Robust Inference for Inverse Stochastic Dominance

Francesco Andreoli

Robust Inference for Inverse Stochastic Dominance

Francesco ANDREOLI

Living Conditions Department, Luxembourg Institute of Socio-Economic Research (LISER), L-4366
Esch-sur-Alzette/Belval Campus, Luxembourg (*francesco.andreoli@liser.lu*)

The notion of inverse stochastic dominance is gaining increasing support in risk, inequality, and welfare analysis as a relevant criterion for ranking distributions, which is alternative to the standard stochastic dominance approach. Its implementation rests on comparisons of two distributions' quantile functions, or of their multiple partial integrals, at fixed population proportions. This article develops a novel statistical inference model for inverse stochastic dominance that is based on the influence function approach. The proposed method allows model-free evaluations that are limitedly affected by contamination in the data. Asymptotic normality of the estimators allows to derive tests for the restrictions implied by various forms of inverse stochastic dominance. Monte Carlo experiments and an application promote the qualities of the influence function estimator when compared with alternative dominance criteria.

KEY WORDS: Income distribution; Influence function; Inequality; Lorenz curve; Nonparametric methods; Welfare.

1. INTRODUCTION

One objective of social welfare analysis is to define and implement criteria to rank distributions representing, for instance, risky prospects or realizations of income, consumption, and wealth. The major challenge is to define dominance criteria that are robust to the choice of the evaluation method adopted to assess and compare the degree of risk or inequality implied by the distributions. Motivated by welfare concerns, the literature has brought about the notion of standard stochastic dominance, which has become a workhorse in empirical distribution analysis as statistical inference techniques have been made increasingly available (Bishop, Chakraborti, and Thistle 1989; Anderson 1996; Davidson and Duclos 2000; Barrett and Donald 2003; Linton, Maasoumi, and Whang 2005).

The notion of inverse stochastic dominance (ISD) introduced by Muliere and Scarsini (1989) identifies criteria for ranking distributions that are distinct from standard stochastic dominance. The ISD is linked to the rank-dependent (nonexpected) utility approach pioneered by Yaari (1987), which resolves important paradoxes in decision theory (Quiggin 1993), and has natural applications in finance (Wang and Young 1998), as well as in inequality (Yaari 1988; Zoli 2002) and welfare analysis (Sen 1974; Aaberge 2009; Aaberge, Havnes, and Mogstad 2013). Connections with ISD are also found in program evaluation studies based on quantile regression methods (Firpo, Fortin, and Lemieux 2009; Andreoli, Havnes, and Lefranc 2014).

The most demanding notion of ISD, denoted ISD at order one or *rank dominance* (Saposnik 1981), holds whenever the *Pen's Parade* chart (i.e., the quantile function graph) of the dominant distribution lies nowhere below, and at least some point above, the corresponding chart of the dominated distribution. Weaker criteria of ISD involve comparisons of multiple partial integrals of the distributions' quantile functions at every abscissa corresponding to a population proportion. For instance, ISD at order two is implemented by comparing the integrals of the quantile functions, taken over the space of population proportions.

These give the *generalized Lorenz curves* of the distributions (Gastwirth 1971), with all the welfare implications that generalized Lorenz dominance bears (Kolm 1969; Atkinson 1970; Shorrocks 1983). Further integrations yield conditions for even weaker forms of ISD, which guarantee robust, yet more conclusive, welfare rankings of the distributions. Despite economic and statistical advantages of ISD over standard stochastic dominance, there is limited scope for ISD in empirical analysis, as a consistent statistical framework for testing ISD is still missing.

This article develops a statistical inference model for testing ISD at higher orders. The methodology consists in calculating the ordinates of the recursive integrals of the generalized Lorenz curves for the full set of abscissae implied by the empirical distribution functions, and then using these estimates to compute tractable formulations of the covariances between ordinates of these curves. The implementation of this approach is based on the linear decomposition of the ISD estimators into their *influence functions*, measuring the extent to which the estimator is influenced by an infinitesimal amount of contamination in the data (Cowell and Victoria-Feser 2002; Barrett and Donald 2009). This is done by (i) linearly decomposing the sample estimator of the multiple partial integrals of the generalized Lorenz curves evaluated at various population proportions into the corresponding influence functions; (ii) retrieving an empirically tractable estimator of these influence functions; (iii) obtaining analytical expressions of the underlying asymptotic covariances, which depend on the estimated influence functions. Standard algebra shows that asymptotic normality is granted at $\sqrt{n}$ convergence rate. Thus, Wald-type joint test statistics can be used to test the hypothesis of equality, ISD, and lack of dominance at any order.

There are proposals of estimators for ISD at order one and two that are model-free (Beach and Davidson 1983; Bishop,

Chakraborti, and Thistle 1989) and/or robust to the complex design of the sample (Kovacevic and Binder 1997; Zheng 1999; Zheng 2002). Aaberge (2006) proposed related estimators in a model-dependent context. The influence function approach yields, instead, model-free estimators for ISD at any order. For instance, the covariance estimators for quantile and generalized Lorenz curves proposed by Beach and Davidson (1983) are shown to coincide with the influence function estimators for ISD at order one and two, respectively, if simple random sampling is assumed.

The ISD criteria at order one and two yield equivalent welfare implications as first-degree and second-degree stochastic dominance. The inference strategies proposed in the literature for these forms of standard stochastic dominance rest, nevertheless, on comparisons at fixed population proportions (Beach, Davidson, and Slotsve 1994; Abadie 2002; Barrett, Donald, and Bhattacharya 2014), as required by ISD implementation.

To rank distributions when their generalized Lorenz curves cross it is wise to choose criteria that are finer than generalized Lorenz dominance, but still consistent with it, that is, that are implied by ISD at order two. Extending comparisons to ISD at order three and above is an interesting possibility, since the order of ISD identifies restrictions on the class of rank-dependent evaluation functions upon which welfare dominance is evaluated (Maccheroni, Muliere, and Zoli 2005). Refinements of standard stochastic dominance (Fishburn 1976; Fishburn and Vickson 1978; Le Breton and Peluso 2009) yield welfare implications that are distinct, although equally valid from a normative standpoint, from those of ISD.

Empirical analysis seems, nonetheless, to advocate for the use of the ISD criterion. In fact, when two generalized Lorenz curves cross, the empirical practice is to give priority to welfare considerations implied by one or few inequality indicators, rather than resorting to finer dominance criteria. Almost ubiquitously, the focus is on the *Gini inequality coefficient* and/or on its *single-parameter S-Gini* extensions (Weymark 1981; Donaldson and Weymark 1983), which are strictly connected to the generalized Lorenz curve (Yitzhaki 1982) and for which asymptotic results are available (Barrett and Pendakur 1995; Barrett and Donald 2009). The ranking of distributions based upon the Gini index, however, is not necessarily consistent with standard stochastic dominance at orders higher than the second (Newbery 1970; Dardanoni and Lambert 1988). Rather, if the generalized Lorenz curve of the distribution with higher Gini coefficient intersects the generalized Lorenz curve of another distribution from above, it is always the case that the former inverse stochastic dominates at order three the latter (Zoli 1999).

Additional motivations supporting the ISD model are illustrated in the rest of the article, which is organized as follows. The normative background of ISD is examined in Section 2. The asymptotic results for the influence function approach for ISD are developed in Section 3, where it is also shown that, under reasonable assumptions, the influence functions estimates are bounded. The proposed estimator is therefore robust, in the sense that its value cannot drastically change with an infinitesimal amount of contamination in the data (Cowell and Victoria-Feser 2002). Extensions to the complex survey design setting are discussed in Section 3.4. Section 4 develops joint tests for various forms of ISD. The properties of these tests are investi-

gated in the context of a Monte Carlo study (Section 5), based on parametric models for distributions of disposable income and durable consumption in the U.S., and in an empirical evaluation of equality of opportunity in France (Section 6), where higher order ISD relations are tested to uncover patterns of unfair advantage in labor market outcomes of workers with heterogenous backgrounds of origin. In both assessments, the influence function estimator's performances are compared with the bootstrap estimator for ISD and with standard stochastic dominance tests (Davidson and Duclos 2000; Barrett and Donald 2003). The simulation study, in particular, supports the use of the influence function estimator for ISD as the most powerful testing criterion when the sample size is not too small, and especially when data are contaminated. Section 7 concludes.

2. ROBUST WELFARE ANALYSIS AND ISD

Let Y be a random variable with cumulative distribution function (c.d.f.) F and inverse (quantile) distribution function $F^{-1}(p) = \inf\{y \in \mathbb{R}_+ : F(y) \geq p\}$, for $p \in [0, 1]$. The distribution F represents, for instance, the distribution of income in the population. Following Gastwirth (1971), the integral function $GL^F(p) = \int_0^p F^{-1}(t)\,dt$ defines the generalized Lorenz curve (GL) of F. Denote $\Lambda_2^F(p) = GL^F(p)$, the recursive integrals of this function in the space of population proportions $p \in [0, 1]$ give $\Lambda_k^F(p) = \int_0^p \Lambda_{k-1}^F(t)\,dt$, for any $k \geq 2$. Integration by part reveals that Λ_k^F is a linear transformation of the distribution's quantiles:

$$\Lambda_k^F(p) = \frac{1}{(k-2)!} \int_0^p (p-t)^{k-2} F^{-1}(t)dt. \qquad (1)$$

The relations between the quantile function $(k = 1)$ and its recursive integrals at orders $k = 2$ and $k = 3$ are represented in Figure 1. At order one, the so-called Pen's Parade "of dwarfs and giants incomes" of F gives the level of income attained by the poorest $p100\%$ of the population. At order two, the generalized Lorenz curve coordinates correspond to the expected income attained by the poorest $p100\%$. At order three, the integral of
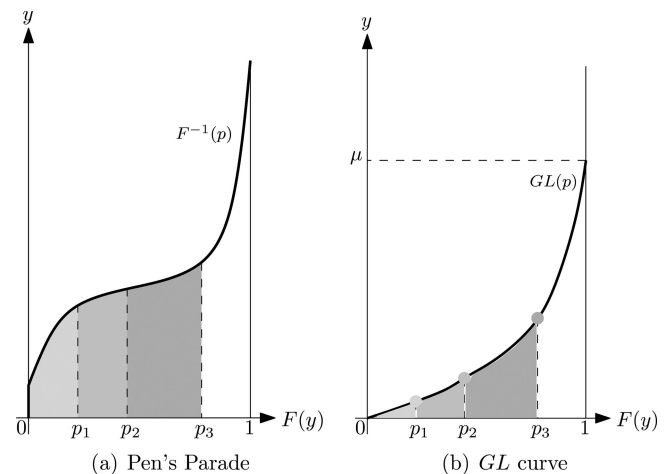

Figure 1. The quantile function curve and the generalized Lorenz curve, along with its integrals. Gray dots on the curve in panel (b) correspond to the value of the integrals of the curves in panel (a), computed at different population shares p_1, p_2, and p_3.

the generalized Lorenz curve trades-off expectation of income attained by the poorest $p100\%$ with income inequality. As the order of integration grows, more weight is given to inequality among the poor.

Muliere and Scarsini (1989) defined a situation where the graph of Λ_k^F lies nowhere below the graph of Λ_k^G, that is, $\Lambda_k^F(p) \geq \Lambda_k^G(p)$, $\forall p \in [0, 1]$, as F *inverse stochastic dominates* G at some order k, denoted $F \succcurlyeq_{\text{ISD}k} G$. The ISD1 criterion stands for rank dominance (Saposnik 1981), while ISD2 denotes generalized Lorenz dominance (Shorrocks 1983). Although ISD at orders one and two provide equivalent normative implications as standard first- and second-degree stochastic dominance, the two criteria do not coincide at higher orders of dominance.

The ISD relation induces a partial order of distributions parameterized by k, which has a normative appeal for social welfare analysis. It is in fact related to the Yaari's (1987) rank-dependent representation of social evaluation functions (SEF). Any SEF W in the broadest class of rank-dependent SEF, denoted $\mathcal{R}$, can be written as a weighted average of realizations:

$$W(F) = \int_0^1 w(p)F^{-1}(p)dp,$$

where the weighting (distortion) function $w(p)$ is such that $w(p) \geq 0$, $\forall p \in [0, 1]$, and $\widetilde{w}(p) = \int_0^p w(t)dt \in [0, 1]$ with $\widetilde{w}(1) = 1$. When $w(p) = 1$ for all p, the SEF is simply the expectation of F. Otherwise, $w(p)$ provides a distortion of the probability p of observing an income lower than $F^{-1}(p)$, thus incorporating value judgements about the role of low and high income realizations on overall welfare. Restrictions on these evaluations have been introduced in the form of assumptions on the alternate sign of high-order derivatives of $\widetilde{w}(p)$ (Maccheroni, Muliere, and Zoli 2005; Aaberge 2009). The order of these derivatives, denoted by an integer $k \geq 2$, defines subsets of $\mathcal{R}$. So, if $\mathcal{R}^k$ gathers all SEF in $\mathcal{R}$ where k assumptions have been made on the sign of the derivatives of the weighting function up to order k, and $\mathcal{R}^l$ is instead obtained by making $l > k$ assumptions on the first l derivatives of the weighting function, then $\mathcal{R}^l \subset \mathcal{R}^k \subset \mathcal{R}$.

Muliere and Scarsini (1989) provided the normative foundation for ISD, by showing that $F \succcurlyeq_{\text{ISD}k} G$ if and only if $W(F) \geq W(G)$ for all $W \in \mathcal{R}^k$. Since ISDk refines ISD2, ISDk is always consistent with generalized Lorenz dominance and $F \succcurlyeq_{\text{ISD}k} G$ implies $F \succcurlyeq_{\text{ISD}l} G$, for all $l > k$, but not the reverse. As a consequence, when GL curves cross and second-degree stochastic dominance is rejected, refinements are still possible by studying ISD3, which gathers agreement about the preferred distribution among all SEF in $\mathcal{R}^3$.

The ISD3 criterion is also coherent with the empirical practice, which, in the presence of intersecting GL curves, seems to focus on ordering distributions through the Gini inequality coefficient. The Gini coefficient is a member of a particular family of SEF in $\mathcal{R}^k$, the Generalized Single Parameter S-Gini SEF (Donaldson and Weymark 1983). Denote $W_k(p, F)$ as one of these SEF, measuring social welfare for the poorest $p100\%$ of the population in F, such that

$$W_k(p, F) := k \int_0^p (p - t)^{k-1}F^{-1}(t)dt. \quad (2)$$

The SEF related to the Gini coefficient is simply $W_2(1, F)$ (Yaari 1988). Hence, on the one hand, $F \succcurlyeq_{\text{ISD}3} G$ implies that the Gini coefficient of F is smaller than that of G (when both distributions have the same mean income), while, on the other hand, Gini coefficients computed at every proportion p of the poorest population in both distributions can be used to establish $F \succcurlyeq_{\text{ISD}3} G$ (Zoli 1999). Using the fact that $W_{k-1}(p, F)$ in (2) is proportional to (1) by a factor $1/(k - 1)!$, it is possible to show that the relation between inequality measurement and ISD extends to any order $k \geq 3$.

Remark 1. $F \succcurlyeq_{\text{ISD}k} G$ if and only if $W_{k-1}(p, F) \geq W_{k-1}(p, G)$, $\forall p \in [0, 1]$.

This result reveals a clear parallel between the way in which comparisons of ISD and of standard stochastic dominance at orders three and above are implemented, although the two concepts remain distinct beyond order two. In fact, standard stochastic dominance is implemented by checking that, at every poverty line, the dominant distribution displays less poverty (as measured by the Foster, Greer, and Thorbecke (1984) poverty index) than the dominated one. Davidson and Duclos (2000) exploited poverty gaps from predetermined income thresholds to produce tests of stochastic dominance. In this vein, I propose tests for ISDk making use of estimates of Λ_k^F and Λ_k^G, for any pair of distributions F and G, at selected populations proportions. Since the asymptotic results for the proposed estimator crucially depend upon its influence functions decomposition, I call it the *influence function (IF) estimator* for ISDk.

3. ESTIMATORS FOR ISDK AND LIMITING DISTRIBUTIONS

3.1 The Influence Function IF Estimator

The derivation of the IF estimator relies on a key result in distributional theory. Denote $T(\widehat{H})$ a scalar valued functional of some empirical process $\widehat{H}$ that is defined on $[0, 1]$. Using Hadamard differentiability of the linear functional T, it can be shown that

$$\sqrt{n}\left(T(\widehat{H}) - T(H)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n T_H'\left(\phi_i(., \widehat{H})\right) + o(1), \quad (3)$$

where the iid random variables $\phi_i(., \widehat{H})$ are referred to as the *influence functions* of $\widehat{H}$, and give the effect of an observation i on the estimator $\widehat{H}$ of the underlying process H. The value of the influence function at $p \in [0, 1]$ is $\phi_i(p, \widehat{H})$ (for references, see Barrett and Donald 2009).

Consider a sequence of realizations $y_1, \ldots, y_i, \ldots, y_n$ of n-independent random variables identically distributed as F. Denote the inverse of F as F^{-1} and its generalized Lorenz curve as GL^F. Their empirical counterparts are denoted as $\widehat{F}, \widehat{F}^{-1}$, and $\widehat{\text{GL}}^F$, respectively. The empirical counterpart of Λ_k^F in (1) is denoted as $\widehat{\Lambda}_k^F$. To avoid cumbersome notation, the explicit reference to F in the superscript is dropped, unless disambiguation is needed.

The estimator $\widehat{\Lambda}_k$ is tied to $\widehat{GL}$ by the following integral function:

$$\widehat{\Lambda}_k(p) \,=\, \frac{1}{(k-3)!} \int_0^p (p-t)^{k-3} \widehat{GL}(t)dt. \qquad (4)$$

Since the generalized Lorenz curve is a continuous linear map defined on the space of population shares, it can always be decomposed into its influence functions as in (3). This follows by the fact that $\widehat{GL}$ is a linear transformation of the quantile function, so the process $\sqrt{n}\left(\widehat{GL}(p) - GL(p)\right)$ consists in a linear transformation of the Bahadur's (1966) representation of quantiles. By setting $\widehat{\Lambda}_k(p) = T(\widehat{H})$ where $\widehat{H} = \widehat{GL}$, also the process $\sqrt{n}\left(\widehat{\Lambda}_k(p) - \Lambda_k(p)\right)$ can be represented as a sum of iid variables plus a residual term that vanishes asymptotically:

$$\sqrt{n}\left(\widehat{\Lambda}_k(p) - \Lambda_k(p)\right) = \frac{1}{(k-3)!} \int_0^p (p-t)^{k-3} \sqrt{n}$$

$$(\widehat{GL}(t) - GL(t))dt$$

$$= \frac{1}{(k-3)!} \int_0^p (p-t)^{k-3} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_i(t, \widehat{GL})\right) dt + o(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{(k-3)!} \int_0^1 (p-t)^{k-3} \phi_i(t, \widehat{GL}) dt + o(1) \qquad (5)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_i(p, \widehat{\Lambda}_k) + o(1). \qquad (6)$$

The n iid random variables $\phi_i(p, \widehat{\Lambda}_k)$ are the influence functions of $\widehat{\Lambda}_k(p)$. Their formula is implicitly given by (5). The influence function measures the impact that a given observation i would have on the estimator if the realized income of i were drawn from the true population distribution F with probability $1 - \varepsilon$ and with probability ε from a distribution $C^{(i)}(y)$, assigning a unit point mass to income y_i (Cowell and Victoria-Feser 2002). The hypothetical contaminated distribution is denoted as $F_\varepsilon^{(i)}(y) = (1 - \varepsilon)F(y) + \varepsilon\, C^{(i)}(y)$, where $C^{(i)}(y) := \mathbf{1}(y \geq y_i)$ and $\mathbf{1}(.)$ is the indicator function. The parameter ε captures the importance of the contamination. For an infinitesimal amount of contamination, define

$$\phi_i(p, \widehat{\Lambda}_k) \,=\, \frac{d}{d\,\varepsilon} \widehat{\Lambda}_k^{F_\varepsilon^{(i)}}(p)\Big|_{\varepsilon \to 0}.$$

An estimator for a stochastic order is robust, that is, its value cannot drastically change with an infinitesimal amount of contamination in the data, if its influence function is bounded. Cowell and Victoria-Feser (1996) had shown that poverty indicators, upon which standard stochastic dominance relations are constructed, are generally robust, while inequality indicators (comprising the Gini index) are generally not. I show that under similar conditions, the ISDk estimator $\widehat{\Lambda}_k(p)$ is robust.

Lemma 1. If the domain of F is bounded from below then $\phi_i(p, \widehat{\Lambda}_k)$ is bounded.

The condition is always satisfied in cases where, for instance, ISD is evaluated over income, wealth, or consumption distributions admitting nonnegative realizations. Analytical results reported in the supplemental Appendix show that the bias due to the contamination vanishes as k grows large, implying that

the asymptotic covariances of the influence functions estimates are also bounded and measured with limited bias in the presence of contaminated data.

The next proposition establishes the asymptotic distributional behavior of the IF estimator for ISDk.

Proposition 1. Suppose that the estimator $\widehat{\Lambda}_k(p)$ for $k = 3, 4, \ldots$ is obtained from a sample of size n drawn from F, which is strictly monotonic and defined over a support that is bounded from below. Then $\sqrt{n}\left(\widehat{\Lambda}_k(p) - \Lambda_k(p)\right)$ converges in distribution to a normal distribution with mean zero and covariance kernel given by

$$\sigma_k(p, p') \,=\, \mathrm{E}\left[\phi_i(p, \widehat{\Lambda}_k)\phi_i(p', \widehat{\Lambda}_k)\right]. \qquad (7)$$

Proof. Given that $\widehat{\Lambda}_k$ is a linear functional transformation of the quantile functions of F, the representation of quantiles by Bahadur (1966) directly implies that $\sqrt{n}\left(\widehat{\Lambda}_k(p) - \Lambda_k(p)\right)$ converges asymptotically to a sum of iid random variables with zero expectations, for any $p \in [0, 1]$. Hence, the central limit theorem applies. If F is bounded from below, the influence functions of $\widehat{\Lambda}_k(p)$ are also bounded, and the process displays finite asymptotic covariance kernels. Noticing that the influence functions are iid random variables, it follows that $\mathrm{E}[\phi_i(p, \widehat{\Lambda}_k)\phi_{i'}(p', \widehat{\Lambda}_k)] = 0$ for any $i \neq i'$ and $\mathrm{E}[\phi_i(p, \widehat{\Lambda}_k)\phi_i(p', \widehat{\Lambda}_k)] = \mathrm{E}[\phi_{i'}(p, \widehat{\Lambda}_k)\phi_{i'}(p', \widehat{\Lambda}_k)]$ for all i, i'. There are n of these equalities, implying that the covariance kernel can be written as

$$\mathrm{E}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_i(p, \widehat{\Lambda}_k) \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_i(p', \widehat{\Lambda}_k)\right]$$

$$= \mathrm{E}\left[\phi_i(p, \widehat{\Lambda}_k)\phi_i(p', \widehat{\Lambda}_k)\right]. \qquad (8)$$

$\square$

It is common practice in empirical analysis to estimate a set of ordinates of the process $\widehat{\Lambda}_k(p)$, corresponding to a set of prechosen m abscissae indexed by $\{p_j | j = 1, \ldots, m\}$ with $0 < p_1 < \cdots < p_m \leq 1$. The implied ordinates can be collected in an $m \times 1$ vector denoted $\widehat{\mathbf{\Lambda}}_k = (\widehat{\Lambda}_k(p_1), \ldots, \widehat{\Lambda}_k(p_m))^\top \in \mathbb{R}_+^m$, with $\mathbf{\Lambda}_k$ being the corresponding vector in the population. The IF decomposition in (5), as well as its robustness properties, extend to vector notation:

$$\sqrt{n}\left(\widehat{\mathbf{\Lambda}}_k - \mathbf{\Lambda}_k\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\phi}_i + o(1), \qquad (9)$$

where $\boldsymbol{\phi}_i = (\phi_i(p_1, \widehat{\Lambda}_k), \ldots, \phi_i(p_m, \widehat{\Lambda}_k))^\top$ and the term $o(1)$ should be understood in the proper dimensionality. It follows from Proposition 1 that the vector $\sqrt{n}(\widehat{\mathbf{\Lambda}}_k - \mathbf{\Lambda}_k)$ is asymptotically m-variate normal with covariance $\mathbf{\Sigma}_k = \mathrm{E}\left[\boldsymbol{\phi}_i \cdot \boldsymbol{\phi}_i^\top\right]$. The element of $\mathbf{\Sigma}_k$ corresponding to abscissae $(p_j, p_{j'})$ is equal to $\sigma_k(p_j, p_{j'})$ in (7). A direct consequence is that, for any $k \geq 3$, $\widehat{\mathbf{\Lambda}}_k$ is asymptotically distributed as $\mathcal{N}(\mathbf{\Lambda}_k, \mathbf{\Sigma}_k/n)$.

3.2 Sample Implementation of the IF Estimator

Consider a sample of size n where realizations are denoted $y_1, \ldots, y_n$. To save notation, the subscript i denotes both an observation and the position it occupies in the ranking of realizations arranged by increasing magnitude, so that

$y_0 \leq y_1 \leq \cdots \leq y_i \leq \cdots \leq y_n$ with y_0 an inferior bound. The weights, often representing the inverse probability of selection from the population, are denoted $\varpi_i \geq 0$ and are indexed according to the sample units: $\varpi_1, \ldots, \varpi_n$. Results are first derived within the simple random sampling structure and then extended to complex survey design.

The empirical c.d.f. $\widehat{F}$ estimated at any point y is a step function with increments $\widehat{\pi}_i = \varpi_i / \sum_i \varpi_i$ associated with each observation. Its ordinates are denoted by

$$\widehat{p}_i \ = \ \sum_{j=1}^{i} \widehat{\pi}_j.$$

If in the sample there are no ties, that is, $y_i < y_{i+1}$ for all i's, then $\widehat{F}(y) = \widehat{p}_i$ for any $y \in [y_i, y_{i+1})$. If there are τ ties in the sample among $y_i = y_{i+1} = \cdots = y_{i+\tau}$, then $\widehat{F}(y) = \widehat{p}_{i+\tau}$ whenever $y_{i-1} < y = y_i = y_{i+1} = \cdots = y_{i+\tau}$. With this representation, it is possible to associate quantiles to observed incomes. The empirical quantile function $\widehat{F}^{-1}(p)$ at population proportion p is

$$\widehat{F}^{-1}(p) \ = \ \sum_{i=1}^{n} y_i \mathbf{1}(\widehat{p}_{i-1} < p \leq \widehat{p}_i), \tag{10}$$

where $\widehat{F}(y_0) = p_0 = 0$. Thus, for $p \in (\widehat{p}_{i-1}, \widehat{p}_i]$ the quantile function takes value y_i and it is well defined even in the case of ties in the sample. Using the fact that $\int_0^1 = \sum_{i=1}^{n} \int_{p_{i-1}}^{p_i}$ with $p_0 = 0$, the sample estimator of the GL curve for $p \in [\widehat{p}_{j-1}, \widehat{p}_j)$ can be defined as follows:

$$\widehat{GL}(p) := \int_0^p \widehat{F}^{-1}(t)dt = \sum_{i=1}^{n} \int_{\widehat{p}_{i-1}}^{\widehat{p}_i} \widehat{F}^{-1}(t)dt\mathbf{1}(\widehat{p}_i \leq p)$$

$$+ y_j(p - \widehat{p}_{j-1})$$

$$= \sum_{i=1}^{n} y_i \int_{\widehat{p}_{i-1}}^{\widehat{p}_i} dt \mathbf{1}(\widehat{p}_i \leq p) + y_j(p - \widehat{p}_{j-1})$$

$$\approx \sum_{i=1}^{n} y_i \widehat{\pi}_i \, \mathbf{1}(\widehat{p}_i \leq p). \tag{11}$$

In a similar way, it is possible to define a consistent estimator of the integrals of the GL curve in (1). The estimator is denoted as $\widehat{\Lambda}_k(p)$. For $p \in [\widehat{p}_{j-1}, \widehat{p}_j)$, it corresponds to

$$\widehat{\Lambda}_k(p) := \frac{(k-1)}{(k-1)!} \left\{ \sum_{i=1}^{n} \int_{\widehat{p}_{i-1}}^{\widehat{p}_i} (p-t)^{k-2} \widehat{F}^{-1}(t)dt\mathbf{1}(\widehat{p}_i \leq p) \right.$$

$$\left. + y_j \int_{\widehat{p}_{j-1}}^{p} (p-t)^{k-2}dt \right\}$$

$$= \frac{1}{(k-1)!} \left\{ \sum_{i=1}^{n} y_i \int_{\widehat{p}_{i-1}}^{\widehat{p}_i} (k-1)(p-t)^{k-2}dt\mathbf{1}(\widehat{p}_i \leq p) \right.$$

$$\left. + y_j \int_{\widehat{p}_{j-1}}^{p} (k-1)(p-t)^{k-2}dt \right\}$$

$$\approx \frac{1}{(k-1)!} \sum_{i=1}^{n} y_i \left\{ (p - \widehat{p}_{i-1})^{k-1} \right.$$

$$\left. - (p - \widehat{p}_i)^{k-1} \right\} \mathbf{1}(\widehat{p}_i \leq p). \tag{12}$$

The approximations in (11) and (12) are asymptotically valid, since the correction terms vanish as the sample size grows. Furthermore, notice that the estimator is linear in incomes, and therefore the presence of ties in the sample does not give rise to computational issues.

The consistent estimator of the asymptotic variance-covariances of $\widehat{\Lambda}_k$, $\Sigma_k / n = \frac{1}{n}E[\phi_i \cdot \phi_i^\top]$, can be obtained by replacing the expectation for its sample counterpart as follows:

$$\widehat{\Sigma}_k = \frac{1}{\sum_i \varpi_i} \sum_{i=1}^{n} \varpi_i \, \phi_i \cdot \phi_i^\top. \tag{13}$$

At abscissae $(p_j, p_{j'})$, the estimator writes:

$$\widehat{\sigma}_k(p_j, p_{j'}) \ = \ \frac{1}{\sum_i \varpi_i} \sum_{i=1}^{n} \varpi_i \, \phi_i(p_j, \widehat{\Lambda}_k)\phi_i(p_{j'}, \widehat{\Lambda}_k).$$

To compute the empirical covariances, it is necessary to estimate the influence functions $\phi_i(p_j, \widehat{\Lambda}_k)$ at every abscissa p_j. The consistent estimators of the influence functions can be obtained by plugging the influence function estimator of the generalized Lorenz curve $\widehat{GL}(p)$, denoted as $\widehat{\phi}_i(p, \widehat{GL})$, into the definition of $\widehat{\Lambda}_k(p)$ in (5). Using the influence function algebra, one can show that

$$\widehat{\phi}_i(p, \widehat{GL}) \ = \ \left(p\widehat{F}^{-1}(p) - \widehat{GL}(p) \right) - \mathbf{1}\left(y_i \leq \widehat{F}^{-1}(p) \right)$$

$$\times \left(\widehat{F}^{-1}(p) - y_i \right), \tag{14}$$

as in Cowell and Victoria-Feser (2002) and Barrett and Donald (2009). Equation (14) defines a model-free estimator of the influence function, meaning that its values can be computed from a sufficiently large sample of observations and no parametric assumptions on the underlying distribution functions have to be made. Plugging (14) into (5) gives the empirical counterpart of the influence function of $\widehat{\Lambda}_k$, which can be further decomposed into three elements as $\widehat{\phi}_i(p, \widehat{\Lambda}_k) = \sum_{h=1}^{3} I_h$, where

$$I_1 = \frac{1}{(k-3)!} \int_0^1 (p-t)^{k-3} t \, \widehat{F}^{-1}(t)\mathbf{1}(t \leq p)dt,$$

$$I_2 = -\frac{1}{(k-3)!} \int_0^1 (p-t)^{k-3} \widehat{GL}(t)\mathbf{1}(t \leq p)dt,$$

$$I_3 = -\frac{1}{(k-3)!} \int_0^1 (p-t)^{k-3} \mathbf{1}\left(y_i \leq \widehat{F}^{-1}(t) \right)$$

$$\times \left(\widehat{F}^{-1}(t) - y_i \right)\mathbf{1}(t \leq p)dt.$$

The empirical estimator of I_h with $h = 1, 2, 3$ can be derived using similar approximations as in (11) and (12). These estimators are asymptotically unbiased and, using integrations by parts, can be written as weighted averages across sample units:

$$\widehat{I}_1 \approx \frac{1}{(k-1)!} \sum_{j=1}^{n} y_j \left\{ (k-1)\Delta_{k-2}(\widehat{p}_j) + \Delta_{k-1}(1) \right\} \mathbf{1}(\widehat{p}_j \leq p),$$

$$\widehat{I}_2 \approx -\frac{1}{(k-1)!} \sum_{j=1}^{n} \left\{ (k-1)\Delta_{k-2}(\widehat{GL}(\widehat{p}_j)) + y_j \Delta_{k-1}(1) \right\} \mathbf{1}(\widehat{p}_j \leq p),$$

$$\widehat{I}_3 \approx -\frac{(k-1)}{(k-1)!} \sum_{j=1}^{n} (y_j - y_i)\mathbf{1}(y_i \leq y_j)\left\{ \Delta_{k-2}(1) \right\}\mathbf{1}(\widehat{p}_j \leq p),$$

where

$$\Delta_\kappa\big(g(\widehat{p}_j)\big) := \big(p - \widehat{p}_{j-1}\big)^\kappa \cdot g(\widehat{p}_{j-1}) - \big(p - \widehat{p}_j\big)^\kappa \cdot g(\widehat{p}_j)$$

with κ a positive integer, and g can be either the estimator of the GL curve (by setting $g(\widehat{p}_j) := \widehat{GL}(\widehat{p}_j)$, which yields $\Delta_\kappa(\widehat{GL}(\widehat{p}_j))$), the estimator of the c.d.f. at income y_j (by setting $g(\widehat{p}_j) := \widehat{p}_j$, which yields $\Delta_\kappa(\widehat{p}_j)$), or a constant (by setting $g(\widehat{p}_j) := 1$, which yields $\Delta_\kappa(1)$).

Replacing the estimator in (11) into $\widehat{I}_h$ leads to the asymptotically consistent estimator of $\widehat{\phi}_i(p, \widehat{\Lambda}_k)$. These estimators must be separately computed for all observations of the sample and for all quantiles implied by the chosen grid. The estimator of the asymptotic covariance $\widehat{\boldsymbol{\Sigma}}_k$ is the weighted sample covariance between the influence function estimates.

3.3 The IF Estimator for ISD1 and ISD2

Under simple random sampling, the validity of Proposition 1 extends to the IF estimators for rank dominance (ISD1) and for generalized Lorenz dominance (ISD2). Hence, the quantiles and generalized Lorenz curves estimators are asymptotically normal with covariance kernels $\sigma_1(p, p') = E\big[\phi_i(p, \widehat{F}^{-1})\phi_i(p', \widehat{F}^{-1})\big]$ and $\sigma_2(p, p') = E\big[\phi_i(p, \widehat{GL})\phi_i(p', \widehat{GL})\big]$, respectively. In this case, the influence function of a quantile of F is

$$\widehat{\phi}_i(p, \widehat{F}^{-1}) = \big(p - \mathbf{1}(\widehat{F}^{-1}(p) \geq y_i)\big) / \widehat{f}(\widehat{F}^{-1}(p)),$$

while the influence function of the generalized Lorenz curve is given in (14). This representation turns out to be related to well-known estimators for rank and generalized Lorenz dominance.

Remark 2. Suppose that for a set of population abscissae $\{p_j | j = 1, \dots m\}$, the distribution function F is continuous and derivable in p_j with density function $f(F^{-1}(p_j)) > 0$ for every j. Then, for $p_j \geq p_{j'}$, $\sigma_1(p_j, p_{j'})$ and $\sigma_2(p_j, p_{j'})$ coincide with the covariance estimators in Beach and Davidson (1983), Lemma 1 and Theorem 1, respectively.

Aside from particular situations spotted in Cowell and Victoria-Feser (2002), which are unlikely to occur in empirical welfare analysis, when the domain of F is bounded from below and from above the influence functions of $\widehat{F}^{-1}$ and of $\widehat{GL}$ are bounded, meaning that the estimators proposed by Beach and Davidson (1983) are robust to contamination. This makes the IF estimator the most natural candidate for extending the robust, model-free estimator for generalized Lorenz dominance to higher-order ISD assessments.

3.4 Implementation With Complex Sampling Design

In empirical analysis, the simple random sampling assumption is highly unrealistic. Most economic data display a stratified, clustered, or multistage design. Assumptions about the sampling method may lead to different estimation procedures for the IF estimator. This is shown, for instance, in Zheng (2002) where consistent estimators for the covariance between ordinates of the Lorenz curve are derived under nonsimple random sampling design. I derive here the consistent estimator for the asymptotic covariance in Proposition 1 under the stratified single-stage sample design, a case also discussed in the empirical exercise.

To extract a stratified single-stage sample, a population is first divided into strata $s = 1, \dots, S$, and then a set of clusters is selected in a simple random manner from each stratum s. There are N_s clusters in stratum s of the population, from which n_s clusters are randomly selected. In a survey design context, each cluster $j = 1, \dots, n_s$ represents a primary sampling unit (PSU) that has a population of size N_{sj}, with $N = \sum_s \sum_j N_{sj}$ the total population. Only n_{sj} observations are drawn from each PSU. An observation is denoted with $i = 1, \dots, n_{sj}$, and has a survey weight of ϖ_{sji}. In the presence of nonsimple sampling design, the random vector $\widehat{\boldsymbol{\Lambda}}_k - \boldsymbol{\Lambda}_k$ is equivalent in distribution to the random vector:

$$\sum_{s=1}^{S} \sum_{j=1}^{n_s} \theta_{sj} \sum_{i=1}^{n_{sj}} \frac{\varpi_{sji}}{N_{sj}} \boldsymbol{\phi}_{sji},$$

where $\theta_{sj} = N_{sj}/N$ and $\boldsymbol{\phi}_{sji}$ is the influence function of the vector $\widehat{\boldsymbol{\Lambda}}_k$ for an observation i in cluster j of stratum s, obtained from the overall distribution. If the PSU are drawn independently both within and across strata, then $\widehat{\boldsymbol{\Lambda}}_k$ preserves its m-variate asymptotic normality, and its asymptotic covariance can be computed from the IF covariances (see Proposition 1).

To estimate the asymptotic covariance, denote the sample estimator of N_{sj} as $\widehat{N}_{sj} = \sum_{i=1}^{n_{sj}} \varpi_{sji}$ and that of the overall population as $\widehat{N} = \sum_{s=1}^{S} \sum_{j=1}^{n_s} \widehat{N}_{sj}$. The asymptotic covariance matrix estimator with nonsimple random sample design of the data is the covariance of the influence function realizations across PSU, multiplied by a correction term, which depends upon the strata and cluster dimensions:

$$\widehat{\boldsymbol{\Sigma}}_k = \sum_{s=1}^{S} \left(\frac{N_s}{n_s} - 1\right) \frac{n_s}{n_s - 1} \sum_{j=1}^{n_s} \widehat{\theta}_{sj}^2 \left[\big(\widehat{\boldsymbol{\phi}}_{sj} - \widehat{\boldsymbol{\phi}}_s\big) \cdot \big(\widehat{\boldsymbol{\phi}}_{sj} - \widehat{\boldsymbol{\phi}}_s\big)^\top\right],$$

where $\widehat{\theta}_{sj} = \widehat{N}_{sj}/\widehat{N}$ and $\widehat{\boldsymbol{\phi}}_{sj}$ is the weighted average of the influence function for each PSU (s) while $\widehat{\boldsymbol{\phi}}_s$ is the sample mean of the influence functions in stratum s, that is,

$$\widehat{\boldsymbol{\phi}}_{sj} = \sum_{i=1}^{n_{sj}} \frac{\varpi_{sji}}{\widehat{N}_{sj}} \widehat{\boldsymbol{\phi}}_{sji}; \quad \widehat{\boldsymbol{\phi}}_s = \frac{1}{n_s} \sum_{j=1}^{n_s} \widehat{\boldsymbol{\phi}}_{sj}.$$

To obtain asymptotic consistency, the sampling design requires both N_s and n_s to be large. Estimation under more complex (multi-stage) sample design are covered in Zheng (2002) and Deville (1999). The extension of the results presented above in that setting is rather direct.

3.5 Alternative Estimators

Alternatively to the IF estimator, I present the *bootstrap estimator* for ISDk. It consists in bootstrapping a sufficiently large number of times the empirical counterpart of $\boldsymbol{\Lambda}_k$ from the original sample. The bootstrap covariance of these parameters is the covariance of the bootstrapped estimates. Let $\mathcal{Y}$ be the original sample of size n drawn from the distribution F. Bootstrap computations are conditional on $\mathcal{Y}$. Let a random sample of size n^* drawn with replacement from $\mathcal{Y}$ be denoted as $\mathcal{Y}^b = \{y_i^b\}_{i=1}^{n^*}$, with empirical distribution $\widehat{F}_b$. It is often the case that $n^* = n$. For every sub-sample $\mathcal{Y}^b$, calculate the estimator $\widehat{\Lambda}_k^{F_b}(p_j)$ for a finite set of m abscissae. By repeatedly drawing random samples from $\mathcal{Y}$, say B times, and calculating for each of the sub-samples

the values taken by $\widehat{\Lambda}_k^{F_b}(p_j)$ at each abscissa, one obtains a $B \times m$ matrix of data. The $m \times m$ empirical covariance matrix computed from these data gives the bootstrap estimator $\boldsymbol{\Sigma}_k^{\mathrm{BS}}/n$, where

$$\frac{\widehat{\sigma}_k^{\mathrm{BS}}(p_j, p_{j'})}{n} := \frac{1}{B-1}\left(\sum_{b=1}^{B} \widehat{\Lambda}_k^{F_b}(p_j)\widehat{\Lambda}_k^{F_b}(p_{j'}) \right.$$
$$\left. -B\, \widehat{\Lambda}_k(p_j)\, \widehat{\Lambda}_k(p_{j'}) \right). \qquad (15)$$

The application of the bootstrap estimator only requires the calculation of a vector of m ordinates of $\widehat{\Lambda}_k^{F_b}$ at every resampling stage, although in general it does not offer a refinement of the asymptotic approximation illustrated in Proposition 1.

Other estimators, connected to the IF approach, have been proposed in the literature. For instance, Aaberge (2006) discussed an estimator for the covariances between the partial integrals of the generalized Lorenz curve taken at different population proportions, and have proposed tests for comparing quantiles of the distributions and implied Gini indices. Aaberge, Havnes, and Mogstad (2013) studied the asymptotic properties of a Kolmogorov–Smirnov test for upward stochastic dominance criteria, which is obtained by representing the quantile function as a Gaussian continuous process. The implementation of these test statistics, however, relies on appropriate estimates of the population density functions. Reliable nonparametric estimators are difficult to obtain when the sample size is relatively small, and additional assumptions over the domain of realizations (such as boundedness) have to be imposed.

4. NULL HYPOTHESIS AND TEST STATISTICS FOR ISDk

For two distributions F and G, the null hypothesis of equality in distributions, that is, $F \succcurlyeq_{\mathrm{ISD}k} G$ *and* $G \succcurlyeq_{\mathrm{ISD}k} F$, is equivalent to test $\Lambda_k^G(p) = \Lambda_k^F(p)$ at every p. The null hypothesis of ISDk, that is, $F \succcurlyeq_{\mathrm{ISD}k} G$, can be tested against an unrestricted one, comprising the nondominance case. The alternative case, placing nondominance at the null, is of interest when the researcher is confident in claiming ISD only when there is strong evidence in its favor. In all these cases, the hypotheses are not formulated on k directly, but rather they postulate that an ISD relation at order k holds between the two distributions.

Following Anderson (1996), Dardanoni and Forcina (1998, 1999), and Davidson and Duclos (2000), I define tests for ISDk based on a finite number of m abscissae. As a consequence, every null hypothesis can be formulated through joint hypothesis on a vector of parameters $\boldsymbol{\Lambda}_k^F - \boldsymbol{\Lambda}_k^G$. Let $\boldsymbol{\Lambda}_k$ be the $2m \times 1$ vector obtained by staking the vectors $\boldsymbol{\Lambda}_k^F$ and $\boldsymbol{\Lambda}_k^G$. Its sample counterpart is denoted as $\widehat{\boldsymbol{\Lambda}}_k$ and is estimated from samples of size n^F and n^G, respectively, where $n = n^F + n^G$ indicates the pooled sample size. The relative size of the samples is denoted by $r^F = n^F/n$ and $r^G = n^G/n$. Let $\mathbf{R} = (\mathbf{I}_m, -\mathbf{I}_m)$ be the $m \times 2m$ *differences* matrix, with $\mathbf{I}_m$ indicating the $m \times m$ identity matrix. Define the parametric vector of differences, $\boldsymbol{\delta}_k \in \mathbb{R}^m$, as $\boldsymbol{\delta}_k = \mathbf{R}\boldsymbol{\Lambda}_k$.

Under the assumption that F and G are generated by independent processes, the asymptotic normality of the IF estima-

tor allows to establish that $\sqrt{n}\,\widehat{\boldsymbol{\delta}}_k = \sqrt{n}\,\mathbf{R}\,\widehat{\boldsymbol{\Lambda}}_k$ is asymptotically distributed as $\mathcal{N}\left(\sqrt{n}\,\mathbf{R}\,\boldsymbol{\Lambda}_k,\,\boldsymbol{\Omega}\right)$ for $k \geq 1$, where $\widehat{\boldsymbol{\delta}}_k$ denotes the sample counterpart of $\boldsymbol{\delta}_k$, and

$$\boldsymbol{\Omega} = \mathbf{R}\,\mathrm{diag}\left(\frac{\boldsymbol{\Sigma}_k^F}{r^F}, \frac{\boldsymbol{\Sigma}_k^G}{r^G} \right)\mathbf{R}^{\top}.$$

An asymptotically valid estimator of $\boldsymbol{\Omega}$, denoted as $\widehat{\boldsymbol{\Omega}}$, is obtained by plugging the empirical counterpart of the influence function estimator in Proposition 1 in place of $\boldsymbol{\Sigma}_k^F$ and of $\boldsymbol{\Sigma}_k^G$.

The null hypotheses are formulated as m linear constraints on $\boldsymbol{\delta}_k$. The *null hypothesis of equality* can be confronted with an unrestricted alternative, indicating the case in which some equalities at given population proportions only occur as a result of intersections.

$$H_0^k: \ \boldsymbol{\delta}_k = \mathbf{0}; \qquad H_1^k: \ \boldsymbol{\delta}_k \neq \mathbf{0}.$$

Under the asymptotic normality of $\widehat{\boldsymbol{\delta}}_k$, the null hypothesis can be assessed by a Wald-type test statistic $T_1^k := n\,\widehat{\boldsymbol{\delta}}_k^{\top}\,\widehat{\boldsymbol{\Omega}}^{-1}\,\widehat{\boldsymbol{\delta}}_k$, which is χ_m^2 distributed. The decision rule can be formulated in terms of p-values as "Reject H_0^k if $\widehat{p}_k < \alpha$."

The *null hypothesis of dominance* consists in comparing $\boldsymbol{\delta}_k$ nonnegative with the unrestricted alternative, that is,

$$H_0^k: \ \boldsymbol{\delta}_k \in \mathbb{R}_+^m; \qquad H_1^k: \ \boldsymbol{\delta}_k \in \mathbb{R}^m.$$

Under the null H_0^k, consider the following test statistic:

$$T_2^k = \min_{\boldsymbol{\delta}_k \in \mathbb{R}_+^m} \left\{ n\,(\widehat{\boldsymbol{\delta}}_k - \boldsymbol{\delta}_k)^{\top}\,\widehat{\boldsymbol{\Omega}}^{-1}\,(\widehat{\boldsymbol{\delta}}_k - \boldsymbol{\delta}_k) \right\}.$$

Under the asymptotic normality of $\widehat{\boldsymbol{\delta}}_k$, Kodde and Palm (1986) showed that T_2^k is asymptotically distributed as a mixture of χ^2 distributions:

$$T_2^k \sim \overline{\chi}^2 := \sum_{j=0}^{m} w\left(m, m-j, \widehat{\boldsymbol{\Omega}}\right) \Pr\left(\chi_j^2 \geq c\right),$$

where $w(m, m-j, \widehat{\boldsymbol{\Omega}})$ denotes the probability that $m-j$ elements of $\boldsymbol{\delta}_k$ are strictly positive. To estimate $w\left(m, m-j, \widehat{\boldsymbol{\Omega}}\right)$, I draw $10,000$ m-variate normal vectors with mean zero and covariance matrix $\widehat{\boldsymbol{\Omega}}$. Then, I compute the proportion of vectors with $m-j$ positive elements. Kodde and Palm (1986) provided a tabulation of the lower (lb_α) and upper (ub_α) bounds of the rejection region for the null H_0^k at standard confidence levels α. The decision rule becomes "Reject (accept) H_0^k if $T_2^k > ub_\alpha$ $(T_2^k < lb_\alpha)$." In all cases in-between, the usual decision rule based on the p-value applies. To test the reverse dominance order, that is, $G \succcurlyeq_{\mathrm{ISD}k} F$, it is sufficient to replace $-\widehat{\boldsymbol{\delta}}_k$ and $-\boldsymbol{\delta}_k$ in the calculation of T_2^k. Tests for $F \succcurlyeq_{\mathrm{ISD}k} G$ against restricted alternatives can be derived from Dardanoni and Forcina (1999), where the asymptotic distributions of tests for equality and strong dominance are compounded.

Finally, the *null hypothesis of nondominance* encompasses cases in which either $G \succcurlyeq_{\mathrm{ISD}k} F$, or the graphs of $\widehat{\Lambda}_k^F(p)$ and $\widehat{\Lambda}_k^G(p)$ intersect. Opposing this hypothesis to an alternative of strong ISDk implies that the researcher is willing to conclude for ISD only if there is a strong evidence in its support, which is a defendable perspective in evaluation studies:

$$H_0^k: \ \boldsymbol{\delta}_k \notin \mathbb{R}_+^m; \qquad H_1^k: \ \boldsymbol{\delta}_k \in \mathbb{R}_+^m.$$

Following Dardanoni and Forcina (1999), the test statistic under H_0^k correspond to a collection of standard-normal distributed statistics $Z_{p_j} = \sqrt{n}(\widehat{\Lambda}_k^F(p_j) - \widehat{\Lambda}_k^G(p_j))/\sqrt{\widehat{\boldsymbol{\Omega}}_{jj}}$, $j = 1, \ldots, m$, where $\widehat{\boldsymbol{\Omega}}_{jj}$ is the estimator of $\boldsymbol{\Omega}$ at abscissae (p_j, p_j). A test statistic can be derived in more compact notation as

$$T_3^k = \min_{p \in \{p_j \mid j=1,\ldots,m\}} \{Z_p\} \sim \mathcal{N}(0, 1).$$

Rejection of the null is based on a unilateral test with critical values taken from the standard normal tabulation. Equivalently, H_0^k is rejected only if the graph of the lower bound of the confidence interval of $\widehat{\boldsymbol{\delta}}_k$ lies entirely above the horizontal axis.

As pointed out by Davidson and Duclos (2000) in the context of standard stochastic dominance analysis, assessing dominance at a finite number of thresholds might rise the problem of test inconsistency. This is true for standard stochastic dominance, where dominance is inferred on the basis of income thresholds potentially gathering no observational mass in the close neighborhood. Even contamination at the bottom tail of a distribution might lead to crossings that extend over a number of income thresholds. The tests for ISDk, instead, allow to perform comparisons at fixed population proportions. In this way, sample coverage is always granted. When the grid is very fine (a parameter that is controlled by the researcher), the ISDk test is likely to measure differences in transformations of realized income at a continuum of population shares in the sample. This is close to the logic underpinning a Kolmogorov–Smirnov statistic for quantiles transformations. Barrett and Donald (2003) studied a consistent test for standard stochastic dominance based on the same principle, but evaluating transformations of the c.d.f. at every income level on a bounded realizations support. Their result relies on the fact that the asymptotic distribution for this test statistic involves a Brownian Bridge process. Using the Bahadur (1966) representation of quantiles, similar methods can be developed in the context of ISD analysis. For instance, Barrett, Donald, and Bhattacharya (2014) had developed consistent tests for assessing Lorenz dominance at every population proportion. The Monte Carlo study hereafter provides intuitions on the effect of increasing the quantity of population shares where ISD is tested. It also builds comparisons with standard stochastic dominance tests.

5. MONTE CARLO RESULTS

The size and power properties of the estimators discussed so far are assessed through a series of Monte Carlo experiments. Each experiment involves tests for ISD at order one, two, and three, the relevant case where standard stochastic dominance analysis and ISD analysis differ. The Monte Carlo experiment provides intuitions on the behavior of the different estimators when the sample size is relatively small, and allows to draw conclusions about the effect of increasing the sample size and of manipulating the number of threshold at which dominance is assessed. The study is based on reliable models of real income and durable consumption distributions in the United States, already validated in the literature.

Various estimators have been used to test the null hypothesis that a distribution F dominates another distribution G at some

order k, versus an unrestricted alternative. The Monte Carlo experiment consists in simulating 1000 independent sample draws from parametric models of F and G. The design of each parametric model is inspired by Barrett and Donald (2009), whereby each simulated draw i of a random variable Y_i at a given simulation stage is generated by a lognormal distribution, that is, $Y_i = \exp(\sigma Z_i + \mu)$, where Z_i is a realization of an $\mathcal{N}(0, 1)$ random variable and (σ, μ) are the dispersion and location parameters. Each experiment involves the simulation of three samples of size 100, 500, and 1500, respectively. For each experiment, the null hypothesis is tested for $k \in \{1, 2, 3\}$ (both in strict dominance and equality forms) using different estimators for the asymptotic covariances, while setting the number of abscissae to $m \in \{5, 10, 20\}$. These abscissae correspond to increments in population proportions of, respectively, 20%, 10%, and 5%. The parameters μ and σ are chosen so that a specific dominance relation holds in the population. For each simulated sample, a series of indicators informing about acceptance or rejection of a given null hypothesis are recorded and then results are reported as averages of these indicators across all Monte Carlo iterations.

Three cases are investigated here. In the first case, I evaluate the *size and power* of the ISD3 test based on the IF estimator, and I compare it to the behavior of the bootstrap estimator. The first objective is to check the size of the tests by recording the proportion of simulated draws where the null is rejected by the data at a nominal size of 5%, knowing that the null is true and cannot be rejected when tested on the population (using one million observations). I consider $F \succcurlyeq_{\text{ISD3}} G$ to be the null hypothesis. Following Barrett and Donald (2009), I assume that each income draw from F is representative of the gross individual-equivalent income in the United States from March 1998 CPS data, where $\mu_F = 9.85$ and $\sigma_F = 0.6$. I consider instead that the data drawn from G are generated using $\mu_G = 9.85$ and $\sigma_G = 0.7$. The difference between F and G approximates the change of gross individual-equivalent income over the 1980s and 1990s. Graphical analysis shows that the generalized Lorenz curves of F crosses that of G from above (hence $F \succcurlyeq_{\text{ISD2}} G$ does not hold). Potentially, F is the distribution yielding higher social welfare for all SEF sufficiently averse to inequality.

The second objective of the Monte Carlo study is to check the power of the ISD tests. This is done by recording the proportion of simulated draws in which the null is rejected, knowing that the alternative is true. In this case, I consider two lognormal distributions F' and G' such that $F' \succcurlyeq_{\text{ISD}k} G'$ for some $k > 3$ but not $F' \succcurlyeq_{\text{ISD3}} G'$. Again, the parameterization is as in Barrett and Donald (2009) and gives the most likely guess for the distribution of per-capita nondurable expenditures in the United States over the 1990s. Hence, $\mu_{F'} = 6.37$ and $\sigma_{F'} = 0.48$ while $\mu_{G'} = 6.4$ and $\sigma_{G'} = 0.55$. In this case, the generalized Lorenz curves of F' and G' cross once but ISD3 does not hold in the population.

Detailed results about the size (case 1) and power (case 2) of the ISD tests based on different estimators are reported in Table 1. For relatively small samples, the IF estimator for ISD is shown to have the correct size and power in the baseline case. The size of the tests based on the ISD3-IF estimator are generally smaller than 0.10 when inference is made on five abscissae, while there is no clear pattern of variation with respect to the size of the samples. The size of the dominance test increases

Table 1. Monte Carlo rejection rates for ISD, nominal size 0.05 (baseline)

	$n = 100$				$n = 500$				$n = 1500$			
	ISD1	ISD2	ISD3-IF	ISD3-BS	ISD1	ISD2	ISD3-IF	ISD3-BS	ISD1	ISD2	ISD3-IF	ISD3-BS
Case 1—Size comparisons for the true null H_0: $F \succcurlyeq_{ISD3} G$												
5	0.056	0.062	0.051	0.083	0.157	0.170	0.104	0.147	0.474	0.460	0.077	0.102
10	0.044	0.118	0.240	0.181	0.157	0.240	0.236	0.250	0.531	0.597	0.188	0.198
20	0.022	0.258	0.403	0.281	0.091	0.276	0.296	0.310	0.451	0.648	0.223	0.231
Case 2—Power comparisons for the true null H_0: $F' \succcurlyeq_{ISDk} G'$ for $k > 3$												
5	0.099	0.107	0.115	0.129	0.346	0.330	0.289	0.327	0.773	0.731	0.414	0.444
10	0.103	0.192	0.366	0.244	0.332	0.416	0.502	0.535	0.810	0.872	0.706	0.716
20	0.048	0.366	0.527	0.321	0.235	0.470	0.621	0.597	0.759	0.886	0.745	0.747

NOTE: The Monte Carlo experiments are based on simulations from the lognormal distribution. The specification for each experiment is as follows: case 1—$Y_F = \mathcal{N}(9.85, 0.6)$ versus $\ln Y_G = \mathcal{N}(9.85, 0.7)$; case 2—$\ln Y_{F'} = \mathcal{N}(6.37, 0.48)$ versus $\ln Y_{G'} = \mathcal{N}(6.4, 0.55)$. Each experiment involves 1000 Monte Carlo iterations. For the BS estimator, 200 repetitions are used.

slightly with the number of abscissae. This is not surprising, given that the test becomes more demanding in terms of comparisons while the number of observations is held fixed, leading to higher likelihood of rejection of the true null hypothesis of ISD3. When the number of abscissae is set to 20, there is evidence of the negative association between size of the sample and size of the test, which evolves from 0.403 when the sample size is small (100 observations) to 0.223 when the sample size is larger (1500 observations). The discriminatory power of the test (case 2) is small when the number of abscissae is set to 5, but it grows rapidly to acceptable levels (in general larger than 0.7) when the sample size is of 1500 observations. The inference for ISD3 based on the bootstrap estimator (ISD3-BS) yields very similar results as for the IF estimator. In small samples, the bootstrap estimator shows, nevertheless, somehow larger size and smaller power compared to the IF estimator.

In the second case, I evaluate the size and power of the ISD3 estimators in the presence of data contamination. Results presented in the supplemental Appendix reveal similar patterns as in the baseline scenario, thus showing that the IF estimator is substantially robust to artificial contamination of the data.

In the third and final case, I contrast the size and power of the IF estimator reported in Table 1 with tests for second- and third-degree stochastic dominance. For these cases, the focus is on the tests by Davidson and Duclos (2000) (denoted as DD), implementing comparisons of distributions at 5, 10, and 20 evenly spaced income thresholds, and on the consistent test by Barrett and Donald (2003) (denoted as BD). The BD test is a Kolmogorov–Smirnov statistic of the difference between recursive partial integrals of the c.d.f. taken over a fine grid of the income realizations domain. When evaluated in the population, both tests suggest that F third-degree (but not second-degree) stochastic dominates G, while F' does not stochastic dominate G' neither at second- nor at third-degree. Both tests reject the null of equality at conventional levels of significance.

Table 2 reports the Monte Carlo study results for the DD and the BD tests. The analysis of third-degree stochastic dominance (SD3 in the table) reveals that the size of the SD3-DD estimator is generally larger than the nominal 5%, and that the power is acceptable only when the number of income thresholds is set to 20. The SD3-BD test has the correct size but very low power, generally smaller than 0.32. A comparative analysis reveals that

the IF estimator for ISD3 outperforms the SD3-DD estimator when the sample size is large. Both estimators lead to tests that have larger size than the SD3-BD test, but also substantially larger power. In the supplemental Appendix, I document that these patterns persist even in case of contamination, and that the IF estimator outperforms the DD estimator in the ability of distinguishing a genuine cross in the curves used to assess ISD3 or SD3 (which implies that the underlying distributions can be ranked at some higher order) from a situations where they are statistically indistinguishable. This is an important issue in evaluation studies, where the impossibility of rejecting the equality null hypothesis might prevent the evaluator from further investigating higher order welfare effects.

Tables 1 and 2 are also useful to assess the power of tests taking generalized Lorenz dominance as the null. In this case, the tests for ISD2 and for SD2 have equivalent normative implications but rely on substantially different implementation methods. The former is implemented by checking generalized Lorenz dominance at fixed population ranks, the latter at fixed incomes. Confronting power levels in Table 1, case 1, with the respective records in Table 2, it emerges that the DD estimator is somehow more powerful than the Beach and Davidson (1983) estimator for ISD2, which in turn dominates the BD method when the number of population proportions' abscissae is above 5. This pattern is substantially preserved when contamination is artificially introduced. Again, I find that in samples of size 500 or less, the ISD2 test is substantially more discriminatory in rejecting the false null of statistical equality of generalize Lorenz curves than the alternative SD2 estimators, although this difference vanishes in samples of larger size.

6. ILLUSTRATION: EQUALITY OF OPPORTUNITY IN FRANCE

This section provides an illustrative application of ISD to check for robustness of *equality of opportunity* (EOP hereafter) assessments. The EOP principle for income acquisition posits that differences in the background of origin across individuals should not predict their labor market income prospects, a monetary measure of the individuals' opportunity set.

To formalize this notion, consider the situation where, from an ethical standpoint, the population can be divided into two groups

Table 2. Monte Carlo rejection rates for standard stochastic dominance, nominal size 0.05 (baseline)

	$n = 100$				$n = 500$				$n = 1500$			
	SD2-DD	SD3-DD	SD2-BD	SD3-BD	SD2-DD	SD3-DD	SD2-BD	SD3-BD	SD2-DD	SD3-DD	SD2-BD	SD3-BD
Case 1—Size comparisons for the true null H_0: $F \succeq_{SD3} G$												
5	0.078	0.085	0.109	0.054	0.130	0.129	0.204	0.059	0.262	0.089	0.374	0.058
10	0.120	0.241	0.109	0.054	0.279	0.358	0.204	0.059	0.659	0.328	0.374	0.058
20	0.567	0.527	0.109	0.054	0.473	0.414	0.204	0.059	0.755	0.348	0.374	0.058
Case 2—Power comparisons for the true null H_0: $F' \succeq_{SDk} G'$ for $k > 3$												
5	0.097	0.096	0.155	0.095	0.205	0.164	0.418	0.193	0.420	0.273	0.775	0.322
10	0.535	0.530	0.155	0.095	0.507	0.528	0.418	0.193	0.820	0.666	0.775	0.322
20	0.741	0.653	0.155	0.095	0.874	0.677	0.418	0.193	0.969	0.746	0.775	0.322

NOTE: The Monte Carlo experiments are based on simulations from the lognormal distribution. The specification for each experiment is as follows: cases 1—$\ln Y_F = \mathcal{N}(9.85, 0.6)$ versus $\ln Y_G = \mathcal{N}(9.85, 0.7)$; cases 2—$\ln Y_{F'} = \mathcal{N}(6.37, 0.48)$ versus $\ln Y_{G'} = \mathcal{N}(6.4, 0.55)$. Each experiment involves 1000 Monte Carlo iterations. BD p-values are obtained via 250 simulations of uncorrelated normal vectors evaluated at 50 equally spaced income thresholds. The symbol $\succeq_{SD}$ stands for stochastic dominates.

gathering individuals with either background a or b. These are addressed to as individual circumstances $c \in \{a, b\}$. The income y_i of individual i is, then, the result of the interplay between her circumstances and other components. Building on this setting, Roemer (1998) clarified that under reasonable assumptions the empirical labor income distribution $\widehat{F}_c$ conditional on the circumstances c serves as a valid proxy of the opportunities faced in the labor market by individuals belonging to this group. In this context, Lefranc, Pistolesi, and Trannoy (2009) proposed welfare-based criteria to assess when EOP is satisfied on the data. In their view, EOP prevails if there is no agreement among inequality averse social evaluation functions in preferring a society where incomes are distributed according to F_a rather than F_b (the distributions in the respective sub-populations) or vice-versa. That is, $F_c \succeq_{ISD2} F_{c'}$ for $c \neq c'$, implying that the economic advantage enjoyed by people with circumstances c over c' cannot be easily established. When EOP prevails, the cases $F_c \succeq_{ISD3} F_{c'}$ and, more generally, $F_c \succeq_{ISDk} F_{c'}$ with k arbitrary large are equivalent. However, there is larger agreement on the existence of an unjust advantage of c over c' whenever ISD3 holds rather than when ISDk for $k > 3$ holds. Hence, if EOP is not rejected by the data, ISD3 becomes a natural test for the robustness of the EOP statement coming from the violation of ISD2.

Practically, a test for EOP robustness requires (i) to partition the sample into groups defined by circumstances a and b, (ii) to estimate the groups' specific conditional distributions $\widehat{F}_a$ and $\widehat{F}_b$ and (iii) to check the minimal order k at which it is not possible to reject $F_a \succeq_{ISDk} F_b$ or $F_b \succeq_{ISDk} F_a$ or both. The focus of this section will be limited to ISD3 comparisons.

I make use of the French LFS—Labor Force Survey data (*Enquête Emploi*) provided by INSEE to estimate the labor income prospects of French workers made conditional on their background of origin. The *circumstance a* gathers all French workers whose parents are either non-French or were occupied as manual workers or farmers. The *circumstance b* is instead associated with the middle class parental background, gathering artisans, small entrepreneurs, and nonmanual workers. The residual class, gathering individuals whose father was employed as a white collar, manager, or professional is not considered in this study for expositional sake. The analysis by Lefranc, Pistolesi, and Trannoy (2009), based on the same data, shows

that the opportunity profile associated with this class always dominates at ISD2 the profiles of the other classes.

To estimate the opportunity profiles of the two groups, I make use of monthly labor income realizations observed for relatively homogenous cohorts of individuals born between 1958 and 1962 whose fathers' characteristics are observed. The investigation is restricted to French LFS waves 2004, 2006, 2008, and 2010. Picking up information every 2 years allows to deal with the panel rotation mechanism of the French LFS, that after 2003 is of 1 year and a half (i.e., one-sixth of the sample is replaced every trimester). The pooled estimating sample consists of 2326 French workers (1810 observations are associated with circumstance a and 516 to circumstance b).

Figure 2 reports the Pen's Parades and the generalized Lorenz curves of the two circumstances' income distributions, after that the data have been depurated of years of survey fixed effects. The two income profiles cannot be ranked according to ISD1 or ISD2 since their respective quantile functions and generalized Lorenz curves cross at least once. It seems clear, however, that the poorest workers in group a enjoy a higher advantage than the poorest workers in group b, while the order swaps as soon as the income deciles grow. Although group a is a priori expected to be the most disadvantaged one, due to the parental background characteristics it represents, it turns out that it is ranked as the advantaged one by all social evaluation functions that give enough weight to the poorest realizations. This suggests that the correct dominance relations to be verified on the data are $F_a \succeq_{ISDk} F_b$ for $k = 1, 2, 3$. The statistical behavior of various estimators is studied.

The French LFS data are *areolar*: they are not drawn directly from a selection of households or individuals, but from a selection of geographical areas made up of 20 adjacent households on average. Then, information on earnings for workers aged 15 to 65 within each area is collected in the survey. The clustered sampling scheme of the French LFS is then taken into account when computing the covariance structure of the influence functions estimator for income deciles (10 abscissae). Figure 3 reports the gaps $\widehat{GL}^{F_a} - \widehat{GL}^{F_b}$ and $\widehat{\Lambda}_3^{F_a} - \widehat{\Lambda}_3^{F_b}$ for the two distributions under scrutiny, along with the estimates standard errors. Considering the GL dominance case (panel (a) of the figure), the confidence interval calculated for at least one population proportion falls completely below the horizontal axis,
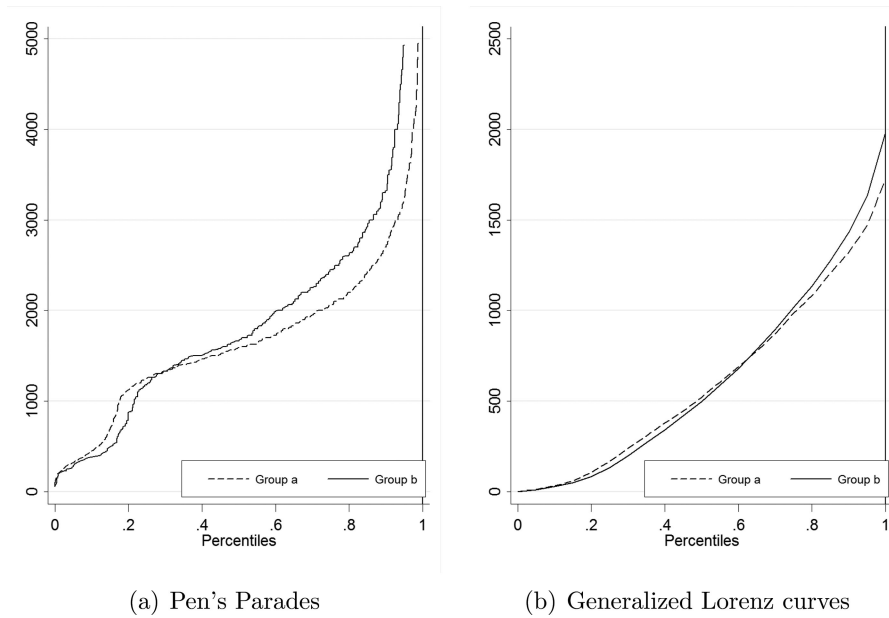
(a) Pen's Parades

(b) Generalized Lorenz curves

Figure 2. ISD1 and ISD2 relations between circumstance *a* and circumstance *b*.
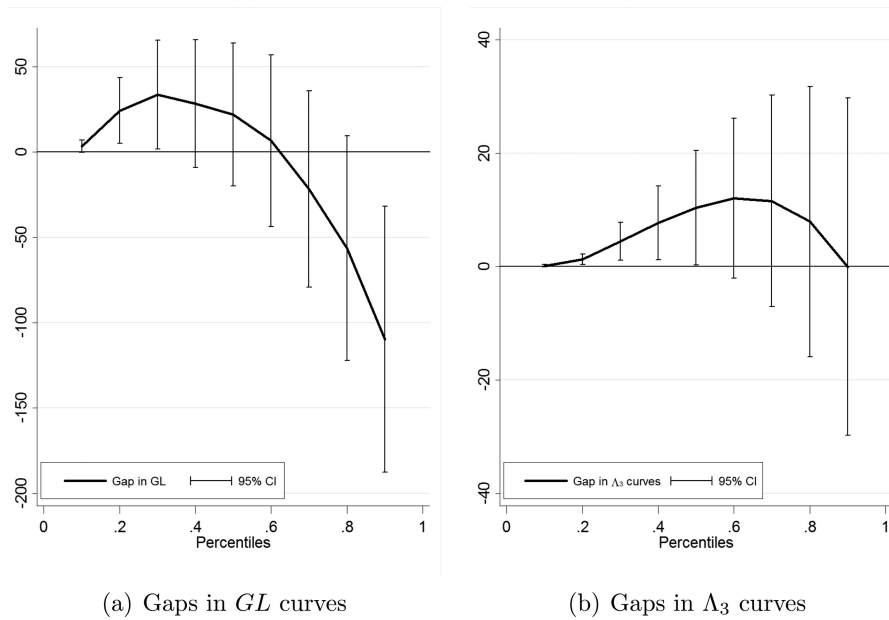

(a) Gaps in GL curves

(b) Gaps in Λ_3 curves

Figure 3. Gaps among GL and Λ_3 curves, with 95% CI based on the IF estimator.

clearly indicating that the null hypothesis of a crossing cannot be rejected. A similar conclusion cannot be drawn immediately from the comparisons reported in panel (b) of the same figure, since the presence of intersections may simply be a symptom of a weak form of dominance.

Table 3 reports income quantiles, generalized Lorenz curves coordinates, Λ_3 coordinates for selected deciles of $\widehat{F}_a$ and $\widehat{F}_b$. The survey design of French LFS is taken into consideration when computing the GL curves standard errors, as well as for the IF and the bootstrap estimator for ISD3. The table shows that, independently of the evaluation method used to infer ISD3, pairwise comparisons of estimated coefficients at fixed deciles lead to inconclusive results: in many cases the differences in Pen's Parades, in generalized Lorenz curves and in integrals of

the generalized Lorenz curves at a given decile are not statistically different from zero, or their signs do not concord across deciles. Joint tests for the equality and dominance null hypotheses are therefore preferred.

The Wald-type test statistics and their simulated *p*-values (T_1^3 for the equality in distributions null hypothesis and T_2^3 for the ISD3 null hypothesis) are reported in Table 4. The joint tests show that rank and generalized Lorenz equality/dominance null hypotheses are rejected at 1% confidence level. Thus, EOP between types *a* and *b* cannot be rejected by the data, since there is no unambiguous advantage of one type compared to the other. To evaluate the robustness of this finding, I investigate ISD comparisons at order three. The IF estimator, leading to the preferred inference strategy, shows that ISD3 and equality

Table 3. Statistics and standard errors for $\widehat{F}_a \succcurlyeq_{\text{ISD}k} \widehat{F}_b$

	Wage, circumstance a (1)		Wage, circumstance b (2)		Difference (1)–(2)
	Coeff.	(SE)	Coeff.	(SE)	
ISD at order 1 (Rank dominance)					
10%	446.3	(22.65)	378.8	(52.33)	67.44
30%	1327.3	(17.32)	1327.3	(52.84)	0.00
50%	1593.9	(16.62)	1666.5	(48.76)	−72.63
70%	1950.4	(23.73)	2253.0	(67.33)	−302.66[*]
90%	2700.0	(53.08)	3319.7	(150.47)	−619.74[*]
ISD at order 2 (Generalized Lorenz dominance)					
10%	31.5	(1.22)	28.0	(1.75)	3.41
30%	230.1	(8.66)	196.5	(16.91)	33.62
50%	522.0	(10.85)	500.1	(22.47)	21.86
70%	871.9	(13.54)	893.6	(31.48)	−21.71
90%	1321.1	(17.52)	1430.8	(42.92)	−109.71[*]
ISD at order 3 (IF estimator)					
10%	1.3	(0.22)	1.2	(0.41)	0.17
30%	24.1	(1.05)	19.6	(2.43)	4.49
50%	98.4	(2.24)	88.0	(5.57)	10.38
70%	236.4	(3.88)	224.9	(10.06)	11.57
90%	453.2	(6.18)	453.2	(16.46)	−0.01
ISD at order 3 (Bootstrap estimator)					
10%	1.3	(0.05)	1.2	(0.09)	0.17
30%	24.1	(1.08)	19.6	(1.66)	4.49[*]
50%	98.4	(2.91)	88.0	(5.34)	10.38
70%	236.4	(5.17)	224.9	(10.20)	11.57
90%	453.2	(8.04)	453.2	(16.63)	−0.01
Mean	1727.0	(31.72)	1984.9	(70.05)	−257.86[*]
Sample size:	1810		516		–

NOTE: French LFS data, waves 2004, 2006, 2008, and 2010. The IF and bootstrap estimators are based on estimates by deciles. Standard errors reported in parenthesis account for the clustering structure of the sample. The bootstrap estimator is based on 200 repetitions. [*] denotes significantly different from zero at 5% on the basis of a two-tailed test.

Table 4. Wald tests and p-values for various dominance relations of F_a over F_b

Null hypothesis	Equality (1)		Dominance (2)	
	Test statistic	p-value	Test statistic	p-value
Rank dominance	55.13	0.000	41.35	0.000
Generalized Lorenz dominance	57.84	0.000	57.05	0.000
ISD at order 3:				
- IF estimator	235.55	0.000	90.91	0.000
- Bootstrap estimator	54.30	0.000	158.20	0.000
Standard stochastic dominance at order 3, various tests:				
- Davidson and Duclos (2000)	50.16	0.000	683.43	0.000
- Barrett and Donald (2003)	.	.	.	0.000

NOTE: French LFS data, waves 2004, 2006, 2008, and 2010. Wald-type tests for equality and dominance based on distributions deciles (for rank dominance, generalized Lorenz dominance, and ISD3) and on 10 evenly spaced income thresholds (for Davidson and Duclos (2000) tests of standard stochastic dominance). The null hypothesis of dominance cannot be rejected (is rejected) at the significance level α if the Wald test is smaller (larger) than the corresponding *lower bound lb_α* (*upper bound ub_α*). For $\alpha = 10\%$, 5%, 1%, the corresponding lower bounds are $lb_\alpha = 1.642$, 2.706, 5.412 and the corresponding upper bounds are $ub_\alpha = 14.067$, 16.274, 20.972 (see Kodde and Palm 1986). In the remaining cases, decisions are based on p-values. Reported p-values of the Barrett and Donald (2009) test for the random variable $\overline{S}_3^{F_a,F_b}$ are simulated.

dominance null hypothesis. Also this criterion allows to reject third-degree stochastic dominance.

The empirical application highlights two facts. The first fact is that the ISD3 test based on the IF estimator produces evaluations that are perfectly aligned with the prediction of Davidson and Duclos (2000) tests. This coherency signals that both dominance criteria are able to detect that the negative gaps between the GL curves of $\widehat{F}_a$ and $\widehat{F}_b$ overcompensate the positive gaps (indicating dominance) at the bottom. The second fact is that the ISD3 test based on predictions at population deciles behaves comparatively as good as the consistent test for stochastic dominance in showing that EOP assessments are robust.

7. CONCLUDING REMARKS

Inverse stochastic dominance is a convenient tool for assessing when there is consensus, in a well-defined class of social evaluation functions, in ranking an income distribution as socially preferred to another. Applications in income (re)distribution analysis, policy evaluation, and risk assessment have clarified the importance of this tool. This article provides estimators that can be used to produce inference for inverse stochastic dominance comparisons, and shows how the restrictions implied by various forms of inverse stochastic dominance can be tested on the data.

The preferred estimator for inverse stochastic dominance is grounded on the influence functions decomposition of the recursive integrals of the generalized Lorenz curve. This methodology gives, as a special case, the estimators for rank and generalized Lorenz dominance developed by Beach and Davidson (1983) and provides their natural extension to less demanding comparisons of distributions at fixed population proportions. Monte Carlo experiments show that the influence function estimator outperforms the alternative bootstrap estimator both in

null hypotheses are rejected by the data at any conventional level of significance. According to the IF method, the equality of opportunity evaluation is robust, since it is not possible, at order three, to determine the advantaged group. Interestingly, the bootstrap estimator (which is computationally less intensive to obtain) leads to similar conclusions as the IF estimator.

Table 4 also displays the evaluations of third-degree standard stochastic dominance, an alternative criterion for assessing the robustness of the EOP evaluation. Wald-type tests based on Davidson and Duclos (2000) estimators under the equality and dominance null hypotheses have been obtained by partitioning the income domain at 10 evenly spaced income thresholds, where poverty levels are evaluated. As shown in the table, the test rejects both dominance and equality at high significance. This is coherent with the predictions from ISD3 tests, implying that testing on a small number of income abscissae still allows to capture the patterns of the gaps in the GL curves represented in Figure 3(a). The consistent test for standard stochastic dominance at order three by Barrett and Donald (2003) leads to similar conclusions. The test is operationalized by partitioning the income domain on a fine grid (100 thresholds) and computing simulated p-values of the Kolmogorov–Smirnov test for the

term of size and power, provided that the relevant estimators are computed for a sufficiently fine grid of abscissae corresponding to population proportions. Furthermore, the size and power of inverse stochastic dominance comparisons based on the influence function estimator are virtually unaffected by the presence of contamination in the data. This is an important feature for empirical income distribution analysis. Experiments involving standard stochastic dominance tests also seem to favor the adoption of the influence function estimator, although these criteria are not comparable from a normative standpoint.

One important caveat to these simulation results is that they are all based on comparisons of pairs of distributions that are generated by independent processes. This is empirically unattractive when, for instance, actual and counterfactual distributions from a policy experiment have to be compared, or when observations are serially correlated. Linton, Maasoumi, and Whang (2005) proposed estimates of Kolmogorov–Smirnov tests' critical values in the context of non-iid samples of correlated prospects. Interesting avenues for future research are to use the influence function methodology for testing ISD over the whole set of population proportions implied by the data, as well as to expand these results to the case where the estimating samples are correlated.

SUPPLEMENTARY MATERIALS

Proofs, simulation results, and Stata routines: A supplemental Appendix collects the proofs of Lemma 1 and Remark 2 and a detailed account of the Monte Carlo experiment results. Stata routines "ISDtest" and "SDtest," reporting code to perform ISD tests and selected standard stochastic dominance tests, along with Monte Carlo simulation routines and French LFS analysis records, are made available. (Zipped archive file)

ACKNOWLEDGMENTS

REFERENCES

Aaberge, R. (2006), "Asymptotic Distribution Theory of Empirical Rank-Dependent Measures of Inequality," in *Advances in Statistical Modeling and Inference—Essays in Honor of Kjell A. Doksum*, ed. V. Nair, Singapore: World Scientific, chapter 24, pp. 495–512. [147,152]

———— (2009), "Ranking Intersecting Lorenz Curves," *Social Choice and Welfare*, 33, 235–259. [146,148]

Aaberge, R., Havnes, T., and Mogstad, M. (2013), "A Theory for Ranking Distribution Functions," IZA Discussion Paper No. 7738. [146,152]

Abadie, A. (2002), "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292. [147]

Anderson, G. (1996), "Nonparametric Tests of Stochastic Dominance in Income Distributions," *Econometrica*, 64, 1183–1193. [146,152]

Andreoli, F., Havnes, T., and Lefranc, A. (2014), "Equalization of Opportunity: Definitions, Implementable Conditions and Application to Early-Childhood Policy Evaluation," IZA Working Paper No. 8503. [146]

Atkinson, T. B. (1970), "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 244–263. [146]

Bahadur, R. R. (1966), "A Note on Quantiles in Large Samples," *The Annals of Mathematical Statistics*, 37, 577–580. [149,153]

Barrett, G. F., and Donald, S. G. (2003), "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 71–104. [146,147,153,154,157]

———— (2009), "Statistical Inference With Generalized Gini Indices of Inequality, Poverty, and Welfare," *Journal of Business & Economic Statistics*, 27, 1–17. [146,147,148,150,153,157]

Barrett, G. F., Donald, S. G., and Bhattacharya, D. (2014), "Consistent Nonparametric Tests for Lorenz Dominance," *Journal of Business & Economic Statistics*, 32, 1–13. [147,153]

Barrett, G. F., and Pendakur, K. (1995), "The Asymptotic Distribution of the Generalized Gini Indices of Inequality," *The Canadian Journal of Economics/Revue canadienne d'Economique*, 28, 1042–1055. [147]

Beach, C. M., and Davidson, R. (1983), "Distribution-Free Statistical Inference With Lorenz Curves and Income Shares," *The Review of Economic Studies*, 50, 723–735. [146,151,154,157]

Beach, C. M., Davidson, R., and Slotsve, G. A. (1994), "Distribution-Free Statistical Inference for Inequality Dominance With Crossing Lorenz Curves," QED Working Paper No. 912. [147]

Bishop, J. A., Chakraborti, S., and Thistle, P. D. (1989), "Asymptotically Distribution-Free Statistical Inference for Generalized Lorenz Curves," *The Review of Economics and Statistics*, 71, 725–727. [146]

Cowell, F. A., and Victoria-Feser, M.-P. (1996), "Robustness Properties of Inequality Measures," *Econometrica*, 64, 77–101. [149]

———— (2002), "Welfare Rankings in the Presence of Contaminated Data," *Econometrica*, 70, 1221–1233. [146,147,149,150,151]

Dardanoni, V., and Forcina, A. (1998), "A Unified Approach to Likelihood Inference on Stochastic Orderings in a Nonparametric Context," *Journal of the American Statistical Association*, 93, 1112–1123. [152]

———— (1999), "Inference for Lorenz Curve Orderings," *Econometrics Journal*, 2, 49–75. [152]

Dardanoni, V., and Lambert, P. (1988), "Welfare Rankings of Income Distributions: A Role for the Variance and Some Insights for Tax Reform," *Social Choice and Welfare*, 5, 1–17. [147]

Davidson, R., and Duclos, J.-Y. (2000), "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica*, 68, 1435–1464. [146,147,148,152,153,154,157]

Deville, J. (1999), "Estimation de Variance Pour Des Statistiques et Des Estimateurs Complexes: Linéarisation et Techniques Des Résidus," *Techniques d'enquête*, 25, 219–230. [151]

Donaldson, D., and Weymark, J. A. (1983), "Ethically Flexible Gini Indices for Income Distributions in the Continuum," *Journal of Economic Theory*, 29, 353–358. [147,148]

Firpo, S., Fortin, N. M., and Lemieux, T. (2009), "Unconditional Quantile Regressions," *Econometrica*, 77, 953–973. [146]

Fishburn, P. C. (1976), "Continua of Stochastic Dominance Relations for Bounded Probability Distributions," *Journal of Mathematical Economics*, 3, 295–311. [147]

Fishburn, P. C., and Vickson, R. G. (1978), "Theoretical Foundations of Stochastic Dominance," "Theoretical Foundations of Stochastic Dominance," in *Stochastic Dominance: An Approach to Decision Making Under Risk*, eds. G. Whitmore and M. Findlay, Lanham, MD: Lexington Books, pp. 39–114. [147]

Foster, J., Greer, J., and Thorbecke, E. (1984), "A Class of Decomposable Poverty Measures," *Econometrica*, 52, 761–766. [148]

Gastwirth, J. L. (1971), "A General Definition of the Lorenz Curve," *Econometrica*, 39, 1037–1039. [146,147]

Kodde, D. A., and Palm, F. C. (1986), "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica*, 54, 1243–1248. [152,157]

Kolm, S.-C. (1969), *The Optimal Production of Social Justice (Public Economics)*, London: McMillan, pp. 145–200. [146]

Kovacevic, M. S., and Binder, D. A. (1997), "Variance Estimation for Measures of Income Inequality and Polarization: The Estimating Equations Approach," *Journal of Official Statistics*, 13, 41–58. [147]

Le Breton, M., and Peluso, E. (2009), "Third-Degree Stochastic Dominance and Inequality Measurement," *The Journal of Economic Inequality*, 7, 249–268. [147]

Lefranc, A., Pistolesi, N., and Trannoy, A. (2009), "Equality of Opportunity and Luck: Definitions and Testable Conditions, With an Application to Income in France," *Journal of Public Economics*, 93, 1189–1207. [155]

Linton, O., Maasoumi, E., and Whang, Y.-J. (2005), "Consistent Testing for Stochastic Dominance Under General Sampling Schemes," *The Review of Economic Studies*, 72, 735–765. [146,158]

Maccheroni, F., Muliere, P., and Zoli, C. (2005), "Inverse Stochastic Orders and Generalized Gini Functionals," *METRON*, 63, 529–559. [147,148]

Muliere, P., and Scarsini, M. (1989), "A Note on Stochastic Dominance and Inequality Measures," *Journal of Economic Theory*, 49, 314–323. [146,148]

Newbery, D. (1970), "A Theorem on the Measurement of Inequality," *Journal of Economic Theory*, 2, 264–266. [147]

Quiggin, J. (1993), *Generalized Expected Utility Theory: The Rank-Dependent Model*, Boston, MA: Kluwer. [146]

Roemer, J. (1998), *Equality of Opportunity*, Cambridge: Harvard University Press. [155]

Saposnik, R. (1981), "Rank-Dominance in Income Distributions," *Public Choice*, 36, 147–151. [146,148]

Sen, A. (1974), "Informational Bases of Alternative Welfare Approaches: Aggregation and Income Distribution," *Journal of Public Economics*, 3, 387–403. [146]

Shorrocks, A. F. (1983), "Ranking Income Distributions," *Economica*, 50, 3–17. [146,148]

Wang, S. S., and Young, V. R. (1998), "Ordering Risks: Expected Utility Theory Versus Yaari's Dual Theory of Risk," *Insurance: Mathematics and Economics*, 22, 145–161. [146]

Weymark, J. A. (1981), "Generalized Gini Inequality Indices," *Mathematical Social Sciences*, 1, 409–430. [147]

Yaari, M. E. (1987), "The Dual Theory of Choice Under Risk," *Econometrica*, 55, 95–115. [146,148]

——— (1988), "A Controversial Proposal Concerning Inequality Measurement," *Journal of Economic Theory*, 44, 381–397. [146,148]

Yitzhaki, S. (1982), "Stochastic Dominance, Mean Variance, and Gini's Mean Difference," *The American Economic Review*, 72, 178–185. [147]

Zheng, B. (1999), "Statistical Inferences for Testing Marginal Rank and (Generalized) Lorenz Dominances," *Southern Economic Journal*, 65, 557–570. [147]

——— (2002), "Testing Lorenz Curves With Non-Simple Random Samples," *Econometrica*, 70, 1235–1243. [147,151]

Zoli, C. (1999), "Intersecting Generalized Lorenz Curves and the Gini Index," *Social Choice and Welfare*, 16, 183–196. [147,148]

——— (2002), "Inverse Stochastic Dominance, Inequality Measurement and Gini Indices," *Journal of Economics*, 9, 119–161. [146]