq_2 Y$. Differently, Property 7 deals just with the class $\mathcal{T}_p$ and shows that functionals of this kind measure the deviations from $\mathcal{H}_0$ in a monotone way, that is, smaller (larger) values of p downsize (emphasize) deviations, represented by the function $(I - Z)_+$. Proposition 1 generalizes Lemma 2 of Barrett et al. (2014), which deals with the special cases of $\mathcal{T}_1$ and $\mathcal{T}_\infty$. They introduced tests for the Lorenz dominance by applying $\mathcal{T}$ to the difference between the (scaled) Lorenz curves, that is $\mathcal{T}(L_G/\mu_G - L_F/\mu_F)$. One may extend their approach to SSD by considering $\mathcal{T}(L_G - L_F)$ (see, e.g., Zhuang et al., 2023). However, in this article, we propose leveraging $\mathcal{T}(I - Z)$, which has some advantages over $\mathcal{T}(L_G - L_F)$. For instance, $I - Z$

is scale-free by properties of the LPP. On the contrary, if X and Y are multiplied by a positive scale factor $c > 0$, then the difference between the unscaled Lorenz curves becomes $c(L_G - L_F)$. Moreover, $|L_G - L_F| < \max(\mu_F, \mu_G)$ whereas, for any F and G, $|I - Z|$ is always upper-bounded by 1.

3 | ESTIMATION OF THE LPP

3.1 | Sampling assumptions

Let $\mathcal{X} = \{X_1,\dots,X_n\}$ and $\mathcal{Y} = \{Y_1,\dots,Y_m\}$ be i.i.d. random samples from F and G, respectively. As in Barrett et al. (2014), we will deal with two different sampling schemes: independent sampling and matched pairs. In the first scheme, the two samples $\mathcal{X}$ and $\mathcal{Y}$ are independent of each other, and sample sizes n and m may differ. In contrast, in the matched-pairs scheme, $n = m$ and we have n i.i.d. pairs $\{(X_1, Y_1),\dots,(X_n, Y_n)\}$ drawn from a bivariate distribution with F and G as marginal CDFs. For both sampling schemes, we will consider the asymptotic regime in which $n \to \infty$, $\lim_{n\to\infty} nm/(n+m) = \infty$ and $\lim_{n\to\infty} n/(n+m) = \lambda \in [0,1]$. These assumptions are quite standard in the literature (see, e.g., Barrett et al., 2014; Sun & Beare, 2021). For $\lambda \in (0,1)$, this implies that, as n diverges, m also goes to infinity with the same order, as the ratio n/m tends to $\lambda/(1-\lambda)$, which is finite and strictly positive. However, λ can also take one of the endpoints when one of the sample sizes grows faster than the other.

3.2 | Empirical LPP

The abovementioned random samples $\mathcal{X}$ and $\mathcal{Y}$ yield the empirical CDFs

$$F_n(x) = (1/n)\sum_{i=1}^{n} \mathbb{1}(X_i \le x) \quad \text{and} \quad G_m(x) = (1/m)\sum_{j=1}^{m} \mathbb{1}(Y_j \le x),$$

respectively. We denote with $X_{(k)}$ and $Y_{(k)}$ the order statistics of rank k from $\mathcal{X}$ and $\mathcal{Y}$, and their sample means with $\overline{X}_n$ and $\overline{Y}_m$, respectively. Using the plugin method, the empirical counterparts of L_F and L_G^{-1} are L_{F_n} and $L_{G_m}^{-1}$, where $L_{F_n}(p) = \int_0^p F_n^{-1}(t)dt$ for $p \in [0,1]$, L_{G_m} is defined similarly, and $L_{G_m}^{-1}$ is the inverse of L_{G_m}. Coherently with our definition of L_G^{-1}, we let $L_{G_m}^{-1}(p) = 1$ for $p > \overline{Y}_m$. Note that L_{F_n} coincides with the empirical unscaled Lorenz curve (Shorrocks, 1983), that is a piecewise linear function joining the points $(k/n, (1/n)\sum_{i=0}^{k} X_{(i)})$, for $k = 0,\dots,n$, with $X_{(0)} := 0$.

Alternatively, L_F and L_G^{-1} can also be estimated using the following step functions,

$$\widetilde{L}_{F_n}(p) = \begin{cases} 0 & p = 0, \\ (1/n)\displaystyle\sum_{i=1}^{\lceil np \rceil} X_{(i)} & p \in (0,1], \end{cases}$$

where $\lceil \cdot \rceil$ is the ceiling function, and

$$\widetilde{L}_{G_m}^{-1}(p) = \begin{cases} 0 & p \in [0, Y_{(1)}/m), \\ j/m & p \in [(1/m)\displaystyle\sum_{k=1}^{j} Y_{(k)}, (1/m)\sum_{k=1}^{j+1} Y_{(k)}), \quad 1 \le j \le m-1, \\ 1 & p \ge \overline{Y}_m. \end{cases}$$

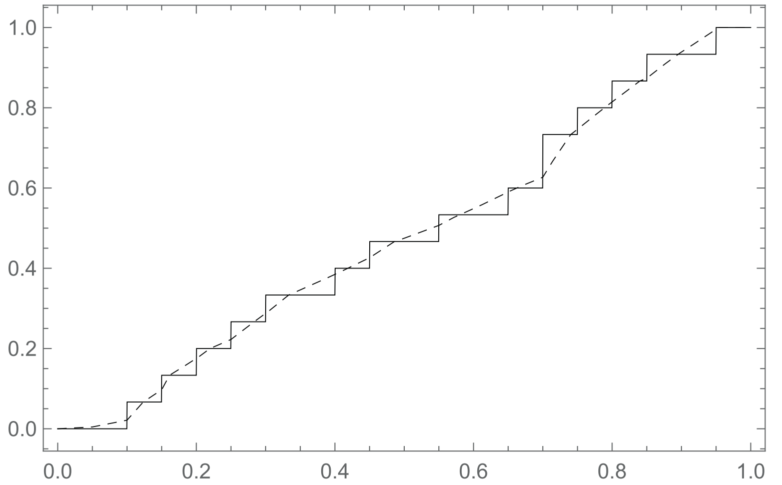

FIGURE 2 $\widetilde{Z}_{n,m}$ (solid) versus $Z_{n,m}$ (dashed) for $n = 20$ and $m = 15$.

$\widetilde{L}_{G_m}$ and $\widetilde{L}^{-1}_{F_m}$ are defined similarly. Here, $\widetilde{L}^{-1}_{G_m}$ is the right-continuous generalized inverse of $\widetilde{L}_{G_m}$, that is, $\widetilde{L}^{-1}_{G_m}(p) = \inf\{u : \widetilde{L}_{G_m}(u) > p\}$. Note that L_{F_n} and $\widetilde{L}_{F_n}$ coincide at i/n, for $i = 1, \ldots, n$, and, likewise, $L^{-1}_{G_m}$ and $\widetilde{L}^{-1}_{G_m}$ coincide at $(1/m)\sum_{k=1}^{j} Y_{(k)}$, for $j = 1, \ldots, m$, so these alternative empirical versions of L_F and L^{-1}_G are clearly asymptotically equivalent.

According to the different empirical versions of L_F and L^{-1}_G, we may obtain different estimators of Z. One may consider $Z_{n,m} = L^{-1}_{G_m} \circ L_{F_n}$, which is a continuous piecewise linear function, or alternatively $\widetilde{Z}_{n,m} = \widetilde{L}^{-1}_{G_m} \circ \widetilde{L}_{F_n}$, which is a left-continuous step function with jumps at the points $\{i/n : i = 1, \ldots, n\}$, taking values in $\{j/m : j = 1, \ldots, m\}$. In an economic framework, $\widetilde{L}^{-1}_{G_m}(p)$ gives the relative frequency of observations from Y whose average level of income is less than p. Therefore, $\widetilde{Z}_{n,m}(i/n)$ returns the relative frequency of observations from Y whose average level of income is less than that of the poorest i individuals from X, that is, $(1/n)\sum_{k=1}^{i} X_{(k)}$. Note that the value of $Z_{n,m}$ at its "node" points i/n does not generally coincide with the value of $\widetilde{Z}_{n,m}$ at its jump points. The difference between these two functions is depicted in Figure 2. In this article, we will use $\widetilde{Z}_{n,m}$ or $Z_{n,m}$ as is more convenient, since the two are asymptotically equivalent. In fact, the sup-distance among $\widetilde{Z}_{n,m}$ and $Z_{n,m}$ tends to zero as n and m diverge, as established in the following proposition.

Proposition 2. *For any $n, m > 0$, $\|\widetilde{Z}_{n,m} - Z_{n,m}\|_\infty \leq 1/m$. Consequently, as $m \to \infty$, $\|\widetilde{Z}_{n,m} - Z_{n,m}\|_\infty \to 0$.*

In our asymptotic scenario, when $n \to \infty$ we also have $m \to \infty$, hence the second part of Proposition 2 holds. Moreover, based on the strong uniform consistency of Lorenz curve estimators and their inverse functions (Csörgö et al., 2013; Goldie, 1977), we can prove the strong uniform consistency of $Z_{n,m}$ and $\widetilde{Z}_{n,m}$.

Proposition 3. *As $n, m \to \infty$, $Z_{n,m} \to Z$ and $\widetilde{Z}_{n,m} \to Z$ a.s. and uniformly in* $[0, 1]$.

4 | WEAK CONVERGENCE OF THE LPP PROCESS

The empirical process associated with Z, henceforth referred to as the LPP process, may be useful to characterize the limit distribution of the test statistic under the null hypothesis of SSD. In this section, we study the asymptotic properties of the LPP process, defined as

$$\mathcal{Z}_n(p) = \sqrt{r_n}(Z_{n,m}(p) - Z(p)), \qquad p \in [0,1],$$

where $r_n = nm/(n+m)$, and let $\nu_n = 1 \wedge L_{F_n}^{-1}(\overline{Y}_m)$ be the empirical counterpart of $\nu = 1 \wedge L_F^{-1}(\mu_G)$. For $\nu < 1$ we know that $Z(t) = 1$ when $t \in (\nu, 1]$. In this case we have $\|\mathcal{Z}_n \mathbb{1}(\nu, 1]\|_\infty \to 0$ a.s., since also $Z_{n,m}(p) = 1$ for $t \in (\nu_n, 1]$ and $\nu_n \to \nu$ a.s. In other words, the interval $(\nu_n, 1]$ contains no information. Accordingly, we are particularly interested in the asymptotic behavior of $\mathcal{Z}_n$ restricted to $[0, \nu_n]$, namely $\mathcal{Z}_n \mathbb{1}[0, \nu_n]$. Weak convergence of the LPP process can be derived under the following assumptions.

Assumption 1. Both F and G are continuously differentiable with strictly positive density, and have a finite moment of order $2+\epsilon$ for some $\epsilon > 0$. Moreover, $F(0) = G(0) = 0$.

Assumption 2. There exists some number $c > 0$ such that $G^{-1}(0^+) = c$.

The latter assumption does not represent a limitation in terms of applicability. In fact, if $G^{-1}(0^+) = 0$, one can apply the test to the shifted samples $\mathcal{X} + \delta$ and $\mathcal{Y} + \delta$, for some arbitrarily small $\delta > 0$, recalling that $X \geq_2 Y$ if and only if $X + \delta \geq_2 Y + \delta$. In our simulations we set $\delta = 10^{-4}$, obtaining results that are almost indistinguishable from those under $\delta = 0$. However, as the unscaled Lorenz curve is not translation invariant, the outcome of any test based on it (such as, e.g., Andreoli, 2018; Zhuang et al., 2023) may depend on the shift δ. Actually, in our experiments, we noted that larger values of δ may even improve the power of our tests.

The following theorem establishes the weak convergence of $\mathcal{Z}_n$, leveraging some recent results in Kaji (2018) that enable the derivation of the Hadamard differentiability of the map from CDFs to quantile functions (see also Lemma A.18 in Weitkamp et al. (2024)), and the corresponding weak convergence of the quantile process in the L_1 norm, under Assumption 1. As discussed in Sun and Beare (2021, section 2.4), this extends the applicability of earlier Hadamard differentiability conditions, based on stronger distributional assumptions such as bounded support (Van der Vaart & Wellner, 1996, Lemma 3.9.23). Because the LPP is obtained using the composition of L_G^{-1} with L_F, Assumption 2 is used in our proof to ensure that the derivative of L_G^{-1}, that is, $(L_G^{-1})' = 1/G^{-1}(L_G^{-1})$, is bounded. This yields the Hadamard differentiability of the composition map $\zeta(L_F, L_G^{-1}) = L_G^{-1} \circ L_F$ using Lemmas 3.9.25 and 3.9.27 of Van der Vaart and Wellner (1996). Finally, the weak convergence of $\mathcal{Z}_n$ follows the functional delta method (Van der Vaart & Wellner, 1996, Sect. 3.9).

Let $\mathcal{B}$ be a centered Gaussian element of $C[0,1] \times C[0,1]$ with covariance function $Cov(\mathcal{B}(x_1, y_1), \mathcal{B}(x_2, y_2)) = C(x_1 \wedge x_2, y_1 \wedge y_2) - C(x_1, y_1)C(x_2, y_2)$. Under the independent-sampling scheme, $C(x_1, y_1) = x_1 y_1$ is the product copula, whereas, under the matched-pairs scheme, C is the copula associated with the pair (X_i, Y_i), $i = 1, \ldots, n$. Now, let $\mathcal{B}_1(x_1) = \mathcal{B}(x_1, 1)$ and $\mathcal{B}_2(x_2) = \mathcal{B}(1, x_2)$. The random elements $\mathcal{B}_1$ and $\mathcal{B}_2$ are Brownian bridges that are independent under the independent-sampling scheme, but may be dependent under the matched-pairs one.

Theorem 1. *Under Assumptions 1 and 2 and both independent-sampling and matched-pairs schemes, we have* $\sqrt{r_n}(Z_{n,m} - Z) \rightsquigarrow \mathcal{Z}\mathbb{1}[0, \nu]$ *in* $C[0,1]$, *where*

$$\mathcal{Z}(p) = \frac{\sqrt{1-\lambda}\int_0^p \mathcal{B}_1(t)dF^{-1}(t) - \sqrt{\lambda}\int_0^{Z(p)} \mathcal{B}_2(t)dG^{-1}(t)}{G^{-1} \circ Z(p)}.$$

The dependence structure between $\mathcal{X}$ and $\mathcal{Y}$ affects the variance of the limit process $\mathcal{Z}$, which is given by

$$\begin{aligned} Var(\mathcal{Z}(p)) = \frac{1}{(G^{-1}\circ Z(p))^2}\Bigg[&(1-\lambda)Var\left(\int_0^p \mathcal{B}_1(u)dF^{-1}(u)\right) + \lambda Var\left(\int_0^{Z(p)} \mathcal{B}_2(u)dG^{-1}(u)\right) \\ &- 2\sqrt{\lambda(1-\lambda)}Cov\left(\int_0^p \mathcal{B}_1(u)dF^{-1}(u), \int_0^{Z(p)} \mathcal{B}_2(u)dG^{-1}(u)\right)\Bigg] \\ = \frac{1}{(G^{-1}\circ Z(p))^2}\Bigg[&(1-\lambda)\int_0^p (u-u^2)dF^{-1}(u) + \lambda\int_0^{Z(p)} (u-u^2)dG^{-1}(u) \\ &- 2\sqrt{\lambda(1-\lambda)}\int_0^{Z(p)}\int_0^p (C(u,v)-uv)dF^{-1}(u)dG^{-1}(v)\Bigg]. \end{aligned}$$

Recalling a notion of *positive dependence* in Chapter 9 of Shaked and Shantikumar (2007), the copula C is larger than the copula C' in the *positive quadrant dependence* (PQD) order if $C(u,v) \geq C'(u,v)$, for every $u, v \in [0,1]$. Given F and G, the integrand of the last summand, $C(u,v) - uv$, gets larger (smaller) when the copula is larger (smaller) in the PQD order, while it is null under independence. This implies that $Var(\mathcal{Z}(p))$ is smaller (larger) when the dependence structure is positive (negative).

It is also interesting to observe that, if $X =_d Y$, the result of Theorem 1 boils down to

$$\begin{aligned} \sqrt{r_n}(Z_{n,m}(t) - t) \rightsquigarrow &\frac{1}{F^{-1}(t)}\int_0^t \left(\sqrt{1-\lambda}\mathcal{B}_1(u) - \sqrt{\lambda}\mathcal{B}_2(u)\right)dF^{-1}(u) \\ = &\frac{1}{F^{-1}(t)}\int_0^t \widetilde{\mathcal{B}}(u)dF^{-1}(u), \end{aligned}$$

in $C[0,1]$, where $\widetilde{\mathcal{B}}$ is the Brownian bridge defined as $\widetilde{\mathcal{B}} = \sqrt{1-\lambda}\mathcal{B}_1 - \sqrt{\lambda}\mathcal{B}_2$.

Finally, note that, using the asymptotic equivalence implied by Proposition 2, all the results in this section still hold if one replaces $Z_{n,m}$ with $\widetilde{Z}_{n,m}$.

5 | ASYMPTOTIC PROPERTIES OF THE TEST

As discussed in Section 2.3, deviations from $\mathcal{H}_0 : X \geq_2 Y$ can be measured via the test statistic $\widehat{\mathcal{T}}_n = \sqrt{r_n}\,\mathcal{T}(I - Z_{n,m})$. Intuitively, we reject $\mathcal{H}_0$ if $\widehat{\mathcal{T}}_n$ is large enough. However, since the null hypothesis is nonparametric, the main issue is how to determine the distribution of $\widehat{\mathcal{T}}_n$, or alternatively of an upper bound for $\widehat{\mathcal{T}}_n$, under $\mathcal{H}_0$. Following the approach of Barrett et al. (2014), it is easily seen that, under $\mathcal{H}_0$, the test statistic $\sqrt{r_n}\,\mathcal{T}(I - Z_{n,m})$ is dominated by $\sqrt{r_n}\,\mathcal{T}(Z - Z_{n,m})$, which therefore can be used to simulate p-values or critical values via bootstrap, thus ensuring that the size of the test is asymptotically bounded by some arbitrarily small probability α. Using the continuous mapping theorem, $\sqrt{r_n}\,\mathcal{T}(Z - Z_{n,m})$ is asymptotically distributed as $\mathcal{T}(\mathcal{Z})$, allowing us to derive large-sample properties of the test. The limit behavior of $\widehat{\mathcal{T}}_n$ under the null and the alternative hypotheses is established in the following lemma.

Lemma 1.

1. *Under* $\mathcal{H}_0$, $\sqrt{r_n}\,\mathcal{T}(I - Z_{n,m}) \leq \sqrt{r_n}\,\mathcal{T}(Z - Z_{n,m}) \rightsquigarrow \mathcal{T}(\mathcal{Z})$. *Moreover, for any* $\alpha < 1/2$, *the* $(1-\alpha)$ *quantile of* $\mathcal{T}(\mathcal{Z})$ *is positive, finite, and unique.*

2. *Under* $\mathcal{H}_1$, $\sqrt{r_n}\ \mathcal{T}(I - Z_{n,m}) \to_p \infty$.

In practice, the limit distribution of $\sqrt{r_n}\mathcal{T}(Z - Z_{n,m})$ under the null hypothesis may be approximated using a bootstrap approach, as discussed in the next subsection.

5.1 | Bootstrap decision rule

Let us denote the bootstrap estimators of the empirical CDFs F_n and G_m with F_n^* and G_m^*, respectively, defined as

$$F_n^*(x) = (1/n)\sum_{i=1}^{n} M_i^{(1)}\mathbb{1}(x \le X_i), \qquad G_m^*(x) = (1/m)\sum_{j=1}^{m} M_j^{(2)}\mathbb{1}(x \le Y_j),$$

where $M^{(1)} = (M_1^{(1)}, \dots, M_n^{(1)})$ and $M^{(2)} = (M_1^{(2)}, \dots, M_m^{(2)})$ are independent of the data and are drawn from a multinomial distribution according to the chosen sampling scheme. Each $M_i^{(1)}$ ($M_j^{(2)}$) counts how many times observation X_i (Y_j) is resampled (Van der Vaart & Wellner, 1996, p. 180). In particular, under the independent-sampling scheme, $M^{(1)}$ and $M^{(2)}$ are independently drawn from multinomial distributions with uniform probabilities over n and m trials, respectively. Under the matched-pairs scheme, we have $M^{(1)} = M^{(2)}$ drawn from the multinomial distribution with uniform probabilities over $n = m$ trials, which means that we sample (with replacement) pairs of data, from the n pairs $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$. Correspondingly, by applying the definitions in Section 2, we obtain the bootstrap estimators of the unscaled Lorenz curves, denoted with $L_{F_n^*}$ and $L_{G_m^*}$, as well as the inverse $L_{G_m^*}^{-1}$, and we define $Z_{n,m}^* = L_{G_m^*}^{-1} \circ L_{F_n^*}$. As is shown below, the random process $\sqrt{r_n}\ \mathcal{T}(Z_{n,m} - Z_{n,m}^*)$ has the same limiting distribution $\mathcal{T}(\mathcal{Z})$, as $\sqrt{r_n}\ \mathcal{T}(Z - Z_{n,m})$. Therefore, bootstrap p-values are determined by

$$p = P\{\sqrt{r_n}\ \mathcal{T}(Z_{n,m} - Z_{n,m}^*) > \sqrt{r_n}\ \mathcal{T}(I - Z_{n,m})\},$$

and can be approximated, based on K bootstrap replicates, by

$$p \approx (1/K)\sum_{k=1}^{K} \mathbb{1}\{\sqrt{r_n}\ \mathcal{T}(Z_{n,m} - Z_{k;n,m}^*) > \sqrt{r_n}\ \mathcal{T}(I - Z_{n,m})\},$$

where $Z_{k;n,m}^*$ is the kth resampled realization of $Z_{n,m}^*$. As usual, the test rejects $\mathcal{H}_0$ if $p < \alpha$. The asymptotic behavior of the test is addressed using the following proposition.

Proposition 4. *Under Assumptions 1 and 2 and the sampling schemes in Section 3.1*,

1. *If* $\mathcal{H}_0$ *is true*, $\lim_{n\to\infty} P\{\text{reject } \mathcal{H}_0\} \le \alpha$;
2. *If* $\mathcal{H}_0$ *is false*, $\lim_{n\to\infty} P\{\text{reject } \mathcal{H}_0\} = 1$.

6 | EXTENSION TO FRACTIONAL-DEGREE SD

An important topic in SD theory is represented by SD relations "between" FSD and SSD. This is motivated by the fact that FSD is a strong requirement, but, on the other hand, SSD corresponds to total risk aversion, which is quite restrictive in some cases (Müller et al., 2017). There are different ways to define classes of orders that interpolate between FSD and SSD, and each leads to a different family of SD relations, typically parameterized by a real number that represents the strength of the dominance. The first attempt in this direction is ascribable to Fishburn (1980), who used fractional-degree integration to interpolate the classic kth order SD at all integer orders $k \geq 1$. More recently, Müller et al. (2017), Huang et al. (2020), and Lando and Bertoli-Barsotti (2020) proposed different parameterizations, with different interpretations and properties, which coincide with classic SD only at orders 1 and 2. In this section, we introduce a simple but very general family of fractional-degree orders, which have the advantage that they can be easily tested using the LPP method discussed earlier. This family can be defined as follows.

Let $\mathcal{U}$ be the family of increasing absolutely continuous functions u over the non-negative half line. Under an economic perspective, u may be understood as a utility function, assigning values to monetary outcomes. For some $u \in \mathcal{U}$, we say that X dominates Y with respect to *u-transformed stochastic dominance* (u-TSD), and write $X \geq_u^T Y$, if $u(X) \geq_2 u(Y)$. TSD has been studied by Meyer (1977), who denoted it as SSD with respect to u, and by Huang et al. (2020), who focused on a particular parametric choice of u. Since u-TSD represents SSD between the transformed random variables $u(X)$ and $u(Y)$, then it can be simply expressed and tested through the LPP of $u(X)$ and $u(Y)$.

The behavior of TSD clearly depends on the choice of u. To understand this, let $u, \tilde{u} \in \mathcal{U}$ be two transformation functions defined on the same interval. Generalizing Chan et al. (1990), we say that u is *more convex* than $\tilde{u}$ and write $u \geq_c \tilde{u}$ iff $u \circ \tilde{u}^{-1}$ is convex. The following theorem shows that TSD can be equivalently expressed in terms of expected utilities, thus generalizing Theorem 1 of Huang et al. (2020).

Theorem 2. *$X \geq_u^T Y$ if and only if $\mathbb{E}(\phi(X)) \geq \mathbb{E}(\phi(Y))$, for every increasing utility ϕ such that $u \geq_c \phi$.*

It is easy to see that, if u and ϕ are twice differentiable, the condition $u \geq_c \phi$ is equivalent to $\rho_\phi(x) \geq \rho_u(x), \forall x$, where $\rho_g(x) = g''(x)/g'(x)$ is the Arrow-Pratt index of absolute risk aversion associated with the utility function g. Moreover, the following general properties hold.

Theorem 3.

1. *If $u_1 \geq_c u_2$ then $X \geq_{u_1}^T Y \Rightarrow X \geq_{u_2}^T Y$;*
2. *$X \geq_1 Y$ if and only if $X \geq_u^T Y, \forall u \in \mathcal{U}$.*

Intuitively, the degree of convexity of the function u determines the strength of the SD relation, and SSD is obtained by taking u to be the identity function, whereas FSD is obtained when u is "infinitely steep".

Families of utility functions within $\mathcal{U}$ can be obtained easily by composing the quantile function and the CDF of two absolutely continuous random variables. For example, one may consider the class of utility functions studied by Huang et al. (2020) and given by $u_c(x) = \exp((1/c - 1)x)$, for $c \in (0, 1)$. Since this article deals with tests for non-negative random variables, we focus on a simpler choice, that is $u_\theta(x) = x^\theta$, with $\theta \geq 0$. Correspondingly, hereafter we denote the ordering relation $X \geq_{u_\theta}^T Y$ with $X \geq_{1+1/\theta}^T Y$, thus yielding a continuum of SD relations that get stronger and stronger as θ grows. Using Theorem 3, this order is characterized by those utility functions that have an Arrow-Pratt index larger than or equal to $(\theta - 1)/x$.

Since $X \geq_{1+1/\theta}^T Y$ is equivalent to $X^\theta \geq_2 Y^\theta$, a test for $\mathcal{H}_0^{1+1/\theta} : X \geq_{1+1/\theta}^T Y$ versus $\mathcal{H}_1^{1+1/\theta} : X \ngeq_{1+1/\theta}^T Y$ is readily obtained by applying our method to the LPP of the transformed random samples. In particular, we consider the *generalized LPP*, given by $\widetilde{Z}_{n,m}^\theta = (\widetilde{L}_{G_m}^\theta)^{-1} \circ \widetilde{L}_{F_n}^\theta$, where $\widetilde{L}_{F_n}^\theta$ and

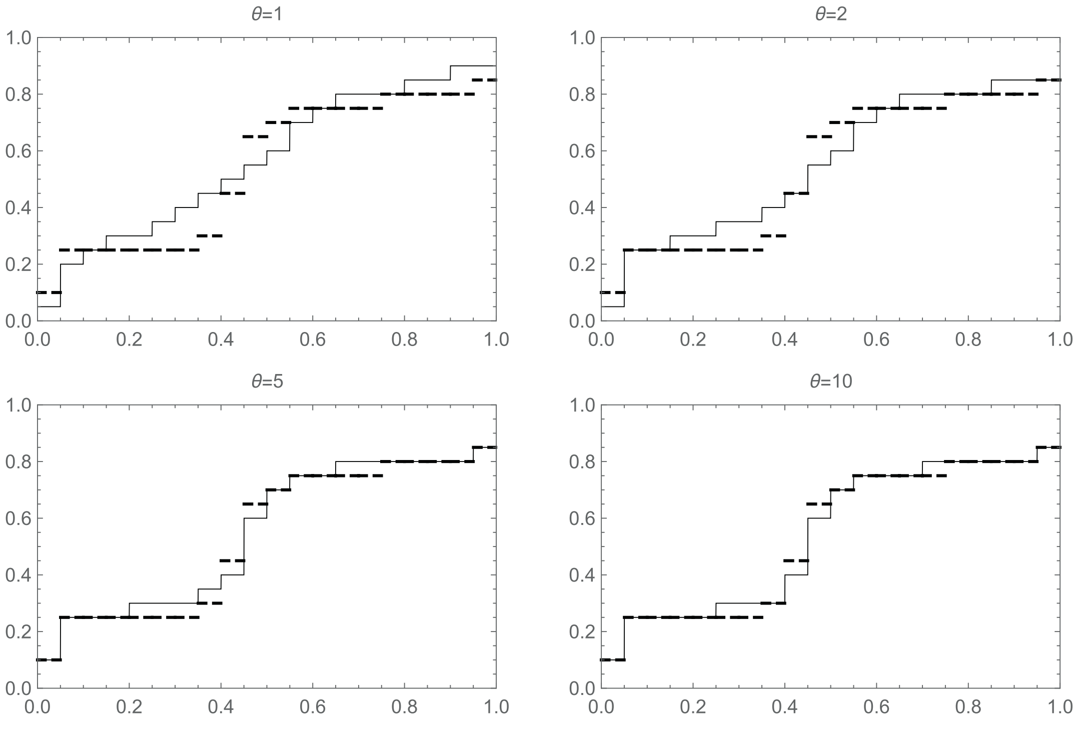

FIGURE 3 The P–P plot (dashed) of two samples of size $n = m = 20$ versus the generalized LPP $\widetilde{Z}^{\theta}_{n,m}$ (solid), for $\theta = 1, 2, 5, 10$. In this example, the plots coincide for $\theta \geq 48$.

$\widetilde{L}^{\theta}_{G_m}$ are the empirical (step-valued) unscaled Lorenz curves corresponding to the transformed samples $\{X^{\theta}_i : i = 1, \ldots, n\}$, and $\{Y^{\theta}_j : j = 1, \ldots, m\}$. $\widetilde{Z}^{\theta}_{n,m}$ is a generalized P–P plot, in that it coincides with $\widetilde{Z}_{n,m}$ for $\theta = 1$. More interestingly, we prove that, as $\theta \to \infty$, $\widetilde{Z}^{\theta}_{n,m}$ tends to the classic P–P plot of the nontransformed samples, that is to $G_m \circ F_n^{-1}$, as depicted in Figure 3. In particular, one may always find some θ large enough such that the two P–P plots coincide, meaning that our tests may be also applied to FSD, expressed as $G \circ F^{-1}(x) \geq x$; see, for example, Davidov and Herman (2012) and Beare and Clarke (2022) for FSD tests based on the P–P plot. In fact, this idea is coherent with the intuition that, for $\theta \to \infty$, the stochastic inequality $X \geq^{T}_{1+1/\theta} Y$ reduces to $X \geq_1 Y$, as formally established in the following theorem.

Theorem 4.

1. *As $\theta \to \infty$, the condition for $X \geq^{T}_{1+1/\theta} Y$, that is*

$$\int_{-\infty}^{x} F(t) du_{\theta}(t) \leq \int_{-\infty}^{x} G(t) du_{\theta}(t), \qquad \forall x,$$

   tends to the condition for $X \geq_1 Y$, that is $F(x) \leq G(x), \forall x$.
2. *There exists some θ_0 such that, for $\theta \geq \theta_0$, the generalized LPP coincides with the classic P–P plot, that is, $\widetilde{Z}^{\theta}_{n,m} = G_m \circ F_n^{-1}$.*

To test FSD as a limit case of TSD, one should choose a value of θ that ensures the result above. However, if θ is too large, computations may be difficult, depending on the precision of the software used. We recommend using $\theta = 50$, which corresponds to testing $\geq {}^{T}_{1.02}$, for a good approximation of FSD.

7 | SIMULATIONS

We perform numerical analyses to investigate the finite-sample properties of the proposed tests. In all simulations, we consider a significance level $\alpha = 0.1$, and run 500 experiments, with 500 bootstrap replicates for each experiment. For simplicity, we set $n = m$, so henceforth we will drop the subscript m. Namely, we consider $n = m = 50, 100, 200, 500, 1000$. The shift δ is set to 10^{-4}, as discussed in Section 4. All computations have been performed in R, and the code is openly available at https://github.com/siriolegramanti/SSD.

In light of Proposition 2, instead of Z_n we use $\widetilde{Z}_n$, which can be computed faster. Accordingly, we focus on two different test statistics, namely $\mathcal{T}_\infty(I - \widetilde{Z}_n)$ and $\mathcal{T}_1(I - \widetilde{Z}_n)$, as defined in Section 2.3. For $n = m$, $\mathcal{T}_\infty$ and $\mathcal{T}_1$ can be rewritten, respectively, as

$$\mathcal{T}_\infty(I - \widetilde{Z}_n) = \max_i \left(\frac{i}{n} - \widetilde{Z}_n\left(\frac{i}{n}\right) \right), \qquad \mathcal{T}_1(I - \widetilde{Z}_n) = \frac{1}{n}\sum_{i=1}^{n} \Psi\left(\frac{2i-1}{2n}\right),$$

where $\Psi(t) = (t - \widetilde{Z}_n(t))_+$. In Section 7.2.4 we also consider test statistics $\mathcal{T}_p$ for a generic $p > 1$.

Our results are compared with those obtained from the tests of Barrett and Donald (2003), which represent the state of the art for SSD tests. In particular, Barrett and Donald (2003) propose three bootstrap-based tests, based on a least favorable configuration, denoted as KSB1, KSB2, and KSB3, which differ just for the bootstrap method employed to simulate the p-values. We focus on KSB3 since it is based on the approach that is most similar to ours. Moreover, KSB3 seems to provide the best results compared to KSB1 and KSB2 as far as SSD is concerned; see tables II-A and II-B in Barrett and Donald (2003). KSB3 is computed as follows. First, we estimate $F^{[2]}(x) = \int_0^x F(t)dt$ with $F_n^{[2]}(x) = \int_0^x F_n(t)dt$ and $G^{[2]}(x) = \int_0^x G(t)dt$ with $G_n^{[2]}(x) = \int_0^x G_n(t)dt$. Accordingly, the test statistic is given by $\sqrt{n}\sup_{x\geq 0}(G_n^{[2]}(x) - F_n^{[2]}(x))$. The p-values are computed by simulating the distribution of $\sqrt{n}\sup_{x\geq 0}((G_n^{[2]*}(x) - G_n^{[2]}(x)) - (F_n^{[2]*}(x) - F_n^{[2]}(x)))$, where $F_n^{[2]*}$ and $G_n^{[2]*}$ are bootstrap versions of $F_n^{[2]}$ and $G_n^{[2]}$, respectively. In particular, the sup is approximated using a grid of evenly spaced values $t_1 < \ldots < t_r$, where t_1 and t_r are the smallest and the largest values in the pooled sample, respectively. As for the number of grid points, we set $r = 100$ as in Barrett and Donald (2003), but we did not notice substantial differences when r increases.

Note that one pair of distributions gives rise to two different hypothesis tests. In fact, one may test $\mathcal{H}_0 : F \geq_2 G$ versus $\mathcal{H}_1 : F \ngeq_2 G$, but also the reversed hypotheses, denoted as $\mathcal{H}_0^R : G \geq_2 F$ versus $\mathcal{H}_1^R : G \ngeq_2 F$. Except for the trivial case $F = G$, if Z does not cross the identity we may have that $\mathcal{H}_0$ is true while $\mathcal{H}_0^R$ is false, or vice versa; differently, if Z crosses the identity, $\mathcal{H}_0$ and $\mathcal{H}_0^R$ are both false.

7.1 | Size properties

To investigate the behavior of the tests under the null hypothesis, we simulate samples from the Weibull family, denoted by $W(a, b)$, with CDF $F_W(x; a, b) = 1 - \exp\{-(x/b)^a\}$. Since the mean of a $W(a, b)$ is b/q_a, where $q_a = 1/\Gamma(1 + 1/a)$, we let $F \sim W(a, q_a)$, for $a = 1.0, 1.1, 1.2, 1.3$, and fix $G \sim W(1, 1)$. All these distributions have mean 1, and in all these cases $\mathcal{H}_0$ holds. Clearly, for $a = 1$

we have $F = G$, whereas the dominance of F over G becomes stronger, and more apparent, for larger values of a.

The results in Tables A1, A2a, A3a, and A4a confirm that the proposed tests, both with $\mathcal{T}_\infty$ and $\mathcal{T}_1$, behave as described in Proposition 4, part 1. Namely, the rejection rate tends to be bounded by $\alpha = 0.1$ under $\mathcal{H}_0$. More specifically, we observe that the rejection rate of the proposed tests tends to α when $F = G$ (see Table A1), while it tends to 0 when F strictly dominates G (see Tables A2a, A3a, and A4a). The rejection rate for the KSB3 test by Barrett and Donald (2003) is also asymptotically bounded by α but, when the dominance is stronger, it is still about α for $n = 1000$. For such a sample size, the rejection rate of both the proposed tests has already reached 0.

7.2 | Power properties

We now investigate the behavior of the tests under $\mathcal{H}_1$. Namely, we focus on cases where F is dominated by G, so that $\mathcal{H}_0$ should be rejected quite easily since Z is always below the identity. As we discuss in Section 7.2.1, the three tests considered behave quite similarly in such cases. We also focus on critical cases in which neither of the two distributions dominates the other, and therefore Z crosses the identity. In particular, the most critical situation for our class of tests is when Z is above the identity everywhere but on a small interval (see Figure 4). The simulation results in Sections 7.2.2 and 7.2.3 show that, in some of the most difficult cases, $\mathcal{T}_1$ and KSB3 struggle to reject $\mathcal{H}_0$, whereas the proposed $\mathcal{T}_\infty$ test stands out as the most reliable.

7.2.1 | Weibull distribution

Using the same distributions as in Section 7.1, except for the case $F = G$, we have that $F >_2 G$ (strictly) and therefore $G \not\geq_2 F$. In these cases, Z is always above the identity. The results, reported

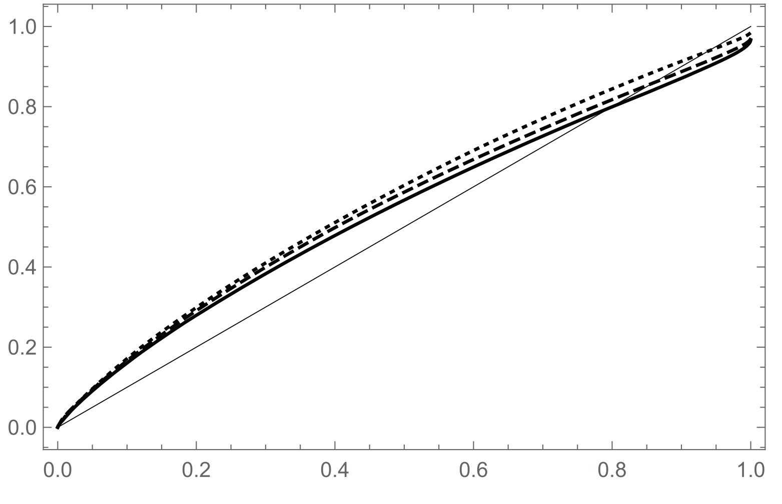

FIGURE 4 The behavior of Z for $q = 1.2$ (solid), $q = 1.5$ (dashed) and $q = 1.8$ (dotted) in the Singh-Maddala case. Especially for $q = 1.8$, it becomes very hard to detect deviations from $\mathcal{H}_0$. In the reverse cases, the LPPs are just the inverse functions of these.

in Tables A2(b), A3(b) and A4(b), show that the power of the tests increases with the sample size. In particular, $\mathcal{T}_1$ seems to outperform $\mathcal{T}_\infty$ for smaller sample sizes, while both the proposed $\mathcal{T}_1$ and $\mathcal{T}_\infty$ tests provide larger power compared to the KSB3 test by Barrett and Donald (2003).

7.2.2 | Lognormal mixture versus lognormal distribution

As a more critical example, we focus on a special case considered by Barrett and Donald (2003, Case 5). Here, F is a mixture of lognormal distributions, $F = 0.9F_{LN}(0.85, 0.4) + 0.1F_{LN}(0.4, 0.4)$, whereas $G = F_{LN}(0.86, 0.6)$. These CDFs cross multiple times, and also Z crosses the identity from below so that $F \ngeq_2 G$ but also $G \ngeq_2 F$. In other words, both $\mathcal{H}_0$ and $\mathcal{H}_0^R$ are false. In the latter case, the null hypothesis is hard to reject, because Z crosses the identity from above, and it exceeds the identity just in a small subset of the unit interval. Note that Barrett and Donald (2003) just apply their test to $\mathcal{H}_0$ versus $\mathcal{H}_1$, overlooking the reverse situation $\mathcal{H}_0^R$ versus $\mathcal{H}_1^R$. As illustrated in Table A5, $\mathcal{T}_1$ exhibits quite a poor performance with the sample sizes considered (to increase its power up to 0.68, we need to reach $n = 5000$), while $\mathcal{T}_\infty$ performs better but with lower power than KSB3. Conversely, KSB3 has a really poor performance in rejecting $\mathcal{H}_0^R$, while the proposed $\mathcal{T}_\infty$ and $\mathcal{T}_1$ tests provide a large power in this critical setting.

7.2.3 | Singh–Maddala Distribution

As a third case, let us consider the Singh–Maddala distribution, denoted as SM(a, q, b), with CDF $F_{\text{SM}}(x; a, q, b) = 1 - [1 + (x/b)^a]^{-q}$. In all the following scenarios, the scale parameter b is set to 1 and hence omitted, while the two shape parameters a and q vary. As in Section 7.2.2, we generate scenarios in which Z crosses the identity. In particular, we target the worst-case scenarios for our proposed tests by setting $F \sim \text{SM}(1.5, q)$ and $G \sim \text{SM}(1, q)$, for $q = 1.2, 1.5, 1.8$. As shown in Figure 4, larger values of q correspond to cases in which it is harder to detect the difference between Z and the identity, especially using $\mathcal{T}_1$. Tables A6a, A7a, and A8a show that KSB3 delivers larger power compared to our tests in such critical cases. In particular, while the performance of $\mathcal{T}_\infty$ significantly improves for larger samples and lower q, the power of $\mathcal{T}_1$ is constantly close to 0, even for $n = 1000$ and $q = 1.2$. In light of part 7 of Proposition 1, this is due to the fact that $\mathcal{T}_1$ downsizes the deviations from the null, which are hardly classified as "large", at least with the sample sizes considered. In such cases, a test statistic $\mathcal{T}_p$ with larger p may be more effective, as will be discussed in the next section. However, when applied to the reverse hypotheses $\mathcal{H}_0^R$ and $\mathcal{H}_1^R$, the proposed tests $\mathcal{T}_\infty$ and $\mathcal{T}_1$ exhibit good performance, with rejection rates significantly increasing with n; see Tables A6b, A7b and A8b. On the contrary, KSB3 struggles to detect nondominance and its power remains close to zero, even for large samples.

7.2.4 | Behavior of T_p

As discussed in Section 2.3, one may consider a general statistic $\mathcal{T}_p$, for $p \geq 1$. Our simulation focus on the choices $\mathcal{T}_1$ and $\mathcal{T}_\infty$ because they are simple to compute, besides being related to some important statistics. For a general p, the computation can be more complicated and may require numerical integration. However, we can approximate $\mathcal{T}_p$ by replacing the identity function I with the step function I_n, with constant jumps of size $1/n$ at each point k/n, for $k = 1,\ldots, n$. For this

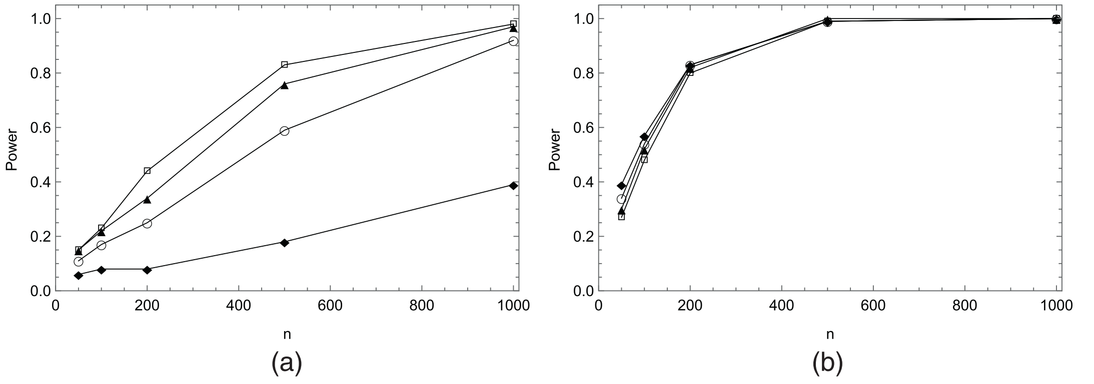

FIGURE 5 Rejection rates of $\mathcal{T}_p$ under independent $F \sim \text{SM}(1.5, 1.2)$, $G \sim \text{SM}(1, 1.2)$, for $p = 2$ (◆), $p = 5$ (○), $p = 10$ (▲), $p = 50$ (□). (a) Test for $\mathcal{H}_0 : F \geq_2 G$ (false); (b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false).

approximation, we can easily compute

$$\mathcal{T}_p(I_n - \widetilde{Z}_n) = \left(\frac{1}{n} \sum_{i=1}^{n} \left(\frac{i}{n} - \widetilde{Z}_n\left(\frac{i}{n}\right) \right)_+^p \right)^{1/p}.$$

The performance of $\mathcal{T}_p$ depends on the distributions considered. In general, we note that a larger value of p may yield a test statistic that is more sensitive when the distance between the LPP and the identity is small. In Figure 5, we illustrate the behavior of $\mathcal{T}_p$ for $p \in \{2, 5, 20, 50\}$, $F \sim \text{SM}(1.5, 1.2)$ and $G \sim \text{SM}(1, 1.2)$, which is the same pair of distributions considered in Table A8. When testing the null hypothesis $\mathcal{H}_0$, we may encounter problems in detecting deviations from the null, as discussed earlier. However, it can be seen that larger values of p correspond to a larger power, often with a large improvement, especially when moving from $p = 2$ to $p = 5$. For the reversed case $\mathcal{H}_0^R$, the results are quite different. In this case, smaller values of p correspond to larger power, however, all the statistics have a somewhat similar (and good) performance, which becomes indistinguishable for $n \geq 500$. This is logical because this case is easier to detect. Comparing these results with those reported in Table A8 we can confirm that $\mathcal{T}_\infty$ and $\mathcal{T}_1$ have the largest power when testing $\mathcal{H}_0$ and $\mathcal{H}_0^R$, respectively. Also note that the performance of $\mathcal{T}_{50}$ is very close to that of $\mathcal{T}_\infty$, under both scenarios.

7.3 | Paired samples

To simulate dependent samples we first draw a sample $\{(Z_i^1, Z_i^2) : i = 1, \ldots, n\}$ from a bivariate normal distribution, with standard marginals and correlation coefficient ρ. Then, by transforming the data via the standard normal CDF Φ, we obtain a dependent sample from a bivariate distribution with uniform marginals $\{U_i^1 = \Phi(Z_i^1) : i = 1, \ldots, n\}$ and $\{U_i^2 = \Phi(Z_i^2) : i = 1, \ldots, n\}$. Finally, a dependent sample from a bivariate distribution with margins F and G is obtained as $\{(F^{-1}(U_i^1), G^{-1}(U_i^2)) : i = 1, \ldots, n\}$. In particular, we consider $\rho = -0.75, -0.5, -0.25, +0.25, +0.5, +0.75$. As in the previous subsections, we compare our results with those of KSB3. Note that, although Barrett and Donald (2003) assume independence to prove the consistency properties of such a test, our simulations reveal that KSB3 exhibits a good performance even in the dependent case.

In this paired setting, we consider the same Weibull distributions as in Sections 7.1 and 7.2.1, focusing on the case $a = 1.1$. The results in Tables A9–A14 show that a stronger positive dependence generally leads to a smaller type-I error probability, under the null hypothesis, and larger rejection rates under the alternative. The situation is reversed for negative dependence, which negatively affects the performance of all the tests considered. These results can be explained using the expression of the variance of $\mathcal{Z}$, derived in Section 4, which suggests that positive (negative) dependence yields a more (less) efficient estimate of Z, and accordingly, a better (worse) performance of the test, at least asymptotically.

7.4 | Test for FSD

As discussed in Section 6, our methodology also allows to test TSD, including an approximation of FSD, obtained as $\geq^{T}_{1+1/\theta}$ with $\theta \to \infty$. We then apply the method described in Section 6 to the same Singh-Maddala distributions studied in Section 7.2.3. Since in these cases SSD does not hold, we have that, *a fortiori*, the FSD null hypothesis, denoted as $\mathcal{H}_0^1 : F \geq_1 G$, is also false. This hypothesis can be tested using a sufficiently large value of θ, as discussed in Section 6. In particular, we set $\theta = 50$, which corresponds to approximating the FSD null hypothesis, $\mathcal{H}_0^1$, with $\mathcal{H}_0^{1.02}$. Our method is compared with the FSD version of the KSB3 test described in Barrett and Donald (2003). In contrast to the KSB3 test for SSD, this latter test may be shown to be consistent even in the case of unbounded support.

All the tests considered tend to provide a larger simulated power compared to the SSD case. This is logical since FSD is more stringent than SSD, and therefore, for the same pairs of distributions, it is easier to detect violations of FSD rather than of SSD. The results in Tables A15–A17 show that KSB3 tends to provide larger power than our $\mathcal{T}_\infty$ and $\mathcal{T}_1$ tests under $\mathcal{H}_1^1 : F \not\geq_1 G$. On the contrary, under the reverse alternative $(\mathcal{H}_1^1)^R : G \not\geq_1 F$, KSB3 exhibits a worse performance than our $\mathcal{T}_\infty$ and $\mathcal{T}_1$, also showing an unexpected behavior, in that its rejection rates first increase and then decrease as n grows.

8 | CONCLUDING REMARKS

In this article, we proposed leveraging the LPP as a new tool to detect deviations from SSD in the case of non-negative random variables. The same approach can be used to test TSD, hence including FSD as a limit case. The asymptotic properties in Section 5 and the simulation results in Section 7 show that our family of tests can be a valid alternative to the established tests based on the difference between integrals of CDFs, such as those in Barrett and Donald (2003). In particular, the KSB3 test is outperformed by our proposed sup-based test $\mathcal{T}_\infty$ in most of the cases analyzed, sometimes with a remarkable gap.

Among the tests proposed, our simulations reveal that the sup-based test $\mathcal{T}_\infty$ is also overall more reliable than the integral-based $\mathcal{T}_1$, which has a lower power in the most critical cases. However, both tests can be useful. In fact, in light of Proposition 1 part 7, and according to our numerical results in Sections 7.2.2 and 7.2.3, $\mathcal{T}_\infty$ performs better than $\mathcal{T}_1$ when deviations from $\mathcal{H}_0$ are subtle, while $\mathcal{T}_1$ provides higher power than $\mathcal{T}_\infty$ when deviations are more apparent. Therefore, in applications, it could be useful to use both tests and compare the p-values. It is also worth noting that the power of our proposed tests improves when the samples are positively correlated, and deteriorates under negative correlation.

In general, the advantage of using the LPP instead of integrals of CDFs is that it can be approximated uniformly, which allows to establish asymptotic properties without requiring a compact support; moreover, the LPP has a different sensitivity in detecting violations of SSD, compared to other methods. With regard to our assumptions, we deal with non-negative random variables because, otherwise, the unscaled Lorenz curves are not monotone (therefore not invertible), and the LPP plot cannot be defined. However, in practice, for finite-sample sizes, the tests may be applied also when negative observations occur. By location-invariance of SSD, this can be done just by shifting the samples, adding $-x^* + \delta$, where $x^* < 0$ is the minimum value of the pooled sample and $\delta > 0$ is an arbitrarily small constant. Finally, the power of our tests may be further improved by combining the same proposed test statistics with different and less conservative bootstrap schemes. The latter represents an interesting direction for future work.

ACKNOWLEDGMENTS

We are grateful to the anonymous referees and the associate editor for their careful reading and useful suggestions.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

FUNDING STATEMENT

This research was supported using the Italian funds ex-MURST60%.

ORCID

Tommaso Lando https://orcid.org/0000-0003-4288-0264

REFERENCES

Andreoli, F. (2018). Robust inference for inverse stochastic dominance. *Journal of Business & Economic Statistics*, *36*(1), 146–159.

Barrett, G. F., & Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, *71*(1), 71–104.

Barrett, G. F., Donald, S. G., & Bhattacharya, D. (2014). Consistent nonparametric tests for Lorenz dominance. *Journal of Business & Economic Statistics*, *32*(1), 1–13.

Beare, B. K., & Clarke, J. D. (2022). Modified Wilcoxon-Mann-Whitney tests of stochastic dominance. arXiv preprint, arXiv:2210.08892.

Beare, B. K., & Moon, J.-M. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory*, *31*(3), 471–492.

Chan, W., Proschan, F., & Sethuraman, J. (1990). Convex-ordering among functions, with applications to reliability and mathematical statistics. *Lecture Notes-Monograph Series*, *16*, 121–134.

Corless, R. M., Gonnet, G. H., Hare, D. E., Jeffrey, D. J., & Knuth, D. E. (1996). On the Lambert W function. *Advances in Computational Mathematics*, *5*, 329–359.

Csörgö, M., Csörgö, S., & Horváth, L. (2013). *An asymptotic theory for empirical reliability and concentration processes* (Vol. *33*). Springer Science & Business Media.

Davidov, O., & Herman, A. (2012). Ordinal dominance curve based inference for stochastically ordered distributions. *Journal of the Royal Statistical Society, Series B (Methodology)*, *74*(5), 825–847.

Davidson, R., & Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, *68*(6), 1435–1464.

Donald, S. G., & Hsu, Y.-C. (2016). Improving the power of tests of stochastic dominance. *Econometric Reviews*, *35*(4), 553–585.

Fishburn, P. C. (1980). Continua of stochastic dominance relations for unbounded probability distributions. *Journal of Mathematical Economics*, *7*(3), 271–285.

Goldie, C. M. (1977). Convergence theorems for empirical Lorenz curves and their inverses. *Advances in Applied Probability*, *9*(4), 765–791.

Hadar, J., & Russell, W. R. (1969). Rules for ordering uncertain prospects. *The American Economic Review*, *59*(1), 25–34.

Hanoch, G., & Levy, H. (1969). The efficiency analysis of choices involving risk. *The Review of Economic Studies*, *36*(3), 335–346.

Hsieh, F., & Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics*, *24*(1), 25–40.

Huang, R. J., Tzeng, L. Y., & Zhao, L. (2020). Fractional degree stochastic dominance. *Management Science*, *66*(10), 4630–4647.

Kaji, T. (2018). *Essays on asymptotic methods in econometrics*. Ph.D. thesis. Massachusetts Institute of Technology.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.

Lando, T., Arab, I., & Oliveira, P. E. (2023). Transform orders and stochastic monotonicity of statistical functionals. *Scandinavian Journal of Statistics*, *50*(3), 1183–1200.

Lando, T., & Bertoli-Barsotti, L. (2020). Distorted stochastic dominance: A generalized family of stochastic orders. *Journal of Mathematical Economics*, *90*, 132–139.

Lehmann, E. L., & Rojo, J. (1992). Invariant directional orderings. *Annals of Statistics*, *20*(4), 2100–2110.

Linton, O., Maasoumi, E., & Whang, Y.-J. (2005). Consistent testing for stochastic dominance under general sampling schemes. *The Review of Economic Studies*, *72*(3), 735–765.

Linton, O., Song, K., & Whang, Y.-J. (2010). An improved bootstrap test of stochastic dominance. *Journal of Econometrics*, *154*(2), 186–202.

Meyer, J. (1977). Second degree stochastic dominance with respect to a function. *International Economic Review*, *18*, 477–487.

Muliere, P., & Scarsini, M. (1989). A note on stochastic dominance and inequality measures. *Journal of Economic Theory*, *49*(2), 314–323.

Müller, A., Scarsini, M., Tsetlin, I., & Winkler, R. L. (2017). Between first- and second-order stochastic dominance. *Management Science*, *63*(9), 2933–2947.

Schmid, F., & Trede, M. (1996). Testing for first-order stochastic dominance: A new distribution-free test. J. R. Stat. Soc. Ser. D: The. *Stat*, *45*(3), 371–380.

Shaked, M., & Shantikumar, J. G. (2007). *Stochastic orders*. Springer.

Shorrocks, A. F. (1983). Ranking income distributions. *Economica*, *50*(197), 3–17.

Sun, Z., & Beare, B. K. (2021). Improved nonparametric bootstrap tests of Lorenz dominance. *Journal of Business & Economic Statistics*, *39*(1), 189–199.

Tang, C.-F., Wang, D., & Tebbs, J. M. (2017). Nonparametric goodness-of-fit tests for uniform stochastic ordering. *Annals of Statistics*, *45*(6), 2565.

Van der Vaart, A., & Wellner, J. (1996). *Weak convergence and empirical processes with applications to statistics*. Springer Science & Business Media.

Wang, S. S., & Young, V. R. (1998). Ordering risks: Expected utility theory versus Yaari's dual theory of risk. *Insurance: Mathematics & Economics*, *22*(2), 145–161.

Weitkamp, C. A., Proksch, K., Tameling, C., & Munk, A. (2024). Distribution of distances based object matching: Asymptotic inference. *Journal of the American Statistical Association*, *119*(545), 538–551.

Whang, Y.-J. (2019). *Econometric analysis of stochastic dominance: Concepts, methods, tools, and applications*. Cambridge University Press.

Whitmore, G., & Findlay, M. (1978). *Stochastic dominance: An approach to decision-making under risk*. Heath.

Zhuang, W., Yao, S., & Qiu, G. (2023). Test of dominance relations based on kernel smoothing method. *Journal of Nonparametric Statistics*, *35*(4), 685–708.

How to cite this article: Lando, T., & Legramanti, S. (2024). A new class of nonparametric tests for second-order stochastic dominance based on the Lorenz P–P plot. *Scandinavian Journal of Statistics*, 1–33. https://doi.org/10.1111/sjos.12761

APPENDIX A. TABLES

TABLE A1 Rejection rates for $\mathcal{H}_0 : F \geq_2 G$ (true) under independent $F, G \sim W(1,1)$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
50	0.09	0.15	0.10
100	0.10	0.15	0.11
200	0.12	0.15	0.11
500	0.09	0.10	0.09
1000	0.08	0.09	0.09

TABLE A2 Rejection rates under independent $F \sim W(1.1, q_{1.1})$ and $G \sim W(1,1)$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (true)			
50	0.08	0.10	0.12
100	0.04	0.06	0.11
200	0.03	0.02	0.10
500	0.02	0.00	0.11
1000	0.00	0.00	0.09
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.19	0.31	0.12
100	0.22	0.29	0.11
200	0.27	0.36	0.11
500	0.50	0.54	0.12
1000	0.72	0.72	0.18

TABLE A3 Rejection rates under independent $F \sim W(1.2, q_{1.2})$ and $G \sim W(1,1)$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (true)			
50	0.04	0.05	0.12
100	0.04	0.02	0.13
200	0.02	0.00	0.12
500	0.01	0.00	0.11
1000	0.00	0.00	0.10
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.29	0.45	0.14
100	0.37	0.50	0.13
200	0.56	0.66	0.15
500	0.87	0.88	0.26
1000	0.99	0.99	0.48

TABLE A4 Rejection rates under independent $F \sim W(1.3, q_{1.3})$ and $G \sim W(1, 1)$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (true)			
50	0.03	0.03	0.12
100	0.02	0.00	0.12
200	0.01	0.00	0.11
500	0.01	0.00	0.11
1000	0.00	0.00	0.08
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.41	0.57	0.16
100	0.56	0.67	0.18
200	0.81	0.85	0.28
500	0.99	0.99	0.51
1000	1.00	1.00	0.88

TABLE A5 Rejection rates under independent $F = 0.9F_{LN}(0.85, 0.4) + 0.1F_{LN}(0.4, 0.9)$, $G = F_{LN}(0.86, 0.6)$. Note: in case (a), the empirical power of $\mathcal{T}_1$ reaches 0.68 for $n = 5000$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (false)			
50	0.25	0.17	0.43
100	0.43	0.15	0.59
200	0.58	0.10	0.74
500	0.89	0.08	0.98
1000	0.99	0.08	1.00
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.16	0.34	0.03
100	0.17	0.29	0.01
200	0.26	0.30	0.01
500	0.51	0.41	0.01
1000	0.84	0.68	0.02

TABLE A6 Rejection rates under independent $F \sim \mathrm{SM}(1.5, 1.8)$, $G \sim \mathrm{SM}(1, 1.8)$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (false)			
50	0.02	0.00	0.38
100	0.04	0.00	0.52
200	0.05	0.00	0.63
500	0.11	0.00	0.88
1000	0.23	0.00	0.97
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.56	0.77	0.06
100	0.76	0.87	0.03
200	0.97	0.98	0.02
500	1.00	1.00	0.01
1000	1.00	1.00	0.03

TABLE A7 Rejection rates under independent $F \sim \mathrm{SM}(1.5, 1.5)$, $G \sim \mathrm{SM}(1, 1.5)$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (false)			
50	0.07	0.01	0.51
100	0.10	0.01	0.66
200	0.16	0.00	0.83
500	0.44	0.00	0.96
1000	0.81	0.00	1.00
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.43	0.68	0.04
100	0.63	0.77	0.01
200	0.92	0.94	0.00
500	1.00	1.00	0.00
1000	1.00	1.00	0.00

TABLE A8 Rejection rates under independent $F \sim \mathrm{SM}(1.5, 1.2)$, $G \sim \mathrm{SM}(1, 1.2)$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (false)			
50	0.16	0.05	0.64
100	0.24	0.03	0.79
200	0.45	0.01	0.92
500	0.85	0.01	0.99
1000	0.99	0.01	1.00
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.27	0.52	0.03
100	0.48	0.61	0.00
200	0.77	0.83	0.01
500	0.99	0.99	0.00
1000	1.00	1.00	0.00

TABLE A9 Rejection rates under dependent $F \sim W(1.1, q_{1.1})$ and $G \sim W(1, 1)$; $\rho = -0.75$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (true)			
50	0.09	0.10	0.11
100	0.04	0.06	0.09
200	0.04	0.05	0.10
500	0.03	0.02	0.08
1000	0.01	0.00	0.07
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.16	0.25	0.11
100	0.17	0.28	0.11
200	0.24	0.30	0.10
500	0.36	0.40	0.11
1000	0.53	0.57	0.14

TABLE A10 Rejection rates under dependent $F \sim W(1.1, q_{1.1})$ and $G \sim W(1,1)$; $\rho = -0.5$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (true)			
50	0.08	0.10	0.11
100	0.04	0.06	0.09
200	0.04	0.04	0.10
500	0.02	0.02	0.08
1000	0.01	0.00	0.07
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.17	0.27	0.11
100	0.19	0.29	0.11
200	0.24	0.31	0.11
500	0.39	0.42	0.12
1000	0.60	0.61	0.15

TABLE A11 Rejection rates under dependent $F \sim W(1.1, q_{1.1})$ and $G \sim W(1,1)$; $\rho = -0.25$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (true)			
50	0.07	0.10	0.11
100	0.05	0.06	0.09
200	0.04	0.04	0.10
500	0.02	0.01	0.08
1000	0.01	0.00	0.07
(b) Test for $\mathcal{H}_0^R : G \geq_e q_2 F$ (false)			
50	0.18	0.27	0.11
100	0.19	0.30	0.12
200	0.28	0.34	0.11
500	0.42	0.45	0.13
1000	0.64	0.68	0.17

TABLE A12 Rejection rates under dependent $F \sim W(1.1, q_{1.1})$ and $G \sim W(1,1)$; $\rho = +0.25$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (true)			
50	0.06	0.09	0.11
100	0.04	0.06	0.11
200	0.04	0.02	0.09
500	0.01	0.01	0.10
1000	0.01	0.00	0.09
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.23	0.36	0.13
100	0.25	0.37	0.12
200	0.35	0.42	0.13
500	0.57	0.61	0.17
1000	0.79	0.79	0.22

TABLE A13 Rejection rates under dependent $F \sim W(1.1, q_{1.1})$ and $G \sim W(1,1)$; $\rho = +0.5$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (true)			
50	0.06	0.10	0.10
100	0.04	0.05	0.11
200	0.03	0.02	0.10
500	0.01	0.00	0.10
1000	0.00	0.00	0.08
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.24	0.43	0.14
100	0.28	0.44	0.13
200	0.42	0.52	0.14
500	0.69	0.74	0.20
1000	0.89	0.91	0.29

TABLE A14 Rejection rates under dependent $F \sim W(1.1, q_{1.1})$ and $G \sim W(1, 1)$; $\rho = +0.75$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $\mathcal{H}_0 : F \geq_2 G$ (true)			
50	0.06	0.11	0.10
100	0.02	0.04	0.11
200	0.02	0.01	0.10
500	0.01	0.00	0.09
1000	0.00	0.00	0.07
(b) Test for $\mathcal{H}_0^R : G \geq_2 F$ (false)			
50	0.30	0.57	0.15
100	0.42	0.59	0.15
200	0.61	0.72	0.18
500	0.88	0.91	0.33
1000	0.98	0.99	0.50

TABLE A15 FSD test. Rejection rates for independent $F \sim \text{SM}(1.5, 1.2)$, $G \sim \text{SM}(1, 1.2)$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $H_0^1 : F \geq_1 G$ (false)			
50	0.11	0.14	0.24
100	0.22	0.16	0.41
200	0.44	0.21	0.64
500	0.86	0.42	0.92
1000	1.00	0.82	0.95
(b) Test for $(H_0^1)^R : G \geq_1 F$ (false)			
50	0.03	0.29	0.19
100	0.10	0.38	0.19
200	0.31	0.53	0.14
500	0.92	0.92	0.05
1000	1.00	1.00	0.00

TABLE A16 FSD test. Rejection rates for independent $F \sim \text{SM}(1.5, 1.5)$, $G \sim \text{SM}(1, 1.5)$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $H_0^1 : F \geq_1 G$ (false)			
50	0.08	0.09	0.15
100	0.11	0.06	0.24
200	0.20	0.05	0.39
500	0.64	0.10	0.85
1000	0.96	0.29	0.97
(b) Test for $(H_0^1)^R : G \geq_1 F$ (false)			
50	0.08	0.44	0.32
100	0.19	0.56	0.44
200	0.52	0.80	0.52
500	0.99	0.99	0.39
1000	1.00	1.00	0.22

TABLE A17 FSD test. Rejection rates for independent $F \sim \text{SM}(1.5, 1.8)$, $G \sim \text{SM}(1, 1.8)$.

n	$\mathcal{T}_\infty$	$\mathcal{T}_1$	KSB3
(a) Test for $H_0^1 : F \geq_1 G$ (false)			
50	0.03	0.02	0.07
100	0.04	0.02	0.13
200	0.09	0.01	0.24
500	0.36	0.01	0.60
1000	0.77	0.02	0.92
(b) Test for $(H_0^1)^R : G \geq_1 F$ (false)			
50	0.10	0.59	0.47
100	0.30	0.70	0.63
200	0.74	0.91	0.78
500	0.99	1.00	0.78
1000	1.00	1.00	0.67

APPENDIX B. PROOFS

Calculations of Example 1. The unscaled Lorenz curve of F is

$$L_F(p) = b\left(\Gamma\left(1 + \frac{1}{a}\right) - \Gamma\left(1 + \frac{1}{a}, -\log(1-p)\right)\right),$$

while

$$L_G(p) = p + (1-p)\log(1-p).$$

It is well known (e.g. Goldie, 1977) that L_G can be expressed as $L_G(p) = M_G \circ G^{-1}(p)$, with

$$M_G(x) = \int_0^x t\, dG(t) = 1 - e^{-x}(x+1), \qquad x \geq 0.$$

Noting that $M_G(x) \le \mu_G = 1$ for any $x \ge 0$, this function can be inverted using the Lambert W_{-1} function (Corless et al., 1996), that is $M_G^{-1}(t) = -1 - W_{-1}((t-1)/e)$. Accordingly,

$$L_G^{-1}(t) = G \circ M_G^{-1}(t) = 1 - \exp\left(1 + W_{-1}\left(\frac{t-1}{e}\right)\right).$$

Finally, by composition, we obtain the expression of Z in Example 1. ■

Proof of Proposition 1.

1. This follows from the properties of the L^p norm.
2. If $v_2(x) \le 0, \forall x \in [0,1]$ then $v_1(x) - v_2(x) \ge v_1(x), \forall x \in [0,1]$ which implies $(v_1(x) - v_2(x))_+^p \ge (v_1(x))_+^p, \forall x \in [0,1]$ and therefore $\|(v_1 - v_2)_+\|_p \ge \|(v_1)_+\|_p$ by monotonicity of integrals.
3. The proof is the same as in Lemma 2 of Barrett et al. (2014) and relies on the fact that $v_1 \in C[0,1]$.
4. Minkowski's inequality implies that, for some pair of functions $u, v \in C[0,1]$, $\|u\|_p = \|(u-v)+v\|_p \le \|u-v\|_p + \|v\|_p$, so that $\|u\|_p - \|v\|_p \le \|u-v\|_p$, and similarly, $\|u-v\|_p \ge \|v\|_p - \|u\|_p$; therefore, $|\|u\|_p - \|v\|_p| \le \|u-v\|_p$. Then

$$|\|(v_1)_+\|_p - \|(v_2)_+\|_p| \le \|(v_1)_+ - (v_2)_+\|_p \le \|v_1 - v_2\|_p \le \|v_1 - v_2\|_\infty,$$

   where the second inequality follows from the fact that, for every $x \in [0,1]$, $|(v_1(x))_+ - (v_2(x))_+| \le |v_1(x) - v_2(x)|$.
5. The proof follows from absolute homogeneity of the L^p norm.
6. Let $\beta \in [0,1]$. By convexity of the function $(\cdot)_+$, Minkowski's inequality, and absolute homogeneity of the L^p norm,

$$\begin{aligned}\mathcal{T}_p(\beta(v_2) + (1-\beta)v_1) &= \|(\beta v_2 + (1-\beta)v_1)_+\|_p \le \|\beta(v_2)_+ + (1-\beta)(v_1)_+\|_p \\ &\le \beta\|(v_2)_+\|_p + (1-\beta)\|(v_1)_+\|_p = \beta\mathcal{T}_p(v_2) + (1-\beta)\mathcal{T}_p(v_1).\end{aligned}$$

7. This follows from basic properties of L^p norms. ■

Proof of Proposition 2. $\widetilde{Z}_{n,m} - Z_{n,m}$ can be expressed as

$$\widetilde{L}_{G_m}^{-1} \circ \widetilde{L}_{F_n} - L_{G_m}^{-1} \circ L_{F_n} = (\widetilde{L}_{G_m}^{-1} \circ \widetilde{L}_{F_n} - \widetilde{L}_{G_m}^{-1} \circ L_{F_n}) + (\widetilde{L}_{G_m}^{-1} \circ L_{F_n} - L_{G_m}^{-1} \circ L_{F_n}).$$

For the first summand, which is the difference between two step functions, we have $\widetilde{L}_{G_m}^{-1} \circ \widetilde{L}_{F_n}(p) \ge \widetilde{L}_{G_m}^{-1} \circ L_{F_n}(p)$ for every $p \in [0,1]$, since $\widetilde{L}_{F_n}(p) \ge L_{F_n}(p)$ for every $p \in [0,1]$. Moreover, $\widetilde{L}_{G_m}^{-1} \circ \widetilde{L}_{F_n}(k/n) = \widetilde{L}_{G_m}^{-1} \circ L_{F_n}(k/n)$ for $k = 0, \ldots, n$, while, within each interval $((k-1)/n, k/n)$, the difference $\widetilde{L}_{G_m}^{-1} \circ \widetilde{L}_{F_n}(p) - \widetilde{L}_{G_m}^{-1} \circ L_{F_n}(p)$ is bounded from above by the height of the jumps of $\widetilde{L}_{G_m}^{-1}$, that is, $1/m$. For the latter summand, $\widetilde{L}_{G_m}^{-1} \circ L_{F_n} - L_{G_m}^{-1} \circ L_{F_n} \in [-1/m, 0]$, since clearly $L_{G_m}^{-1} \circ L_{F_n}$ is the linear interpolator of the jump points of the step function $\widetilde{L}_{G_m}^{-1} \circ L_{F_n}$. Hence, the result follows. ■

Proof of Proposition 3. As proved in Theorem 10.1 and Theorem 13.2 of Csörgö et al. (2013), $\widetilde{L}_{G_m}^{-1}$ and $\widetilde{L}_{F_n}$ converge strongly and uniformly to L_G^{-1} and L_F, respectively. Since L_G^{-1} is uniformly continuous in $[0, \infty)$ and $\sup_{p\in[0,1]} |L_{F_n}(p) - L_F(p)| \to 0$ almost surely, we obtain that $\sup_{p\in[0,1]} |L_G^{-1}\circ\widetilde{L}_{F_n}(p) - L_G^{-1}\circ L_F(p)| \to 0$ almost surely. Then, for every $p \in (0, 1)$,

$$|\widetilde{L}_{G_m}^{-1}\circ\widetilde{L}_{F_n}(p) - L_G^{-1}\circ L_F(p)| \le |\widetilde{L}_{G_m}^{-1}\circ\widetilde{L}_{F_n}(p) - L_G^{-1}\circ\widetilde{L}_{F_n}(p)| + |L_G^{-1}\circ\widetilde{L}_{F_n}(p) - L_G^{-1}\circ L_F(p)|$$
$$\le \sup_{p\in[0,1]} |\widetilde{L}_{G_m}^{-1}(p) - L_G^{-1}(p)| + \sup_{p\in[0,1]} |L_G^{-1}\circ\widetilde{L}_{F_n}(p) - L_G^{-1}\circ L_F(p)|.$$

Since both terms in the right-hand side converge to 0 with probability 1, we obtain that $\widetilde{Z}_{n,m}$ converges strongly and uniformly to Z in $[0, 1]$. By Proposition 2, $\|\widetilde{Z}_{n,m} - Z_{n,m}\|_\infty \to 0$ for $n \to \infty$ and $m \to \infty$, therefore the same property is satisfied by $Z_{n,m}$. ∎

Proof of Theorem 1. Let $\mathbb{L}$ be the space of maps $z : [0, \infty) \to \mathbb{R}$ with $\lim_{x\to-\infty} z(x) = 0$ and $\lim_{x\to\infty} z(x) = 1$, and the norm $\|z\|_{\mathbb{L}} = \max\{\|z\|_\infty, \|1 - z\|_1\}$. As shown by Kaji (2018), under Assumption 1, the map $\phi(F) = F^{-1}$ from CDFs to quantile functions is Hadamard differentiable at F, tangentially to the set $\mathbb{L}_0$ of continuous functions in $\mathbb{L}$, with derivative map

$$\phi'_F(z) = -(z\circ F^{-1})(F^{-1})'.$$

The linear map $\psi(F^{-1}) = \int_0^\cdot F^{-1}(t)dt$ coincides with its Hadamard derivative. Accordingly, using the chain rule (Van der Vaart & Wellner, 1996, Lemma 3.9.3), the composition map $\psi\circ\phi : F \to L_F$ is also Hadamard differentiable at F tangentially to $\mathbb{L}_0$, with derivative

$$(\psi\circ\phi)'_F(z) = \psi'_{\phi(F)}\circ\phi'_F(z) = -\int_0^\cdot z\circ F^{-1}(p)dF^{-1}(p).$$

Now, observe that

$$\begin{pmatrix} \sqrt{n}(F_n - F) \\ \sqrt{m}(G_m - G) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathcal{B}_1\circ F \\ \mathcal{B}_2\circ G \end{pmatrix} \text{ in } \mathbb{L}\times\mathbb{L},$$

as shown in Lemma 5.1 of Sun and Beare (2021). Then, the functional delta method (Van der Vaart & Wellner, 1996, Theorem 3.9.13) implies the joint weak convergence

$$\begin{pmatrix} \sqrt{n}(L_{F_n} - L_F) \\ \sqrt{m}(L_{G_m} - L_G) \end{pmatrix} = \begin{pmatrix} \sqrt{n}(\psi\circ\phi(F_n) - \psi\circ\phi(F)) \\ \sqrt{m}(\psi\circ\phi(G_m) - \psi\circ\phi(G)) \end{pmatrix} \rightsquigarrow \begin{pmatrix} (\psi\circ\phi)'_F(\mathcal{B}_1\circ F) \\ (\psi\circ\phi)'_G(\mathcal{B}_2\circ G) \end{pmatrix}$$
$$= \begin{pmatrix} -\int_0^\cdot \mathcal{B}_1(t)dF^{-1}(t) \\ -\int_0^\cdot \mathcal{B}_2(t)dG^{-1}(t) \end{pmatrix} =: \begin{pmatrix} \mathcal{L}_F \\ \mathcal{L}_G \end{pmatrix} \text{ in } C[0,1]\times C[0,1]. \quad \text{(B1)}$$

Now, consider the process $\sqrt{m}(L_{G_m}^{-1}(t) - L_G^{-1}(t))$, for $t \in [0, \mu_G]$. L_G is increasing and continuous on $[0, 1]$, therefore the inverse function L_G^{-1} is increasing and continuous on $[0, \mu_G]$, moreover the derivative $L'_G = G^{-1}$ is strictly positive in the unit interval, since Assumption 2 entails that $G^{-1}(0^+) = c > 0$. Then, using the inverse map theorem (Van der Vaart & Wellner, 1996, Lemma 3.9.20) the map $\eta : L_G \to L_G^{-1}$ is Hadamard differentiable at L_G, tangentially to the set of bounded functions on [0, 1], with derivative

$$\eta'_{L_G}(z) = -\frac{z \circ L_G^{-1}}{L'_G \circ L_G^{-1}} = -\frac{z \circ L_G^{-1}}{G^{-1} \circ L_G^{-1}}.$$

Since $r_n/n \to 1-\lambda$ and $r_n/m \to \lambda$, by (B1) and the functional delta method, the above result implies

$$\sqrt{r_n}\begin{pmatrix} L_{F_n} - L_F \\ L_{G_m}^{-1} - L_G^{-1} \end{pmatrix} \rightsquigarrow \begin{pmatrix} \lambda \mathcal{L}_F \\ (1-\lambda)\eta'_{L_G}(\mathcal{L}_G) \end{pmatrix} =: \begin{pmatrix} \lambda \mathcal{L}_F \\ (1-\lambda) \mathcal{C}_G \end{pmatrix} \text{ in } C[0,1] \times C[0, \mu_G], \tag{B2}$$

where $\mathcal{C}_G$ is defined as

$$\mathcal{C}_G(t) = \frac{\int_0^{L_G^{-1}(t)} \mathcal{B}_2(p) dG^{-1}(p)}{G^{-1} \circ L_G^{-1}(t)}, \qquad t \in [0, \mu_G].$$

Now, consider the maps $\pi = \psi \circ \phi : F \to L_F$, $\theta : G \to L_G^{-1}$, and the composition map $\zeta : C[0,1] \times C[0, \mu_G] \to C[0, \nu]$ defined by $\zeta(\pi, \theta)(x) = \theta \circ \pi(x)$. The Hadamard derivative of θ at L_F is $\theta'_{L_F}(\alpha) = ((L_G^{-1})' \circ L_F)\alpha$ by Lemma 3.9.25 of Van der Vaart and Wellner (1996), since, by Assumption 2, $(L_G^{-1})' = 1/G^{-1} \circ L_G^{-1} \leq 1/c$ is bounded and continuous in $[0, \mu_G]$. Therefore we can apply Lemma 3.2.27 of Van der Vaart and Wellner (1996), which establishes that ζ is Hadamard differentiable at (π, θ), tangentially to the set $C[0,1] \times UC[0, \mu_G]$, where $UC[0, \mu_G]$ is the family of uniformly continuous functions on $[0, \mu_G]$, with derivative

$$\zeta'_{\pi,\theta}(\alpha, \beta)(x) = \beta \circ \pi(x) + \theta'_{\pi(x)}(\alpha(x)) = \beta \circ L_F(x) + \theta'_{L_F}(\alpha(x)) = \beta \circ L_F(x) + ((L_G^{-1})' \circ L_F)\alpha(x).$$

Now, since $Z = \zeta(L_F, L_G^{-1})$, using (B2), the functional delta method and the Hadamard differentiability of the composition map ζ give

$$\begin{aligned} \sqrt{r_n}(Z_{n,m} - Z)\mathbb{1}[0, \nu] &= \sqrt{r_n}(\zeta(L_{F_n}, L_{G_m}^{-1}) - \zeta(L_F, L_G^{-1}))\mathbb{1}[0, \nu] \\ &\rightsquigarrow \zeta'_{L_F, L_G^{-1}}(\lambda \mathcal{L}_F, (1-\lambda)\mathcal{C}_G) = \sqrt{\lambda}\mathcal{C}_G \circ L_F + \sqrt{1-\lambda}\frac{\mathcal{L}_F}{G^{-1} \circ L_G^{-1} \circ L_F} \\ &= \frac{-\sqrt{1-\lambda}\int_0^{\cdot} \mathcal{B}_1 dF^{-1}(p) + \sqrt{\lambda}\int_0^Z \mathcal{B}_2(p) dG^{-1}(p)}{G^{-1} \circ Z} \quad \text{in } C[0, \nu], \end{aligned}$$

which implies the statement, since $\mathcal{Z}_n \mathbb{1}(\nu, 1] \rightsquigarrow 0$. ∎

Proof of Lemma 1. Bear in mind that $\mathcal{Z}\mathbb{1}[0, \nu]$ is a mean zero Gaussian process since it is obtained by integrating and normalizing Gaussian processes. Under $\mathcal{H}_0$, $p - Z(p) \leq 0, \forall p \in [0, 1]$. Hence

$$\sqrt{r_n}\, \mathcal{T}(I - Z_{n,m}) \leq \sqrt{r_n}\, \mathcal{T}(Z - Z_{n,m}) = \mathcal{T}(\sqrt{r_n}(Z - Z_{n,m})) \rightsquigarrow \mathcal{T}(\mathcal{Z}),$$

where the last step follows from the continuous mapping theorem, since the map $\mathcal{T}$ satisfies $|\mathcal{T}(u) - \mathcal{T}(v)| \leq \|u - v\|_\infty$, where u, v are continuous functions on the unit interval, as proved in Proposition 1. The $(1-\alpha)$ quantile of the distribution of $\mathcal{T}(\mathcal{Z})$ is positive, finite, and unique because $\mathcal{Z}$ is a mean zero Gaussian process, so the proof follows the same arguments used in the proof of Lemma 4 in Barrett et al. (2014). Because, by Proposition 3, $Z_{n,m}$ converges strongly and uniformly to Z, under $\mathcal{H}_1$ we have $\mathcal{T}(I - Z_{n,m}) \to_p \mathcal{T}(I - Z) > 0$. Finally, multiplying by $\sqrt{r_n}$, we obtain the second part of the statement. ∎

Proof of Proposition 4. As proved in Lemma 5.2 of Sun and Beare (2021),

$$\begin{pmatrix} \sqrt{n}(F_n^* - F_n) \\ \sqrt{m}(G_n^* - G_n) \end{pmatrix} \overset{as*}{\underset{M}{\rightsquigarrow}} \begin{pmatrix} \mathcal{B}_1 \circ F \\ \mathcal{B}_2 \circ G \end{pmatrix} \text{ in } \mathbb{L} \times \mathbb{L},$$

where $\overset{as*}{\underset{M}{\rightsquigarrow}}$ denotes weak convergence conditional on the data a.s.; see Kosorok (2008, p. 20). The proof of Theorem 1 establishes the Hadamard differentiability of the maps $\psi \circ \phi : F \to L_F$ and $\eta \circ \psi \circ \phi : G \to L_G^{-1}$, so that the functional delta method for the bootstrap implies

$$\sqrt{r_n}\begin{pmatrix} L_{F_n^*} - L_{F_n} \\ L_{G_n^*}^{-1} - L_{G_n}^{-1} \end{pmatrix} \overset{P}{\underset{M}{\rightsquigarrow}} \begin{pmatrix} \lambda \mathcal{L}_F \\ (1-\lambda)\mathcal{C}_G \end{pmatrix} \text{ in } C[0,1] \times C[0, \mu_G],$$

where $\overset{P}{\underset{M}{\rightsquigarrow}}$ denotes weak convergence conditional on the data in probability; see Kosorok (2008, p. 20). By the Hadamard differentiability of the composition map $\zeta(L_F, L_G^{-1}) = L_G^{-1} \circ L_F$, the functional delta method for bootstrap implies $\sqrt{r_n}(Z_{n,m}^* - Z_{n,m}) \overset{P}{\underset{M}{\rightsquigarrow}} \mathcal{Z}$, which entails that $\mathcal{T}(\sqrt{r_n}(Z_{n,m} - Z_{n,m}^*)) \overset{P}{\underset{M}{\rightsquigarrow}} \mathcal{T}(-\mathcal{Z}) =_d \mathcal{T}(\mathcal{Z})$ by the continuous mapping theorem. The test rejects the null hypothesis if the test statistic exceeds the bootstrap threshold $c_n^*(\alpha) = \inf\{y : P(\sqrt{r_n}\, \mathcal{T}(Z_{n,m} - Z_{k;n,m}^*) > y | \mathcal{X}, \mathcal{Y}) \leq \alpha\}$, but the weak convergence result implies $c_n^*(\alpha) \to_p c(\alpha) = \inf\{y : P(\mathcal{T}(\mathcal{Z}) > y) \leq \alpha\}$. Hence, Lemma 1 yields the result. ∎

Proof of Theorem 2. Integrating by substitution, we can see that $X \geq_u^T Y$ if and only if $u(X) \geq_2 u(Y)$, since $P(u(X) \leq t) = F_X \circ u^{-1}(t)$, and similarly for Y. Hence, by setting $\phi = g \circ u$, the proof follows from the classic characterization of SSD, since $\mathbb{E}(g \circ u(X)) \geq \mathbb{E}(g \circ u(Y))$, for any increasing concave function g. ∎

Proof of Theorem 3. Point 1 follows from the fact that $u_1(X) \geq_2 u_1(Y)$ implies $\mathbb{E}(u_2 \circ u_1^{-1} \circ u_1(X)) = \mathbb{E}(u_2(X)) \geq \mathbb{E}(u_2(Y))$, because the composition $u_2 \circ u_1^{-1}$ is concave by construction. The "only if" part of point 2 is trivial. The "if" part follows from the characterization of FSD, taking into account that the equivalent condition of Theorem 2, that is, $\mathbb{E}(\phi(X)) \geq \mathbb{E}(\phi(Y)), \forall \phi \leq_c u$, for every $u \in \mathcal{U}$, implies that such an inequality holds just for every increasing $\phi \in \mathcal{U}$. Since any increasing function may be approximated by a sequence of functions in $\mathcal{U}$, we have $X \geq_1 Y$. ∎

Proof of Theorem 4.

1. Since $X^\theta = u_\theta(X)$ has CDF $F \circ u_\theta^{-1}$, and similarly Y^θ has CDF $G \circ u_\theta^{-1}$, a change of variable shows that $X \geq^T_{1+1/\theta} Y$ can be expressed as

$$\int_{-\infty}^{x} F(t)du_\theta(t) \leq \int_{-\infty}^{x} G(t)du_\theta(t), \qquad \forall x. \tag{B3}$$

Integrating by parts and by substitution, we obtain that, for both $H = F$ and $H = G$,

$$\mathcal{I}_H^\theta(x) = \int_0^x H(t)du_\theta(t) = u_\theta(x)H(x) - \int_0^x u_\theta(t)dH(t) = u_\theta(x)H(x) - \int_0^{H(x)} u_\theta \circ H^{-1}(y)dy.$$

Hence, as $\theta \to \infty$,

$$\frac{\mathcal{I}_H^\theta(x)}{u_\theta(x)} = H(x) - \int_0^{H(x)} \frac{u_\theta \circ H^{-1}(y)}{u_\theta(x)}dy = H(x) - \int_0^{H(x)} \left(\frac{H^{-1}(y)}{x}\right)^\theta dy \to H(x),$$

by the Lebesgue dominated convergence theorem, recalling that $H^{-1}(y)/x \leq 1$ as $y \leq H(x)$. Now, it is readily seen that $X \geq^T_{1+1/\theta} Y$ if and only if $\mathcal{I}_F^\theta(x)/u_\theta(x) \leq \mathcal{I}_G^\theta(x)/u_\theta(x)$ for any x, which implies the result.

2. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ be ordered realizations from X and Y, respectively. For $i = 1, \ldots, n$, and $j = 1, \ldots, m-1$, $\widetilde{Z}_{n,m}^\theta(i/n)$ returns j/m if

$$(1/n)\sum_{k=1}^{i} x_k^\theta \in \left[(1/m)\sum_{k=1}^{j} y_k^\theta, (1/m)\sum_{k=1}^{j+1} y_k^\theta\right),$$

which is equivalent to

$$x_i\left[\frac{1}{n}\left(\sum_{k=1}^{i-1}\left(\frac{x_k}{x_i}\right)^\theta + 1\right)\right]^{1/\theta} \in \left[y_j\left[\frac{1}{m}\left(\sum_{k=1}^{j-1}\left(\frac{y_k}{y_j}\right)^\theta + 1\right)\right]^{1/\theta}, y_{j+1}\left[\frac{1}{m}\left(\sum_{k=1}^{j}\left(\frac{y_k}{y_{j+1}}\right)^\theta + 1\right)\right]^{1/\theta}\right].$$

The terms in square brackets can be arbitrarily close to 1 by the choice of θ. Then, there exists some θ_0 such that, for $\theta \geq \theta_0$, $(1/n)\sum_{k=1}^{i} x_k^\theta \in [(1/m)\sum_{k=1}^{j} y_k^\theta, (1/m)\sum_{k=1}^{j+1} y_k^\theta)$ if and only if $x_i \in [y_j, y_{j+1})$. That is, for any $i = 1, \ldots, n$ and $\theta \geq \theta_0$, $\widetilde{Z}_{n,m}^\theta(i/n)$ returns j/m if $x_i \in [y_j, y_{j+1})$, which coincides with the P–P plot $G_m \circ F_n^{-1}$. ■