Improved nonparametric bootstrap tests of Lorenz dominance

Zhenting Sun\textsuperscript{1} and Brendan K. Beare\textsuperscript{2}

\textsuperscript{1}National School of Development, Peking University
\textsuperscript{2}Department of Economics, University of California, San Diego

May 19, 2019

Abstract

One income or wealth distribution is said to Lorenz dominate another when the Lorenz curve for the former is nowhere below that of the latter, indicating a (weakly) more equitable allocation of resources. Existing tests of the null of Lorenz dominance based on pairs of samples of income or wealth achieve the nominal rejection rate asymptotically when the two Lorenz curves are equal, but are conservative at other null configurations. We propose new nonparametric bootstrap tests of Lorenz dominance based on preliminary estimation of a contact set. Our tests achieve the nominal rejection rate asymptotically on the boundary of the null; that is, when Lorenz dominance is satisfied, and the Lorenz curves coincide on some interval. Numerical simulations indicate that our tests enjoy substantially improved power compared to existing procedures at relevant sample sizes.

\textsuperscript{1}We thank Roy Allen, Qihui Chen, Zheng Fang, James Hamilton, Tetsuya Kaji, Ivana Komunjer, Andres Santos, Xiaoxia Shi, Yixiao Sun and Yinchu Zhu for helpful comments, and especially Alexis Toda for advice on the design of the numerical simulations. This paper is a revised version of the second chapter of the first author’s doctoral thesis at UC San Diego.
1 Introduction

Lorenz curves are widely used for the analysis of economic inequality. A Lorenz curve is a function of the distribution of wealth (or income) across a population, which graphs the cumulative proportion of total wealth by cumulative proportion of the population ordered from poorest to richest. In practice, people are interested in comparing the Lorenz curves between different populations. If one Lorenz curve is below another one, the wealth in the former population is more unequally distributed toward the rich. We use the concept of Lorenz dominance to formalize the comparison of two Lorenz curves: distribution A Lorenz dominates distribution B if the Lorenz curve for A is nowhere below that for B.

Because of the economic importance of Lorenz dominance, methods of statistically testing for Lorenz dominance are of interest. Bishop et al. (1991a,b) employed pair-wise multiple comparisons of sample Lorenz ordinates to test for differences between Lorenz curves and then determine Lorenz dominance. Dardanoni and Forcina (1999) and Davidson and Duclos (2000) proposed tests of Lorenz dominance at a chosen set of points. Barrett et al. (2014) pointed out that these tests are potentially inconsistent because they limit attention to a small fixed set of grid points. They proposed a new class of consistent non-parametric tests for testing the null hypothesis of Lorenz dominance, which are analogous to tests of stochastic dominance proposed by McFadden (1989) and elaborated and extended by Barrett and Donald (2003). The tests are constructed based on a general functional $\mathcal{F}$ applied to $\hat{\phi}$, a function on $[0,1]$ equal to the difference between two empirical Lorenz curves. Two specific functionals used to construct test statistics are $S$, which computes the supremum of $\hat{\phi}$, and $I$, which integrates $\hat{\phi}$ over the region where $\hat{\phi}$ is positive. The $I$-based test statistic was first proposed by Bhattacharya (2007).

A pair of distributions satisfying the null of Lorenz dominance is said to be on the boundary of the null whenever the corresponding Lorenz curves coincide over some interval. To obtain critical values, Barrett et al. (2014) employ a bootstrap procedure that leads to a test with limiting rejection rate equal to the nominal level when the two Lorenz curves are equal, and below the nominal level elsewhere in the null. If we are at a point on the boundary of the null where the Lorenz curves are not equal, then their test has limiting rejection rate below the nominal level, and thus lacks power against nearby points in the alternative. Our main contribution is an alternative construction of bootstrap critical values for the test statistics of Barrett et al. (2014) that achieves a limiting rejection rate equal to the nominal level over the boundary of the null,
thereby improving power. Numerical simulations indicate that the improvement to power can be large. The main driving force behind the improved power is a contact set estimator which excludes regions of the unit interval from critical value calculations when the data indicate that the Lorenz curves are unlikely to touch in those regions.

The primary technical obstacle to obtaining a valid bootstrap approximation over the boundary of the null is that $\mathcal{F}$ typically fails to be Hadamard differentiable in this region, which is known to imply inconsistency of the bootstrap (Dümbgen, 1993). By applying recent results of Fang and Santos (2019) on bootstrap inference under nondifferentiability, we show that a modified bootstrap procedure based on preliminary estimation of a contact set can deliver consistent approximation over the boundary of the null. Our power-improving modification to the tests of Lorenz dominance proposed by Barrett et al. (2014) can be viewed as analogous to the modifications made by Linton et al. (2010) and Donald and Hsu (2016) to the tests of stochastic dominance proposed by Barrett and Donald (2003), or to the modification made by Beare and Shi (2019) to the tests of density ratio ordering proposed by Carolan and Tebbs (2005) and Beare and Moon (2015), or to the modifications made by Seo (2018) to the tests of stochastic monotonicity and conditional stochastic dominance proposed by Delgado and Escanciano (2012, 2013).

Our asymptotic results draw on important work by Kaji (2018), who has established weak convergence of the empirical quantile process and bootstrap empirical quantile process in the $L^1$-semimetric under mild technical conditions. Such convergence implies weak convergence of the empirical Lorenz process and bootstrap empirical Lorenz process in the uniform metric, greatly facilitating our analysis.

Throughout this paper, we let $\ell^\infty[0, \infty)$ denote the collection of bounded real valued functions on $[0, \infty)$ equipped with the uniform metric, we let $L^1(0, 1)$ be the space of integrable real valued functions on $(0, 1)$ equipped with the $L^1$-semimetric, and we let $C[0, 1]$ be the space of continuous real valued functions on $[0, 1]$ equipped with the uniform metric. We let $\Rightarrow$ denote Hoffman-Jørgensen weak convergence in a semimetric space or product of semimetric spaces.
2 Hypothesis tests of Lorenz dominance

2.1 Hypothesis formulation

Suppose that $F_1 : [0, \infty) \to \mathbb{R}$ and $F_2 : [0, \infty) \to \mathbb{R}$ are the cumulative distribution functions (CDFs) of the two income (or wealth, etc.) distributions we are concerned with. We impose the following regularity conditions on $F_1$ and $F_2$.

**Assumption 2.1.** For $j = 1, 2$, the CDF $F_j$ satisfies $F_j(0) = 0$ and is continuously differentiable on the interior of its support, with strictly positive derivative. Moreover, each $F_j$ has finite $(2 + \epsilon)$-th absolute moment for some $\epsilon > 0$.

Let $Q_1$ and $Q_2$ denote the quantile functions corresponding to $F_1$ and $F_2$ respectively; that is,

$$Q_j(p) = \inf \{ x \in [0, \infty) : F_j(x) \geq p \}, \quad p \in [0, 1].$$

(2.1)

When $F_j$ has finite first moment $\mu_j$, as it does under Assumption 2.1, the quantile function $Q_j$ must be integrable, and its integral satisfies

$$\int_0^1 Q_j(p) \, dp = \mu_j.$$

In this case we may define the Lorenz curve corresponding to $F_j$ by

$$L_j(p) = \frac{1}{\mu_j} \int_0^p Q_j(t) \, dt, \quad p \in [0, 1].$$

(2.2)

Note that every Lorenz curve is a convex CDF on $[0, 1]$.

**Definition 2.1.** Given two CDFs $F_1 : [0, \infty) \to \mathbb{R}$ and $F_2 : [0, \infty) \to \mathbb{R}$ with finite first moments, we say that $F_1$ weakly Lorenz dominates $F_2$ if the Lorenz curve $L_1$ for $F_1$ is nowhere below the Lorenz curve $L_2$ for $F_2$; that is, $L_1(p) - L_2(p) \geq 0$ for all $p \in [0, 1]$. We generally omit the modifier weakly for brevity.

The hypotheses we seek to discriminate between are

$$H_0 : L_2(p) \leq L_1(p) \text{ for all } p \in [0, 1],$$

$$H_1 : L_2(p) > L_1(p) \text{ for some } p \in [0, 1].$$

The null hypothesis $H_0$ is satisfied when $F_1$ weakly Lorenz dominates $F_2$, while the alternative hypothesis $H_1$ is satisfied when such dominance does not occur.

2.2 Sampling frameworks

Following Barrett et al. (2014), we consider two alternative frameworks for sampling from $F_1$ and $F_2$. In both frameworks, for $j = 1, 2$ we observe an
independent and identically distributed (iid) sample \( \{ X^j_i \}_{i=1}^{n_j} \) drawn from \( F_j \).

In the first sampling framework, called \textit{independent sampling}, we assume that the two samples are independent of one another, and we allow the sample sizes \( n_1 \) and \( n_2 \) to differ. In the development of asymptotics, we treat the sample sizes \( n_1 \) and \( n_2 \) as functions of an underlying index \( n \in \mathbb{N} \), such that as \( n \to \infty \) we have

\[
\frac{n_1 n_2}{n_1 + n_2} \to \infty \quad \text{and} \quad \frac{n_1}{n_1 + n_2} \to \lambda \in [0, 1]. \tag{2.3}
\]

In the second sampling framework, called \textit{matched pairs}, we require the two samples to include the same number of observations \( n = n_1 = n_2 \), and we require the pairs \( \{(X^1_i, X^2_i)\}_{i=1}^n \) to be iid. Dependence between paired observations is permitted; we denote the bivariate copula characterizing this dependence by \( C \). We will require \( C \) to have maximal correlation strictly less than one (see e.g. Beare, 2010, Def. 3.2). In this framework we set \( \lambda = 1/2 \).

\textbf{Assumption 2.2.} The iid samples \( \{ X^1_i \}_{i=1}^{n_1} \) and \( \{ X^2_i \}_{i=1}^{n_2} \) drawn from \( F_1 \) and \( F_2 \) satisfy one of the following conditions.

(i) \textit{(Independent sampling.)} \( \{ X^1_i \}_{i=1}^{n_1} \) and \( \{ X^2_i \}_{i=1}^{n_2} \) are mutually independent, and the sample sizes \( n_1 \) and \( n_2 \) satisfy (2.3).

(ii) \textit{(Matched pairs.)} The sample sizes \( n_1 \) and \( n_2 \) satisfy \( n_1 = n_2 = n \), the pairs \( \{(X^1_i, X^2_i)\}_{i=1}^n \) are iid, and the bivariate copula \( C \) for those pairs has maximal correlation strictly less than one.

\subsection{2.3 Construction of test statistics}

For \( j = 1, 2 \), define the empirical CDF

\[
\hat{F}_j(x) = \frac{1}{n_j} \sum_{i=1}^{n_j} I(X^j_i \leq x), \quad x \in [0, \infty),
\]

the empirical quantile function

\[
\hat{Q}_j(p) = \inf \left\{ x \in [0, \infty) : \hat{F}_j(x) \geq p \right\}, \quad p \in [0, 1],
\]

and the empirical Lorenz curve

\[
\hat{L}_j(p) = \frac{1}{\mu_j} \int_0^p \hat{Q}_j(t) dt, \quad p \in [0, 1],
\]
where $\hat{\mu}_j$ is the sample mean of $\{X_j^i\}_{i=1}^{n_j}$. The difference between empirical Lorenz curves is denoted by

$$\hat{\phi}(p) = \hat{L}_2(p) - \hat{L}_1(p), \quad p \in [0, 1],$$

and the corresponding population quantity is denoted by

$$\phi(p) = L_2(p) - L_1(p), \quad p \in [0, 1].$$

In this paper, all “hatted” quantities are estimated from one or both samples and are implicitly indexed by $n_1$ and/or $n_2$ (or, equivalently, by $n$).

The test statistics we consider are those of Barrett et al. (2014). These test statistics are of the form

$$T_n = \frac{n_1 n_2}{n_1 + n_2},$$

and $F : C[0, 1] \to \mathbb{R}$ is some functional which, loosely speaking, measures the size of the positive part of $\hat{\phi}$. We assume the following property of $F$. (More will be assumed later.)

**Assumption 2.3.** The functional $F : C[0, 1] \to \mathbb{R}$ is such that, for any $h \in C[0, 1]$,

(i) if $h(p) \leq 0$ for all $p \in [0, 1]$ and $h(p) = 0$ for some $p \in [0, 1]$, then $F(h) = 0$;

(ii) if $h(p) > 0$ for some $p \in (0, 1)$, then $F(h) > 0$.

Lorenz curves increase from zero to one, so we always have $\phi(0) = \phi(1) = 0$. Therefore, under Assumption 2.3, the null hypothesis $H_0$ is satisfied if and only if $F(\phi) = 0$, while the alternative hypothesis $H_1$ is satisfied if and only if $F(\phi) > 0$.

Following Barrett et al. (2014), we will focus mostly on two particular choices of $F$, denoted by $S$ and $I$. For $h \in C[0, 1]$, these functionals are given by

$$S(h) = \sup_{p \in [0, 1]} h(p), \quad I(h) = \int_0^1 \max\{h(p), 0\} \, dp. \quad (2.4)$$

Clearly both $S$ and $I$ satisfy Assumption 2.3.

### 2.4 Asymptotic analysis

#### 2.4.1 Weak convergence of empirical Lorenz processes

As a first step to studying the asymptotic behavior of our test statistics, we seek to establish suitable joint weak convergence of the empirical Lorenz processes
$n_1^{1/2}(\hat{L}_1 - L_1)$ and $n_2^{1/2}(\hat{L}_2 - L_2)$. Our approach to the problem is somewhat different to that taken by Barrett et al. (2014), and makes use of recent work by Kaji (2018) establishing weak convergence of empirical quantile processes in the $L^1$-semimetric.

We commence with a statement of the joint weak convergence of the empirical processes $n_1^{1/2}(\hat{F}_1 - F_1)$ and $n_2^{1/2}(\hat{F}_2 - F_2)$. Let $B$ be a centered Gaussian random element of $C[0, 1]^2$ with covariance kernel

$$\text{Cov}(B(u, v), B(u', v')) = C(u \wedge u', v \wedge v') - C(u, v)C(u', v'),$$

(2.5)

where under Assumption 2.2(i) (independent sampling) $C$ is the product copula $C(u, v) = uv$, and under Assumption 2.2(ii) (matched pairs) $C$ is the unique copula function for the pair $(X_1^i, X_2^i)$. Let $B_1$ and $B_2$ be the centered Gaussian random elements of $C[0, 1]$ given by $B_1(u) = B(u, 1)$ and $B_2(u) = B(1, u)$. Note that $B_1$ and $B_2$ are Brownian bridges that are independent under independent sampling, but may be dependent in the matched pairs sampling framework. In either case, from the classical Donsker theorem it is straightforward to deduce the weak convergence

$$\left( n_1^{1/2}(\hat{F}_1 - F_1) \right) \Rightarrow \left( B_1 \circ F_1 \right) \quad \text{in } \ell^\infty[0, \infty) \times \ell^\infty[0, \infty).$$

(2.6)

We would like to apply the functional delta method to deduce from (2.6) joint weak convergence of the empirical Lorenz processes. This is a surprisingly tricky problem, and has been solved only recently by Kaji (2018). Kaji’s insight was to first strengthen the weak convergence in (2.6) so that it obtains under a norm stronger than the uniform norm in each coordinate; see Lemma A.1 for details. This can be done when $F_1$ and $F_2$ have finite $(2 + \epsilon)$-th moment, as they do under Assumption 2.1. With this strengthening of (2.6), it becomes possible to verify a Hadamard differentiability condition on the mapping from CDFs to quantile functions that leads, through an application of the functional delta method, to the weak convergence

$$\left( n_1^{1/2}(\hat{Q}_1 - Q_1) \right) \Rightarrow \left( -Q_1' \cdot B_1 \right) \quad \text{in } L^1(0, 1) \times L^1(0, 1).$$

(2.7)

Here, $Q_j'$ is the derivative of $Q_j$. The demonstration of the weak convergence of empirical quantile processes in $L^1(0, 1)$ provided by Kaji (2018) represents a significant advance over earlier results establishing weak convergence in $\ell^\infty(0, 1)$,
which required unpleasant technical conditions ruling out random variables that are unbounded or have densities not bounded away from zero on their support. Kaji (2018) also permits random variables to take a finite number of values with positive probability, but we do not adopt that level of generality here. With (2.7) in hand, another routine application of the functional delta method, which we relegate to Appendix A along with the proofs of further lemmas and propositions to be stated, leads to the following result establishing joint weak convergence of the empirical Lorenz processes.

**Lemma 2.1.** Under Assumptions 2.1 and 2.2, we have

\[
\left(\frac{n^{1/2}}{2}(\hat{L}_1 - L_1) \right) \overset{d}{\to} \left(\frac{L_1}{L_2}\right) \quad \text{in} \quad C[0,1] \times C[0,1],
\]

(2.8)

where \( L_1 \) and \( L_2 \) are random elements of \( C[0,1] \) given by

\[
L_j(p) = -\int_0^p L_j''(t)B_j(t)dt + L_j(p) \int_0^1 L_j''(t)B_j(t)dt, \quad p \in [0,1].
\]

(2.9)

Moreover, we have

\[
T_n^{1/2}(\hat{\phi} - \phi) \overset{d}{\to} \bar{L} := \lambda^{1/2}L_2 - (1-\lambda)^{1/2}L_1 \quad \text{in} \quad C[0,1].
\]

(2.10)

Lemma 2.1 is similar to Lemma 3 of Barrett et al. (2014) (which also includes a statement of the almost sure uniform convergence of the empirical Lorenz curve). Barrett et al. (2014) appeal to Bhattacharya (2007) for sufficient Hadamard differentiability to deduce the weak convergence of empirical Lorenz processes from the functional delta method. However, Bhattacharya (2007) claims only to establish Hadamard differentiability of the mapping from CDFs to Lorenz curves when \( C[0,1] \), the codomain of that mapping, is equipped with the \( L^1 \)-metric rather than the uniform metric. The implied weak convergence in (2.10) thus holds only in the weaker \( L^1 \)-metric, and is insufficient to obtain the limit distribution of functionals of \( T_n^{1/2}(\hat{\phi} - \phi) \) that are not continuous under the \( L^1 \)-metric, such as the supremum functional \( S \). The new results of Kaji (2018) allow us to close this gap and confirm that (2.10) does indeed hold in the uniform metric.

Much older demonstrations of the weak convergence of empirical Lorenz processes, not using the functional delta method, have been provided by Goldie (1977) and Csörgö et al. (1986) under slightly different technical conditions. The proof using the results of Kaji (2018) and the functional delta method has
the advantage of being easily adapted to our matched pairs sampling framework, or to other sampling frameworks under which joint weak convergence of the two empirical processes obtains, such as serially dependent sampling under a suitable mixing condition. It also provides an immediate justification for bootstrap approximations via the functional delta method for the bootstrap. Recent work by Beutner and Zähle (2010, 2016) using an adaptation of the delta method to derive limit theory for risk functionals (in particular, average value-at-risk, which is closely related to the Lorenz curve) from the weak convergence of empirical processes under a weighted norm seems to be related to that of Kaji (2018), and may provide an alternative path to establishing our results, but we have not worked out the details.

2.4.2 Asymptotic distribution of test statistics

Under Assumption 2.3, the null hypothesis of Lorenz dominance is satisfied if and only if $F(\phi) = 0$. In this case our test statistic satisfies

$$T_n^{1/2} F(\hat{\phi}) = T_n^{1/2} (F(\hat{\phi}) - F(\phi)).$$  \hfill (2.11)

In view of (2.11) and the weak convergence of $T_n^{1/2} (\hat{\phi} - \phi)$ established in Lemma 2.1, it appears that the functional delta method may provide a natural approach to obtaining the limit distribution of our test statistic $T_n^{1/2} F(\hat{\phi})$ under $H_0$. Standard accounts of the functional delta method (Kosorok, 2008) would insist in our setting that the functional $F$ be Hadamard differentiable as a map from $C[0,1]$ to $\mathbb{R}$. This is a prohibitively strong requirement that rules out both $S$ and $I$ as candidate functionals. Fortunately, the functional delta method remains valid under a weaker smoothness condition known as Hadamard directional differentiability. This was shown independently by Shapiro (1991) and Dümbgen (1993), and has been exploited in a number of recent contributions to econometrics including Beare and Moon (2015), Kaido (2016), Beare and Fang (2017), Seo (2018), Beare and Shi (2019) and Fang and Santos (2019).

**Definition 2.2.** Let $D$ and $E$ be normed spaces. A map $F : D \to E$ is said to be Hadamard directionally differentiable at $\phi \in D$ if there is a map $F'_\phi : D \to E$ such that

$$\lim_{n \to \infty} \left\| \frac{F(\phi + t_n h_n) - F(\phi) - F'_\phi(h_n)}{t_n} \right\|_E = 0,$$ \hfill (2.12)

for all sequences $\{h_n\} \subset D$ and $\{t_n\} \subset \mathbb{R}_+$ such that $t_n \downarrow 0$ and $h_n \to h \in D$. 

9
Hadamard directional differentiability differs from Hadamard differentiability in that we only consider sequences \( \{t_n\} \) converging to zero from above, and we do not require the approximating map \( F'_\phi \) to be linear. Though \( F'_\phi \) may be nonlinear, it will always be continuous and positive homogeneous of degree one (Shapiro, 1990). We will see shortly in Remarks 2.1 and 2.2 that Hadamard directional differentiability is satisfied by both of the functionals \( S \) and \( I \). In developing asymptotic theory for test statistics based on an arbitrary functional \( F \), we shall assume directly that Hadamard directional differentiability is satisfied.

**Assumption 2.4.** The functional \( F : C[0,1] \to \mathbb{R} \) is Hadamard directionally differentiable at \( \phi \in C[0,1] \), with directional derivative \( F'_\phi : C[0,1] \to \mathbb{R} \).

**Remark 2.1.** Define the set
\[
\Psi(\phi) = \arg\max_{p \in [0,1]} \phi(p).
\]
Lemma S.4.9 of Fang and Santos (2019) establishes that the functional \( S \) satisfies Assumption 2.4, with directional derivative
\[
S'_\phi(h) = \sup_{p \in \Psi(\phi)} h(p), \quad h \in C[0,1].
\]

**Remark 2.2.** Define the sets
\[
B_0(\phi) = \{p \in [0,1] : \phi(p) = 0\} \quad \text{and} \quad B_+(\phi) = \{p \in [0,1] : \phi(p) > 0\}.
\]
Lemma S.4.5 of Fang and Santos (2019) establishes that the functional \( I \) satisfies Assumption 2.4, with directional derivative
\[
I'_\phi(h) = \int_{B_+(\phi)} h(p)dp + \int_{B_0(\phi)} \max\{h(p), 0\}dp, \quad h \in C[0,1].
\]

The next result follows from Lemma 2.1 by applying the more general version of the functional delta method due to Shapiro (1991) and Dümbgen (1993).

**Proposition 2.1.** Under Assumptions 2.1, 2.2 and 2.4, we have
\[
T_n^{1/2}(F(\hat{\phi}) - F(\phi)) \Rightarrow F'_\phi(\bar{L}) \quad \text{in} \ \mathbb{R}.
\]
Suppose further that \( F \) satisfies Assumption 2.3. If \( H_0 \) is satisfied then we have
\[
T_n^{1/2}F(\hat{\phi}) \Rightarrow F'_\phi(\bar{L}) \quad \text{in} \ \mathbb{R},
\]
whereas if $H_1$ is satisfied then $T_n^{1/2} F(\hat{\phi})$ diverges in probability to infinity.

Proposition 2.1 establishes the asymptotic distribution of the test statistic $T_n^{1/2} F(\hat{\phi})$ at all points in the null. Barrett et al. (2014) did not provide such a characterization. Rather, they observed that if $F$ is monotone and positive homogeneous of degree one then the inequality $T_n^{1/2} F(\hat{\phi}) \leq F(T_n^{1/2} (\hat{\phi} - \phi))$ is valid everywhere in the null, and holds with equality when $\phi = 0$. The limit distribution of the right-hand side of the inequality can be obtained from Lemma 2.1 by applying the continuous mapping theorem, provided that $F$ is suitably continuous. Their critical value is obtained using a bootstrap approximation to this limit distribution, and yields a test with limiting rejection rate equal to the nominal level when $\phi = 0$, and no greater than the nominal level elsewhere in the null. In Section 3 we will propose an alternative bootstrap scheme that approximates the relevant upper quantile of the limit distribution $F(\hat{\phi}(L))$ directly, rather than that of an upper bound. The aim is to improve power.

3 Modified bootstrap procedure

3.1 Construction of bootstrap critical values

We propose to obtain a bootstrap critical value for the test statistic $T_n^{1/2} F(\hat{\phi})$ in the following way. First, we compute from the data an estimate $\hat{F}_\phi'$ of $F'_\phi$, to be discussed in more detail below. Next, we independently generate a large number $N$ of bootstrap statistics of the form $\hat{F}_\phi'(T_n^{1/2} (\hat{\phi}^* - \hat{\phi}))$. This differs from the procedure of Barrett et al. (2014), who generate bootstrap statistics of the form $F(T_n^{1/2} (\hat{\phi}^* - \hat{\phi}))$. To obtain a test with nominal level $\alpha$ we choose as our critical value the $[N(1 - \alpha)]$-th largest of $N$ bootstrap statistics. The estimated functional $\hat{F}_\phi'$ does not vary over bootstrap samples. The bootstrap quantity $\hat{\phi}^*$ is constructed as in Barrett et al. (2014). First we construct bootstrap versions of $\hat{F}_1$ and $\hat{F}_2$ by setting

$$\hat{F}_j^*(x) = \frac{1}{n_j} \sum_{i=1}^{n_j} W_{i,n_j}^j I(X_i^j \leq x), \quad x \in [0, \infty).$$

In the independent sampling framework the weights $W_{n_1}^1 = (W_{1,n_1}^1, \ldots, W_{n_1,n_1}^1)$ and $W_{n_2}^2 = (W_{1,n_2}^2, \ldots, W_{n_2,n_2}^2)$ are drawn independently of the data and of one another from the multinomial distribution with probabilities spread evenly over the categories $1, \ldots, n_1$ and $1, \ldots, n_2$ respectively. In the matched pairs sampling framework we set $W_{n_1}^1 = W_{n_2}^2$, and draw this vector independently
of the data from the multinomial distribution with probabilities spread evenly over the categories \(1, \ldots, n\). In either framework, we then compute bootstrap empirical quantile functions

\[
\hat{Q}^*_j(p) = \inf\{x \in [0, \infty) : \hat{F}^*_j(x) \geq p\}, \quad p \in [0, 1],
\]

and bootstrap empirical Lorenz curves

\[
\hat{L}^*_j(p) = \int_0^p \hat{Q}^*_j(t)dt / \int_0^1 \hat{Q}^*_j(t)dt, \quad p \in [0, 1],
\]

and set \(\hat{\phi}^* = \hat{L}^*_2 - \hat{L}^*_1\).

The estimated functional \(\hat{F}^*_\phi\) needs to consistently approximate the directional derivative \(F^*_\phi\) in a sense to be made precise in Section 3.2. Naturally, the choice of \(\hat{F}^*_\phi\) depends on the choice of \(F\) used to construct our test statistic. We propose specific estimated functionals \(\hat{S}'_\phi\) and \(\hat{I}'_\phi\) to be used when \(F\) is chosen to be \(S\) or \(I\) respectively. These estimated functionals depend on a positive tuning parameter \(\tau_n\) and a regularized estimator \(\hat{V}(p)\) of the variance of \(T_n^{1/2}\phi(p)\), to be discussed in more detail below. Define the estimated contact set

\[
\hat{B}_0(\phi) = \left\{p \in [0, 1] : \left|T_n^{1/2}\hat{\phi}(p)\right| \leq \tau_n\hat{V}(p)^{1/2} \right\}. \tag{3.1}
\]

We propose to estimate \(S'_\phi\) with the functional

\[
\hat{S}'_\phi(h) = \sup_{p \in \hat{B}_0(\phi)} h(p), \quad h \in C[0, 1].
\]

We propose to estimate \(I'_\phi\) with the functional

\[
\hat{I}'_\phi(h) = \int_{\hat{B}_0(\phi)} \max\{h(p), 0\} dp, \quad h \in C[0, 1].
\]

Note that \(\hat{B}_0(\phi)\) is used to estimate both of the sets \(B_0(\phi)\) and \(\Psi(\phi)\). Since the Lorenz curves \(L_1\) and \(L_2\) always touch at zero and one, if the null hypothesis is satisfied then the maximum value achieved by \(\phi\) is zero; thus \(\Psi(\phi) = B_0(\phi)\), and a single estimator suffices. The set \(B_+(\phi)\) appearing in the Hadamard directional derivative of \(I\) is always empty when the null hypothesis is satisfied, so there is no need for us to estimate it.

The following formula for the variance of \(\bar{L}(p)\), the limit in distribution of \(T_n^{1/2}(\hat{\phi}(p) - \phi(p))\) obtained in Lemma 2.1, motivates our choice of \(\hat{V}(p)\).
Proposition 3.1. Under Assumptions 2.1 and 2.2, $\bar{L}(p)$ has variance equal to

$$\text{Var}\left(\frac{(1-\lambda)^{1/2}}{\mu_1}(L_1(p)X^1 - Q_1(p) \land X^1) - \frac{\lambda^{1/2}}{\mu_2}(L_2(p)X^2 - Q_2(p) \land X^2)\right)$$

for each $p \in [0, 1]$.

Motivated by the variance formula in Proposition 3.1, we propose the following (nonregularized) plug-in variance estimators for $T_1^{1/2}\hat{\phi}(p)$: under independent sampling, we set $\tilde{V}(p)$ equal to the sum of the sample variances of the two samples

$$\left\{\frac{T_n^{1/2}}{\hat{\mu}_j n_j^{1/2}}(\hat{L}_j(p)X^1_j - \hat{Q}_j(p) \land X^1_j)\right\}_{i=1}^{n_j}, \quad j = 1, 2,$$

and in the matched pairs sampling framework we set $\tilde{V}(p)$ equal to the sample variance of

$$\left\{\frac{T_n^{1/2}}{\hat{\mu}_1 n^{1/2}}(\hat{L}_1(p)X^1_1 - \hat{Q}_1(p) \land X^1_1) - \frac{T_n^{1/2}}{\hat{\mu}_2 n^{1/2}}(\hat{L}_2(p)X^2_1 - \hat{Q}_2(p) \land X^2_1)\right\}_{i=1}^{n}.$$

Our regularized variance estimator is then given by $\hat{V}(p) = \tilde{V}(p) \lor \nu$, where $\nu$ is a small positive constant, say $\nu = 0.001$. The point of regularization is to ensure that the threshold $\tau_n \hat{V}(p)^{1/2}$ used in the construction of the estimated contact set in (3.1) does not approach zero at $p = 0$ and $p = 1$. A similarly regularized estimator is used by Beare and Shi (2019) to estimate a contact set in a related testing problem.

The form of the contact set estimator in (3.1) lends a nice interpretation to the tuning parameter $\tau_n$: each point $p \in (0, 1)$ is included in our estimated contact set if $T_n^{1/2}\hat{\phi}(p)$ is no more than $\tau_n$ estimated standard deviations from zero. Thus we are effectively using pointwise confidence intervals to select points. In the asymptotic analysis presented in Section 3.2 we will require $\tau_n$ to increase to infinity at a rate slower than $T_n^{1/2}$. In practice, what this means is that $\tau_n$ must be large enough for our estimated contact set to include the true contact set with high probability, but not so large that we include points that the data indicate we may confidently exclude. For a test with nominal level 0.05 and a moderate sample size, a reasonable guess for a good choice of $\tau_n$ might therefore be somewhere in the vicinity of 2 or 3. We provide further discussion of the selection of $\tau_n$ in practice in Section 4.2.
3.2 Asymptotic analysis

When studying the asymptotic behavior of bootstrap procedures it is useful to employ a conditional version of weak convergence in which we condition on the data and allow the bootstrap weights to vary. In our setting, the data are our samples \( \{ X_1^i \}_{i=1}^{n_1} \) and \( \{ X_2^i \}_{i=1}^{n_2} \), and the bootstrap weights are the pair of multinomial random vectors \( W_n := (W_{n_1}^1, W_{n_2}^2) \). We write \( \overset{P}{\rightarrow} \) to denote weak convergence conditional on the data in probability in the sense of Kosorok (2008, pp. 19–20). The following result establishes that, in this sense, the bootstrap process \( T_{1/2}^n (\hat{\phi}^* - \hat{\phi}) \) consistently approximates the weak limit of \( T_{1/2}^n (\tilde{\phi} - \phi) \). It can be deduced using the functional delta method for the bootstrap by applying results of Kaji (2018).

Lemma 3.1. Under Assumptions 2.1 and 2.2, we have

\[
T_{1/2}^n (\hat{\phi}^* - \hat{\phi}) \overset{P}{\rightarrow} \bar{\mathcal{L}} \quad \text{in} \ C[0,1]. \tag{3.2}
\]

Given the conditional weak convergence established in Lemma 3.1, when do we expect the conditional law of \( \hat{\mathcal{F}}_{\phi}^\prime (T_{1/2}^n (\hat{\phi}^* - \hat{\phi})) \) to consistently approximate \( \mathcal{F}_{\phi}^\prime (\bar{\mathcal{L}}) \), the limit distribution of our test statistic under \( H_0 \)? This question can be answered by applying results of Fang and Santos (2019). The following high level assumption on \( \hat{\mathcal{F}}_{\phi}^\prime \) implies Assumption 4 in their paper.

Assumption 3.1. The estimated functional \( \hat{\mathcal{F}}_{\phi}^\prime : C[0,1] \rightarrow \mathbb{R} \) satisfies, for every compact \( K \subseteq C[0,1] \), the property

\[
P \left( \sup_{h \in K} |\hat{\mathcal{F}}_{\phi}^\prime (h) - \mathcal{F}_{\phi}^\prime (h)| > \epsilon \right) \rightarrow 0.
\]

Assumption 3.1 may be tricky to verify, depending on the choice of \( \mathcal{F} \) and \( \hat{\mathcal{F}}_{\phi}^\prime \).

Our next result establishes that our proposed estimators \( \hat{\mathcal{S}}_{\phi}^\prime \) and \( \hat{\mathcal{I}}_{\phi}^\prime \) both satisfy Assumption 3.1 if the tuning parameter \( \tau_n \) is chosen to increase to infinity more slowly than \( T_{1/2}^n \).

Proposition 3.2. Suppose that Assumptions 2.1 and 2.2 are satisfied, and that \( \tau_n \rightarrow \infty \) and \( T_{n}^{-1/2} \tau_n \rightarrow 0 \) as \( n \rightarrow \infty \). Then the estimated functionals \( \hat{\mathcal{S}}_{\phi}^\prime \) and \( \hat{\mathcal{I}}_{\phi}^\prime \) satisfy the condition placed on \( \hat{\mathcal{F}}_{\phi}^\prime \) in Assumption 3.1.

The next result establishes that, if \( \hat{\mathcal{F}}_{\phi}^\prime \) is chosen to satisfy Assumption 3.1, then the distribution of our bootstrap statistic \( \hat{\mathcal{F}}_{\phi}^\prime (T_{1/2}^n (\hat{\phi}^* - \hat{\phi})) \) conditional on the data consistently approximates the weak limit \( \mathcal{F}_{\phi}^\prime (\bar{\mathcal{L}}) \) appearing in Proposition 2.1. It is proved by applying Theorem 3.2 of Fang and Santos (2019).
Proposition 3.3. Under Assumptions 2.1, 2.2, 2.4 and 3.1, we have

\[
\hat{F}'_{\phi}(T_{1/2}^1(\hat{\phi}^* - \hat{\phi})) \overset{P}{\rightarrow} F_{\phi}(\bar{L}) \quad \text{in } \mathbb{R}.
\] (3.3)

Let \( \hat{c}_{1-\alpha} \) denote the \((1 - \alpha)\)-quantile of the bootstrap law of \( \hat{F}'_{\phi}(T_{1/2}^1(\hat{\phi}^* - \hat{\phi})) \):

\[
\hat{c}_{1-\alpha} = \inf \left\{ c \in \mathbb{R} : P \left( \hat{F}'_{\phi}(T_{1/2}^1(\hat{\phi}^* - \hat{\phi})) \leq c \mid \{X_1^{11}_{i=1}, \{X_1^{22}_{i=1}\} \right\} \geq 1 - \alpha \right\}.
\] (3.4)

In practice we approximate \( \hat{c}_{1-\alpha} \) by computing the \([N(1 - \alpha)]\)-th largest of \( N \) independently generated bootstrap statistics, with \( N \) chosen as large as is computationally convenient. The decision rule for our test is

\[
\text{Reject } H_0 \text{ if } T_{n}^{1/2} F(\hat{\phi}) > \hat{c}_{1-\alpha}.
\] (3.5)

The next result characterizes the asymptotic rejection probabilities of our test.

Proposition 3.4. Suppose that Assumptions 2.1, 2.2, 2.3, 2.4 and 3.1 are satisfied.

(i) If \( H_0 \) is true, and the CDF of \( F(\bar{L}) \) is continuous and strictly increasing at its \( 1 - \alpha \) quantile, then \( P(T_{n}^{1/2} F(\hat{\phi}) > \hat{c}_{1-\alpha}) \to \alpha \).

(ii) If \( H_0 \) is false, then \( P(T_{n}^{1/2} F(\hat{\phi}) > \hat{c}_{1-\alpha}) \to 1 \).

Proposition 3.4 establishes that our test consistently rejects arbitrary violations of the null hypothesis, and delivers a limiting rejection rate equal to nominal size at all null configurations for which the CDF of \( F(\bar{L}) \) is continuous and strictly increasing at its \( 1 - \alpha \) quantile. For typical choices of \( F \) and with \( \alpha \in (0, 1/2) \), these configurations will be those for which \( F(\bar{L}) \) is not degenerate at zero. The next result establishes that this is the case for the functionals \( S \) and \( I \).

Proposition 3.5. Suppose that \( \alpha \in (0, 1/2) \), that Assumptions 2.1, 2.2 and 2.4 are satisfied, and that \( H_0 \) is satisfied. Let \( \bar{L} \) be the random element of \( C[0, 1] \) appearing in Lemma 2.1. Then:

(i) Either the CDF of \( S_{\phi}(\bar{L}) \) is continuous and strictly increasing at its \( 1 - \alpha \) quantile, or \( S_{\phi}(\bar{L}) \) is degenerate at zero.

(ii) Either the CDF of \( I_{\phi}(\bar{L}) \) is continuous and strictly increasing at its \( 1 - \alpha \) quantile, or \( I_{\phi}(\bar{L}) \) is degenerate at zero.
The null configurations for which \( F_{\phi}^\prime(\bar{L}) \) is not degenerate at zero constitute what Linton et al. (2010), in a closely related context, refer to as the boundary of the null. For the test based on the functional \( S \), the boundary of the null excludes the case where the Lorenz curves \( L_1 \) and \( L_2 \) touch only at zero and one. For the test based on the functional \( I \), the boundary of the null excludes the case where \( L_1 \) and \( L_2 \) touch only on a set of measure zero. At null configurations that are not on the boundary, so that \( F_{\phi}^\prime(\bar{L}) \) is degenerate at zero, Proposition 2.1 tells us that our test statistic converges in probability to zero, while Proposition 3.3 tells us that our bootstrap critical value converges in probability to zero. It is not clear from our results how the rejection rate of our test will behave asymptotically in this case. This is a common theoretical limitation of tests based on the machinery of Fang and Santos (2019), and also of tests based on generalized moment selection (Andrews and Soares, 2010; Andrews and Shi, 2013), where it manifests at null configurations at which no moment inequalities bind. The usual resolution is to replace the bootstrap critical value \( \hat{c}_{1-\alpha} \) with \( \max\{\hat{c}_{1-\alpha}, \eta\} \) or \( \hat{c}_{1-\alpha} + \eta \), where \( \eta \) is some small positive constant, say \( \eta = 10^{-6} \). This prevents the bootstrap critical value from converging in probability to zero alongside the test statistic, thereby forcing the limiting rejection rate of the test to be zero at null configurations that are not on the boundary. See, for instance, Donald and Hsu (2016, p. 13). We have found in numerical simulations that our test is conservative at null configurations not on the boundary even if we set \( \eta \) equal to zero.

4 Numerical simulations

4.1 Simulation design

We ran a number of Monte Carlo simulations to investigate the finite sample properties of our test. Here we confine attention to the independent sampling framework. Simulations for the matched pairs sampling framework, which produced similar results, are reported in Appendix B. In each simulation we set \( n_1 = n_2 = 2000 \) (a modest size for survey data) and based critical values on \( N = 1000 \) bootstrap samples, with nominal significance level \( \alpha = 0.05 \). Rejection rates were computed over 10000 experimental replications. The contact set estimator for our test was constructed using five different tuning parameter values: \( \tau_n = 1, 2, 3, 4, \infty \). Setting \( \tau_n = \infty \) yields the test of Barrett et al. (2014).

To endow our simulations with a degree of realism we used income distributions belonging to the double Pareto parametric family. Reed (2001, 2003) and
Toda (2012) provide compelling theoretical and empirical evidence for income distributions being well approximated by members of this family. The double Pareto distribution is continuous, with probability density function

\[
f(x) = \begin{cases} 
\frac{\alpha \beta}{\alpha + \beta} x^{-\alpha - 1} & \text{for } x \geq M, \\
\frac{\alpha \beta}{\alpha + \beta} M^{-\beta} x^{-\beta - 1} & \text{for } 0 \leq x < M.
\end{cases}
\]

Here, \( M > 0 \) is a scale parameter, which we may set equal to one without loss of generality since we are concerned with Lorenz curves, and \( \alpha, \beta > 0 \) are parameters governing the shape of the distribution to the right and left of \( M \). We write \( X \sim dP(\alpha, \beta) \) to indicate that a random variable \( X \) has the double Pareto distribution with \( M \) normalized to one and shape parameters \( \alpha, \beta \). This distribution satisfies Assumption 2.1 provided that \( \alpha > 2 \).

### 4.2 Size control and tuning parameter selection

The first simulations we ran were designed to investigate size control and the selection of the tuning parameter \( \tau_n \). We focus on the case where \( X^1 \sim X^2 \sim dP(\alpha, \beta) \), so that the two Lorenz curves coincide, and report rejection rates for a range of values for \( \alpha \) and \( \beta \). Reed (2003, Table 1) reports estimates of \( \alpha \) and \( \beta \) for the income distributions of the US (1997), Canada (1996), Sri Lanka (1981) and Bohemia (1933). (The estimate of \( \alpha \) for the US is misprinted as 22.43 instead of 2.43.) He obtains estimates of \( \alpha \) ranging from 2.09 to 4.16, and of \( \beta \) ranging from 0.79 to 8.4. Using these numbers as a guide, we computed rejection rates for integer values of \( \alpha \) between 2 and 5 and of \( \beta \) between 1 and 8. They are reported in Table 4.1.

It is apparent from Table 4.1 that the test of Barrett et al. (2014), which corresponds to setting \( \tau_n = \infty \), is effective in controlling size. This is true even when \( \alpha = 2 \), which violates Assumption 2.1. As we reduce \( \tau_n \), the rejection rates must weakly increase. However, in all cases considered, the rejection rates with \( \tau_n = 3 \) are identical to those with \( \tau_n = \infty \). It is not until we reduce \( \tau_n \) to 2 that we start to see the rejection rates rise, and this is only for the test based on the functional \( \mathcal{I} \); when using \( \mathcal{S} \), control of size is lost only when we reduce \( \tau_n \) to 1. Based on these results (and on other unreported simulations with sample sizes as small as 200), for tests with a nominal level of 0.05 we suggest using \( \tau_n = 2 \) with the \( \mathcal{S} \) functional or \( \tau_n = 3 \) with the \( \mathcal{I} \) functional. These choices may need to be increased at sample sizes much larger than 2000.
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\tau_n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>12.2</td>
<td>12.4</td>
<td>12.2</td>
<td>12.1</td>
<td>12.1</td>
<td>12.1</td>
<td>12.1</td>
<td>12.1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>5.1</td>
<td>4.8</td>
<td>4.8</td>
<td>4.7</td>
<td>4.8</td>
<td>4.8</td>
<td>4.8</td>
<td>4.9</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>5.1</td>
<td>4.8</td>
<td>4.8</td>
<td>4.7</td>
<td>4.8</td>
<td>4.8</td>
<td>4.8</td>
<td>4.9</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>5.1</td>
<td>4.8</td>
<td>4.8</td>
<td>4.7</td>
<td>4.8</td>
<td>4.8</td>
<td>4.8</td>
<td>4.9</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>5.1</td>
<td>4.8</td>
<td>4.8</td>
<td>4.7</td>
<td>4.8</td>
<td>4.8</td>
<td>4.8</td>
<td>4.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\tau_n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>9.8</td>
<td>10.2</td>
<td>10.3</td>
<td>10.4</td>
<td>10.5</td>
<td>10.6</td>
<td>10.8</td>
<td>10.7</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>4.5</td>
<td>4.4</td>
<td>4.5</td>
<td>4.5</td>
<td>4.6</td>
<td>4.5</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>4.5</td>
<td>4.4</td>
<td>4.5</td>
<td>4.5</td>
<td>4.6</td>
<td>4.5</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>4.5</td>
<td>4.4</td>
<td>4.5</td>
<td>4.5</td>
<td>4.6</td>
<td>4.5</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>4.5</td>
<td>4.4</td>
<td>4.5</td>
<td>4.5</td>
<td>4.6</td>
<td>4.5</td>
<td>4.5</td>
<td>4.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\tau_n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>17.5</td>
<td>17.4</td>
<td>17.4</td>
<td>17.6</td>
<td>17.5</td>
<td>17.6</td>
<td>17.6</td>
<td>17.7</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>5.7</td>
<td>5.6</td>
<td>5.5</td>
<td>5.4</td>
<td>5.3</td>
<td>5.5</td>
<td>5.5</td>
<td>5.5</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>5.4</td>
<td>5.3</td>
<td>5.3</td>
<td>5.2</td>
<td>5.2</td>
<td>5.3</td>
<td>5.3</td>
<td>5.3</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>5.4</td>
<td>5.3</td>
<td>5.3</td>
<td>5.2</td>
<td>5.2</td>
<td>5.3</td>
<td>5.3</td>
<td>5.3</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>5.4</td>
<td>5.3</td>
<td>5.3</td>
<td>5.2</td>
<td>5.2</td>
<td>5.3</td>
<td>5.3</td>
<td>5.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\tau_n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>16.5</td>
<td>16.5</td>
<td>16.9</td>
<td>16.9</td>
<td>16.8</td>
<td>16.8</td>
<td>16.9</td>
<td>17.0</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>5.6</td>
<td>5.5</td>
<td>5.6</td>
<td>5.4</td>
<td>5.4</td>
<td>5.4</td>
<td>5.3</td>
<td>5.3</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>4.9</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>4.8</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>4.9</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>4.8</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>4.9</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>4.8</td>
<td>5.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\tau_n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>16.8</td>
<td>16.5</td>
<td>16.6</td>
<td>16.7</td>
<td>16.6</td>
<td>16.6</td>
<td>16.7</td>
<td>16.7</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>6.0</td>
<td>5.7</td>
<td>5.6</td>
<td>5.7</td>
<td>5.8</td>
<td>5.7</td>
<td>5.7</td>
<td>5.6</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>5.1</td>
<td>5.0</td>
<td>5.0</td>
<td>5.1</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>5.1</td>
<td>5.0</td>
<td>5.0</td>
<td>5.1</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>5.1</td>
<td>5.0</td>
<td>5.0</td>
<td>5.1</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
<td>5.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\tau_n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>16.5</td>
<td>16.4</td>
<td>16.3</td>
<td>16.6</td>
<td>16.5</td>
<td>16.6</td>
<td>16.8</td>
<td>16.8</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>6.1</td>
<td>5.9</td>
<td>5.9</td>
<td>5.8</td>
<td>5.8</td>
<td>5.9</td>
<td>5.7</td>
<td>5.7</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>5.2</td>
<td>5.1</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.2</td>
<td>5.2</td>
<td>5.1</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>5.2</td>
<td>5.1</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.2</td>
<td>5.2</td>
<td>5.1</td>
</tr>
<tr>
<td>$\infty$</td>
<td></td>
<td>5.2</td>
<td>5.1</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.2</td>
<td>5.2</td>
<td>5.1</td>
</tr>
</tbody>
</table>

Table 4.1: Null rejection rates with $X^1 \sim X^2 \sim dP(\alpha, \beta)$ and independent samples of size $n_1 = n_2 = 2000$. Rejection rates are in bold when they exceed the corresponding rate obtained with $\tau_n = \infty$ by more than 0.1 percentage point.
4.3 Power comparison

The next simulations we ran involved power comparisons. In each iteration the data \( \{X^1_i\}^n_{i=1} \) were generated as independent copies of \( X^1 \sim dP(3, 1.5) \), and the data \( \{X^2_i\}^n_{i=1} \) were generated as independent copies of \( X^2(\beta) \sim dP(2.1, \beta) \), whose law is parametrized by \( \beta \). We allowed \( \beta \) to vary between 1.5 and 4 in increments of 0.1. The Lorenz curves corresponding to \( X^1 \), \( X^2_{(1.5)} \) and \( X^2_4 \) are displayed in Figure 4.1. The Lorenz curve for \( X^2_{(1.5)} \) is everywhere equal to or less than the Lorenz curve for \( X^1 \), so that the null hypothesis that \( X^1 \) Lorenz dominates \( X^2_{(1.5)} \) is satisfied. The Lorenz curve for \( X^2_4 \) crosses that for \( X^1 \), so that the null hypothesis is not satisfied. In fact, \( X^1 \) does not Lorenz dominate \( X^2(\beta) \) whenever \( \beta > 1.5 \).

Using data from the Current Population Survey from 2000–2009, Toda (2012, Fig. 2) finds that the US income distribution is well approximated by the \( dP(\alpha, \beta) \) distribution with \( \alpha \approx 2.5 \) and \( \beta \approx 1.5 \). Thus, the Lorenz curve for \( X^1 \) (resp. \( X^2_{(1.5)} \)) is similar to that for the US income distribution, but with a somewhat lighter (resp. heavier) concentration of wealth toward the upper quantiles. As we increase \( \beta \) from 1.5 to 4, the Lorenz curve for \( X^2(\beta) \) shifts so that income inequality among the lower income quantiles is reduced, and neither \( X^1 \) nor \( X^2_{(\beta)} \) Lorenz dominate the other.

The results of our simulations are displayed in Figure 4.2. Here we see immediately that power improves substantially as the tuning parameter \( \tau_n \) is reduced. This is particularly true for the test based on the functional \( I \). When \( \tau_n = \infty \) the test based on the functional \( S \) exhibits substantially more power than the test based on the functional \( I \), but as \( \tau_n \) is reduced this ordering is
Figure 4.2: Power with $X^1 \sim dP(3, 1.5)$ and $X^2_{(\beta)} \sim dP(2.1, \beta)$ as a function of the parameter $\beta$. Going from top to bottom in each panel, the five power curves correspond to our test with $\tau_n = 1, 2, 3, 4$, and the test of Barrett et al. (2014). Samples are independent with $n_1 = n_2 = 2000$.

reversed. Using our recommended tuning parameters of $\tau_n = 2$ with the $S$ functional and $\tau_n = 3$ with the $I$ functional, we obtain slightly more power with the $I$ functional.

A Proofs and supplementary lemmas

Let $L$ be the space of maps $h : [0, \infty) \to \mathbb{R}$ that have right-limits $h(\infty) := \lim_{x \to \infty} h(x)$ and are bounded and satisfy $\int_0^\infty |h(x) - h(\infty)|dx < \infty$. Equip $L$ with the norm

$$
\|h\|_L := \left( \sup_{x \in [0, \infty)} |h(x)| \right) \lor \left( \int_0^\infty |h(x) - h(\infty)|dx \right).
$$

The following lemma is used in the proof of Lemma 2.1.

Lemma A.1. Under Assumptions 2.1 and 2.2, we have

$$
\left( \frac{n_1^{1/2}(\hat{F}_1 - F_1)}{n_2^{1/2}(\hat{F}_2 - F_2)} \right) \rightsquigarrow \left( \frac{B_1 \circ F_1}{B_2 \circ F_2} \right) \quad \text{in } \mathbb{L} \times \mathbb{L}. \quad (A.1)
$$

Proof of Lemma A.1. Proposition 1.2 of Kaji (2018) gives us the convergences

$$
n_1^{1/2}(\hat{F}_1 - F_1) \rightsquigarrow B_1 \circ F_1 \quad \text{and} \quad n_2^{1/2}(\hat{F}_2 - F_2) \rightsquigarrow B_2 \circ F_2 \quad \text{in } \mathbb{L}.
$$
Under Assumption 2.2(i) (independent sampling) we immediately deduce that (A.1) holds. Under Assumption 2.2(ii) (matched pairs) it remains for us to show that the convergence holds jointly. In view of Lemmas 7.12 and 7.14 of Kosorok (2008), the weak convergence of the sequences \( n^{1/2}(\hat{F}_1 - F_1) \) and \( n^{1/2}(\hat{F}_2 - F_2) \) in \( L \) implies that the sequence \( (n^{1/2}(\hat{F}_1 - F_1), n^{1/2}(\hat{F}_2 - F_2)) \) in \( L \times L \) is asymptotically measurable and asymptotically tight. Thus Prohorov’s theorem (Kosorok, 2008, Thm. 7.13) implies that every subsequence of our sequence in \( L \times L \) has a further subsequence that converges weakly. All such weak limits must be equal to \((B_1 \circ F_1, B_2 \circ F_2)\), because otherwise we could not have the weak convergence to that limit in (2.6) obtained using the (weaker) uniform norm instead of \( \| \cdot \|_L \). (Weak limits are unique; see Kosorok, 2008, p. 108.) Thus we obtain the claimed weak convergence in \( L \times L \).

**Proof of Lemma 2.1.** The joint weak convergence of empirical quantile processes in (2.7) follows from Lemma A.1 by applying the functional delta method (see e.g. Kosorok, 2008, Thm. 2.8), using Theorem 1.3 of Kaji (2018) to obtain suitable Hadamard differentiability of the map from distribution functions to quantile functions. To show that (2.7) implies the joint weak convergence of empirical Lorenz processes in (2.8), we again apply the functional delta method. This entails verifying suitable Hadamard differentiability of the map from quantile functions to Lorenz curves. Let \( \mathcal{C}[0, 1] \) be the collection of functions in \( C[0, 1] \) that are not equal to zero at one, and define the maps \( \mathcal{G} : L^1(0, 1) \to C[0, 1] \) and \( \mathcal{H} : \mathcal{C}[0, 1] \subset C[0, 1] \to C[0, 1] \) by

\[
\mathcal{G}(Q) = \int_0^1 Q(t)dt \quad \text{and} \quad \mathcal{H}(R) = \frac{R(\cdot)}{R(1)}.
\]

The map \( \mathcal{G} \) is linear, and thus its own Hadamard derivative. For any \( R \in \mathcal{C}[0, 1] \) and any sequences \( \{h_n\} \subset C[0, 1] \) and \( \{t_n\} \subset \mathbb{R} \setminus \{0\} \) such that \( t_n \to 0 \) and \( h_n \to h \in C[0, 1] \) as \( n \to \infty \), and \( R + t_nh_n \in \mathcal{C}[0, 1] \) for all \( n \), we have

\[
\frac{\mathcal{H}(R + t_nh_n) - \mathcal{H}(R)}{t_n} = \frac{1}{t_n} \left( \frac{R(\cdot) + t_nh_n(\cdot)}{R(1) + t_nh_n(1)} - \frac{R(\cdot)}{R(1)} \right) \cdot h_n(\cdot) = \frac{h_n(1)R(\cdot)}{R(1) + t_nh_n(1)} - \frac{h_n(1)R(\cdot)}{R(1)^2 + t_nh_n(1)R(1)} \\
\to h(\cdot) - \frac{h(1)R(\cdot)}{R(1)^2} \quad \text{(A.2)}
\]

in the uniform metric. Thus \( \mathcal{H} \) is Hadamard differentiable, with derivative \( \mathcal{H}'_R(h) \) given by the limit in (A.2). The composition \( \mathcal{H} \circ \mathcal{G} \) is the map from
quantile functions to Lorenz curves whose Hadamard differentiability we seek to establish. The chain rule for Hadamard derivatives (Kosorok, 2008, Lem. 6.19) implies its differentiability at any \( Q \in L^1(0,1) \) with nonzero integral, with derivative

\[
(\mathcal{H} \circ \mathcal{G})'_Q(h) = \frac{1}{\mu} \int_0^1 h(t)dt - \frac{1}{\mu^2} \int_0^1 h(t)dt \int_0^t Q(t)dt,
\]

where we define \( \mu = \int_0^1 Q(t)dt \). Thus, by applying the functional delta method to the weak convergence (2.7), we obtain

\[
\left( \frac{n_1^{1/2}}{n_1^{1/2} + n_2^{1/2}} \left( \hat{L}_1 - L_1 \right) \right) \rightsquigarrow \left( \frac{n_1^{1/2}}{n_1^{1/2} + n_2^{1/2}} \left( (\mathcal{H} \circ \mathcal{G})'_Q(\hat{Q}_1) - (\mathcal{H} \circ \mathcal{G})(Q_1) \right) \right) \Rightarrow \left( \frac{L_1}{L_1^T} \right) \tag{A.3}
\]

in \( C[0,1] \times C[0,1] \times [0,1] \), as claimed.

It remains to establish (2.10). To achieve this we augment (A.3) to obtain

\[
\left( \frac{n_1^{1/2}}{n_1^{1/2} + n_2^{1/2}} (\hat{L}_1 - L_1) \right) \rightsquigarrow \left( \frac{n_1^{1/2}}{n_1^{1/2} + n_2^{1/2}} (\mathcal{H} \circ \mathcal{G})'_Q(\hat{Q}_1, -Q_1^1 \cdot B_1) \right) \Rightarrow \left( \frac{L_1}{L_1^T} \right)
\]

in \( C[0,1] \times C[0,1] \times [0,1] \), as claimed.

For \( j = 1, 2 \) and \( p, t \in (0,1) \), define

\[
H_{j,p}(t) = \frac{1}{\mu_j} \left( L_j(p)Q_j(t) - Q_j(t \wedge p) \right).
\]

Note that \( H_{j,p}(\cdot) \) is square-integrable under Assumption 2.1. The following
lemma is used in the proofs of Propositions 3.1 and 3.5.

**Lemma A.2.** Under Assumptions 2.1 and 2.2, the Lorenz processes $L_1$ and $L_2$ satisfy $\text{Var}(L_1(p)) = \text{Var}(H_{1,p}(U))$, $\text{Var}(L_2(p)) = \text{Var}(H_{2,p}(V))$, and

$$\text{Cov}(L_1(p), L_2(p)) = \text{Cov}(H_{1,p}(U), H_{2,p}(V))$$

for each $p \in (0, 1)$, where $(U, V)$ is a pair of random variables with joint CDF given by the copula $C$.

**Proof of Lemma A.2.** In view of (2.5), the covariances between the Brownian bridges $B_1$ and $B_2$ satisfy

$$\text{Cov}(B_1(s), B_2(t)) = C(s, t) - st, \quad s, t \in [0, 1],$$

(A.4)

where under Assumption 2.2(i) (independent sampling), $C$ is the product copula, and under Assumption 2.2(ii) (matched pairs), $C$ is the bivariate copula common to the pairs $(X^1_i, X^2_i)$. For $j = 1, 2$ and $p, t \in (0, 1)$, define

$$h_{j,p}(t) = \frac{1}{\mu_j}(L_j(p) - 1(t \leq p))Q'_j(t).$$

Note that almost sure integrability of $h_{j,p}(\cdot)B_j(\cdot)$ follows from the weak convergence $n^{1/2}(\hat{Q}_j - Q_j) \rightsquigarrow -Q'_jB_j$ in $L^1(0, 1)$ established by Kaji (2018). Note also that $L_j(p) = \int_0^1 h_{j,p}(t)B_j(t)dt$. Therefore, applying (A.4) and Fubini’s theorem, we obtain

$$\text{Cov}(L_1(p), L_2(p)) = \int_0^1 \int_0^1 h_{1,p}(s)h_{2,p}(t)(C(s, t) - st)dsdt.$$

The function $H_{j,p}$ is the antiderivative of $h_{j,p}$: it satisfies $H_{j,p}(\cdot) = \int_0^* h_{j,p}(t)dt$. A generalization of Hoeffding’s lemma due to Lo (2017, Thm. 3.1) – see also Cuadras (2002) and Beare (2009) – thus implies that

$$\text{Cov}(H_{1,p}(U), H_{2,p}(V)) = \int_0^1 \int_0^1 h_{1,p}(s)h_{2,p}(t)(C(s, t) - st)dsdt.$$

This proves our claimed covariance formula. From this we obtain the claimed variance formulas for $L_1(p)$ and $L_2(p)$ by setting $F_1 = F_2$ and setting $C$ equal to the Fréchet-Hoeffding upper bound, so that $H_{1,p}(U) = H_{2,p}(V)$ almost surely. (Note that the derivation of our covariance formula was valid for any copula $C$, including the Fréchet-Hoeffding upper bound.)
Proof of Proposition 3.1. The CDFs \(F_1\) and \(F_2\) are continuous under Assumption 2.1, so \((F_1(X^1), F_2(X^2))\) is a pair of random variables with joint CDF given by the copula \(C\). It therefore follows from Lemma A.2 that

\[
\text{Var}(L_j(p)) = \text{Var}(H_j,p(F_j(X^j))) = \text{Var}\left(\frac{1}{\mu_j}(L_j(p)X^j - Q_j(p)\land X^j)\right)
\]

for \(j = 1, 2\), and

\[
\text{Cov}(L_1(p), L_2(p)) = \text{Cov}(H_1,p(F_1(X^1)), H_2,p(F_2(X^2)))
\]

\[
= \text{Cov}\left(\frac{1}{\mu_1}(L_1(p)X^1 - Q_1(p)\land X^1), \frac{1}{\mu_2}(L_2(p)X^2 - Q_2(p)\land X^2)\right).
\]

The desired result follows easily.

Let \(\Rightarrow \) denote weak convergence conditional on the data almost surely in the sense of Kosorok (2008, pp. 19–20). The next lemma is used to prove Lemma 3.1.

Lemma A.3. Under Assumptions 2.1 and 2.2, we have

\[
\begin{pmatrix}
\frac{n_1^{1/2}}{\mu_1}(\hat{F}^*_1 - \hat{F}_1) \\
\frac{n_2^{1/2}}{\mu_2}(\hat{F}^*_2 - \hat{F}_2)
\end{pmatrix}
\Rightarrow \begin{pmatrix}
B_1 \circ F_1 \\
B_2 \circ F_2
\end{pmatrix}
\text{ in } L \times L.
\] (A.5)

Proof of Lemma A.3. Standard results on bootstrapping empirical processes (see e.g. Kosorok, 2008, Thm. 2.7) imply the weak convergence

\[
\begin{pmatrix}
\frac{n_1^{1/2}}{\mu_1}(\hat{F}^*_1 - \hat{F}_1) \\
\frac{n_2^{1/2}}{\mu_2}(\hat{F}^*_2 - \hat{F}_2)
\end{pmatrix}
\Rightarrow \begin{pmatrix}
B_1 \circ F_1 \\
B_2 \circ F_2
\end{pmatrix}
\text{ in } \ell^\infty(0, \infty) \times \ell^\infty(0, \infty).
\] (A.6)

Our task is to strengthen the norm in which this weak convergence obtains. Lemmas 1.G.1, 1.G.2 and 1.G.3 of Kaji (2018), and the surrounding discussion, establish that

\[
n_1^{1/2}(\hat{F}^*_1 - \hat{F}_1) \Rightarrow \frac{P}{W}B_1 \circ F_1 \quad \text{and} \quad n_2^{1/2}(\hat{F}^*_2 - \hat{F}_2) \Rightarrow \frac{P}{W}B_2 \circ F_2 \text{ in } L.
\] (A.7)

In fact, the conclusion of Lemma 1.G.3 of Kaji (2018) can be strengthened to hold outer almost surely rather than in outer probability by arguing as in the first paragraph of the proof of Theorem 2.9.7 of van der Vaart and Wellner (1996) to show that (in Kaji's notation) \(E_\xi \|Z_n\|_{\ell^\infty_{n,d}} \rightarrow 0\) outer almost surely as \(n \rightarrow \infty\) followed by \(\delta \downarrow 0\). With this strengthening, Kaji's results imply the
stronger convergences

\[ n_{1/2}^1 (\hat{F}^* - \hat{F}_1) \xrightarrow{\text{as}} W_b \circ F_1 \quad \text{and} \quad n_{1/2}^2 (\hat{F}^*_2 - \hat{F}_2) \xrightarrow{\text{as}} W_b \circ F_2 \quad \text{in} \ L. \tag{A.8} \]

Under Assumption 2.2(i) (independent sampling) we immediately deduce that (A.5) holds. Under Assumption 2.2(ii) (matched pairs) it remains for us to show that the convergence holds jointly. To achieve this we will argue as in the proof of Lemma A.1 while holding the data fixed. Holding \( \{(X^{1}_i, X^{2}_i)\}_{i=1}^{\infty} \) fixed at any point outside a set of outer probability zero, we deduce from (A.8) that the sequence \( (n_{1/2}^1 (\hat{F}^* - \hat{F}_1), n_{1/2}^2 (\hat{F}^*_2 - \hat{F}_2)) \) in \( L \times L \) is asymptotically measurable and asymptotically tight (Kosorok, 2008, Lems. 7.12, 7.14). We then obtain (A.5) by applying Prohorov’s theorem and observing that no subsequence of our sequence in \( L \times L \) may have a weak limit different to \( (B_1 \circ F_1, B_2 \circ F_2) \) without contradicting the weak convergence in (A.6) obtained using the (weaker) uniform norm instead of \( \| \cdot \|_L \).

Proof of Lemma 3.1. The conditional weak convergence

\[ \left( \begin{array}{c} n_{1/2}^1 (\hat{Q}_1^* - \hat{Q}_1) \\ n_{1/2}^2 (\hat{Q}_2^* - \hat{Q}_2) \end{array} \right) \xrightarrow{\text{P}} \left( \begin{array}{c} -Q_1^* \cdot B_1 \\ -Q_2^* \cdot B_2 \end{array} \right) \text{ in } L^1((0,1) \times L^1(0,1)) \]

may be obtained from Lemma A.3 by applying the functional delta method for the bootstrap (Kosorok, 2008, Thm. 2.9), using Theorem 1.3 of Kaji (2018) to obtain suitable Hadamard differentiability of the map from distribution functions to quantile functions. In view of the Hadamard differentiability of the map from quantile functions to Lorenz curves established in the proof of Lemma 2.1, another application of the functional delta method for the bootstrap yields

\[ \left( \begin{array}{c} n_{1/2}^1 (\hat{L}_1^* - \hat{L}_1) \\ n_{1/2}^2 (\hat{L}_2^* - \hat{L}_2) \end{array} \right) \xrightarrow{\text{P}} \left( \begin{array}{c} L_1 \\ L_2 \end{array} \right) \text{ in } C[0, 1] \times C[0, 1]. \tag{A.9} \]

From this we may deduce the claimed conditional weak convergence by applying a conditional version of the continuous mapping theorem (Kosorok, 2008, Prop. 10.7) with the map \( \theta \) defined at the end of the proof of Lemma 2.1. \( \square \)

For \( \delta > 0 \), define \( B_\delta(\phi) = \{ p \in [0, 1] : |\phi(p)| \leq \delta \} \). The following lemma is used in the proof of Proposition 3.2.

Lemma A.4. Suppose that Assumptions 2.1 and 2.2 are satisfied, and that \( \tau_n \to \)
∞ and $T_n^{-1/2} \tau_n \to 0$ as $n \to \infty$. If $H_0$ is true, then for any $\delta > 0$ we have

$$P \left( B_0(\phi) \subseteq \overline{B_0} \subseteq B_0(\phi) \right) \to 1.$$  

**Proof of Lemma A.4.** We first show that $\sup_{p \in [0,1]} \tilde{V}(p)$ is $O_{a.s.}(1)$. Since $\hat{L}_j(p)X_i^j$ and $\hat{Q}_j(p) \wedge X_i^j$ are both nonnegative and $|\hat{L}_j(p)| \leq 1$, we have

$$|\hat{L}_j(p)X_i^j - \hat{Q}_j(p) \wedge X_i^j| \leq (\hat{L}_j(p)X_i^j) \lor (\hat{Q}_j(p) \wedge X_i^j) \leq X_i^j,$$

for $j = 1, 2$ and $p \in [0,1]$. Using this bound and the strong law of large numbers, it is simple to show that $\tilde{V}$ satisfies

$$\sup_{p \in [0,1]} \tilde{V}(p) \leq \frac{T_n}{\mu_1^2 n^2} \sum_{i=1}^{n_1} (X_i^1)^2 + \frac{T_n}{\mu_2^2 n^2} \sum_{i=1}^{n_2} (X_i^2)^2 \to 1 - \lambda \mu_1^2 \mu_2^2 \frac{n}{2} E(X^1)^2 + \lambda \mu_1^2 \mu_2^2 \frac{n}{2} E(X^2)^2$$

almost surely under Assumption 2.2(i) (independent sampling), or

$$\sup_{p \in [0,1]} \tilde{V}(p) \leq \frac{1}{2n} \sum_{i=1}^{n} \left( \frac{1}{\mu_1^2} X_i^1 + \frac{1}{\mu_2^2} X_i^2 \right)^2 \leq \frac{1}{\mu_1^2 n} \sum_{i=1}^{n} (X_i^1)^2 + \frac{1}{\mu_2^2 n} \sum_{i=1}^{n} (X_i^2)^2 \to \frac{n}{2} E(X^1)^2 + \frac{n}{2} E(X^2)^2$$

almost surely under Assumption 2.2(ii) (matched pairs). Thus $\sup_{p \in [0,1]} \tilde{V}(p)$ is $O_{a.s.}(1)$ as claimed.

Let $\| \cdot \|$ denote the uniform norm on $C[0,1]$. The set $\overline{B_0(\phi)}$ will contain $B_0(\phi)$ if $|T_n^{1/2}(\phi(p) - \phi(p))| \leq \tau_n \tilde{V}(p)^{1/2}$ for all $p \in [0,1]$. Since $\tilde{V}(p) \geq \nu$, this will certainly be true if $\|T_n^{1/2}(\phi - \phi')\| \leq \tau_n \nu^{1/2}$. Therefore,

$$P \left( B_0(\phi) \subseteq \overline{B_0(\phi)} \right) \geq P \left( \|T_n^{1/2}(\phi - \phi')\| \leq \tau_n \nu^{1/2} \right) \to 1,$$

with the convergence to one following from the fact that $\tau_n \nu^{1/2} \to \infty$ while $T_n^{1/2}(\phi - \phi) \sim \tilde{L}$ in $C[0,1]$ by Lemma 2.1.

Let $\varepsilon_n = T_n^{-1/2} \tau_n \sup_{p \in [0,1]} \tilde{V}(p)$. Since $\sup_{p \in [0,1]} \tilde{V}(p)$ is $O_{a.s.}(1)$, we have $\varepsilon_n \to 0$ almost surely. The set $B_0(\phi)$ will contain $B_0(\phi)$ if $|\phi(p)| \leq \delta$ whenever $|T_n^{1/2}(\phi(p))| \leq \tau_n \tilde{V}(p)^{1/2}$. If $\varepsilon_n < \delta$, then this occurs if $\phi$ is everywhere within $\delta - \varepsilon_n$ of $\phi$. Since $\varepsilon_n \to 0$ almost surely, for sufficiently large $n$ we therefore have

$$P \left( \overline{B_0(\phi)} \subseteq B_0(\phi) \right) \geq P \left( \|T_n^{1/2}(\phi - \phi')\| \leq \tau_n \nu^{1/2}(\delta - \varepsilon_n) \right) \to 1,$$
with the convergence to one following from the fact that \( T_n^{1/2}(\delta - \varepsilon_n) \to \infty \) almost surely while \( T_n^{1/2}(\hat{\phi} - \phi) \sim \mathcal{L} \) in \( C[0, 1] \) by Lemma 2.1.

**Proof of Proposition 3.2.** It is easy to see that our estimated functionals satisfy the Lipschitz conditions

\[
|\hat{S}_\phi'(h_1) - \hat{S}_\phi'(h_2)| \leq \|h_1 - h_2\| \quad \text{and} \quad |\hat{T}_\phi'(h_1) - \hat{T}_\phi'(h_2)| \leq \|h_1 - h_2\|
\]

for \( h_1, h_2 \in C[0, 1] \). Therefore, Lemma S.3.6 of Fang and Santos (2019) implies that a sufficient condition for \( \hat{S}_\phi' \) and \( \hat{T}_\phi' \) to satisfy their Assumption 4 is that, for any \( \epsilon > 0 \),

\[
P \left( \left| \hat{S}_\phi'(h) - S_\phi'(h) \right| > \epsilon \right) \to 0 \quad \text{and} \quad P \left( \left| \hat{T}_\phi'(h) - T_\phi'(h) \right| > \epsilon \right) \to 0 \quad (A.10)
\]

for each \( h \in C[0, 1] \). Moreover, since \( n^{1/2}(\hat{\phi}^* - \hat{\phi}) \) is a Borel measurable map into the separable space \( C[0, 1] \), Assumption 4 of Fang and Santos (2019) is equivalent to our Assumption 3.1; see their Remark 3.3. Thus we need only verify (A.10).

To verify the first part of (A.10), fix \( h \in C[0, 1] \) and \( \epsilon > 0 \), and choose \( \delta > 0 \) small enough that \( |h(p) - h(q)| < \epsilon \) whenever \( |p - q| < \delta \). Next choose \( \eta > 0 \) small enough that \( |p - q| < \delta \) for some \( q \in B_\eta(\phi) \) whenever \( p \in B_\eta(\phi) \). Observe that if

\[
B_\eta(\phi) \subseteq \hat{B}_\eta(\phi) \subseteq B_\eta(\phi),
\]

then it must be the case that

\[
\left| \hat{S}_\phi'(h) - S_\phi'(h) \right| \leq \sup_{p \in B_\eta(\phi)} h(p) - \sup_{p \in B_\eta(\phi)} h(p) \leq \epsilon.
\]

The first part of (A.10) now follows from Lemma A.4.

To verify the second part of (A.10), fix \( h \in C[0, 1] \) and \( \epsilon > 0 \), and choose \( \delta > 0 \) small enough that

\[
\|h\| \int_0^1 1(0 < |\phi(p)| \leq \delta)dp \leq \epsilon,
\]

which is possible by the dominated convergence theorem. Observe that if

\[
B_\eta(\phi) \subseteq \hat{B}_\eta(\phi) \subseteq B_\delta(\phi),
\]

27
then it must be the case that
\[ \left| \hat{I}_h \phi(h) - I_\phi(h) \right| \leq \|h\| \int_0^1 \mathbb{1}(0 < |\phi(p)| \leq \delta) dp \leq \epsilon. \]

The second part of (A.10) now follows from Lemma A.4. \( \square \)

**Proof of Proposition 3.3.** Conditional weak convergence follows from Theorem 3.2 of Fang and Santos (2019). We need only verify Assumptions 1–4 of their result. Their Assumption 1 is implied by our Assumption 2.4. Their Assumption 2 is implied by our Lemma 2.1. Their Assumption 3 is implied by our Lemma 3.1. Their Assumption 4 is implied by our Assumption 3.1 (see their Remark 3.3). \( \square \)

**Proof of Proposition 3.4.** Part (i) follows from Theorem 3.3 of Fang and Santos (2019), whose local analysis reduces to a statement about pointwise size control when we set \( \lambda = 0 \). The first four of their assumptions were verified in our proof of Proposition 3.3, and the fifth is redundant when \( \lambda = 0 \). Part (ii) is true because \( F(\hat{\phi}) \to F(\phi) > 0 \) in probability as \( n \to \infty \) under \( H_1 \) (due to the uniform consistency of \( \hat{\phi} \) implied by Lemma 2.1) and \( \hat{c}_{1-\alpha} \) is bounded in probability as \( n \to \infty \) (a consequence of Proposition 3.3). \( \square \)

The following lemma is used in the proof of Proposition 3.5.

**Lemma A.5.** Under Assumptions 2.1 and 2.2, the variance of \( \mathcal{L}(p) \) is strictly positive for all \( p \in (0, 1) \).

**Proof.** Under Assumption 2.2(i) (independent sampling) \( \mathcal{L}_1(p) \) and \( \mathcal{L}_2(p) \) are independent. Therefore, since each of them has strictly positive variance for \( p \in (0, 1) \) by Lemma A.2, their weighted difference \( \mathcal{L}(p) \) trivially also has strictly positive variance. It remains to establish that \( \mathcal{L}(p) \) has strictly positive variance under Assumption 2.2(ii) (matched pairs). Since \( C \) has maximal correlation strictly less than one, we must have
\[
\text{Cov}(H_{1,p}(U), H_{2,p}(V)) < \sqrt{\text{Var}(H_{1,p}(U))\text{Var}(H_{2,p}(V))}. \tag{A.11}
\]
We thus deduce from Lemma A.2 that
\[
\text{Cov}(\mathcal{L}_1(p), \mathcal{L}_2(p)) < \sqrt{\text{Var}(\mathcal{L}_1(p))\text{Var}(\mathcal{L}_2(p))}, \tag{A.12}
\]
meaning that the correlation between \( \mathcal{L}_1(p) \) and \( \mathcal{L}_2(p) \) must be strictly less than one. The weighted difference \( \mathcal{L}(p) \) therefore cannot have zero variance. \( \square \)
Proof of Proposition 3.5. We first observe that since $\bar{L}$ is Gaussian and the directional derivatives $S'_\phi$ and $I'_\phi$ are continuous and convex, Theorem 11.1 of Davydov et al. (1998) can be used to show that the CDFs of $S'_\phi(\bar{L})$ and $I'_\phi(\bar{L})$ are continuous everywhere except perhaps at zero, and that if either CDF assigns probability less than one to zero, then it is strictly increasing on $(0, \infty)$. Thus if either CDF is not continuous and strictly increasing at its $1-\alpha$ quantile, then it must assign probability of at least $1-\alpha$ to zero.

To demonstrate claim (a), observe that if the set $\Psi(\phi)$ includes some point $p_0 \not\in \{0, 1\}$, then

$$P\left(S'_\phi(\bar{L}) > 0\right) \geq P\left(\bar{L}(p_0) > 0\right) = \frac{1}{2},$$

with the final equality following from Lemma A.5 and the fact that $\bar{L}(p_0)$ is a centered Gaussian random variable. Thus the CDF of $S'_\phi(\bar{L})$ can assign probability of no greater than one half to zero. Since $1-\alpha > 1/2$, we conclude that the CDF must be continuous and strictly increasing at its $1-\alpha$ quantile. On the other hand, if $\Psi(\phi)$ does not include any point $p_0 \not\in \{0, 1\}$, then clearly $S'_\phi(\bar{L})$ is degenerate at zero.

To demonstrate claim (b), suppose that $I'_\phi(\bar{L})$ is not degenerate at zero. Since we have assumed $H_0$ to be satisfied, we have $B_+(\phi) = \emptyset$, and so $B_0(\phi)$ must be a set of positive measure. Thus the Lebesgue density theorem ensures the existence of $p_0 \in B_0 \cap (0, 1)$ such that the set $(p_0 - \epsilon, p_0 + \epsilon) \cap B_0$ has positive measure for all $\epsilon > 0$. Since $\bar{L}(p)$ is continuous in $p$, if $\bar{L}(p_0) > 0$ then we must have $\bar{L} > 0$ on $(p_0 - \epsilon, p_0 + \epsilon)$ for some $\epsilon > 0$, implying that

$$I'_\phi(\bar{L}) \geq \int_{(p_0 - \epsilon, p_0 + \epsilon) \cap B_0} \bar{L}(p)dp > 0.$$

Thus we have

$$P\left(I'_\phi(\bar{L}) > 0\right) \geq P\left(\bar{L}(p_0) > 0\right) = \frac{1}{2},$$

with the final equality following from Lemma A.5 and the fact that $\bar{L}(p_0)$ is a centered Gaussian random variable. Thus the CDF of $I'_\phi(\bar{L})$ can assign probability of no greater than one half to zero, and since $1-\alpha > 1/2$, we conclude that the CDF must be continuous and strictly increasing at its $1-\alpha$ quantile. \qed
B  Further numerical simulations

The numerical simulations reported in Section 4 pertained to the independent sampling framework. Here we report analogous simulations for the matched pairs sampling framework. The simulation design is the same as described in Section 4.1, except that dependence between pairs was induced by linking the random variables $X_1^i$ and $X_2^i$ with a Gaussian copula with parameter $\rho = 0.25, 0.5, 0.75$. In Tables B.1, B.2 and B.3 we report results analogous to those reported in Table 4.1, and in Figure B.1 we report results analogous to those reported in Figure 4.2. Qualitatively, the results for the matched pairs sampling framework are similar to those for the independent sampling framework. Increasing the dependence between pairs appears to reduce the null rejection rates with smaller tuning parameter choices.

References


<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\tau_n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$12.4$</td>
<td>$12.3$</td>
<td>$12.3$</td>
<td>$12.2$</td>
<td>$12.2$</td>
<td>$12.3$</td>
<td>$12.3$</td>
<td>$12.3$</td>
<td>$12.3$</td>
</tr>
<tr>
<td>2</td>
<td>$5.1$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
</tr>
<tr>
<td>3</td>
<td>$5.1$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
</tr>
<tr>
<td>4</td>
<td>$5.1$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
</tr>
<tr>
<td>$\infty$</td>
<td>$5.1$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.9$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$F = S$</th>
<th>$\beta$</th>
<th>9.9</th>
<th>10.3</th>
<th>10.3</th>
<th>10.2</th>
<th>10.2</th>
<th>10.1</th>
<th>10.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$8.6$</td>
<td>$9.0$</td>
<td>$9.4$</td>
<td>$9.6$</td>
<td>$9.7$</td>
<td>$9.7$</td>
<td>$9.7$</td>
<td>$9.7$</td>
</tr>
<tr>
<td>2</td>
<td>$4.7$</td>
<td>$4.5$</td>
<td>$4.5$</td>
<td>$4.5$</td>
<td>$4.5$</td>
<td>$4.5$</td>
<td>$4.5$</td>
<td>$4.6$</td>
</tr>
<tr>
<td>3</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.6$</td>
</tr>
<tr>
<td>4</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.6$</td>
</tr>
<tr>
<td>$\infty$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.7$</td>
<td>$4.6$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$F = I$</th>
<th>$\beta$</th>
<th>17.8</th>
<th>17.3</th>
<th>17.2</th>
<th>17.1</th>
<th>17.1</th>
<th>17.0</th>
<th>17.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$16.5$</td>
<td>$16.6$</td>
<td>$16.3$</td>
<td>$16.3$</td>
<td>$16.4$</td>
<td>$16.3$</td>
<td>$16.1$</td>
<td>$16.1$</td>
</tr>
<tr>
<td>2</td>
<td>$5.6$</td>
<td>$5.3$</td>
<td>$5.3$</td>
<td>$5.2$</td>
<td>$5.1$</td>
<td>$5.1$</td>
<td>$5.1$</td>
<td>$5.1$</td>
</tr>
<tr>
<td>3</td>
<td>$5.1$</td>
<td>$4.9$</td>
<td>$4.9$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
</tr>
<tr>
<td>4</td>
<td>$5.1$</td>
<td>$4.9$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
</tr>
<tr>
<td>$\infty$</td>
<td>$5.1$</td>
<td>$4.9$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
<td>$4.8$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$F = I$</th>
<th>$\beta$</th>
<th>15.2</th>
<th>15.7</th>
<th>15.9</th>
<th>15.9</th>
<th>15.7</th>
<th>15.7</th>
<th>15.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$5.5$</td>
<td>$5.5$</td>
<td>$5.5$</td>
<td>$5.5$</td>
<td>$5.4$</td>
<td>$5.5$</td>
<td>$5.5$</td>
<td>$5.5$</td>
</tr>
<tr>
<td>2</td>
<td>$5.0$</td>
<td>$5.1$</td>
<td>$5.0$</td>
<td>$5.1$</td>
<td>$5.1$</td>
<td>$5.2$</td>
<td>$5.3$</td>
<td>$5.2$</td>
</tr>
<tr>
<td>3</td>
<td>$5.0$</td>
<td>$5.1$</td>
<td>$5.0$</td>
<td>$5.1$</td>
<td>$5.1$</td>
<td>$5.2$</td>
<td>$5.3$</td>
<td>$5.2$</td>
</tr>
<tr>
<td>$\infty$</td>
<td>$5.0$</td>
<td>$5.1$</td>
<td>$5.0$</td>
<td>$5.1$</td>
<td>$5.1$</td>
<td>$5.2$</td>
<td>$5.3$</td>
<td>$5.2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\tau_n$</th>
<th>15.2</th>
<th>15.6</th>
<th>15.9</th>
<th>16.2</th>
<th>16.1</th>
<th>16.0</th>
<th>15.9</th>
<th>15.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$5.5$</td>
<td>$5.6$</td>
<td>$5.6$</td>
<td>$5.6$</td>
<td>$5.8$</td>
<td>$5.6$</td>
<td>$5.5$</td>
<td>$5.5$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$5.2$</td>
<td>$5.1$</td>
<td>$5.1$</td>
<td>$5.2$</td>
<td>$5.3$</td>
<td>$5.3$</td>
<td>$5.2$</td>
<td>$5.2$</td>
<td></td>
</tr>
<tr>
<td>$\infty$</td>
<td>$5.2$</td>
<td>$5.1$</td>
<td>$5.1$</td>
<td>$5.2$</td>
<td>$5.3$</td>
<td>$5.3$</td>
<td>$5.2$</td>
<td>$5.2$</td>
<td></td>
</tr>
</tbody>
</table>

Table B.1: Null rejection rates with $X^1 \sim X^2 \sim dP(\alpha, \beta)$ and $n = 2000$ matched pairs linked by a Gaussian copula with correlation parameter $\rho = 0.25$. Rejection rates are in bold when they exceed the corresponding rate obtained with $\tau_n = \infty$ by more than 0.1 percentage point.
Table B.2: Null rejection rates with $X^1 \sim X^2 \sim dP(\alpha, \beta)$ and $n = 2000$ matched pairs linked by a Gaussian copula with correlation parameter $\rho = 0.50$. Rejection rates are in bold when they exceed the corresponding rate obtained with $\tau_n = \infty$ by more than 0.1 percentage point.
Table B.3: Null rejection rates with $X^1 \sim X^2 \sim dP(\alpha, \beta)$ and $n = 2000$ matched pairs linked by a Gaussian copula with correlation parameter $\rho = 0.75$. Rejection rates are in bold when they exceed the corresponding rate obtained with $\tau_n = \infty$ by more than 0.1 percentage point.
Figure B.1: Power with $X^1 \sim \text{dP}(3, 1.5)$ and $X^2_{(\beta)} \sim \text{dP}(2.1, \beta)$ as a function of the parameter $\beta$. Going from top to bottom in each panel, the five power curves correspond to our test with $\tau_n = 1, 2, 3, 4$, and the test of Barrett et al. (2014). Samples are $n = 2000$ matched pairs linked by a Gaussian copula with correlation parameter $\rho$. 


