I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13176

Loss Aversion and the Welfare Ranking of Policy Interventions

Sergio Firpo
Antonio F. Galvao
Martyna Kobus
Thomas Parker
Pedro Rosa-Dias

APRIL 2020

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13176

Loss Aversion and the Welfare Ranking of Policy Interventions

Sergio Firpo
Insper and IZA

Antonio F. Galvao
University of Arizona

Martyna Kobus
Polish Academy of Sciences

Thomas Parker
University of Waterloo

Pedro Rosa-Dias
Imperial College London

APRIL 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Loss Aversion and the Welfare Ranking of Policy Interventions*

In this paper we develop theoretical criteria and econometric methods to rank policy interventions in terms of welfare when individuals are loss-averse. The new criterion for "loss aversion-sensitive dominance" defines a weak partial ordering of the distributions of policy-induced gains and losses. It applies to the class of welfare functions which model individual preferences with non-decreasing and loss-averse attitudes towards changes in outcomes. We also develop new statistical methods to test loss aversion-sensitive dominance in practice, using nonparametric plug-in estimates. We establish the limiting distributions of uniform test statistics by showing that they are directionally differentiable. This implies that inference can be conducted by a special resampling procedure. Since point-identification of the distribution of policy-induced gains and losses may require very strong assumptions, we also extend comparison criteria, test statistics, and resampling procedures to a partially-identified case. Finally, we illustrate our methods with an empirical application to welfare comparison of two income support programs.

JEL Classification: C12, C14, I30

Keywords: welfare, loss aversion, policy evaluation, stochastic ordering, directional differentiability

Corresponding author:
Sergio P. Firpo
Insper Institute of Education and Research
R. Quatá, 300
São Paulo – SP 04546-042
Brazil
E-mail: sergiopf1@insper.edu.br

* The authors are grateful to Pedro Carneiro, Hide Ichimura, Radosław Kurek, Essie Maasoumi, Piotr Miłos, Magne Mogstad, Jim Powell, João Santos Silva, Tiemen Woutersen and seminar participants at the University of California Berkeley, University of Arizona, 28th annual meeting of the Midwest Econometrics Group, 35th Meeting of the Canadian Econometric Study Group, and 3rd edition of the Rio-Sao Paulo Econometrics Workshop for useful comments and discussions regarding this paper. Marta Schoch provided excellent research assistance. Computer programs to replicate the numerical analyses are available from the authors. All the remaining errors are ours.

We suffer more, ... when we fall from a better to a worse situation, than we ever enjoy when we rise from a worse to a better.

Adam Smith, The Theory of Moral Sentiments

1 Introduction

The welfare ranking of policy interventions has classically (Atkinson, 1970) been conducted under the Rawlsian principle of the "veil of ignorance": all policies that produce the same marginal outcome distributions are deemed equivalent for the purpose of welfare analysis. From this perspective, individual gains and losses should be irrelevant (Roemer, 1998; Sen, 2000). However, policies often generate heterogeneous effects, potentially giving rise to gains and losses, which can be consequential for several reasons.

More modern approaches to ranking policy interventions greatly emphasize how different individuals are affected by a policy (Heckman and Smith, 1998; Carneiro, Hansen, and Heckman, 2001). A powerful motivation for this lies in the dynamics of political economy. Public support for a policy, and for the authorities that implement it, depend on the balance of gains and losses experienced by different individuals in the electorate. In addition, and in line with the political economy arguments adduced in Carneiro, Hansen, and Heckman (2001), there is mounting empirical evidence corroborating that the electorate exhibits loss aversion — an empirical regularity that has been identified in a wide variety of other contexts (Kahneman and Tversky, 1979; Samuelson and Zeckhauser, 1988; Tversky and Kahneman, 1991; Rabin and Thaler, 2001; Rick, 2011). This aversion to losses among constituents, in turn, drives the actions of policy makers, as documented in contexts as diverse as government support to the steel industry in US trade policy, and the repeal of the Affordable Care Act (Freund and Özden, 2008; Alesina and Passarelli, 2019). In this paper we develop new testable criteria and econometric methods to rank distributions of individual policy effects from a welfare standpoint when individuals exhibit loss aversion. This extends the toolkit available for the evaluating the impact of policy interventions.

Our first contribution is to propose criteria for ranking policies when agents are averse to losses by using the standard welfare function approach (Atkinson, 1970), namely, that policies may be evaluated based on a welfare ranking. We use a ranking based on social *value* functions, which aggregate individual gains and losses evaluated by a cardinal and interpersonally comparable value function, similarly to standard utilitarian welfare ranking. As is well-known, the latter is equivalent to first-order stochastic dominance (FOSD) over distributions of policy outcomes. In a similar spirit, as the first main contribution, we show that the social value function ranking with non-decreasing and loss-averse value functions (Tversky and Kahneman, 1991) is equivalent to a new concept we call loss aversion-sensitive dominance (LASD) over distributions of policy-induced gains and losses. Recall that FOSD requires that the cumulative distribution function of the dominated distribution lies everywhere above the cumulative distribution of the dominant distribution. In contrast, under LASD it must lie sufficiently above the dominant distribution such that potential losses cannot be compensated by potential gains. This is a consequence of loss-aversion. Except for the special case of a *status quo* policy (i.e. a policy of no change) where FOSD and LASD coincide, generally, as we show, LASD can be used to compare policies that are indistinguishable for FOSD.[1]

The LASD criterion relies on gains and losses, which under certain identification conditions could be considered *treatment effects*. It is well known that the point identification of the distribution of treatment effects may require implausible theoretical restrictions such as rank invariance of potential outcomes (Heckman, Smith, and Clements, 1997). We extend our LASD criteria to a partially-identified setting and establish a sufficient condition to rank alternative policies under partial identification of the distributions of their effects. We use Makarov bounds (Makarov, 1982; Rüschendorf, 1982; Frank, Nelsen, and Schweizer, 1987) to bound the distribution of treatment effects when the joint pre- and post-policy outcome distribution is unknown. This provides a testable criterion that can be used in practice, since the marginal distribution functions from samples observed under various treatments can usually be identified and Makarov bounds only rely on marginal information for their identification.

The second contribution of this paper is to develop statistical inference procedures to practically test the loss averse-sensitive dominance condition using sample data. We develop statistical tests for both point-identified and partially-identified distributions of outcomes. The test procedures are designed to assess, uniformly over the two outcome distributions, whether one treatment dominates another in terms of the LASD criterion. Specifically, we suggest Kolmogorov-Smirnov and Cramér-von Mises test statistics that are applied to nonparametric plug-in estimates of the LASD criterion mentioned above. Inference for these statistics uses specially tailored resampling procedures. We show that our procedures control the size of tests uniformly over probability distributions that satisfy the null hypothesis. Our tests are related to the literature on uniform inference for stochastic dominance represented by, e.g., Linton, Song, and Whang (2010); Linton, Maasoumi, and Whang (2005); Barrett and Donald

[1] The literature on stochastic dominance is vast and spans economics and mathematics - we refer the reader to, e.g., Shaked and Shanthikumar (1994) and Levy (2016) for a review. When dominance curves cross, higher order or inverse stochastic dominance criteria have been proposed. The former involves conditions on higher (typically third and fourth) order derivatives of utility function (e.g. Fishburn (1980), Chew (1983) to which Eeckhoudt and Schlesinger (2006) provided interesting interpretation, whereas the latter is related to the rank-dependent theory originally proposed by Weymark (1981) and Yaari (1987, 1988), where social welfare functions are weighted averages of ordered outcomes with weights decreasing with the rank of the outcome (see Aaberge, Havnes, and Mogstad (2018) for a recent refinement of this theory).

(2003) and references cited therein. We contribute to this literature by developing tests for loss averse-sensitive dominance, which are more complex than the standard stochastic dominance. Hence, our tests widen the variety of comparisons available to empirical researchers to other criteria that encode important qualitative features of agent preferences.

Our econometric approach extends existing stochastic dominance testing procedures. Verifying LASD with sample data presents technical challenges for both the point- and partially-identified cases because the criterion that implies LASD of one distribution over another is more complex than the standard FOSD criterion. The function used to map distribution functions to a testable criterion in the point-identified case is nonlinear and ill-behaved. The criterion function has complex pointwise distributions at each point in its support, but a dominance test inherently requires the uniform comparison over distributions, which demands regularity (in the form of differentiability) of the map between the space of distribution functions and the space of criterion functions. However, as such a map the LASD criterion is not smooth. Despite these complications, we show that supremum- or L_2-norm statistics applied to this function are just regular enough that, with some care, resampling can be used to conduct inference. We rely on recent results from Fang and Santos (2019), who built on the work of Dümbgen (1993), to propose an inference procedure that combines standard resampling with an estimate of the way that test statistics depend on underlying data distributions. We contribute to the literature on directionally differentiable test statistics with a new test for LASD. Recent contributions to this literature include, among others, Cattaneo, Jansson, and Nagasawa (2017); Hong and Li (2018); Chetverikov, Santos, and Shaikh (2018); Cho and White (2018); Christensen and Connault (2019); Fang and Santos (2019) and Masten and Poirier (2020). When distributions are only partially identified by bounds, the situation is more challenging, but the problem has a similar solution. This allows us to conduct conservative inference in the partially identified case. Our contribution to this literature is novel because of our focus on uniform tests for dominance in both the point- and partially-identified cases.

Finally, this paper also relates to the strand of literature that develops methods to estimate the optimal treatment assignment policy that maximizes a social welfare function. Recent developments can be found in Manski (2004), Dehejia (2005), Hirano and Porter (2009), Stoye (2009), Bhattacharya and Dupas (2012), Tetenov (2012), Kitagawa and Tetenov (2018, 2019), among others. These papers focus on the decision-theoretic properties and procedures that map empirical data into treatment choices. In this context, our paper is most closely related to Kasy (2016), which focuses on welfare rankings of policies rather than optimal policy choice.

We empirically illustrate the use of our proposed criteria and tests with a welfare comparison of two well-known income support programs using data from Bitler, Gelbach, and Hoynes

(2006). We show that, in the case of a policy with gainers and losers, the use of our loss aversion-sensitive evaluation criteria may lead to a ranking of policy interventions that differs from that obtained when their outcomes are compared using stochastic dominance.

The rest of the paper is organized as follows. Section 2 presents some basic definitions and notation and defines loss aversion-sensitive dominance. Section 3 develops testable criteria for loss aversion-sensitive dominance. Section 4 proposes statistical inference methods for LASD using sample observations. An empirical application appears in Section 5. Finally, Section 6 concludes. One appendix includes auxiliary results and definitions, and a second appendix collects proof of the results in the text.

2 Loss aversion-sensitive dominance

In this section, we propose a novel dominance relation for ordering policies under the assumption that social decision makers consider the distribution of individual gains and losses under different policy scenarios. We call this criterion Loss Aversion-Sensitive Dominance (LASD).

Suppose a random variable X describes individual gains and losses, and X has cumulative distribution function F, and let $\mathscr{F}$ be the set of cumulative distribution functions with bounded support $\mathcal{X}$. We maintain the assumption throughout that $F \in \mathscr{F}$. The bounded support assumption is made to avoid technical conditions on tails of distribution functions.

A decision maker has preferences over X that are represented via a continuous *social value function* (SVF).

Definition 2.1 (Social Value Function)**.** Suppose random variable X has CDF $F \in \mathscr{F}$ and let $W : \mathscr{F} \to \mathbb{R}$ denote the following *social value function*

$$W(F) = \int_{\mathcal{X}} v(x)\mathrm{d}F(x), \tag{1}$$

where $v : \mathcal{X} \to \mathbb{R}$ is called a *value function.*[2]

The social value function defined above is standard in the literature. $W(F)$ is the expected evaluation of the distribution of X by a decision maker who uses value function $v(\cdot)$. The value function $v(\cdot)$ in (1) need not be any agent's actual value function, but simply the utility function that the social planner uses to convert outcomes into an interpersonally-comparable measure of well-being (Gajdos and Weymark, 2012). We depart from the standard assumptions on

[2]Formally speaking we have $W_v(F)$ but we suppress the subscript v for expositional brevity.

the value function in this paper because it is tailored to the evaluation of gains and losses induced by policies. In particular, we assume that v exhibits the following features: (i) agents assign negative value to losses and positive value to gains, (ii) the value function is monotone (increasing), and our key property – (iii) there is asymmetry in gains and losses, namely, losses hurt an agent more than gains of equivalent magnitude make her happy. These properties are formally listed in the next definition.

Definition 2.2 (Properties of the value function)**.** The value function $v : \mathcal{X} \to \mathbb{R}$ satisfies:

1. Disutility of losses and utility of gains: $v(x) \leq 0$ for all $x < 0$, $v(0) = 0$ and $v(x) \geq 0$ for all $x > 0$.
2. Non-decreasing: $v'(x) \geq 0$ for all x.
3. Loss-averse: $v'(-x) \geq v'(x)$ for all $x > 0$.

The properties in Definition 2.2 are typically assumed in Prospect Theory together with the additional requirement of S-shapedness of value function, which we do not consider (see, e.g., p. 279 of Kahneman and Tversky (1979)). Assumptions 1 and 2 are standard concavity and monotone increasing conditions. Assumption 3 expresses the idea that "losses loom larger than corresponding gains" and is a widely accepted definition of loss aversion (Tversky and Kahneman, 1992, p.303). It is a stronger condition than the one considered by Kahneman and Tversky (1979).

The following form of $W(F)$ will be useful in subsequent definitions and results.

Proposition 2.3. *Suppose that $F \in \mathscr{F}$ and v is once differentiable. Then*

$$W(F) = -\int_{-\infty}^{0} v'(x)F(x)\mathrm{d}x + \int_{0}^{\infty} v'(x)(1 - F(x))\mathrm{d}x. \tag{2}$$

Assume that the decision maker's social value function W depends on v which satisfies Definition 2.2, and she wishes to compare random variables X_A and X_B which represent gains and losses under two policies labeled A and B. We use the labels F_A and F_B for the distribution functions of X_A and X_B. The decision maker prefers X_A over X_B if she evaluates F_A as better than F_B using her SVF — specifically, X_A is preferred to X_B if and only if $W(F_A) \geq W(F_B)$, where W is defined in Definition 2.1. This idea is formalized below.

Definition 2.4 (Loss Aversion-Sensitive Dominance)**.** Let X_A and X_B have distribution functions respectively labeled $F_A, F_B \in \mathscr{F}$. If $W(F_A) \geq W(F_B)$ for all value functions v that satisfy Definition 2.2, we say that F_A dominates F_B in terms of *Loss Aversion-Sensitive Dominance*, or LASD for short, and we write $F_A \succeq_{LASD} F_B$.

In the next section we relate this abstract notion to a more concrete condition that depends on the cumulative distribution functions of the outcome distributions, F_A and F_B.

3 Testable criteria for loss aversion-sensitive dominance

In this section we formulate conditions for evaluating distributions of gains and losses. We propose criteria that indicate whether one distribution of gains and losses dominates another in the sense described in Definition 2.4.

For making comparisons between policies A and B, an econometrician can generally observe three relevant distributions. First, suppose that the control or current distribution of agents' outcomes is represented by the random variable Z_0 which has marginal distribution function G_0. Two other random variables, Z_A and Z_B, describe outcomes under policies A and B. Assume their marginal distribution functions are G_A and G_B respectively. However, a decision maker who is sensitive to loss considers differences induced by these prospective policies. The gains and losses due to policies A and B are defined by the random variables $X_A = Z_A - Z_0$ and $X_B = Z_B - Z_0$. The decision maker's goal is to compare policies A and B using the distribution functions F_A and F_B, the distribution functions of X_A and X_B.

The problem with comparing the variables X_A and X_B is well-known in the treatment effects literature: F_A and F_B depend on the joint distribution of (Z_0, Z_A, Z_B), which may not be observable without restrictions imposed by an economic model. In subsection 3.1 we abstract from specific identification conditions and discusses LASD under the assumption that F_A and F_B are identified. In subsection 3.2 we work with a partially identified case where only the marginal distribution functions G_0, G_A and G_B are observable and no restrictions are made to identify F_A and F_B.

3.1 The case of point-identified distributions

The LASD concept in Definition 2.4 requires that one distribution is preferred to another over an entire class of social value functions and is difficult to test directly. The following result relates the LASD concept to a criterion which depends only on marginal distribution functions and orders F_A and F_B according to the class of SVFs allowed in Definition 2.2. In this section we assume that $F_A, F_B \in \mathscr{F}$ are point identified. This may result from a variety of econometric restrictions that deliver identification and are the subject a large literature.

Theorem 3.1. *Suppose that $F_A, F_B \in \mathscr{F}$. The following are equivalent:*

1. $F_A \succeq_{LASD} F_B$.

2. *For all* $x \geq 0$, F_A *and* F_B *satisfy*

$$F_B(-x) - F_A(-x) \geq \max\{0, F_A(x) - F_B(x)\}. \tag{3}$$

3. *For all* $x \geq 0$, F_A *and* F_B *simultaneously satisfy*

$$F_A(-x) - F_B(-x) \leq 0 \tag{4}$$

and

$$(1 - F_A(x)) - F_A(-x) \geq (1 - F_B(x)) - F_B(-x). \tag{5}$$

Theorem 3.1 provides two different conditions that can be used to verify whether one distribution of gains and losses dominates the other in the LASD sense.[3] These criteria compare the outcome distributions by examining how the distribution functions (F_A, F_B) assign probabilities to gains and losses of all possible magnitudes. The particular way that they make a comparison is related to the relative importance of gains and losses. Consider condition (3). For X_B to be dominated, its distribution function must lie above the distribution of X_A *for losses*. X_B can be dominated by X_A even when gains under A are dominated by those under B — that is, when $F_A(x) - F_B(x) \geq 0$ for some $x \geq 0$ — as long as this lack of dominance in gains is compensated by sufficient dominance of X_A over X_B in the losses region. This is a consequence of the asymmetric treatment of gains and losses. On the other hand, consider conditions (4) and (5). Condition (4) is a FOSD condition applied to losses. This is a consequence of loss aversion; note that in the extreme case where only losses matter, we would have (4). Condition (5) is a tail condition on the distributions. It requires that when balancing the probabilities of gains and losses of absolute magnitudes at least as large as x, X_A provides gains to a higher proportion of agents than does X_B. Inequality (3) combines the two inequalities represented by (4) and (5) into a single equation.

It is interesting to note that LASD has one property in common with FOSD, namely, a higher mean is a necessary condition for both types of dominance.

Corollary 3.2. *If* $F_A \succeq_{LASD} F_B$ *then* $\mathrm{E}[X_A] \geq \mathrm{E}[X_B]$.

Note that FOSD cannot rank two distributions that have the same mean — that is, if $F_A \succeq_{FOSD} F_B$ and $\mathrm{E}[X_A] = \mathrm{E}[X_B]$, then $F_A = F_B$. This is not the case for LASD in (3),

[3] LASD is a partial order. Over losses, (4) is a partial order because FOSD is a partial order. For the tail condition (5) checking transitivity we have $(1 - F_A(x)) - F_A(-x) \geq (1 - F_B(x)) - F_B(-x)$, $(1 - F_B(x)) - F_B(-x) \geq (1 - F_C(x)) - F_C(-x)$, and $(1 - F_A(x)) - F_A(-x) \geq (1 - F_C(x)) - F_C(-x)$. If $F_A(-x) - F_B(-x) = 0$ then $F_A(-x) = F_B(-x)$ and using it in (5) gives anti-symmetry.

as the next example demonstrates. Therefore, for example, when comparing two distributions with the same average effect, equation (3) may still be used to differentiate between them.

Example 3.3. Consider the family of uniform distributions on $[-1-y,-y]\cup[y,y+1]$ indexed by $y>0$ and denote the corresponding member distribution functions F_y. The family of such distributions have mean zero and $F_y \succeq_{LASD} F_{y'}$ whenever $y<y'$. Indeed, note that

$$W(F_y) = \frac{1}{2}\left(\int_{-1-y}^{-y} v(z)\mathrm{d}z + \int_{y}^{1+y} v(z)\mathrm{d}z\right)$$

and thus for any v which is loss-averse (see Definition 2.2) we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}y}W(F_y) &= \frac{1}{2}\left(v(-1-y)-v(-y)+v(1+y)-v(y)\right)\\ &= -\int_{-1-y}^{-y} v'(z)\mathrm{d}z + \int_{y}^{1+y} v'(z)\mathrm{d}z\\ &= \int_{y}^{1+y}\left(v'(z)-v'(-z)\right)\mathrm{d}z \le 0.\end{aligned}$$

It is important to note that LASD is a concept that is specialized to the comparison of distributions that represent gains and losses. Standard FOSD is typically applied to the distribution of outcomes in levels without regard to whether the outcomes resulted from gains or losses of agents relative to a pre-policy state — in our notation, G_A and G_B are typically compared with FOSD, instead of F_A and F_B. FOSD applied to post-policy levels may or may not coincide with LASD applied to changes. This means that even when a strong condition such as FOSD holds for final outcomes, if one took into account how agents value gains and losses it may turn out that the dominant distribution is no longer a preferred outcome. One could apply the FOSD rule to compare distributions of income changes, which implies LASD applied to changes, because FOSD applies to a broader class of value functions. However, this type of comparison would ignore agents' loss aversion, the important qualitative feature that LASD accounts for. The following example shows that the analysis of outcomes in levels using FOSD need not correspond to any LASD ordering of outcomes in changes.

Example 3.4. Let Z_0 represent outcomes before policies A or B. Suppose Z_0 is distributed uniformly over $\{0,1,2,3\}$. Policy A assigns post-policy outcomes depending on the realized Z_0 according to the schedule

$$Z_A = \begin{cases} 3 & \text{if } \{Z_0=0\}\\ 2 & \text{if } \{Z_0=1\}\\ 0 & \text{if } \{Z_0=2\}\\ 1 & \text{if } \{Z_0=3\}.\end{cases}$$

Therefore the distribution of $X_A = Z_A - Z_0$ is $P\{X_A = -2\} = 1/2$, $P\{X_A = 1\} = P\{X_A = 3\} = 1/4$. Meanwhile, policy B maintains the status quo: $X_B = Z_B - Z_0 = 0$ with probability 1.

It is straightforward to check that $Z_A \sim Z_B$ thus they dominate each other according to FOSD. However, there is no loss aversion-sensitive dominance between X_A and X_B. Indeed, we can find two value functions that fulfill the conditions of Definition 2.4 but order X_A and X_B differently. For example, take $v_1(x) = x^3$. Then $3 = \int v_1(x)dF_A(x) > \int v_1(x)dF_B(x) = 0$. Next let $v_2(x) = \text{sgn}(x)|x|^{1/3}$. Then $-0.02 \approx \int v_2(x)dF_A(x) < \int v_2(x)dF_B(x) \approx 0$.

In the previous example, policy B left pre-treatment outcomes unchanged, or in other words, maintained a *status quo* condition — we had $X_B = Z_B - Z_0 \equiv 0$. Suppose generally that X_B has a distribution that is degenerate at 0. Then $F_B(x) = 0$ for all $x < 0$ and $F_B(x) = 1$ for all $x \geq 0$. We define this as a *status quo* policy distribution, labelled F_{SQ}. When comparison is between a distribution F_A and F_{SQ}, LASD and standard FOSD are equivalent.

Corollary 3.5. *Suppose that $F_A \in \mathscr{F}$ and $F_B = F_{SQ}$. Then $F_A \succeq_{LASD} F_{SQ} \iff F_A \succeq_{FOSD} F_{SQ}$.*

Remark 3.6. Although in this paper we focus on the distribution of gains and losses, Kőszegi and Rabin (2006) have developed an interesting preference relation in which individuals derive utility from income and also from gains and losses. In particular, their utility function is additively separable in both gains and losses x and income levels z i.e. $\tilde{v}(x, z) = v_G(x) + v_I(z)$, where $x \in \mathbb{R}$ and $z \in [0, \infty)$. Using Kőszegi and Rabin (2006) preferences, policy A dominates policy B if, in our notation, (4) and (5) are satisfied by X_A and X_B along with the additional condition that Z_A dominates Z_B according to FOSD. A proof of this result is given in Appendix B.

3.2 The case of partially-identified distributions

In many situations of interest the cumulative distribution functions of gains and losses, F_A and F_B, are not point identified without a model of the relationship between X_A and X_B. Without information on the dependence between potential outcomes, we can still make some more circumscribed statements with regard to dominance based on bounds for the distribution functions.

A number of authors have considered functions that bound the distribution functions F_A and F_B. Taking X_A as an example, the *Makarov bounds* (Makarov, 1982; Rüschendorf, 1982;

Frank, Nelsen, and Schweizer, 1987) are two functions L and U that satisfy $L(x) \leq F_A(x) \leq U(x)$ for all $x \in \mathbb{R}$, depend only on the marginal distribution functions G_0 and G_A and are pointwise sharp — for any fixed x there exist some Z_0^* and Z_A^* such that the resulting $X_A^* = Z_A^* - Z_0^*$ has a distribution function at x that is equal one of $L(x)$ or $U(x)$. Williamson and Downs (1990) provide convenient definitions for these bound functions. For any two distribution functions G_1, G_2, define the maps

$$L(x, G_1, G_2) = \sup_{u \in \mathbb{R}} (G_2(u) - G_1(u - x))$$
$$U(x, G_1, G_2) = \inf_{u \in \mathbb{R}} (1 + G_2(u) - G_1(u - x)).$$

For convenience define the policy-specific bound functions for F_k, $k \in \{A, B\}$ and all $x \in \mathbb{R}$, which depend on the marginal CDFs G_0 and G_k, by

$$L_k(x) = L(x, G_0, G_k) \tag{6}$$
$$U_k(x) = U(x, G_0, G_k). \tag{7}$$

Using these definitions we obtain a sufficient and a necessary condition for LASD when only bounds of the treatment effects distribution are observable. The next theorem formalizes the result.

Theorem 3.7. *Suppose that $G_0, G_A, G_B \in \mathscr{F}$ and define the bounding functions using formulas (6) and (7) for $k \in \{A, B\}$.*

1. *If for all $x \geq 0$,*
$$L_B(-x) - U_A(-x) \geq \max\{0, U_A(x) - L_B(x)\} \tag{8}$$
then (3) *holds.*

2. *If* (3) *holds then for all $x \geq 0$,*
$$U_B(-x) - L_A(-x) \geq L_A(x) - U_B(x). \tag{9}$$

Theorem 3.7 is and extension of Theorem 3.1 from the point-identified to the partially-identified case. Both Theorems 3.1 and 3.7 will play important parts in the inference procedures discussed in the next Section.

When the comparison is with the *status quo* distribution, the partially identified conditions simplify. Corollary 3.8 below is an extension of Corollary 3.5 to the partially identified case.

Corollary 3.8. *Suppose that $F_B = F_{SQ}$ and that $G_0, G_A \in \mathscr{F}$. Define the bound functions*

U_A and L_A using formulas (6) and (7). Then $U_A(-x) = 0$ for all $x \geq 0 \Rightarrow F_A \succeq_{LASD} F_{SQ}$ and $F_A \succeq_{LASD} F_{SQ} \Rightarrow L_A(-x) = 0$ for all $x \geq 0$.

4 Inferring loss aversion-sensitive dominance

In this section we propose statistical inference methods for the loss aversion-sensitive dominance (LASD) criteria discussed in previous sections. We consider the null and alternative hypotheses

$$\begin{aligned} H_0 &: F_A \succeq_{LASD} F_B \\ H_1 &: F_A \not\succeq_{LASD} F_B. \end{aligned} \tag{10}$$

Under the null hypothesis (10) policy A dominates B in the LASD sense, similar to much of the literature on stochastic dominance — see, for example, Linton, Maasoumi, and Whang (2005); Linton, Song, and Whang (2010). We use the dominance criteria discussed in Theorems 3.1 and 3.7 to design nonparametric tests for H_0. Because the LASD hypothesis is translated into functional inequalities, which we discuss below, tests must be conducted uniformly over all $x \geq 0$. This uniformity in x and features of the LASD conditions present a challenge for inference.

We consider tests for this null hypothesis given sample data observed under two different identification assumptions. We start with the case where the econometrician can directly observe samples $\{X_{Ai}\}_{i=1}^{n_A}$ and $\{X_{Bi}\}_{i=1}^{n_B}$ which represent agents' gains and losses. In other words, we assume that a model has been imposed on the data so that the distribution functions of X_A and X_B are point-identified and their distribution functions can be estimated using the empirical distribution functions from two samples. Next we extend these results to the partially-identified case where no assumption about the joint distribution of potential outcomes under either treatment is assumed. In this case, the econometrician observes three samples, $\{Z_{0i}\}_{i=1}^{n_0}$, $\{Z_{Ai}\}_{i=1}^{n_A}$ and $\{Z_{Bi}\}_{i=1}^{n_B}$, of outcomes under a control or pre-policy state, and outcomes under policies A and B, and tests are based on plug-in estimates for bounds for $X_A = Z_A - Z_0$ and $X_B = Z_B - Z_0$.

We consider distribution functions as members of the space of bounded functions on the support $\mathcal{X} \subseteq \mathbb{R}$, denoted $\ell^\infty(\mathcal{X})$, equipped with the supremum norm, defined for $f : \mathbb{R}^k \to \mathbb{R}^\ell$ by $\|f\|_\infty = \max_j\{\sup_{x\in\mathbb{R}^k} |f_j(x)|\}$. For real numbers x let $(x)^+ = \max\{0, x\}$. Given a sequence of bounded functions $\{f_n\}_n$ and limiting random element f we write $f_n \rightsquigarrow f$ to denote weak convergence in $(\ell^\infty, \|\cdot\|_\infty)$ in the sense of Hoffman-Jørgensen (van der Vaart and Wellner, 1996).

4.1 Inferring dominance from point identified treatment distributions

In this subsection we suppose that the pair of marginal distribution functions $F = (F_A, F_B)$ is identified.

4.1.1 Test statistics

To implement a test of the hypotheses (10) we employ the results of Theorem 3.1 to construct maps of F into criterion functions that are used to detect deviations from the hypothesis H_0. Specifically, recalling that $(x)^+ = \max\{0, x\}$, for the point-identified case we examine maps $T_1 : (\ell^\infty(\mathbb{R}))^2 \to \ell^\infty(\mathbb{R}_+)$ and $T_2 : (\ell^\infty(\mathbb{R}))^2 \to (\ell^\infty(\mathbb{R}_+))^2$, defined for each $x \geq 0$ by

$$T_1(F)(x) = (F_A(x) - F_B(x))^+ + F_A(-x) - F_B(-x) \tag{11}$$

and

$$T_2(F)(x) = \begin{bmatrix} F_A(-x) - F_B(-x) \\ F_A(x) - F_B(x) + F_A(-x) - F_B(-x) \end{bmatrix}. \tag{12}$$

Functions $T_1(F)$ and $T_2(F)$ are designed so that large positive values will indicate a violation of the null. Taking T_1 as an example, Theorem 3.1 states that $W(F_A) \geq W(F_B)$ if and only if $F_B(-x) - F_A(-x) \geq (F_A(x) - F_B(x))^+$ for all $x \geq 0$, so tests can be constructed by looking for x where $T_1(F)(x)$ becomes significantly positive. We will refer to T_j as maps from pairs of distribution functions to another function space, and also refer to them as functions.

The hypotheses (10) can be rewritten in two equivalent forms, depending on whether one uses T_1 or T_2 to transform distribution functions: letting $\mathcal{X} \subseteq \mathbb{R}_+$ be an evaluation set, we have

$$\begin{aligned} H_0^{(1)} &: T_1(F)(x) \leq 0, \quad \text{for all } x \in \mathcal{X}, \\ H_1^{(1)} &: T_1(F)(x) > 0, \quad \text{for some } x \in \mathcal{X} \end{aligned} \tag{13}$$

and

$$\begin{aligned} H_0^{(2)} &: T_2(F)(x) \leq 0_2, \quad \text{for all } x \in \mathcal{X}, \\ H_1^{(2)} &: T_2(F)(x) \not\leq 0_2, \quad \text{for some } x \in \mathcal{X}. \end{aligned} \tag{14}$$

In the second set of hypotheses 0_2 is a two-dimensional vector of zeros and inequalities are taken coordinate-wise.

The next step in testing the hypotheses (13) and (14) is to estimate $T_1(F)$ and $T_2(F)$. Let $\mathbb{F}_n = (\mathbb{F}_{An}, \mathbb{F}_{Bn})$ denote the pair of marginal empirical distribution functions, that is, $\mathbb{F}_{kn}(x) = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{1}\{X_{ki} \leq x\}$ for $k \in \{A, B\}$. These are well-behaved estimators of the components of

F. Letting $n = n_A + n_B$, standard empirical process theory shows that $\sqrt{n}(\mathbb{F}_n - F)$ converges weakly to a Gaussian process under weak assumptions (van der Vaart, 1998, Example 19.6). In order to conduct inference for loss aversion-sensitive dominance, we use plug-in estimators $T_j(\mathbb{F}_n)$ for $j \in \{1, 2\}$. See Remark A.7 in Appendix A for details on the computation of these functions.

In order to detect when $T_j(\mathbb{F}_n)$ is significantly positive, we consider statistics based on a one-sided supremum norm or a one-sided L_2 norm over $\mathcal{X}$. Kolmogorov-Smirnov (i.e., supremum norm) type statistics are

$$V_{1n} = \sqrt{n} \sup_{x \in \mathcal{X}} (T_1(\mathbb{F}_n)(x))^+ \tag{15}$$

$$V_{2n} = \sqrt{n} \max \left\{ \sup_{x \in \mathcal{X}} (T_{21}(\mathbb{F}_n)(x))^+, \sup_{x \in \mathcal{X}} (T_{22}(\mathbb{F}_n)(x))^+ \right\}. \tag{16}$$

Meanwhile Cramér-von Mises (or L_2 norm) test statistics are defined by

$$W_{1n} = \sqrt{n} \left(\int_{\mathcal{X}} \left((T_1(\mathbb{F}_n)(x))^+ \right)^2 \mathrm{d}x \right)^{1/2}, \tag{17}$$

$$W_{2n} = \sqrt{n} \left(\int_{\mathcal{X}} \left((T_{21}(\mathbb{F}_n)(x))^+ \right)^2 + \left((T_{22}(\mathbb{F}_n)(x))^+ \right)^2 \mathrm{d}x \right)^{1/2}. \tag{18}$$

In the sequel, we assume that all functions used in L_2 statistics are square-integrable.

4.1.2 Limiting distributions

We wish to establish the limiting distributions of V_{jn} and W_{jn}, for $j \in \{1, 2\}$, under the null hypothesis $H_0 : F_A \succeq_{LASD} F_B$. This means that we are concerned with the behavior of the empirical criterion function processes $\sqrt{n}(T_j(\mathbb{F}_n) - T_j(F))$, which are random functions.

Two challenges arise when considering these criterion function processes. First, the form of the null hypothesis as a functional inequality to be tested uniformly over $\mathcal{X}$ is a source of irregularity. The assumption that the distribution P satisfies the null hypothesis $F_A \succeq_{LASD} F_B$ implies that the asymptotic distributions of W_j and V_j depend on features of P. This is referred to as *non-uniformity in* P in (Linton, Song, and Whang, 2010; Andrews and Shi, 2013), and requires attention when resampling.

Second, due to the pointwise maximum function in its definition, T_1 is too irregular as a map from the data to the space of bounded functions to establish a limiting distribution for the empirical process $\sqrt{n}(T_1(\mathbb{F}_n) - T_1(F))$ using conventional statistical techniques. In contrast, T_2 is a linear map of F, which implies that $\sqrt{n}(T_2(\mathbb{F}_n) - T_2(F))$ has a well-behaved limiting distribution in $(\ell^\infty(\mathbb{R}_+))^2$.

Despite the above challenges, we show that V_{jn} and W_{jn} (for $j \in \{1, 2\}$) have well-behaved asymptotic distributions, and furthermore, that the limiting random variables satisfy $V_1 \sim V_2$ and $W_1 \sim W_2$. This is an important result because it is the foundation for applying bootstrap techniques for inference. Before stating the formal assumptions and asymptotic properties of the tests, we discuss the two difficulties mentioned above in more detail.

The limiting distributions of V_{jn} and W_{jn} statistics depend on features of the joint probability distribution of (X_A, X_B), which we denote by P. Let $\mathcal{P}_0$ be the set of distributions P such that $F_A \succeq_{LASD} F_B$. These are distributions with marginal distribution functions F such that $T_j(F)(x) \leq 0$ for all $x \geq 0$. To discuss the relationship between these sets of distributions and test statistics, we relabel the two coordinates of the T_2 function as

$$m_1(x) = F_A(-x) - F_B(-x) \tag{19}$$

and

$$m_2(x) = F_A(-x) - F_B(-x) + F_A(x) - F_B(x). \tag{20}$$

When $P \in \mathcal{P}_0$, both $m_1(x) \leq 0$ and $m_2(x) \leq 0$ for all $x \geq 0$.

More detail is required about the behavior of the two coordinate functions to determine the limiting distributions of V_{jn} and W_{jn} statistics. For L_2-norm statistics W_{1n} and W_{2n}, we define the following relevant subdomains of $\mathcal{X}$, which collect the arguments where m_1 or m_2 are equal to zero:

$$\mathcal{X}_0^1(P) = \{x \in \mathcal{X} : m_1(x) = 0\} \tag{21}$$

$$\mathcal{X}_0^2(P) = \{x \in \mathcal{X} : m_2(x) = 0\}. \tag{22}$$

Denote $\mathcal{X}_0(P) \subseteq \mathcal{X}$ as the set of x where $T_1(F)(x) = 0$ or at least one coordinate of $T_2(F)$ equals 0 for probability distribution P. As will be seen below, $\mathcal{X}_0(P)$ is the same for both the T_1 and T_2 functions. Following Linton, Song, and Whang (2010), we call $\mathcal{X}_0(P)$ the contact set for the distribution P. Given the above definitions, under the null hypothesis we can write

$$\mathcal{X}_0(P) = \mathcal{X}_0^1(P) \cup \mathcal{X}_0^2(P).$$

On the other hand, the supremum-norm statistics V_{1n} and V_{2n} need a different family of sets, namely the sets of ϵ-maximizers of m_1 and m_2. For any $\epsilon \geq 0$ and $k \in \{1, 2\}$, let

$$\mathcal{M}^k(\epsilon) = \left\{ x \in \mathcal{X} : m_k(x) \geq \sup_{x \in \mathcal{X}} m_k(x) - \epsilon \right\}. \tag{23}$$

An important subset of $\mathcal{P}_0$ are those P for which test statistics have nontrivial limiting distributions under the null hypothesis — that is, not degenerate at 0, which occurs when there is some x such that $T_j(F)(x) = 0$ (note that there are no x such that $T_j(F)(x) > 0$ when $P \in \mathcal{P}_0$). Define $\mathcal{P}_{00} \subset \mathcal{P}_0$ to be the set of all P such that $\mathcal{X}_0(P) \neq \varnothing$. If $P \in \mathcal{P}_0 \backslash \mathcal{P}_{00}$ then $\mathcal{X}_0(P) = \varnothing$ and because the distribution satisfies the null hypothesis, F_A strictly dominates F_B everywhere and the criterion functions T_j are strictly negative over $\mathcal{X}$. When $P \in \mathcal{P}_0 \backslash \mathcal{P}_{00}$, test statistics have asymptotic distributions that are degenerate at zero because test statistics will detect that policy A is strictly better that B over all of $\mathcal{X}$. When $P \in \mathcal{P}_{00}$, $T_j(F)$ is zero over $\mathcal{X}_0(P)$ and test statistics have a nontrivial asymptotic distribution over $\mathcal{X}_0(P)$. Thus, when $F_A \succeq_{LASD} F_B$, the asymptotic behavior of test statistics depends on whether $P \in \mathcal{P}_{00}$ or $P \in \mathcal{P}_0 \backslash \mathcal{P}_{00}$. Note that when $P \in \mathcal{P}_{00}$, we have $\lim_{\epsilon \searrow 0} \mathcal{M}^k(\epsilon) = \mathcal{X}_0^k(P)$ for whichever coordinate function actually achieves the maximal value zero.

Hadamard differentiability is an analytic tool used to establish the asymptotic distribution of nonlinear maps of the empirical process. Definition A.1 in Appendix A provides a precise statement of the concept. When a map is Hadamard differentiable — for example T_2, which is linear as a map from $(\ell^\infty(\mathbb{R}))^2$ to $(\ell^\infty(\mathbb{R}_+))^2$ and is thus trivially differentiable — the functional delta method can be applied to describe its asymptotic behavior as a transformed empirical process, and a chain rule makes the analysis of compositions of several Hadamard-differentiable maps tractable. Also, the Hadamard differentiability of a map implies resampling is consistent when this map is applied to the resampled empirical process (van der Vaart, 1998, Theorem 23.9) — so, for example, the distribution of resampled criterion processes $\sqrt{n}(T_2(\mathbb{F}_n^*) - T_2(\mathbb{F}_n))$ is a consistent estimate of the asymptotic distribution of $\sqrt{n}(T_2(\mathbb{F}_n) - T_2(F))$ in the space $\ell^\infty(\mathbb{R}_+)$. On the other hand, consider the T_1 map. The pointwise Hadamard directional derivative of $T_1(f)(x)$ at a given $x \geq 0$ in direction $h(x) = (h_A(x), h_B(x))$ is

$$T'_{1f}(h)(x) = \begin{cases} h_A(x) - h_B(x) + h_A(-x) - h_B(-x), & f_A(x) > f_B(x) \\ (h_A(x) - h_B(x))^+ + h_A(-x) - h_B(-x), & f_A(x) = f_B(x) \\ h_A(-x) - h_B(-x), & f_A(x) < f_B(x) \end{cases}. \tag{24}$$

This map, thought of as a map between function spaces, $(\ell^\infty(\mathbb{R}))^2$ and $\ell^\infty(\mathbb{R}_+)$, is not differentiable because the pointwise maximum map is only differentiable at each point x, but not in the codomain $\ell^\infty(\mathbb{R}_+)$. Despite the lack of differentiability of the map $F \mapsto T_1(F)$, we show in Lemma A.3 in Appendix A that $F \mapsto V_1$ and $F \mapsto W_1$ are Hadamard directionally differentiable, which implies these maps are just regular enough that existing statistical methods can be applied to their analysis. Later in this section we apply the resampling technique recently developed in Fang and Santos (2019) along with this directional differentiability to test hypotheses using V_{1n} or W_{1n}.

Having discussed the difficulties in the relationship between distributions and test statistics, we turn to assumptions on the observations. In order to conduct inference using either $T_1(\mathbb{F}_n)$ or $T_2(\mathbb{F}_n)$ we make the following assumptions.

A1 The observations $\{X_{Ai}\}_{i=1}^{n_A}$ and $\{X_{Bi}\}_{i=1}^{n_B}$ are iid samples and independent of each other and are continuously distributed with marginal distribution functions F_A and F_B respectively.

A2 Let the sample sizes n_A and n_B increase in such a way that $n_k/(n_A + n_B) \to \lambda_k$ as $n_A, n_B \to \infty$, where $0 < \lambda_k < 1$ for $k \in \{A, B\}$. Define $n = n_A + n_B$.

Under these assumptions we establish the asymptotic properties of the test statistics under the null and fixed alternatives. Under the above assumptions, there is a Gaussian process $\mathcal{G}_F$ such that $\sqrt{n}(\mathbb{F}_n - F) \rightsquigarrow \mathcal{G}_F$. We denote each coordinate process $\mathcal{G}_{F_A}$ and $\mathcal{G}_{F_B}$, and for convenience define two transformed processes: for each $x \geq 0$ let

$$\mathcal{G}_1(x) = \mathcal{G}_{F_A}(-x) - \mathcal{G}_{F_B}(-x) \tag{25}$$

$$\mathcal{G}_2(x) = \mathcal{G}_{F_A}(x) - \mathcal{G}_{F_B}(x) - \mathcal{G}_{F_A}(-x) + \mathcal{G}_{F_B}(-x). \tag{26}$$

These will be used in the theorem below.

Theorem 4.1. *Make assumptions **A1**-**A2**. Define the limiting Gaussian processes $\mathcal{G}_1$ and $\mathcal{G}_2$ as above. Then:*

1. *Suppose that $P \in \mathcal{P}_{00}$. As $n \to \infty$, $V_{1n} \rightsquigarrow V_1$ and $W_{1n} \rightsquigarrow W_1$, where*

$$V_1 \sim \max\left\{0, \sup_{x \in \mathcal{X}_0^1(P)} \mathcal{G}_1(x) \cdot \mathbf{1}\left\{\sup_{x \in \mathcal{X}} m_1(x) = 0\right\}, \sup_{x \in \mathcal{X}_0^2(P)} \mathcal{G}_2(x) \cdot \mathbf{1}\left\{\sup_{x \in \mathcal{X}} m_2(x) = 0\right\}\right\}$$

and

$$W_1 \sim \left(\int_{\mathcal{X}_0^1(P)} \left((\mathcal{G}_1(x))^+\right)^2 \mathrm{d}x + \int_{\mathcal{X}_0^2(P)} \left((\mathcal{G}_2(x))^+\right)^2 \mathrm{d}x\right)^{1/2}.$$

2. *Suppose that $P \in \mathcal{P}_{00}$. As $n \to \infty$, $V_{2n} \rightsquigarrow V_2$ and $W_{2n} \rightsquigarrow W_2$, where $V_2 \sim V_1$ and $W_2 \sim W_1$.*

3. *Suppose that* $P \in \mathcal{P}_0 \backslash \mathcal{P}_{00}$ *for* $j = 1$ *or* 2. *As* $n \to \infty$, $P\{V_{jn} > \epsilon\} \to 0$ *and* $P\{W_{jn} > \epsilon\} \to 0$ *for all* $\epsilon > 0$.

4. *Suppose that* $P \notin \mathcal{P}_0$. *As* $n \to \infty$, $P\{V_{jn} > c\} \to 1$ *and* $P\{W_{jn} > c\} \to 1$ *for all* $c \geq 0$ *for* $j = 1$ *or* 2.

Theorem 4.1 derives the asymptotic properties of the proposed test statistics. Parts 1 and 2 establish the weak limits of V_{jn} and W_{jn} for $j \in \{1, 2\}$ when the null hypothesis is true. Recall that when $P \in \mathcal{P}_{00}$, $\lim_{\epsilon \searrow 0} \mathcal{M}^k(\epsilon) = \mathcal{X}_0^k(P)$, which is why $\mathcal{M}^k(\epsilon)$ terms are absent in the first part of the theorem. Remarkably, the test statistics using T_1 and T_2 criterion processes have the same asymptotic behavior despite the different appearances of the underlying processes and the irregularity of T_1. Part 3 shows that the statistics are asymptotically degenerate at zero when the contact set is empty, that is, when P lies on the interior of the null region. Part 4 shows that the test statistics diverge when data comes from any distribution that does not satisfy the null hypothesis.

The limiting distributions described in Part 1 of Theorem 4.1 are not standard because the distributions of the test statistics depend on features of P through the $\mathcal{X}_0(P)$ terms in each expression. Therefore, to make practical inference feasible, we suggest the use of resampling techniques below.

4.1.3 Resampling procedures for inference

The proposed test statistics have complex limiting distributions. In this subsection, we present resampling procedures to estimate the limiting distributions of both V_{jn} and W_{jn} for $j \in \{1, 2\}$ under the assumption that $P \in \mathcal{P}_{00}$. Naive use of bootstrap data generating processes in the place of the original empirical process suffers from distortions due to discontinuities in the directional derivatives of the maps that define the distributions of the test statistics. In finite samples the plug-in estimate will not find, for example, the region where $F_A(x) - F_B(x) = 0$, where the derivatives exhibit discontinuous behavior. Our procedure involves making estimates of the derivatives involved in the limiting distribution and a standard exchangeable bootstrap routine, as proposed in Fang and Santos (2019).[4]

In order to estimate contact sets, define a sequence of constants $\{a_n\}$ such that $a_n \searrow 0$ and $\sqrt{n} a_n \to \infty$ and let $\hat{m}_{1n}(x) = \mathbb{F}_{An}(-x) - \mathbb{F}_{Bn}(-x)$ and $\hat{m}_{2n}(x) = \mathbb{F}_{An}(-x) - \mathbb{F}_{Bn}(-x) +$

[4] Given a set of weights $\{W_i\}_{i=1}^n$ that sum to one and are independent of $\{X_i\}_{i=1}^n$, the exchangeable bootstrap measure is a randomly-weighted measure that puts mass W_i at observed sample point X_i for each i. This encompasses, for example, the standard bootstrap, m-of-n bootstrap and wild bootstrap.

$\mathbb{F}_{An}(x) - \mathbb{F}_{Bn}(x)$. Then for W_j statistics define the estimated contact sets by

$$\hat{\mathcal{X}}_0^1 = \{x \in \mathcal{X} : |\hat{m}_{1n}(x)| \leq a_n\} \tag{27}$$
$$\hat{\mathcal{X}}_0^2 = \{x \in \mathcal{X} : |\hat{m}_{2n}(x)| \leq a_n\}. \tag{28}$$

When both sets are empty, replace both estimates by $\mathcal{X}$. Meanwhile, for V_j statistics define estimated ϵ-maximizer sets. For any sequence of constants $\{b_n\}$ such that $b_n \searrow 0$ and $\sqrt{n}b_n \to \infty$, let

$$\hat{\mathcal{M}}^1(b_n) = \{x \in \mathcal{X} : \hat{m}_{1n}(x) \geq \max \hat{m}_{1n}(x) - b_n\}, \tag{29}$$
$$\hat{\mathcal{M}}^2(b_n) = \{x \in \mathcal{X} : \hat{m}_{2n}(x) \geq \max \hat{m}_{2n}(x) - b_n\}. \tag{30}$$

Using these estimates, the distributions of V_1 and W_1 can be estimated from sample data (recall that Part 2 of Theorem 4.1 asserts that these are the same distributions as those of V_2 and W_2). The formulas in part 3 of the steps below are obtained by inserting estimated contact sets and resampled empirical processes in the place of population-level quantities into the functions shown in Part 1 of Theorem 4.1.

Resampling routine to estimate the distributions of V_{jn} and W_{jn} for $j = 1, 2$:

1. If using a Cramér-von Mises statistic, given a sequence of constants $\{a_n\}$, estimate the contact sets $\hat{\mathcal{X}}_0^1$ and $\hat{\mathcal{X}}_0^2$. If using a Kolmogorov-Smirnov statistic, given a sequence of constants $\{b_n\}$, estimate the b_n-maximizer sets of $\hat{m}_{1n}$ and $\hat{m}_{2n}$.

Next repeat the following two steps for $r = 1, \ldots, R$:

2. Construct the resampled processes

$$\mathcal{F}_{r1n}^*(x) = \sqrt{n}\Big(\mathbb{F}_{An}^*(-x) - \mathbb{F}_{Bn}^*(-x) - \mathbb{F}_{An}(-x) + \mathbb{F}_{Bn}(-x)\Big)$$
$$\mathcal{F}_{r2n}^*(x) = \sqrt{n}\Big(\mathbb{F}_{An}^*(-x) - \mathbb{F}_{Bn}^*(-x) - \mathbb{F}_{An}(-x) + \mathbb{F}_{Bn}(-x) + \mathbb{F}_{An}^*(x) - \mathbb{F}_{Bn}^*(x) - \mathbb{F}_{An}(x) + \mathbb{F}_{Bn}(x)\Big)$$

   using an exchangeable bootstrap.

3. Calculate the resampled test statistic. Letting $\hat{k} = \mathrm{argmax}_k\{\sup_{x \geq 0} \hat{m}_{kn}(x)\}$ and $\{c_n\} \searrow$

0 satisfy $\sqrt{n}c_n \to \infty$, calculate

$$V^*_{rn} = \begin{cases} \left(\max_{x\in\hat{\mathcal{M}}^{\hat{k}}(b_n)} \mathcal{F}^*_{r\hat{k}n}(x)\right)^+ & |\max \hat{m}_{1n} - \max \hat{m}_{2n}| > c_n \\ \max\left\{0, \max_{x\in\hat{\mathcal{M}}^1(b_n)} \mathcal{F}^*_{r1n}(x), \max_{x\in\hat{\mathcal{M}}^2(b_n)} \mathcal{F}^*_{r2n}(x)\right\} & |\max \hat{m}_{1n} - \max \hat{m}_{2n}| \leq c_n \end{cases} \tag{31}$$

or

$$W^*_{rn} = \left(\int_{\hat{\mathcal{X}}^1_0} \left((\mathcal{F}^*_{r1n}(x))^+\right)^2 \mathrm{d}x + \int_{\hat{\mathcal{X}}^2_0} \left((\mathcal{F}^*_{r2n}(x))^+\right)^2 \mathrm{d}x\right)^{1/2}. \tag{32}$$

Finally,

4. Let $\hat{q}_{V^*}(1-\alpha)$ and $\hat{q}_{W^*}(1-\alpha)$ be the $(1-\alpha)^{\text{th}}$ sample quantile from the bootstrap distributions of $\{V^*_{rn}\}_{r=1}^R$ or $\{W^*_{rn}\}_{r=1}^R$, respectively, where $\alpha \in (0,1)$ is the nominal size of the tests. Reject the null hypothesis (13) or (14) if V_{jn} and W_{jn} defined in (15)-(18) are, respectively, larger than $\hat{q}_{V^*}(1-\alpha)$ or $\hat{q}_{W^*}(1-\alpha)$.

The resampled statistics are calculated by imposing the null hypothesis and assuming that the region $\mathcal{X}^j_0(P)$ is the only part of the domain that provides a nondegenerate contribution to the asymptotic distribution of the statistic under the null. The two cases of each part in the maximum arise from trying to impose the null behavior on the resampled supremum norm statistics, even when it appears the null is violated based on the value of the sample statistic. A simple alternative way to conduct inference would be to assume the least-favorable null hypothesis that $F_A \equiv F_B$, and to resample using all of $\mathcal{X}$. However, this may result in tests with lower power (Linton, Song, and Whang, 2010) — power loss arises in situations where $\mathcal{X}_0(P) \subset \mathcal{X}$ (strictly), so that the T_j process is only nondegenerate on a subset, while bootstrapped processes that assume $\mathcal{X}_0(P) = \mathcal{X}$ would look over all of $\mathcal{X}$ and result in a stochastically larger bootstrap distribution than the true distribution.

The next result shows that our tests based on the resampling schemes described above have accurate size under the null hypothesis. In order to metrize weak convergence we use test functions from the set BL_1, which denotes Lipschitz functions $\mathbb{R} \to \mathbb{R}$ that have constant 1 and are bounded by 1.

Theorem 4.2. *Make assumptions **A1-A2** and suppose that $P \in \mathcal{P}_{00}$. Let $\hat{q}_{V^*_j}(1-\alpha)$ and $\hat{q}_{W^*_j}(1-\alpha)$ be the $(1-\alpha)^{th}$ sample quantile from the bootstrap distributions as described in the routines above. Then for $j = 1, 2$, the bootstrap is consistent:*

$$\sup_{f\in BL_1} |\mathrm{E}\left[f(V^*_n)|X\right] - \mathrm{E}\left[f(V_1)\right]| = o_P(1)$$

and

$$\sup_{f \in BL_1} |\mathrm{E}[f(W_n^*)|X] - \mathrm{E}[f(W_1)]| = o_P(1),$$

where V_1 and W_1 are defined in Theorem 4.1.

The result in above theorem is stated in terms of the limiting variables V_1 and W_1 and bootstrap analogs. V_1 and W_1, using the functional delta method, are Hadamard directional derivatives of a chain of maps from the marginal distribution functions F to the real line, and the derivatives are most compactly expressed as the definitions in Theorem 4.1.

The bootstrap variables combine conventional resampling with finite-sample estimates of the maps defined in Part 1 of Theorem 4.1, which is a resampling approach proposed in Fang and Santos (2019). Their result is actually more general — it states that with a more flexible estimator V_n^*, we would obtain bootstrap consistency for P in the null and alternative regions. Because our focus is on testing $F_A \succeq_{LASD} F_B$, however, our resampling scheme, and Theorem 4.2, are done under the imposition of the null hypothesis. The resampling consistency result in Theorem 4.2 implies that our bootstrap tests have asymptotically correct size uniformly over probability distributions in the null region, in the same sense as was stressed in Linton, Song, and Whang (2010). A formal statement of this uniformity over $\mathcal{P}_0$ is given in Theorem A.5 in Appendix A. Along with Part 4 of Theorem 4.1 Theorem A.5 additionally implies that our tests are consistent, that is, that their power to detect violations from the null represented by fixed alternative distributions tends to one. This is because the resampling scheme produces asymptotically bounded critical values, while the test statistics diverge under the alternative.

4.2 Inferring dominance from partially-identified treatment distributions

In this section we extend dominance tests to the case that distribution functions F_A and F_B are only partially identified by their Makarov bounds. Suppose that Z_0, Z_A and Z_B are random variables with marginal distribution functions $G = (G_0, G_A, G_B)$, but the joint probability distribution P of the vector (Z_0, Z_A, Z_B) is unknown, so that F_A and F_B are not point identified because they are the unknown distribution functions of $X_A = Z_A - Z_0$ and $X_B = Z_B - Z_0$. Nevertheless, we wish to test the hypotheses in (10), which depend on F_A and F_B.

4.2.1 Test statistics

Recall equations (8) and (9) from Section 3. Restated in terms of the null hypothesis $F_A \succeq_{LASD} F_B$, condition (8) is sufficient to imply the null hypothesis is true, while (9) represents a necessary condition for dominance. Denote by $\mathcal{P}^{suf}$ the set of distributions that satisfy (8) and let $\mathcal{P}^{nec}$ collect all distributions that satisfy (9). Then still using the label $\mathcal{P}_0$ for the set of distributions such that X_A dominates X_B, we have the (strict) inclusions $\mathcal{P}^{suf} \subset \mathcal{P}_0 \subset \mathcal{P}^{nec}$. Given this relation, without any further identification conditions, we look for significant violations of the necessary condition, since $P \notin \mathcal{P}^{nec}$ implies $P \notin \mathcal{P}_0$. This generally results in conservative tests because distributions $P \in \mathcal{P}^{nec} \backslash \mathcal{P}_0$ will also not be rejected, but it avoids overrejection, which would be the result when using the sufficient condition.

To test the null (10) we employ the inequality specified in equation (9) from Theorem 3.7. For each $x \in \mathcal{X}$ let

$$T_3(G)(x) = L_A(-x) + L_A(x) - U_B(-x) - U_B(x), \tag{33}$$

where L_A and U_B are defined in (6) and (7). To see the explicit dependence of T_3 on G, rewrite (33), using the identity $\inf f = -\sup(-f)$ in the definition of U_B as

$$\begin{aligned} T_3(G)(x) = &\sup_{u\in\mathbb{R}}(G_A(u) - G_0(u+x)) + \sup_{u\in\mathbb{R}}(G_A(u) - G_0(u-x)) \\ &- 2 + \sup_{u\in\mathbb{R}}(G_0(u+x) - G_B(u)) + \sup_{u\in\mathbb{R}}(G_0(u-x) - G_B(u)). \end{aligned} \tag{34}$$

As before, T_3 has been written in such a way that a violation of the null hypothesis $F_A \succeq_{LASD} F_B$ is indicated by observing some x such that $T_3(G)(x) > 0$.

The above map shares a similar feature with the T_1 map in the previous section — the marginal (in u) optimization maps are directionally differentiable at each point $x \geq 0$, but $f(u,x) \mapsto \sup_u f(u,x)$ is not Hadamard differentiable as a map from $\ell^\infty(\mathbb{R}\times\mathcal{X})$ to $\ell^\infty(\mathcal{X})$. One solution to this problem is to examine the distribution of test functionals applied to the process, which are Hadamard directionally differentiable (shown in Lemma A.4 in Appendix A).

Given observed samples $\{Z_{ki}\}$ for $k \in \{0, A, B\}$, define the marginal empirical distribution functions $\mathbb{G}_n = (\mathbb{G}_{0n}, \mathbb{G}_{An}, \mathbb{G}_{Bn})$, where $\mathbb{G}_{kn}(z) = \frac{1}{n_k}\sum_i \mathbf{1}\{Z_{ki} \leq z\}$ for $k \in \{0, A, B\}$, and let $\mathbb{L}_{An}$ and $\mathbb{U}_{Bn}$ be the plug-in estimates of the bounds: for each $x \in \mathcal{X}$, let

$$\begin{aligned} \mathbb{L}_{An}(x) &= L(x, \mathbb{G}_{0n}, \mathbb{G}_{An}) \\ \mathbb{U}_{Bn}(x) &= U(x, \mathbb{G}_{0n}, \mathbb{G}_{Bn}), \end{aligned}$$

where the maps L and U were introduced in equations (6) and (7). To estimate T_3 in (33) we use the plug-in estimate $T_3(\mathbb{G}_n)$. As in the previous section, we consider the following Kolmogorov-Smirnov and Cramér-von Mises type test statistics:

$$V_{3n} = \sqrt{n} \sup_{x \in \mathcal{X}} (T_3(\mathbb{G}_n)(x))^+ \tag{35}$$

$$W_{3n} = \sqrt{n} \left(\int_{\mathcal{X}} \left((T_3(\mathbb{G}_n)(x))^+\right)^2 \mathrm{d}x \right)^{1/2}. \tag{36}$$

The next subsections establish limiting distributions for V_{3n} and W_{3n} and suggest a resampling procedure to estimate the distributions.

4.2.2 Limiting distributions

Once again, it is necessary to define the region where the test statistics have nontrivial distributions. Define the contact set for the T_3 criterion function by

$$\mathcal{X}_0^{nec}(P) = \{x \in \mathcal{X} : L_A(-x) + L_A(x) - U_B(-x) - U_B(x) = 0\}.$$

We say that distribution $P \in \mathcal{P}_{00}^{nec}$ when $\mathcal{X}_0^{nec}(P) \neq \varnothing$. As mentioned at the beginning of the section, $\mathcal{P}_{00}^{nec}$ is not the set of P such that $F_A \succeq_{LASD} F_B$, rather those that satisfy this necessary condition, or in other words, $\mathcal{P}_0 \subset \mathcal{P}^{nec}$. There is no obvious connection between $\mathcal{P}_0$ and $\mathcal{P}_{00}^{nec}$ — the P in $\mathcal{P}_{00}^{nec}$ are simply those that lead to nontrivial asymptotic behavior of the T_3 statistic, as will be shown in Theorem 4.3. Next, we define a few functions that are analogous to the m_1 and m_2 used in the point-identified case, and which come from separating equation 34 into four sub-functions. Let $m_1(u,x) = G_A(u) - G_0(u+x)$, $m_2(u,x) = G_A(u) - G_0(u-x)$, $m_3(u,x) = G_B(u) - G_0(u+x)$ and $m_4(u,x) = G_B(u) - G_0(u-x)$. These functions are used to define, for $k = 1, \ldots 4$, for any $x \in \mathcal{X}$ and $\epsilon \geq 0$, the set-valued maps

$$\mathcal{M}^k(x,\epsilon) = \left\{ u \in \mathbb{R} : m_k(u,x) \geq \sup_{u \in \mathbb{R}} m_k(u,x) - \epsilon \right\}. \tag{37}$$

Also for the supremum norm statistic another relevant set of ϵ-maximizers exists: for any $\epsilon \geq 0$, let

$$\mathcal{M}^{nec}(\epsilon) = \left\{ (u,x) \in \mathbb{R} \times \mathcal{X} : \sum_{k=1}^4 m_k(u,x) \geq \sup_{u,x} \sum_{k=1}^4 m_k(u,x) - \epsilon \right\}. \tag{38}$$

Under the null hypothesis that the supremum is zero, $\lim_{\epsilon \searrow 0} \mathcal{M}^{nec}(\epsilon) = \mathcal{X}_0^{nec}$, as seen in the expression for V_3 in the next theorem.

Now we turn to regularity assumptions on the observed data. The only difference between these assumptions and assumptions **A1-A2** is that we must now make assumptions for three samples instead of two.

B1 The observations $\{Z_{0i}\}_{i=1}^{n_0}$, $\{Z_{Ai}\}_{i=1}^{n_A}$ and $\{Z_{Bi}\}_{i=1}^{n_B}$ are iid samples and independent of each other and are continuously distributed with marginal distribution functions G_0, G_A and G_B respectively.

B2 The sample sizes n_0, n_A and n_B increase in such a way that $n_k/(n_0+n_A+n_B) \to \lambda_k$ as $n_0, n_A, n_B \to \infty$, for $k \in \{0, A, B\}$, where $0 < \lambda_k < 1$. Let $n = n_0 + n_A + n_B$.

Before stating the next theorem, it is convenient to make some definitions. Under assumptions **B1-B2**, standard results in empirical process theory show that there is a Gaussian process $\mathcal{G}_G$ such that $\sqrt{n}(\mathbb{G}_n - G) \rightsquigarrow \mathcal{G}_G$ (van der Vaart, 1998, Example 19.6). For each (u, x), denote the transformed empirical processes and their (Gaussian) limits

$$\begin{aligned}
\sqrt{n}(\mathbb{G}_{An}(u) - \mathbb{G}_{0n}(u+x) - G_A(u) + G_0(u+x)) &= \mathbb{G}_{1n}(u,x) \rightsquigarrow \mathcal{G}_1(u,x) \\
\sqrt{n}(\mathbb{G}_{An}(u) - \mathbb{G}_{0n}(u-x) - G_A(u) + G_0(u-x)) &= \mathbb{G}_{2n}(u,x) \rightsquigarrow \mathcal{G}_2(u,x) \\
\sqrt{n}(\mathbb{G}_{0n}(u+x) - \mathbb{G}_{Bn}(u) - G_0(u+x) + G_B(u)) &= \mathbb{G}_{3n}(u,x) \rightsquigarrow \mathcal{G}_3(u,x) \\
\sqrt{n}(\mathbb{G}_{0n}(u-x) - \mathbb{G}_{Bn}(u) - G_0(u-x) + G_B(u)) &= \mathbb{G}_{4n}(u,x) \rightsquigarrow \mathcal{G}_4(u,x)
\end{aligned} \tag{39}$$

Given the above and definitions, the asymptotic behavior of V_{3n} and W_{3n} can be established.

Theorem 4.3. *Under assumptions **B1-B2**:*

1. *Suppose that $P \in \mathcal{P}_{00}^{nec}$. As $n \to \infty$, $V_{3n} \rightsquigarrow V_3$ and $W_{3n} \rightsquigarrow W_3$, where, given the definitions* (39) *and* (37),

$$V_3 = \left(\sup_{x \in \mathcal{X}_0^{nec}(P)} \sum_{k=1}^{4} \lim_{\epsilon \searrow 0} \sup_{u \in \mathcal{M}^k(x,\epsilon)} \mathcal{G}_k(u,x) \right)^+$$

   and

$$W_3 = \left(\int_{\mathcal{X}_0^{nec}(P)} \left(\left(\sum_{k=1}^{4} \lim_{\epsilon \searrow 0} \sup_{u \in \mathcal{M}^k(x,\epsilon)} \mathcal{G}_k(u,x) \right)^+ \right)^2 \mathrm{d}x \right)^{1/2}.$$

2. *Suppose that $P \in \mathcal{P}^{nec} \backslash \mathcal{P}_{00}^{nec}$. Then as $n \to \infty$, $P\{V_3 > \epsilon\} \to 0$ and $P\{W_3 > \epsilon\} \to 0$ for all $\epsilon > 0$.*

3. *Suppose that $P \notin \mathcal{P}^{nec}$. Then as $n \to \infty$, $P\{V_3 > c\} \to 1$ and $P\{W_3 > c\} \to 1$ for all $c \geq 0$.*

The results of this theorem parallel those in Theorem 4.1. The distributions of these test statistics are complex. Therefore a consistent resampling procedure for inference is discussed in the next subsection. The conservatism of these tests is reflected in the second part above. There may be $P \notin \mathcal{P}_0$ such that $P \in \mathcal{P}^{nec} \backslash \mathcal{P}_{00}^{nec}$, meaning the test will not detect that this distribution violates the hypothesis that $F_A \succeq_{LASD} F_B$.

4.2.3 Resampling procedures for inference under partial identification

Now we turn to the issue of conducting practical inference using estimated bound functions and the necessary condition for LASD. As before, resampling can be implemented by estimating the derivatives of either V_3 or W_3. These estimates represent the major difference from the resampling scheme developed in the point identified setting.

The estimates required for tests based on V_{3n} and W_{3n} are similar to those used in the point-identified case. Define a grid of values[5] $\mathbb{X} \subset \mathbb{R}$ and let $\mathbb{X}^+$ be the sub-grid of nonnegative points such that $\mathbb{X}^+ \subset \mathcal{X}$. For a sequence a_n such that $a_n \searrow 0$ and $\sqrt{n} a_n \to \infty$, define the estimate of the contact set

$$\hat{\mathcal{X}}_0^{nec} = \left\{ x \in \mathbb{X}^+ : |\mathbb{L}_{An}(-x) + \mathbb{L}_{An}(x) - \mathbb{U}_{Bn}(-x) - \mathbb{U}_{Bn}(x)| \leq a_n \right\}. \tag{40}$$

When this estimated set is empty, set $\hat{\mathcal{X}}_0^{nec} = \mathbb{X}^+$. The inner maximization step that occurs in the definition of the test statistics requires an estimate of the ϵ-maximizers of each sub-process, that is, estimates of (37) for $k = 1, \ldots 4$. For these sets we also use the same sort of estimator: for $\{b_n\}$ such that $b_n \searrow 0$ and $\sqrt{n} b_n \to \infty$, for each $x \in \mathbb{X}^+$ let

$$\hat{\mathcal{M}}^k(x) = \left\{ u \in \mathbb{X} : \hat{m}_{kn}(u, x) \geq \max_{u \in \mathbb{X}} \hat{m}_{kn}(u, x) - b_n \right\} \tag{41}$$

where the $\hat{m}_{kn}$ are plug-in estimators of m_k. Finally, for a sequence d_n such that $d_n \searrow 0$ and $\sqrt{n} d_n \to \infty$, define the estimator

$$\hat{\mathcal{M}}^{nec} = \left\{ (u, x) \in \mathbb{X} \times \mathbb{X}^+ : \sum_{k=1}^4 \hat{m}_{kn}(u, x) \geq \max_{(u,x) \in \mathbb{X} \times \mathbb{X}^+} \sum_{k=1}^4 \hat{m}_{kn}(u, x) - d_n \right\}. \tag{42}$$

Putting these estimates together, we find the derivative estimates described in the resampling

[5] Otherwise these functions would need to be evaluated over a prohibitive number of points in the support.

scheme below.

Resampling routine to estimate the distributions of V_{3n} and W_{3n}

1. If using a Cramér-von Mises statistic, given a sequence of constants $\{a_n\}$, estimate the contact set $\hat{\mathcal{X}}_0^{nec}$. If using a Kolmogorov-Smirnov statistic, given sequences of constants $\{b_n\}$ and $\{d_n\}$, estimate $\hat{\mathcal{M}}^k(\cdot)$ for $k = 1, \dots 4$ and $\hat{\mathcal{M}}^{nec}$.

Next repeat the following two steps for $r = 1, \dots, R$:

3. Construct the resampled processes $\mathcal{G}_{kn}^* = \sqrt{n}(\mathbb{G}_{kn}^* - \mathbb{G}_{kn})$ using an exchangeable bootstrap.

4. Calculate the resampled test statistic

$$V_{r3n}^* = \left(\max_{x \in \hat{\mathcal{M}}^{nec}} \sum_{k=1}^{4} \max_{u \in \hat{\mathcal{M}}^k(x)} \mathcal{G}_{kn}^*(u, x) \right)^+$$

or

$$W_{r3n}^* = \left(\int_{\hat{\mathcal{X}}_0^{nec}} \left(\left(\sum_{k=1}^{4} \max_{u \in \hat{\mathcal{M}}^k(x)} \mathcal{G}_{kn}^*(u, x) \right)^+ \right)^2 \mathrm{d}x \right)^{1/2}.$$

Finally,

6. Let $\hat{q}_{V_3^*}(1-\alpha)$ and $\hat{q}_{W_3^*}(1-\alpha)$ be the $(1-\alpha)^{\text{th}}$ sample quantile from the bootstrap distributions of $\{V_{r3n}^*\}_{r=1}^R$ or $\{W_{r3n}^*\}_{r=1}^R$, respectively, where $\alpha \in (0, 1)$ is the nominal size of the tests. We reject the null hypothesis (13) if V_{3n} and W_{3n} defined in (35) or (36) are, respectively, larger than $\hat{q}_{V_3^*}(1-\alpha)$ or $\hat{q}_{W_3^*}(1-\alpha)$.

The following theorem guarantees that the resampling scheme is consistent.

Theorem 4.4. *Make assumptions **B1-B2** and suppose that $P \in \mathcal{P}_{00}^{nec}$. Let $\hat{q}_{V_3^*}(1-\alpha)$ and $\hat{q}_{W_3^*}(1-\alpha)$ be the $(1-\alpha)^{th}$ sample quantile from the bootstrap distributions as described in the routines above. Then the bootstrap is consistent:*

$$\sup_{f \in BL_1} |\mathrm{E}\left[f(V_{3n}^*)|X\right] - \mathrm{E}\left[f(V_3)\right]| = o_P(1)$$

and

$$\sup_{f \in BL_1} |\mathrm{E}\left[f(W_{3n}^*)|X\right] - \mathrm{E}\left[f(W_3)\right]| = o_P(1).$$

Like in the point-identified setting, we define a resampling scheme and state Theorem A.6 under the imposition of the hypothesis that $P \in \mathcal{P}_{00}^{nec}$. The testing procedure based on the T_3 criterion function controls size uniformly over $\mathcal{P}^{nec}$, a superset of $\mathcal{P}_0$. The uniform size of the resampling inference scheme over $\mathcal{P}^{nec}$ is stated formally in Theorem A.6 in Appendix A. However, using only a necessary condition for inference comes at a cost, which is the possibility of trivial power against some alternative $P \notin \mathcal{P}_0$. For any $P \in \mathcal{P}^{nec} \backslash \mathcal{P}_0$, the probability of rejecting the null is also less than or equal to α. More generally, results about size and power against various alternatives that can be specified for point identified distributions are not available for the partially identified case. On the other hand, it is remarkable that the test controls size uniformly over the set $\mathcal{P}_0$, which is a set of treatment outcome distributions that cannot be observed directly.

An Online Supplemental Appendix provides Monte Carlo numerical evidence of the finite sample properties of both point- and partially-identified methods. The simulations show that tests have empirical size close to the nominal, and high power against selected alternatives.

5 Empirical application

In this section we illustrate the use of our proposed methods in a policy evaluation context. We contrast our results with a classical stochastic dominance approach. We use household-level data from an experimental evaluation of two federal assistance programs, named Aid to Families with Dependent Children (AFDC) and Jobs First (JF), to analyze the distributional effects of the policies. Bitler, Gelbach, and Hoynes (2006) use these data to document substantial heterogeneity in the impacts of this policy change on recipients' total incomes. The authors focus on this policy because of the availability of experimental data, which provides a clear source of identification.[6] Amongst its main findings, the article shows that this heterogeneity generated income gains and losses in different, sizable groups of recipients.

AFDC was one of the largest federal assistance programs in the United States between 1935 and 1996. It consisted of a means-tested income support scheme for low-income families with dependent children, administered at the state level and funded at the federal level. Following criticism that this program discouraged female labor market participation and perpetuated welfare dependency, AFDC was discontinued in 1996 and replaced, in each state, by more

[6]Bitler, Gelbach, and Hoynes (2006) conduct a test comparing features of households before random assignment and find that they do not differ significantly in terms of observable characteristics. We check additionally that the income distributions were the same before the experiment split households among the two policies. We use a conventional two-sided Cramér-von Mises test for the equality of distributions. The statistic was approximately 0.78 and its p-value was 0.55, implying that before the experiment, the distributions are indistinguishable.

restrictive programs, which generally included strict time limits for the receipt of benefits. In the state of Connecticut, the replacement for AFDC was entitled Jobs First and evaluated experimentally by the Manpower Demonstration and Research Corporation (MDRC). Some participants in the study remained under the AFDC rules, while a randomly selected subgroup was changed to JF rules, essentially implying that these households received more generous transfers but with a stricter time limit on these transfers.

We provide a decision rule, in the form of the LASD partial order, to a hypothetical social decision maker who would like to choose between these programs while taking into account the attitude of a risk-averse household considering enrollment in one of these two programs. Bitler, Gelbach, and Hoynes (2006) focus on quantile treatment effects (QTEs). If QTEs were to be used as a measure of the impact on any individual household in a welfare comparison, it would require the assumption of rank invariance across potential outcome distributions, which would be quite strong. Note that Bitler, Gelbach, and Hoynes (2006) do not make this assumption (see p. 999 for a discussion).

We consider the continue-in-AFDC and move-to-JF samples as sample observations from two policies (equivalent to policies A and B in the previous sections). There are quarterly measures for income, earnings and transfers, but we concentrate only on measures of change in total income, comparing quarterly income before and after the households were randomly assigned to one of the groups. Because assignment is random, we assume that the distribution functions of gains and losses under each policy, F_{JF} and F_{AFDC}, are point-identified by the differences in incomes before and after random assignment.

5.1 LASD using data on changes

To make welfare decisions in terms of gains and losses, we require data in terms of changes, which we construct using several definitions. First, measurements were taken before random assignment (RA) into one of the two programs, and we call these measurements pre-RA observations. All periods after random assignment are labeled post-RA observations. Next, the Jobs First program stopped supporting individuals at what we call the Time Limit (TL), although quarterly income was observed for these households after the time limit. We call pre-TL observations those that were made after random assignment but before the time limit, while post-TL observations are those made after the JF time limit. We summarize the pre/post-RA and pre/post-TL observations in one of two ways — either by averaging income over all quarters in the relevant time span, or by using the final quarter within the time span. Therefore there are four ways of defining income changes based on all the combinations of time limits and measurement summaries.

Changes in household income due to the AFDC and JF policies were defined using one of two methods. First, the natural log of of the average earnings in all post-RA quarters minus the natural log of the average pre-RA quarterly earnings is called the average-RA change. Second, the natural log of the last quarter of post-RA income minus the natural log of the last quarter of pre-RA income is called the last-quarter-RA change. Other changes are defined using data around the Jobs First time limit. The natural log of average post-TL quarterly earnings minus the natural log of average pre-TL quarterly earnings is called the average-TL change. The natural log of the last quarter of post-TL income minus the natural log of the last quarter of pre-TL income is called the last-quarter-TL change.

We conducted formal tests of the hypothesis (10) using W_{2n} statistics (Cramér-von Mises statistics applied to the empirical T_2 process).[7] The results of these tests are presented in left hand side of Table 1. First, we consider the results when changes are defined as across the random assignment. The tests indicate that we cannot reject the hypothesis that $F_{AFDC} \succeq_{LASD} F_{JF}$ unless we measure outcomes using average-RA changes. In that case AFDC does not appear to dominate the JF policy. We also conducted tests of the hypothesis $F_{JF} \succeq_{LASD} F_{AFDC}$. We cannot reject this null hypothesis using either measure. Because in one of these cases both distributions dominate each other, we double-checked using two-sided tests of distributional equality, that is, for the null that $F_{AFDC} \equiv F_{JF}$. Using average income measures the distributions appear to be different, but using last quarter measures, we cannot reject the null that the distributions are indistinguishable. These tests offer some evidence that income changes across random assignment are indistinguishable or better under the JF policy than under the AFDC policy.

Now we consider the case when changes are defined as across the time limit (either using averages or last quarters). In this case, we do not reject the hypothesis that $F_{AFDC} \succeq_{LASD} F_{JF}$, and we reject the hypothesis that $F_{JF} \succeq_{LASD} F_{AFDC}$. This is an indication that the continued support from the AFDC policy effectively supports household incomes across the JF time limit better than the JF policy does — to be expected, since the JF policy provides no more support to any households after the time limit, allowing for a higher probability of losses in household income.

Figure 1 displays the CDFs of gains and losses under the AFDC and JF policies, then the way that the two T_2 coordinate processes compare them — when looking at the coordinates in equation (12), F_A corresponds to F_{AFDC} here, so large positive values correspond to a rejection of the hypothesis $F_{AFDC} \succeq_{LASD} F_{JF}$. This figure uses only average-RA change observations. It can be seen in the second and third panels that the presumable reason that

[7] Results for the other test statistics are qualitatively the same. They are collected in an Online Supplemental Appendix.

	LASD in changes			FOSD in levels		
	$F_{AFDC} \succeq F_{JF}$	$F_{JF} \succeq F_{AFDC}$	equality	$G_{AFDC} \succeq G_{JF}$	$G_{JF} \succeq G_{AFDC}$	equality
avg-RA	3.4335	0.1790		2.8028	0.2240	
p-value	0.0500	0.9055		0.0070	0.8609	
lastQ-RA	0.8000	2.1583	2.1269			
p-value	0.6273	0.3637	0.6413			
avg-TL	0.0858	9.1380		1.2789	2.1285	2.4831
p-value	0.9150	0.0000		0.3387	0.1446	0.2081
lastQ-TL	0.8238	5.8269				
p-value	0.5963	0.0175				

Table 1: This table presents a number of tests that can be used to infer whether the Jobs First (JF) program would be preferred to the Aid to Families with Dependent Children (AFDC) or the opposite. Column titles paraphrase the null hypotheses in the tests. The first three columns use changes in income and the last three columns measure income in levels without regard to pre-policy income. Comparisons made before and after assignment or time limit were measured using the average of all months or using the last quarter. 1999 bootstrap repetitions used in each test.

the AFDC policy does not dominate the JF policy using LASD is because the probability of small losses is higher in the AFDC program and the relation between small gains and small losses is preferable in JF.

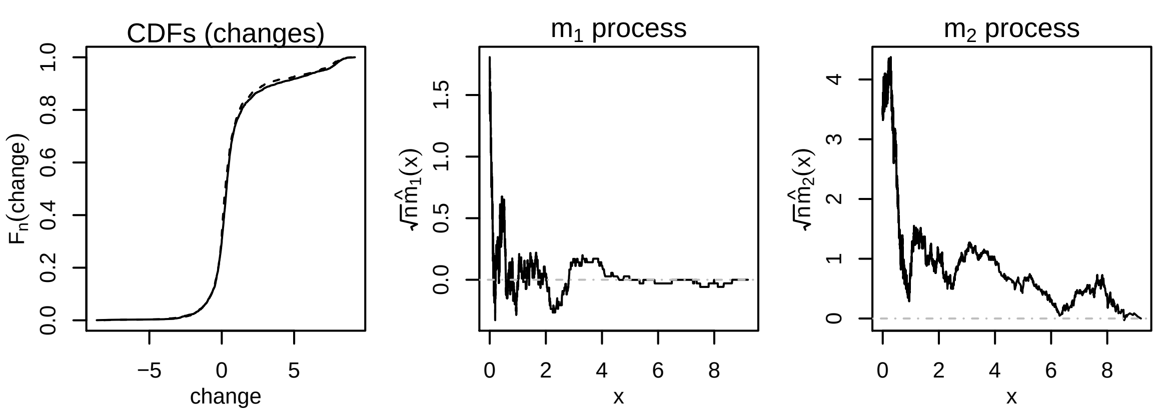

Figure 1: The CDFs of changes in post-RA income and the way that they are turned into $T_2(F)$ coordinate processes. The second and third panels correspond to plug-in estimates of the coordinate functions of equation (12). The large positive values in the second panel drive the rejection of the hypothesis $F_{AFDC} \succeq F_{JF}$ seen in Table 1.

5.2 Tests in levels: first order stochastic dominance

We also conducted an analysis of these data using standard FOSD inference methods. Tests were used to infer dominance of the AFDC or JF policies using post-randomization levels, that is, without regard to pre-randomization state. Income in levels is defined in two ways. Post-RA average income is defined as the natural log of the average income in all post-RA quarters. Post-TL average income is defined as the natural log of the average income in only the post-TL quarters. We conduct tests of the null hypothesis that $G_{AFDC} \succeq_{FOSD} G_{JF}$ or $G_{JF} \succeq_{FOSD} G_{AFDC}$, where the notation G is meant as a reminder that these are marginal final income distributions that do not consider a household's pre-policy income. The results of these tests are presented in right hand side of Table 1.

Using all post-RA quarters, we can reject the hypothesis that $G_{AFDC} \succeq_{FOSD} G_{JF}$, but cannot reject the hypothesis that $G_{JF} \succeq_{FOSD} G_{AFDC}$. Therefore it seems clear that the JF policy dominates the AFDC using final outcome distributions, that is, without regard to the effect that the policies have on any particular household's path from pre- to post-policy income.

When analyzing only the post-TL average income, we cannot reject the hypothesis that $G_{AFDC} \succeq_{FOSD} G_{JF}$ or $G_{JF} \succeq_{FOSD} G_{AFDC}$, although there is weak evidence that the second relation might be violated. We checked a two-sided test for distributional equality, and could not reject that the distributions were indistinguishable. Therefore marginal post-TL income distributions seem indistinguishable while data in changes reveals that households would prefer the AFDC program. The inferences made using data in levels and FOSD can therefore be quite different from those using LASD with data on changes.

The significantly positive part that drives the rejection of the hypothesis $G_{AFDC} \succeq_{FOSD} G_{JF}$ is represented by the spike in the right panel of Figure 2, which is due to the fact that the red AFDC CDF lies significantly above the black JF CDF in the left plot near log income level $x = 8$.

6 Conclusion

Public policies often result in gains for some individuals and losses for others. Evidence shows that the way individuals value such gains and losses is a key determinant of public support for these policies. This in turn, can determine which policies decision makers pursue. Since loss aversion is a well established regularity, how can the welfare associated with alternative

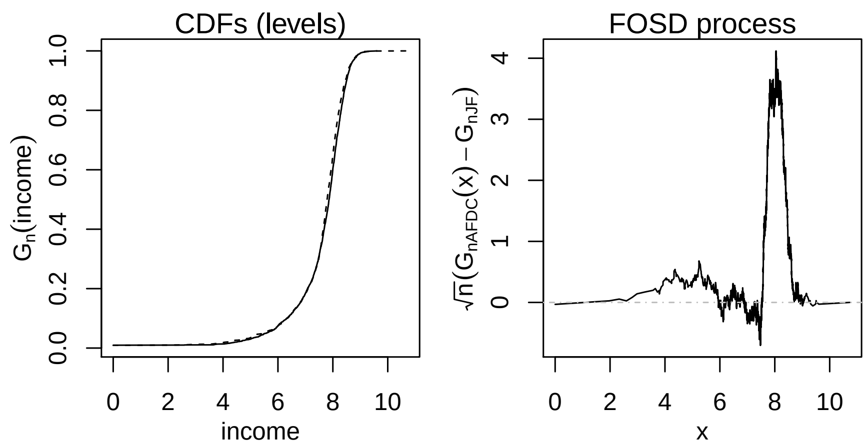

Figure 2: The CDFs of levels of post-RA income and the way that they are used to test first-order stochastic dominance. The large positive values in the second panel drive the rejection of the hypothesis $G_{AFDC} \succeq G_{JF}$ seen in Table 1.

policies be ranked when individuals are loss-averse?

We address this question by defining a social preference relation for distributions of gains and losses caused by a policy: loss aversion-sensitive dominance (LASD). We show that these social preferences are equivalent to criteria that depend solely on distribution functions. The assumption of loss aversion can lead to a welfare ranking of policies that is different from the one that would be brought about if classic utility theory and First-Order Stochastic Dominance were used. We then propose testable conditions for LASD. Because our data come as differences between underlying random variables, we propose a point-identified version of these conditions and also a partially identified analog.

In order to make LASD comparisons using observed data, we propose statistical inference methods to formally test LASD relations in both the point-identified and the partially identified cases. We show that resampling techniques, tailored to specific features of the criterion functions, can be used to conduct inference. Finally, we illustrate our LASD criterion and inference methods with a simple empirical application that uses data from a well known evaluation of a large income support policy in the US. This shows that the ranking of policy options depends crucially on whether changes or levels are used and whether or not one takes individual loss aversion into account.

Appendix

A Results on differentiability, uniform size control and computation

This section includes a definition and short discussion of the Hadamard directional differentiability concept and contains important intermediate results on Hadamard derivatives used to establish the main results in the text. Next we present some results on the control of size over the null region using the proposed resampling methods. Finally, there is one remark regarding the computation of T_1 and T_2 processes (T_3 processes should probably be computed on a grid for the sake of computation time). Proof of the results discussed in this appendix are collected in Appendix B.4.

The Hadamard derivative is a standard tool used to analyze the asymptotic behavior of nonlinear maps in empirical process theory (van der Vaart, 1998, Section 20.2). We provide a definition here for completeness, along with its directional counterpart.

Definition A.1 (Hadamard differentiability)**.** Let $\mathbb{D}$ and $\mathbb{E}$ be Banach spaces and consider a map $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$.

1. ϕ is *Hadamard differentiable* at $f \in \mathbb{D}_\phi$ tangentially to a set $\mathbb{D}_0 \subseteq \mathbb{D}$ if there is a continuous linear map $\phi' : \mathbb{D}_0 \to \mathbb{E}$ such that

$$\lim_{n\to\infty} \left\| \frac{\phi(f + t_n h_n) - \phi(f)}{t_n} - \phi'(h) \right\|_{\mathbb{E}} = 0$$

   for all sequences $\{h_n\} \subset \mathbb{D}$ and $\{t_n\} \subset \mathbb{R}$ such that $h_n \to h \in \mathbb{D}_0$ and $t_n \to 0$ as $n \to \infty$ and $f + t_n h_n \in \mathbb{D}_\phi$ for all n.

2. ϕ is *Hadamard directionally differentiable* at $f \in \mathbb{D}_\phi$ tangentially to a set $\mathbb{D}_0 \subseteq \mathbb{D}$ if there is a continuous map $\phi'_f : \mathbb{D}_0 \to \mathbb{E}$ such that

$$\lim_{n\to\infty} \left\| \frac{\phi(f + t_n h_n) - \phi(f)}{t_n} - \phi'_f(h) \right\|_{\mathbb{E}} = 0$$

   for all sequences $\{h_n\} \subset \mathbb{D}$ and $\{t_n\} \subset \mathbb{R}_+$ such that $h_n \to h \in \mathbb{D}_0$ and $t_n \searrow 0$ as $n \to \infty$ and $f + t_n h_n \in \mathbb{D}_\phi$ for all n.

In both cases of the above definition, ϕ'_f is continuous, with the addition of linearity in the fully-differentiable case (Shapiro, 1990, Proposition 3.1). They also differ in the sequences of admissible $\{t_n\}$, which allows the second definition to encode directions.

Because the pair of marginal distribution functions always occur as the difference $F_A - F_B$, the next few definitions and lemmas are stated for a single function f. For later results, maps will be applied with the function $f = F_A - F_B$. The following maps will be used repeatedly in this section and the proofs for analyzing more complex directionally differentiable maps. Let $\phi : \mathbb{R} \to \mathbb{R}$ be

$$\phi(x) = (x)^+ = \max\{0, x\}, \tag{43}$$

and similarly, define $\psi : \mathbb{R}^2 \to \mathbb{R}$ by

$$\psi(x, y) = \max\{x, y\}. \tag{44}$$

For some domain $\mathcal{X} \subseteq \mathbb{R}^j$ let $\sigma : \ell^\infty(\mathcal{X}) \to \mathbb{R}$ be

$$\sigma(f) = \sup_{x \in \mathcal{X}} f(x). \tag{45}$$

These are all Hadamard directionally differentiable maps. It can be verified that for all $a \in \mathbb{R}$,

$$\phi'_x(a) = \begin{cases} a & x > 0 \\ \max\{0, a\} & x = 0 \\ 0 & x < 0 \end{cases}, \tag{46}$$

while for pairs $(a, b) \in \mathbb{R}^2$,

$$\psi'_{x,y}(a, b) = \begin{cases} a & x > y \\ \max\{a, b\} & x = y \\ b & x < y \end{cases}.$$

For any $\epsilon \geq 0$, let $\mathcal{M}_f(\epsilon) = \{x \in \mathcal{X} : f(x) \geq \sigma(f) - \epsilon\}$ be the set of ϵ-maximizers of f. Cárcamo, Cuevas, and Rodríguez (2019) show that for all directions $h \in \ell^\infty(\mathcal{X})$

$$\sigma'_f(h) = \lim_{\epsilon \searrow 0} \sup_{x \in \mathcal{M}_f(\epsilon)} h(x) \tag{47}$$

and they also give conditions under which the limiting operation can be discarded and the supremum of h can be taken over the set of maximizers of f.

The next lemma shows shows that a weighted L_p norm (for $p > 1$) applied to the positive part of a function is directionally differentiable. Cramér-von Mises statistics are found by setting $p = 2$. The directional differentiability of the L_p norm with $p = 1$ was shown in Lemma S.4.5 of Fang and Santos (2019). Note that this lemma must be shown for the L_p norm applied to the positive-part map, jointly applied to a function f. This is because $f \mapsto (f)^+$ is not differentiable as a map of functions to functions. Nevertheless, the dominated convergence theorem allows one to use pointwise convergence with integrability to find the result.

Lemma A.2. *Suppose $f : \mathcal{X} \subseteq \mathbb{R}^j \to \mathbb{R}^k$ is a bounded and p-integrable function. Let $w : \mathcal{X} \to \mathbb{R}^k_+$ be such that $\int w_i(x)\mathrm{d}x < \infty$ for $i = 1, \dots k$. Let $1 < p < \infty$ and define the one-sided L_p norm of f by*

$$\lambda(f) = \left(\sum_{i=1}^{k} \int_{\mathcal{X}} \left((f_i(x))^+ \right)^p w_i(x) \mathrm{d}x \right)^{1/p}. \tag{48}$$

For $i = 1, \dots k$, define the subdomains $\mathcal{X}^i_- = \{x \in \mathcal{X} : f_i(x) < 0\}$, $\mathcal{X}^i_0 = \{x \in \mathcal{X} : f_i(x) = 0\}$ and $\mathcal{X}^i_+ = \{x \in \mathcal{X} : f_i(x) > 0\}$ and the index collections $\mathcal{I}^0 = \{i \in 1, \dots k : \mu(\mathcal{X}^i_0) > 0\}$ and $\mathcal{I}^+ = \{i \in 1, \dots k : \mu(\mathcal{X}^i_+) > 0\}$, where μ is Lebesgue measure. Then λ is Hadamard directionally differentiable and its derivative for any bounded, p-integrable $h : \mathcal{X} \to \mathbb{R}^k$ is

$$\lambda'_f(h) = \begin{cases} 0 & \mathcal{I}^+ = \mathcal{I}^0 = \varnothing \\ \left(\sum_{i \in \mathcal{I}^0} \int_{\mathcal{X}^i_0} \left((h_i(x))^+ \right)^p w_i(x) \mathrm{d}x \right)^{1/p} & \mathcal{I}^+ = \varnothing, \mathcal{I}^0 \neq \varnothing \\ \frac{1}{\lambda(f)^{p-1}} \sum_{i \in \mathcal{I}^+} \int_{\mathcal{X}^i_+} f_i^{p-1}(x) h_i(x) w_i(x) \mathrm{d}x & \mathcal{I}^+ \neq \varnothing \end{cases}. \tag{49}$$

The above definitions make it easy, if rather abstract, to state the differentiability of the maps from distribution to test statistics that are applied to conduct uniform inference using the T_1 process.

Lemma A.3. *Let $f \in \ell^\infty(\mathcal{X})$ and let*

$$\nu(f) = \sup_{x \in \mathcal{X}} \left((f(x))^+ + f(-x) \right)^+ \tag{50}$$

and, assuming f is square integrable,

$$\omega(f) = \left(\int_{\mathcal{X}} \{ ((f(x))^+ + f(-x))^+ \}^2 \mathrm{d}x \right)^{1/2}. \tag{51}$$

Then ν and ω are Hadamard directionally differentiable, and, letting $f_1(x) = f(-x)$ and $f_2(x) = f(x) + f(-x)$, their derivatives for any direction $h \in \ell^\infty(\mathcal{X})$ are

$$\nu'_f(h) = \left(\phi'_{\psi(\sigma(f_1), \sigma(f_2))} \circ \psi'_{\sigma(f_1), \sigma(f_2)} \right) (\sigma'_{f_1}(h), \sigma'_{f_2}(h)) \tag{52}$$

and, assuming in addition that f, h are square integrable,

$$\omega'_f(h) = \left(\lambda'_{\psi(f_1, f_2)} \circ \psi'_{f_1, f_2} \right) (h, h), \tag{53}$$

where we take the order $p = 2$ and the weight function $w \equiv 1$ in λ'_f defined in (49).

Next we turn to results for the partially identified case. Lemma A.4 provides the theoretical tool needed for the analysis of Kolmogorov-Smirnov-type statistics when using Makarov bounds. First define the abstract map $\theta : (\ell^\infty(\mathcal{U} \times \mathcal{X}))^2 \to \mathbb{R}$ by

$$\theta(f, g) = \sup_{x \in \mathcal{X}} \left(\sup_{u \in \mathcal{U}} f(u, x) + \sup_{u \in \mathcal{U}} g(u, x) \right). \tag{54}$$

For defining the directional derivative of this map at some f and g, we need to consider ϵ-maximizers for any $\epsilon \geq 0$ of these functions in u for each fixed x, which for any $f \in \ell^\infty(\mathcal{U} \times \mathcal{X})$ is the set-valued map

$$\mathcal{M}_f(x, \epsilon) = \left\{ u \in \mathcal{U} : f(u, x) \geq \sup_{u \in \mathcal{U}} f(u, x) - \epsilon \right\}. \tag{55}$$

We reserve one special label for the collection of ϵ-maximizers of the outer maximization problem that defines θ: for any $\epsilon \geq 0$ let

$$\mathcal{M}_\theta(\epsilon) = \{(u, x) \in \mathcal{U} \times \mathcal{X} : f(u, x) + g(u, x) \geq \theta(f, g) - \epsilon\}. \tag{56}$$

Lemma A.4 ahead discusses derivatives of θ, a functional that imposes two levels of maximization with an intermediate addition step, and shows that this operator is directionally differentiable. It is similar to the case of maximizing a bounded bivariate function, and its proof follows that of Theorem 2.1 of Cárcamo, Cuevas, and Rodríguez (2019), which dealt with directional differentiability of the supremum functional applied to a bounded function. The statement is for the sum of only two functions as arguments but it is straightforward to extend to any finite number of functions, as in Theorem 4.3.

Lemma A.4. *Let $\mathcal{U} \subseteq \mathbb{R}^m$ and $\mathcal{X} \subseteq \mathbb{R}^n$. Suppose that $f, g \in \ell^\infty(\mathcal{U} \times \mathcal{X})$, and let θ be the map defined in* (54)*. Then θ is Hadamard directionally differentiable and its derivative at (f, g) for any directions $(h, k) \in (\ell^\infty(\mathcal{U} \times \mathcal{X}))^2$ is*

$$\theta'_{f,g}(h, k) = \lim_{\epsilon \searrow 0} \sup_{x \in \mathcal{M}_\theta(\epsilon)} \left(\sup_{u \in \mathcal{M}_f(x, \epsilon)} h(u, x) + \sup_{u \in \mathcal{M}_g(x, \epsilon)} k(u, x) \right). \tag{57}$$

The behavior of bootstrap tests under the null and alternatives is most easily examined using distributions local to P. We consider sequences of distributions P_n local to the null distribution P such that for a mean-zero, square-integrable function η, P_n have distribution functions F_n (where P has CDF F) that satisfy

$$\lim_{n\to\infty}\int\left(\sqrt{n}\left(\sqrt{\mathrm{d}F_n}-\sqrt{\mathrm{d}F}\right)-\frac{1}{2}\eta\sqrt{\mathrm{d}F}\right)^2\to 0. \tag{58}$$

The behavior of the underlying empirical process under local alternatives satisfies Assumption 5 of Fang and Santos (2019) in a straightforward way (Wellner, 1992, Theorem 1).

Theorem A.5. *Make assumptions **A1-A2** and suppose that $F_A \succeq_{LASD} F_B$. Suppose that $\mathcal{X}$ is convex. Let $\hat{q}_{V_j^*}(1-\alpha)$ and $\hat{q}_{W_j^*}(1-\alpha)$ be the $(1-\alpha)^{th}$ sample quantile from the bootstrap distributions as described in the routines above. Then for $j = 1, 2$,*

1. *When $P \in \mathcal{P}_0$ and $\{P_n\}$ satisfy (58) and $T_j(F_n)(x) \leq 0$ for all $x \geq 0$,*

$$\limsup_{n\to\infty} P_n\left\{V_{jn} > \hat{q}_{V_j^*}(1-\alpha)\right\} \leq \alpha$$

and

$$\limsup_{n\to\infty} P_n\left\{W_{jn} > \hat{q}_{W_j^*}(1-\alpha)\right\} \leq \alpha.$$

2. *When $P \in \mathcal{P}_{00}$ and $\{P_n\}$ satisfy (58) and $T_j(F_n)(x) \leq 0$ for all $x \geq 0$, and the distribution of V or W is increasing at its $(1-\alpha)^{\mathrm{th}}$ quantile,*

$$\lim_{n\to\infty} P_n\left\{V_{jn} > \hat{q}_{V_j^*}(1-\alpha)\right\} = \alpha$$

and

$$\lim_{n\to\infty} P_n\left\{W_{jn} > \hat{q}_{W_j^*}(1-\alpha)\right\} = \alpha.$$

Now we consider using the resampling routine outlined above to test the null hypothesis that $F_A \succeq_{LASD} F_B$ when the distributions are only partially identified. It is no longer possible to guarantee exact rejection probabilities because the test is based on a superset of $\mathcal{P}_0$, but we can still show that the test does not overreject.

Theorem A.6. *Make assumptions **B1-B2**. Also assume that $\mathcal{X}$ is a convex set. Let $\hat{q}_{V_3^*}(1-\alpha)$ and $\hat{q}_{W_3^*}(1-\alpha)$ be the $(1-\alpha)^{th}$ sample quantile from the bootstrap distributions of $\{V_{r3n}^*\}_{r=1}^R$ or $\{W_{r3n}^*\}_{r=1}^R$ as described in the routine above. When the sequence of alternative distributions P_n satisfy (58) and $T_3(F_n)(x) \leq 0$ for all $x \geq 0$,*

$$\limsup_{n\to\infty} P_n\left\{V_{3n} > \hat{q}_{V_3^*}(1-\alpha)\right\} \leq \alpha$$

and

$$\limsup_{n\to\infty} P_n\left\{W_{3n} > \hat{q}_{W_3^*}(1-\alpha)\right\} \leq \alpha.$$

Remark A.7 (A note on computing point-identified criterion functions)**.** Standard empirical distribution functions are used to estimate the marginal distributions F_A and F_B. However, the definitions of the T_1 and T_2 criterion functions contain $F_k(-x)$ terms, making the plug-in $T_j(\mathbb{F}_n)$ left-continuous at some sample observations. Therefore some care must be taken when evaluating them because there may be regions that are relevant for evaluation (i.e., the location of the supremum) that are not attained by any sample observations. This could be dealt with approximately by evaluating the functions on a grid. Instead, we evaluate the function approximately at all the points where it changes its value. For example, let X_n denote the pooled sample (of size $(n_A + n_B)$) of X_A and X_B observations. Then we evaluate T_j at the points $\tilde{X}_n = 0 \cup X_n^+ \cup \{X_n - \epsilon\}^-$, where X_n^+ and X_n^- refer to the positive- and negative-valued elements of the pooled sample X_n and ϵ is a very small amount added to each element of X_n, for example, the square root of the machine's double-precision accuracy. When evaluating the L_2 integrals from an observed sample, the domain can be set to $[0, \tilde{x}_{max}]$, where $\tilde{x}_{max}$ is the largest point in the evaluation set $\tilde{X}_n$, because the integrand is identically zero above that point.

B Proof of results

B.1 Results in Section 2

Proof of Proposition 2.3. Equation (1) implies that

$$W(F) = \int_{\mathbb{R}_-} v(x)\mathrm{d}F(x) + \int_{\mathbb{R}_+} v(x)\mathrm{d}F(x). \tag{59}$$

For the first part of (59) note that

$$\begin{aligned}
\int_{\mathbb{R}_-} v(x)\mathrm{d}F(x) &= \lim_{R\to-\infty} \int_R^0 v(x)\mathrm{d}F(x) \\
&= \lim_{R\to-\infty} \left[v(x)F(x)|_R^0 - \int_R^0 v'(x)F(x)\mathrm{d}x \right] \\
&= -\int_{-\infty}^0 v'(x)F(x)\mathrm{d}x,
\end{aligned}$$

using the normalization $v(0) = 0$ noted in Definition 2.2, the assumed bounded support of F and integration by parts.

Similarly,

$$\begin{aligned}
\int_{\mathbb{R}_+} v(x)\mathrm{d}F(x) &= -\int_{\mathbb{R}_+} v(x)\mathrm{d}(1-F)(x) \\
&= -\lim_{R\to\infty}\int_0^R v(x)\mathrm{d}(1-F)(x) \\
&= -\lim_{R\to\infty}\left[v(x)(1-F(x))|_0^R - \int_0^R v'(x)(1-F(x))\mathrm{d}x\right] \\
&= \int_0^\infty v'(x)(1-F(x))\mathrm{d}x.
\end{aligned}$$

Putting these two parts together yields (2). □

B.2 Proofs of results in Section 3

Proof of Theorem 3.1. Notice that (3) is equivalent to both (4) and (5); in this proof we use the latter two conditions. Using Proposition 2.3 we rewrite $W(F_A) \geq W(F_B)$ as the equivalent condition

$$-\int_{-\infty}^0 v'(z)F_A(z)\mathrm{d}z + \int_0^\infty v'(z)(1-F_A(z))\mathrm{d}z \geq -\int_{-\infty}^0 v'(z)F_B(z)\mathrm{d}z + \int_0^\infty v'(z)(1-F_B(z))\mathrm{d}z.$$

Rearranging terms we find this is equivalent to

$$\int_{-\infty}^0 v'(z)F_B(z)\mathrm{d}z - \int_{-\infty}^0 v'(z)F_A(z)\mathrm{d}z \geq \int_0^\infty v'(z)(1-F_B(z))\mathrm{d}z - \int_0^\infty v'(z)(1-F_A(z))\mathrm{d}z$$

or simply

$$\int_{-\infty}^0 v'(z)(F_B(z)-F_A(z))\mathrm{d}z \geq \int_0^\infty v'(z)(F_A(z)-F_B(z))\mathrm{d}z.$$

This is in turn equivalent to

$$\int_0^\infty v'(-z)(F_B(-z)-F_A(-z))\mathrm{d}z \geq \int_0^\infty v'(z)(F_A(z)-F_B(z)))\mathrm{d}z$$

or

$$-\int_0^\infty v'(-z)(F_A(-z)-F_B(-z))\mathrm{d}z \geq \int_0^\infty v'(z)(F_A(z)-F_B(z)))\mathrm{d}z.$$

Adding $v'(z)(F_A(-z)-F_B(-z))$ to both sides we find this is equivalent to

$$\int_0^\infty (v'(z)-v'(-z))(F_A(-z)-F_B(-z))\mathrm{d}z \geq \int_0^\infty v'(z)(F_A(z)-F_B(z)+F_A(-z)-F_B(-z))\mathrm{d}z. \tag{60}$$

Utilizing the assumptions of loss aversion and non-decreasingness given in Definition 2.2, (4) and (5) are sufficient for (60) to hold for any v. Condition (5) is due to the fact that

$$F_A(x) - F_B(x) + F_A(-x) - F_B(-x) \leq 0 \quad \forall x \geq 0$$

is equivalent to the condition

$$1 - F_A(x) - F_A(-x) \geq 1 - F_B(x) - F_B(-x) \quad \forall x \geq 0.$$

We now show that conditions (4) and (5) are also necessary by means of a contradiction to (60). To this end, assume that there exists some x such that $F_A(-x) - F_B(-x) > 0$. From the fact that the distribution function is right continuous, it follows that there is a neighbourhood (a, b), $b > a \geq 0$, such that for all $x \in (a, b)$, $F_A(-x) - F_B(-x) > 0$. Consider the value function

$$v(x) = \begin{cases} a - b & x \leq -b \\ 0 & x \geq -a \\ x + a & x \in (-b, -a). \end{cases}$$

Note that this v satisfies conditions 1-3 of Definition 2.2. Further, for $x \in (a, b)$, $v'(-x) = 1 > v'(x) = 0$. Therefore

$$\int_0^\infty (v'(z) - v'(-z))(F_A(-z) - F_B(-z))\mathrm{d}z < 0,$$

while

$$\int_0^\infty v'(z)(F_A(z) - F_B(z) + F_A(-z) - F_B(-z))\mathrm{d}z = 0,$$

because $v'(x) = 0$ for $x \geq 0$. This contradicts (60).

The second condition can be proven similarly. Assume that there exists a neighbourhood (a, b), $0 \leq a < b$ such that for all $x \in (a, b)$, $(1 - F_A(x)) - F_A(-x) < (1 - F_B(x)) - F_B(-x)$. Take v non-decreasing and such that $v'(x) = v'(-x)$, for example

$$v(x) = \operatorname{sgn}(x) \times \begin{cases} 0 & |x| \in [0, a] \\ (|x| - a) & |x| \in (a, b) \\ b - a & |x| \in [b, \infty). \end{cases}$$

Using this v we find

$$\int_0^\infty (v'(z) - v'(-z))(F_A(-z) - F_B(-z))\mathrm{d}z = 0$$

while

$$\int_0^\infty v'(z)(F_A(z) - F_B(z) + F_A(-z) - F_B(-z))\mathrm{d}z > 0,$$

which is a contradiction.

□

Proof of Corollary 3.2. Use $v(x) = x$, which belongs to the class of functions used in the first part of Theorem 3.1 and in Definition 2.4. □

Proof of Corollary 3.5. We first notice that $F_A \succeq_{FOSD} F_{SQ}$ is equivalent to the event

$$\{F_A \text{ is supported on } \mathbb{R}_+\}\,. \tag{61}$$

Property (61) easily implies that $F_A \succeq_{LASD} F_{SQ}$, which follows by Property 1 of Definition 2.2. On the other hand one checks that

$$v(x) := \begin{cases} x & x \leq 0 \\ 0 & x > 0 \end{cases}$$

fulfills Definition 2.2. Thus $F_A \succeq_{LASD} F_{SQ}$ implies (61). □

Proof of Remark 3.6. The social value function in this case is the following

$$\int_{\mathbb{R}\times[0,\infty)} v(x) + v(y)\mathrm{d}F(x,y) = \int_0^\infty \int_{-\infty}^0 (v(x) + v(y))f(x,y)\mathrm{d}x\mathrm{d}y$$
$$+ \int_0^\infty \int_0^\infty (v(x) + v(y))f(x,y)\mathrm{d}x\mathrm{d}y.$$

Let us define $\tilde{f}^Y(x,y) = \int_{-\infty}^x f(z,y)\mathrm{d}z$ and let F^X, F^Y denote marginals of, respectively, X, Y. Integrating by parts on the negative domain of x we get

$$\int_0^\infty \left[(v(x) + v(y))\tilde{f}^Y(x,y)|_{-\infty}^0 - \int_{-\infty}^0 v'(x)\tilde{f}^Y(x,y)\mathrm{d}x \right] \mathrm{d}y$$

Knowing that $v(x) = 0$ for $x = 0$ and that $\tilde{f}^Y(x,y) = 0$ for $x = -\infty$ we get

$$\left[\int_0^\infty v(y)\tilde{f}^Y(0,y)\mathrm{d}y\right] - \int_{-\infty}^0 \left[\int_0^\infty v'(x)\tilde{f}^Y(x,y)\mathrm{d}y\right]\mathrm{d}x$$

Performing integration by parts again and noticing that $v'(x)$ is independent of integration area in the second expression, we obtain

$$\left[v(y)F(0,y)|_0^\infty - \int_0^\infty v'(y)F(0,y)\mathrm{d}y\right] - \int_{-\infty}^0 \left[v'(x)\Big(F(x,y)|_0^\infty\Big)\right]\mathrm{d}x$$

$$\left[v(\infty)F^X(0) - \int_0^\infty v'(y)F(0,y)\mathrm{d}y\right] - \int_{-\infty}^0 v'(x)F^X(x)\mathrm{d}x.$$

In the end, we obtain $\left[2v(\infty) - v(\infty)F^X(0) - v(\infty)F^Y(0)\right] - \left[\int_0^\infty v'(y)F^Y(y) - v'(y)F(0,y)\mathrm{d}y\right] - \int_0^\infty \left[v'(x)(F^X(x) - F(x,0))\right]\mathrm{d}x$.
We will now turn to the positive domain of x, thus

$$\int_0^\infty \left[(v(x)+v(y))\tilde{f}^Y(x,y)|_0^\infty - \int_0^\infty v'(x)\tilde{f}^Y(x,y)\mathrm{d}x\right]\mathrm{d}y$$

and

$$\left[\int_0^\infty (v(\infty)+v(y))f^Y(y) - v(y)\tilde{f}^Y(0,y)\mathrm{d}y\right] - \int_0^\infty \left[\int_0^\infty v'(x)\tilde{f}^Y(x,y)\mathrm{d}y\right]\mathrm{d}x.$$

Finally,
$\left[(v(\infty)+v(y))F^Y(y) - v(y)F(0,y)\right]|_0^\infty - \left[\int_0^\infty v'(y)F^Y(y) - v'(y)F(0,y)\mathrm{d}y\right] - \int_0^\infty \left[v'(x)\Big(F(x,y)|_0^\infty\Big)\right]\mathrm{d}x.$

Putting together the negative and the positive side, we obtain $\left[v(\infty)F^X(0) - \int_0^\infty v'(y)F(0,y)\mathrm{d}y\right] - \int_{-\infty}^0 v'(x)F^X(x)\mathrm{d}x + \left[2v(\infty) - v(\infty)F^X(0) - v(\infty)F^Y(0)\right] - \left[\int_0^\infty v'(y)F^Y(y) - v'(y)F(0,y)\mathrm{d}y\right] - \int_0^\infty \left[v'(x)(F^X(x) - F(x,0))\right]\mathrm{d}x$. After simplifying this expression becomes $-\int_{-\infty}^0 v'(x)F^X(x)\mathrm{d}x + \left[2v(\infty) - v(\infty)F^Y(0)\right] - \left[\int_0^\infty v'(y)F^Y(y)\mathrm{d}y\right] - \int_0^\infty \left[v'(x)(F^X(x) - F(x,0))\right]\mathrm{d}x$.
Using the fact that $y \in [0,\infty]$ we get

$$2v(\infty) - \int_{-\infty}^0 v'(x)F^X(x)\mathrm{d}x - \int_0^\infty v'(y)F^Y(y)\mathrm{d}y - \int_0^\infty v'(x)F^X(x)\mathrm{d}x$$

and

$$2v(0) + 2\int_0^\infty v'(x)\mathrm{d}x - \int_{-\infty}^0 v'(x)F^X(x)\mathrm{d}x - \int_0^\infty v'(y)F^Y(y)\mathrm{d}y - \int_0^\infty v'(x)F^X(x)\mathrm{d}x.$$

Knowing that $v(0) = 0$ we have

$$-\int_{-\infty}^{0} v'(x)F^X(x)\mathrm{d}x + \int_0^{\infty} v'(y)(1 - F^Y(y))\mathrm{d}y + \int_0^{\infty} v'(x)(1 - F^X(x))\mathrm{d}x.$$

The only change in comparison to Theorem 3.1 is the addition of the term $\int_0^\infty v'(y)(1 - F^Y(y))\mathrm{d}y$, which looks quite natural knowing that not only gains and losses but also incomes are considered. Applying the first part of the proof of Theorem 3.1 the comparison between distributions F_A and F_B comes down to the following inequality

$$\int_0^{\infty} (v'(x) - v'(-x))(F_A^X(-x) - F_B^X(-x))\mathrm{d}x + \int_0^{\infty} (v'(y)\Big(F_B^Y(y) - F_A^Y(y)\Big)\mathrm{d}y$$
$$\geq \int_0^{\infty} v'(x)(F_A^X(x) - F_B^X(x) + F_A^X(-x) - F_B^X(-x))\mathrm{d}x.$$

In comparison to (60) this inequality includes additionally the comparison of A, B for incomes y. Since $v'(y) \geq 0$ (i.e. utility is increasing with income), assuming (4), (5) and additionally that $F_B^Y(y) - F_A^Y(y) \geq 0$ for all y, that is, A dominates B for incomes according to FOSD, is enough to ensure that A is better than B.

□

Proof of Theorem 3.7. Given the bounds inequality, we have

$$L_B(-x) - U_A(-x) \leq F_B(-x) - F_A(-x) \leq U_B(-x) - L_A(-x)$$

and

$$L_A(x) - U_B(x) \leq F_A(x) - F_B(x) \leq U_A(x) - L_B(x),$$

from which it is clear that (8) is a sufficient condition. As a necessary condition we have (9), as otherwise we would have

$$F_B(-x) - F_A(-x) \leq U_B(-x) - L_A(-x) \leq L_A(x) - U_B(x) \leq F_A(x) - F_B(x).$$

□

Proof of Corollary 3.8. Recall Corollary 3.5 implied that when F_B is a status quo distribution, the FOSD and LASD relations are equivalent. Then $F_A \succeq_{FOSD} F_{SQ}$ implies that $F_A(-x) = 0$ for all $x \geq 0$ because $F_{SQ}(-x) = 0$ for all $x \geq 0$. Therefore a sufficient condition for

$F_A \succeq_{LASD} F_{SQ}$ is that $U_A(-x) = 0$ for all $x \geq 0$. Similarly, if $F_A \succeq_{LASD} F_{SQ}$, equivalent to $F_A \succeq_{FOSD} F_{SQ}$, then it must be the case that $F_A(-x) = 0$ for all $x \geq 0$, implying that $L_A(-x) = 0$ as well. □

B.3 Results in Section 4

Proof of Theorem 4.1. For Part 1 note that if $\mathcal{P} \in \mathcal{P}_{00}$ then by definition, $\mathcal{X}_0^k(P) \neq \varnothing$ for some $k \in \{1, 2\}$ and for all $x \in \mathcal{X}_0^k(P)$, $m_k(x) = 0$. Then the supremum is achieved and $\lim_{\epsilon \searrow 0} \mathcal{M}_k(\epsilon) = \mathcal{X}_0^k(P)$ for at least one coordinate, so that suprema are taken over at least one of $\mathcal{X}_0^1(P)$ and $\mathcal{X}_0^2(P)$ and whichever coordinate satisfies this condition will contribute to the asymptotic distribution. Note that for all $x \in \mathcal{X}_0(P)$, $\sqrt{n}T_1(\mathbb{F}_n)(x) = \sqrt{n}(T_1(\mathbb{F}_n) - T_1(F))(x)$. Lemma A.3 and the null hypothesis, which implies $\mathcal{X}_0^k(P) \neq \varnothing$ for $k \in \{1, 2\}$, imply the result for V_1 and W_1.

To show Part 2, note that T_2 is a linear map of F, and assuming that $\mathcal{X}_0^k(P) \neq \varnothing$ for $k \in \{1, 2\}$, we have that its weak limit (for whichever set is nonempty) is $\sup_{x \in \mathcal{X}_0^k(P)} (T_{2k}(\mathcal{G}_F)(x))^+$ by Lemma A.3. Breaking $\mathcal{X}_0(P)$ into its two subsets and assuming the null hypothesis is true results in the same behavior as the supremum norm statistic from the first part (using the definition of the supremum norm in two coordinates as the maximum of the two suprema). The same reasoning holds for the L_2 statistic in Part 2.

Part 3 follows from the behavior of the test statistics over $\{x \in \mathcal{X} : m_1(x) < 0, m_2(x) < 0\}$ described in Lemma A.3. To show Part 4 for V_{1n} suppose that for some x^*, $T_1(F)(x^*) = \xi > 0$. Then $\sup_{x \in \mathcal{X}} \sqrt{n}T_1(\mathbb{F}_n)(x) \geq \sqrt{n}(T_1(\mathbb{F}_n)(x^*) - T_1(F)(x^*)) + \sqrt{n}\xi$. Then

$$\liminf_{n \to \infty} P\left\{\sup_{x \geq 0} \sqrt{n}T_1(\mathbb{F}_n)(x) > c\right\} \geq \lim_{n \to \infty} P\left\{\sqrt{n}(T_1(\mathbb{F}_n)(x^*) - T_1(F)(x^*)) > c - \sqrt{n}\xi\right\} \to 1,$$

where the last convergence follows from the delta method applied to $\sqrt{n}(\mathbb{F}_n(x^*) - F(x^*))$, which converges in distribution to a tight random variable. The proof for the other statistics is analogous. □

Proof of Theorem 4.2. This theorem is an application of Theorem 3.2 of Fang and Santos (2019). Define the statistics V_1 and W_1 as maps from F to the real line using ν and ω defined in equations (50) and (51) in Lemma A.3, and let their estimators be defined as in part 3 of the resampling scheme. Their Assumptions 1-3 are satisfied either by the definitions of ν and ω and Lemma A.3, the standard convergence result $\sqrt{n}(\mathbb{F}_n - F) \rightsquigarrow \mathcal{G}_F$ (van der Vaart and

Wellner, 1996, Theorem 2.8.4) and the choice of bootstrap weights. We need to show that their Assumption 4 is also satisfied. Write either function as $\|h_1^+\| + \|h_1^+ \vee h_2^+\| + \|h_2^+\|$ using the desired norm. Both norms satisfy a reverse triangle inequality, and using the fact that $|(x)^+ - (y)^+| \leq |x - y|$, the difference for two functions g and h is bounded by $\|g_1 - h_1\| + \|g_1 \vee g_2 - h_1 \vee h_2\| + \|g_2 - h_2\|$. The first difference is bounded by $2\|g - h\|$, and the second and the third are bounded by $4\|g - h\|$. Rewriting equations (31) and (32) as functionals of differential directions h, define

$$\hat{\nu}_n'(h) = \begin{cases} \left(\max_{x \in \hat{\mathcal{M}}^{\hat{k}}(b_n)} h_{\hat{k}}(x)\right)^+ & |\max \hat{m}_{1n} - \max \hat{m}_{2n}| > c_n \\ \max\left\{0, \max_{x \in \hat{\mathcal{M}}^1(b_n)} h_1(x), \max_{x \in \hat{\mathcal{M}}^2(b_n)} h_2(x)\right\} & |\max \hat{m}_{1n} - \max \hat{m}_{2n}| \leq c_n \end{cases}$$

and

$$\hat{\omega}_n'(h) = \left(\int_{\hat{\mathcal{X}}_0^1} \left((h_1(x))^+\right)^2 \mathrm{d}x + \int_{\hat{\mathcal{X}}_0^2} \left((h_2(x))^+\right)^2 \mathrm{d}x\right)^{1/2}. \tag{62}$$

Because both ν and ω are Lipschitz, Lemma S.3.6 of Fang and Santos (2019) implies we need only check that $|\hat{\nu}_n'(h) - \nu_F'(h)| = o_P(1)$ and $|\hat{\omega}_n'(h) - \omega_F'(h)| = o_P(1)$ for each fixed h. This follows from the consistency of the contact set and ϵ-argmax estimators. The consistency of these estimators follow from the uniform law of large numbers for the ϵ-maximizing sets, and the tightness of the limit $\mathcal{G}_F$ for the contact sets, which implies that $\lim_n P\left\{\sqrt{n}\|\mathbb{F}_n - F\|_\infty \leq \sqrt{n}a_n\right\} = 1$. □

Proof of Theorem 4.3. Consider V_3 first. Note that V_{3n} can be rewritten as

$$V_{3n} = \sqrt{n}\sup(T_3(\mathbb{G}_n))^+ = \sqrt{n}\max\{0, \sup T_3(\mathbb{G}_n)\}.$$

Lemma A.4, extended to the four parts of the T_3 process, and the condition that $\mathcal{X}_0^{nec}(P) \neq \varnothing$, implies each of the four inner results. The derivative of the positive-part map discussed in (46), with the hypothesis that $P \in \mathcal{P}_{00}^{nec}$, which implies $\lim_{\epsilon \searrow 0} \mathcal{M}^{nec}(\epsilon) = \mathcal{X}_0^{nec}$, and the chain rule imply the outer part of the derivative and Theorem 2.1 of Fang and Santos (2019) implies the result. For W_{3n} and W_3, the finite-sample integrand converges pointwise for each $x \in \mathcal{X}$ to the limit. By assumption there are no x such that the integrand is positive, which leaves the x in $\mathcal{X}_0^{nec}(P)$ as the nontrivial part of the integral. Because the limit is assumed square-integrable, dominated convergence, Lemma A.2 and Theorem 2.1 of Fang and Santos (2019) imply the result.

For Part 2, note that by hypothesis $\mathcal{X}_0^{nec}(P) = \varnothing$ and there are no x that result in $T_3(G)(x) > 0$. Therefore Theorem 2.1 of Fang and Santos (2019), along with the chain rule, Lemmas A.4 and A.2 and the positive-part map, imply the result. The proof of Part 3 is the same as the analogous part of the proof of Theorem 4.1. □

Proof of Theorem 4.4. For both statistics, Assumptions 1-3 of Fang and Santos (2019) are trivially satisfied (van der Vaart and Wellner, 1996, Theorem 2.8.4) or satisfied by construction in the case of the bootstrap weights. Below we check that their Assumption 4 is also satisfied for both statistics, so that the statement of the theorem follows from their Theorem 3.2.

Consider V_{3n} first, and write the supremum statistic as a function of underlying processes abstractly labeled g: the limiting variable relies (through the delta method) on a map of the form $V_3 = V_3(g) = (\phi'_{\theta(g)} \circ \theta'_g)(h)$, where $g \in (\ell^\infty(\mathbb{R} \times \mathcal{X}))^4$, ϕ'_x is defined in (46) and θ'_g in (57) (extended to four functions as the arguments of the map). V_{3n} uses the sample estimates of these functions. Under the null hypothesis $\theta(g) = 0$, so that we may estimate $\hat{\phi}'_n(x) = (x)^+$, which is Lipschitz because $|(x)^+ - (y)^+| \leq |x - y|$. Writing the formula for the estimate of the derivative of θ for just two functions f and g (since the estimator for four functions can be extended immediately from this case), we have, given sequences $\{b_n\}$ and $\{d_n\}$,

$$\hat{\theta}'(h, k) = \max_{x \in \hat{\mathcal{M}}_\theta} \left(\max_{u \in \hat{\mathcal{M}}_f(x)} h(u, x) + \max_{u \in \hat{\mathcal{M}}_g(x)} k(u, x) \right).$$

This map is Lipschitz in (h, k): given any (f, g) pair, paraphrasing the sets over which maxima are taken and their arguments, we have

$$\begin{aligned}
\left|\hat{\theta}'(h_1, k_1) - \hat{\theta}'(h_2, k_2)\right| &= \left| \max_{\hat{\mathcal{M}}_\theta} \left(\max_{\hat{\mathcal{M}}_f} h_1 + \max_{\hat{\mathcal{M}}_g} k_1 \right) - \max_{\hat{B}} \left(\max_{\hat{\mathcal{M}}_f} h_1 + \max_{\hat{\mathcal{M}}_g} k_1 \right) \right| \\
&\leq \max_{\hat{\mathcal{M}}_\theta} \left| \max_{\hat{\mathcal{M}}_f} h_1 + \max_{\hat{\mathcal{M}}_g} k_1 - \max_{\hat{\mathcal{M}}_f} h_2 - \max_{\hat{\mathcal{M}}_g} k_2 \right| \\
&\leq \max_{\hat{\mathcal{M}}_\theta} \max_{\hat{\mathcal{M}}_f} |h_1 - h_2| + \max_{\hat{\mathcal{M}}_\theta} \max_{\hat{\mathcal{M}}_g} |k_1 - k_2| \\
&\leq 2 \max \{\|h_1 - h_2\|_\infty, \|k_1 - k_2\|_\infty\} \\
&= 2\|(h_1, k_1) - (h_2, k_2)\|_\infty.
\end{aligned}$$

Because all the maps in the chain that defines V_{3n} are Lipschitz, V_{3n} is itself Lipschitz, and therefore Lemma S.3.6 of Fang and Santos (2019) implies that their Assumption 4 holds if $\|(\hat{\phi}'_{\theta(g)} \circ \hat{\theta}'_g)(h) - (\phi'_{\theta(g)} \circ \theta'_g)(h)\| = o_P(1)$ (where the arguments g and h are again elements of $(\ell^\infty(\mathbb{R} \times \mathcal{X}))^4$). This follows from the consistency of the ϵ-maximizer estimates.

Next consider W_{3n}. For this part simplify the inner part to the sum of two functions, f and g, since the result is a simple generalization. Write $W_{3n} = W_{3n}(h, k) = (\hat{\lambda}'_{\mu(f,g)} \circ \hat{\mu}'_{f,g})(h, k)$, where the marginal (in u) maximization map μ is defined for each $x \geq 0$, by $\mu(f,g)(x) = \sup_{\mathcal{U}} f(u,x) + \sup_{\mathcal{U}} g(u,x)$ and for each $x \geq 0$, $\hat{\mu}'_{f,g}(h,k)(x) = \max_{u \in \hat{\mathcal{M}}_f(x)} h(u,x) + \max_{u \in \hat{\mathcal{M}}_g(x)} k(u,x)$ (define $\hat{\mathcal{M}}_f(x)$ and $\hat{\mathcal{M}}_g(x)$ as in (41)). First,

$$\begin{aligned}\left\|\hat{\mu}'(h_1,k_1) - \hat{\mu}'(h_2,k_2)\right\|_\infty &= \sup_{\mathcal{X}} \left| \max_{\hat{\mathcal{M}}_f(x)} h_1 + \max_{\hat{\mathcal{M}}_g(x)} k_1 - \max_{\hat{\mathcal{M}}_f(x)} h_2 - \max_{\hat{\mathcal{M}}_g(x)} k_2 \right| \\ &\leq \|h_1 - h_2\|_\infty + \|k_1 - k_2\|_\infty \\ &\leq 2\|(h_1,k_1) - (h_2,k_2)\|_\infty.\end{aligned}$$

Second, for square integrable f and h consider the estimate, assuming $P \in \mathcal{P}^{nec}$,

$$\hat{\lambda}'(h) = \lambda(h|_{\hat{\mathcal{X}}_0})$$

where $f|_A$ denotes the restriction of the function f to the set A. On $\hat{\mathcal{X}}_0$ the subadditivity of the norm trivially implies that $\hat{\lambda}'$ is Lipschitz there. This implies that $\hat{\lambda}'$ is a Lipschitz map, and in turn that $\hat{\lambda}'_{\mu(f,g)} \circ \hat{\mu}'_{f,g}$ is Lipschitz.

Finally, $\hat{\mu}'_{f,g}(h,k)(x)$ converges for each x the pointwise limit

$$\mu'_{f,g}(h,k)(x) = \lim_{\epsilon \searrow 0} \left(\sup_{u \in \hat{\mathcal{M}}_f(x,\epsilon)} h(u,x) + \max_{u \in \hat{\mathcal{M}}_g(x,\epsilon)} k(u,x) \right).$$

The set estimators $\hat{\mathcal{X}}_0$ and $\hat{\mathcal{X}}_+$ are consistent estimators for $\mathcal{X}_0$ and $\mathcal{X}_+$ using the same argument as above for the supremum norm. Then for square integrable h and k, the dominated convergence theorem implies that for any given f, g,

$$\left|(\hat{\lambda}'_{\mu(f,g)} \circ \hat{\mu}'_{f,g})(h,k) - (\lambda'_{\mu(f,g)} \circ \mu'_{f,g})(h,k)\right| = o_P(1),$$

and Lemma S.3.6 of Fang and Santos (2019) implies the result. □

B.4 Results in Appendix A

Proof of Lemma A.2. Let $\{t_n\}$ be a sequence of positive numbers such that $t_n \searrow 0$ as $n \to \infty$, and let $\{h_n\} \in (\ell^\infty(\mathcal{X}))^k$ be a sequence of bounded, p-integrable functions such that $h_n \to h \in (\ell^\infty(\mathcal{X}))^k$ as $n \to \infty$.

Suppose that for all i and all $x \in \mathcal{X}$, $f_i(x) < 0$, or in other words, $\mathcal{I}^+ = \mathcal{I}^0 = \varnothing$. For any point x there exists some N such that for all $n > N$, $(f_i + t_n h_{ni})^+ = 0$ because $t_n \searrow 0$ and h_i

is bounded. Then dominated convergence implies that the p-th power of the L_p norm satisfies

$$\lim_{n\to\infty}\frac{1}{t_n}\left(\sum_{i=1}^{k}\int_{\mathcal{X}_-^i}\left((f_i(x)+t_nh_{ni}(x))^+\right)^p w_i(x)\mathrm{d}x-\sum_{i=1}^{k}\int_{\mathcal{X}_-^i}\left((f_i(x))^+\right)^p w_i(x)\mathrm{d}x\right)=0.$$

This is also the result for $\lambda(f)$ in this case, which is the difference of these terms each raised to the power $1/p$.

Next suppose $\mathcal{I}^0\neq\varnothing$ and $\mathcal{I}^+=\varnothing$, that is, for some i, $\{\mathcal{X}_0^i\}$ has positive measure but the measure of x that make any coordinate of f positive is zero. Then calculate the differences directly:

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{t_n}\Bigg\{\left(\sum_{i=1}^{k}\int_{\mathcal{X}_0^i}\left((f_i(x)+t_nh_{ni}(x))^+\right)^p w_i(x)\mathrm{d}x\right)^{1/p}-\left(\sum_{i=1}^{k}\int_{\mathcal{X}_0^i}\left((f_i(x))^+\right)^p w_i(x)\mathrm{d}x\right)^{1/p}\Bigg\}\\
=\lim_{n\to\infty}\frac{1}{t_n}\left(t_n^p\sum_{i=1}^{k}\int_{\mathcal{X}_0^i}\left((h_{ni}(x))^+\right)^p w_i(x)\mathrm{d}x\right)^{1/p}\\
=\sum_{i=1}^{k}\int_{\mathcal{X}_0^i}\left((h_i(x))^+\right)^p w_i(x)\mathrm{d}x\end{aligned}$$

using dominated convergence and the p-integrability of h. If the subregions $\{x:f_i(x)<0\}$ have positive measure, they contribute 0 to the limit.

Now suppose that $\mathcal{I}^+$ is not empty, that is, there is at least one i such that $\mathcal{X}_+^i$ has positive measure. Then for each $x\in\mathcal{X}_i^+$ there exists an N such that for $n>N$, $f_i(x)+t_nh_{ni}(x)>0$ for all i. Then for $n>N$, for this x,

$$\begin{aligned}(f_i(x)+t_nh_{ni}(x))^p-f_i^p(x)&=\sum_{j=0}^{p}\binom{p}{j}f_i^j(x)(t_nh_{ni}(x))^{p-j}-f_i^p(x)\\
&=f_i^p(x)+pt_nf_i^{p-1}(x)h_{ni}(x)+O(t_n^2)-f_i^p(x)\\
&=pt_nf_i^{p-1}(x)h_{ni}(x)+O(t_n^2).\end{aligned}$$

This implies that for n large enough, the inner integral, using the calculations from the previous parts to account for the sets where f_i is zero or negative, satisfies

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{t_n}\left\{\sum_{i=1}^{k}\int_{\mathcal{X}}(f_i(x)+t_n h_{ni}(x))^p w_i(x)\mathrm{d}x-\sum_{i=1}^{k}\int_{\mathcal{X}}f_i^p(x)w_i(x)\mathrm{d}x\right\}\\
&\quad=\lim_{n\to\infty}\frac{1}{t_n}\left\{pt_n\sum_{i=1}^{k}\int_{\mathcal{X}_+^i}f_i^{p-1}(x)h_{ni}(x)w_i(x)\mathrm{d}x+O(t_n^2)+O(t_n^p)+0\right\}\\
&\quad=p\sum_{i=1}^{k}\int_{\mathcal{X}_+^i}f_i^{p-1}(x)h_i(x)w_i(x)\mathrm{d}x.
\end{aligned}$$

Using the expansion $(x+th_t)^{1/p}=x^{1/p}+\frac{1}{p}x^{(1-p)/p}th_t+o(|th_t|)$ as $t\to 0$, it can be seen that the Hadamard derivative of $x\mapsto x^{1/p}$ is $\frac{1}{p}x^{(1-p)/p}h$. Therefore the chain rule and integrability of f and h implies that the derivative is

$$\frac{1}{\lambda(f)^{p-1}}\sum_{i=1}^{k}\int_{\mathcal{X}_+^i}f_i^{p-1}(x)h_i(x)w_i(x)\mathrm{d}x.$$

□

Proof of Lemma A.3. For ν write

$$\begin{aligned}
\nu(f)&=\sup_{x\in\mathcal{X}}\left((f(x))^+ + f(-x)\right)^+\\
&=\sup_{x\in\mathcal{X}}\max\left\{0,(f(x))^+ + f(-x)\right\}\\
&=\sup_{x\in\mathcal{X}}\max\left\{0,\max\left\{f(-x),f(x)+f(-x)\right\}\right\}
\end{aligned}$$

and using the definitions of f_1 and f_2 made in the statement of the lemma and changing the order in which the maxima are computed

$$\begin{aligned}
&=\max\left\{0,\max\left\{\sup_{x\in\mathcal{X}}f_1(x),\sup_{x\in\mathcal{X}}f_2(x)\right\}\right\}\\
&=(\phi\circ\psi)(\sigma(f_1),\sigma(f_2)).
\end{aligned}$$

Then using the chain rule (Shapiro, 1990) the derivative is that given in the statement of the lemma. For ω, assume f and h are square integrable and write

$$\begin{aligned}
\omega(f)&=\lambda((f(x))^+ + f(-x))\\
&=\lambda(\max\{f(-x),f(x)+f(-x)\})\\
&=(\lambda\circ\psi)(f_1,f_2).
\end{aligned}$$

Taking a derivative and using the chain rule implies the second expression in the statement of the lemma. □

Proof of Theorem A.5. This is an application of Corollary 3.2 in Fang and Santos (2019), and we only sketch the most important details of the proof. After applying the null hypothesis, the derivatives ν'_F and ω'_F shown in (52) and (53) are both convex. For example, in the expression for ν'_F,

$$\left(\sup_{\mathcal{X}_0^1(P)} (\alpha h_{1A} + (1-\alpha) h_{1B})\right)^+ \leq \alpha \left(\sup_{\mathcal{X}_0^1(P)} h_{1A}\right)^+ + (1-\alpha)\left(\sup_{\mathcal{X}_0^1(P)} h_{1B}\right)^+$$

and similar calculations hold for the other two terms. In the case of ω'_F, for example,

$$\int_{\mathcal{X}_0^1(P)} \left((\alpha h_1 + (1-\alpha) h_2)^+\right)^2 \leq \alpha \int_{\mathcal{X}_0^1(P)} \left((h_1)^+\right)^2 + (1-\alpha) \int_{\mathcal{X}_0^1(P)} \left((h_2)^+\right)^2,$$

where the inequality relies on the nonnegativity of the innermost term and convexity of $x \mapsto x^2$ for $x \geq 0$. Then Theorem 3.3 of Fang and Santos (2019) applies. The second part of the theorem is a special case of the first, when the part of the relationship that leads to nondegenerate behavior is not empty. □

Proof of Lemma A.4. First, let $s_n = t_n^{-1}$ and define the finite differences

$$\Delta_n = \sup_{\mathcal{X}} \left(\sup_{\mathcal{U}} (s_n f + h)(u,x) + \sup_{\mathcal{U}} (s_n g + k)(u,x)\right) - s_n \theta(f,g) \tag{63}$$

so that for any $s_n \nearrow \infty$, we need to show that $\Delta_n \to \theta'_{f,g}(h,k)$ defined in the statement of the theorem.

Fix an $\epsilon > 0$. Then for any $x \notin \mathcal{M}_\theta(\epsilon)$, note that

$$\sup_{\mathcal{U}} (s_n f + h)(u,x) + \sup_{\mathcal{U}} (s_n g + k)(u,x) - s_n \theta(f,g) \leq \sup h + \sup k - s_n \epsilon. \tag{64}$$

Similarly, if $u \notin \mathcal{M}_f(x,\epsilon)$ for any x (the case for u that do not nearly-optimize $g(\cdot,x)$ is symmetric), then also

$$(s_n f + h)(u,x) + \sup_{\mathcal{U}} (s_n g + k)(u,x) - s_n \theta(f,g) \leq \sup h + \sup k - s_n \epsilon \tag{65}$$

for that x. Therefore for any $\epsilon > 0$,

$$\begin{aligned}
&\limsup_n \Delta_n \\
&\quad = \limsup_n \left(\sup_{\mathcal{M}_\theta(\epsilon)} \left(\sup_{\mathcal{M}_f(x,\epsilon)} (s_n f + h)(u,x) + \sup_{\mathcal{M}_g(x,\epsilon)} (s_n g + k)(u,x) \right) - s_n \theta(f,g) \right) \\
&\qquad \le \limsup_n \left(s_n \sup_{\mathcal{M}_\theta(\epsilon)} \left(\sup_{\mathcal{M}_f(x,\epsilon)} f(u,x) + \sup_{\mathcal{M}_g(x,\epsilon)} g(u,x) \right) - s_n \theta(f,g) \right. \\
&\qquad\qquad \left. + \sup_{\mathcal{M}_\theta(\epsilon)} \left(\sup_{\mathcal{M}_f(x,\epsilon)} h(u,x) + \sup_{\mathcal{M}_g(x,\epsilon)} k(u,x) \right) \right) \\
&\qquad\qquad\qquad = \sup_{\mathcal{M}_\theta(\epsilon)} \left(\sup_{\mathcal{M}_f(x,\epsilon)} h(u,x) + \sup_{\mathcal{M}_g(x,\epsilon)} k(u,x) \right), \quad (66)
\end{aligned}$$

so that this inequality holds as $\epsilon \searrow 0$.

Next, for any $\epsilon > 0$ define

$$\bar{t}(\epsilon) = \sup_{\mathcal{M}_\theta(\epsilon)} \left(\sup_{\mathcal{M}_f(x,\epsilon)} h(u,x) + \sup_{\mathcal{M}_g(x,\epsilon)} k(u,x) \right). \quad (67)$$

Because this function is nondecreasing in ϵ, it has a limit as $\epsilon \searrow 0$, so that for any $m \in \mathbb{N}$ there exists an $x_m \in \mathcal{M}_\theta(1/m)$ and (u_m^f, u_m^g) satisfying the inequality

$$h(u_m^f, x_m) + k(u_m^g, x_m) \ge \bar{t}(1/m) - 1/m.$$

Therefore

$$\begin{aligned}
&\bar{t}(1/m) \le h(u_m^f, x_m) + k(u_m^g, x_m) + 1/m \\
&\qquad = s_n f(u_m^f, x_m) + h(u_m^f, x_m) + s_n g(u_m^g, x_m) + k(u_m^g, x_m) \\
&\qquad\qquad + 1/m - s_n(f(u_m^f, x_m) + g(u_m^g, x_m)) \\
&\quad \le \sup_{\mathcal{X}} \left(\sup_{\mathcal{U}} (s_n f + h)(u,x) + \sup_{\mathcal{U}} (s_n g + k)(u,x) \right) - s_n \theta(f,g) + (s_n + 1)/m, \quad (68)
\end{aligned}$$

which implies that

$$\lim_{\epsilon \searrow 0} \sup_{\mathcal{M}_\theta(\epsilon)} \left(\sup_{\mathcal{M}_f(x,\epsilon)} h(u,x) + \sup_{\mathcal{M}_g(x,\epsilon)} k(u,x) \right) = \lim_{m \to \infty} \bar{t}(1/m) \le \Delta_n. \quad (69)$$

□

Proof of Theorem A.6. Start by considering V_3. As in the proof of Theorem 4.2, we simplify the analysis by writing this statistic as a composition of maps that act on just two functional arguments, $(\phi'_{\theta(f,g)} \circ \theta'_{f,g})(h,k)$, where the positive-part map ϕ'_x is defined in (46) and $\theta'_{f,g}$ is, for any $h, k \in \ell^\infty(\mathcal{U} \times \mathcal{X}))$,

$$\theta'_{f,g}(h,k) = \lim_{\epsilon \searrow 0} \sup_{x \in \mathcal{M}_\theta(\epsilon)} \left(\lim_{\epsilon \searrow 0} \sup_{u \in \mathcal{M}_f(x,\epsilon)} h(u,x) + \lim_{\epsilon \searrow 0} \sup_{u \in \mathcal{M}_g(x,\epsilon)} k(u,x) \right),$$

where $\mathcal{M}_f(x,\epsilon)$ and $\mathcal{M}_\theta(\epsilon)$ are defined in (55) and (56).

It can be verified that for a fixed value of $\theta(f,g)$, $\hat{\phi}'_{\theta(f,g)}(x)$ is convex and nondecreasing. Next consider $\theta'_{f,g}$. For any $\epsilon > 0$, consider the map applied to the convex combination of vector-valued functions $\alpha(h_1,k_1) + (1-\alpha)(h_2,k_2)$:

$$\begin{aligned}
&\sup_{\mathcal{M}_\theta(\epsilon)} \left(\sup_{\mathcal{M}_f(x,\epsilon)} (\alpha h_1(u,x) + (1-\alpha)k_1(u,x)) + \sup_{\mathcal{M}_g(x,\epsilon)} (\alpha h_2(u,x) + (1-\alpha)k_2(u,x)) \right) \\
&\leq \sup_{\mathcal{M}_\theta(\epsilon)} \left(\alpha \left(\sup_{\mathcal{M}_f(x,\epsilon)} h_1(u,x) + \sup_{\mathcal{M}_g(x,\epsilon)} k_1(u,x) \right) + (1-\alpha) \left(\sup_{\mathcal{M}_f(x,\epsilon)} h_2(u,x) + \sup_{\mathcal{M}_g(x,\epsilon)} k_2(u,x) \right) \right) \\
&\leq \alpha \sup_{\mathcal{M}_\theta(\epsilon)} \left(\sup_{\mathcal{M}_f(x,\epsilon)} h_1(u,x) + \sup_{\mathcal{M}_g(x,\epsilon)} k_1(u,x) \right) + (1-\alpha) \sup_{\mathcal{M}_\theta(\epsilon)} \left(\sup_{\mathcal{M}_f(x,\epsilon)} h_2(u,x) + \sup_{\mathcal{M}_g(x,\epsilon)} k_2(u,x) \right).
\end{aligned}$$

Therefore, letting $\epsilon \searrow 0$, it can be seen that $\theta'_{f,g}$ is convex. Because V_3 is the composition of a non-decreasing, convex function with a convex function, V_3 is also a convex map of (h,k) to $\mathbb{R}$ (Boyd and Vandenberghe, 2004, eq. 3.11). As mentioned in the text, $\mathcal{P}_0 \subseteq \mathcal{P}^{nec}$. Therefore Corollary 3.2 of Fang and Santos (2019) implies

$$\limsup_{n \to \infty} P_n \left\{ V_{3n} > q_{V_3^*}(1-\alpha) \right\} \leq \alpha.$$

Turn next to W_3. Similarly, write this statistic as a map of pairs of bounded functions to the real line as $W_{3n} = (\lambda'_{\mu(f,g)} \circ \mu'_{f,g})(h,k)$, where for each $x \in \mathcal{X}$,

$$\mu(f,g)(x) = \sup_{\mathcal{U}} f(u,x) + \sup_{\mathcal{U}} g(u,x)$$

and

$$\mu'_{f,g}(h,k)(x) = \lim_{\epsilon \searrow 0} \max_{u \in \mathcal{M}_f(x,\epsilon)} h(u,x) + \lim_{\epsilon \searrow 0} \max_{u \in \mathcal{M}_g(x,\epsilon)} k(u,x),$$

and for any functions $f, h \in \ell^\infty(\mathcal{X})$, $\lambda'_f(h)$ is defined in (49). We show the convexity of this composition directly. Paraphrase $\mu(x) = \mu(f,g)(x)$, and for fixed $\epsilon > 0$,

$$\begin{aligned}
\mu_1'(x) &= \sup_{u \in \mathcal{M}_f(x,\epsilon)} h_1(u,x) + \sup_{u \in \mathcal{M}_g(x,\epsilon)} k_1(u,x) \\
\mu_2'(x) &= \sup_{u \in \mathcal{M}_f(x,\epsilon)} h_2(u,x) + \sup_{u \in \mathcal{M}_g(x,\epsilon)} k_2(u,x) \\
\bar{\mu}'(x) &= \sup_{u \in \mathcal{M}_f(x,\epsilon)} (\alpha h_1 + (1-\alpha)k_1)(u,x) + \sup_{u \in \mathcal{M}_g(x,\epsilon)} (\alpha h_2 + (1-\alpha)k_2)(u,x).
\end{aligned}$$

Finally, let $\mathcal{X}_0$ denote the region where $\mu(x) = 0$. Then Lemma A.2 shows that $\lambda_\mu'(\bar{\mu}') = \lambda(\bar{\mu}'|_{\mathcal{X}_0})$, where $\bar{\mu}'|_{\mathcal{X}_0}$ denotes the restriction of the function $\bar{\mu}'$ to the set $\mathcal{X}_0$. Consider the first term on the right hand side. Inside the integral, it can be seen that

$$\begin{aligned}
0 &\leq \left(\bar{\mu}'(x)\right)^+ \\
&= \left(\sup_{u \in \mathcal{M}_f(x,\epsilon)} (\alpha h_1 + (1-\alpha)h_2)(u,x) + \sup_{u \in \mathcal{M}_g(x,\epsilon)} (\alpha k_1 + (1-\alpha)k_2)(u,x) \right)^+ \\
&\leq \left(\alpha \left(\sup_{u \in \mathcal{M}_f(x,\epsilon)} h_1(u,x) + \sup_{u \in \mathcal{M}_g(x,\epsilon)} k_1(u,x) \right) \right. \\
&\qquad \left. + (1-\alpha) \left(\sup_{u \in \mathcal{M}_f(x,\epsilon)} h_2(u,x) + \sup_{u \in \mathcal{M}_g(x,\epsilon)} k_2(u,x) \right) \right)^+ \\
&= \left(\alpha \mu_1'(x) + (1-\alpha)\mu_2'(x)\right)^+ \\
&\leq \alpha \left(\mu_1'(x)\right)^+ + (1-\alpha) \left(\mu_2'(x)\right)^+.
\end{aligned}$$

Because the integrand is nonnegative, subadditivity of the L_2 norm implies

$$\lambda(\bar{\mu}'|_{\mathcal{X}_0}) \leq \alpha \lambda(\mu_1'|_{\mathcal{X}_0}) + (1-\alpha)\lambda(\mu_2'|_{\mathcal{X}_0}).$$

This inequality holds as $\epsilon \searrow 0$ by the assumed square-integrability of the arguments. Therefore Corollary 3.2 of Fang and Santos (2019) implies

$$\limsup_{n \to \infty} P_n \left\{ W_{3n} > q_{W_3^*}(1-\alpha) \right\} \leq \alpha.$$

□

References

Aaberge, R., T. Havnes, and M. Mogstad (2018): "Ranking Intersecting Distribution Functions," *Journal of Applied Econometrics*, forthcoming.

Alesina, A., and F. Passarelli (2019): "Loss Aversion, Politics and Redistribution," *American Journal of Political Science*, 63, 936–947.

Andrews, D. W., and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

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

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

Bhattacharya, D., and P. Dupas (2012): "Inferring Welfare Maximizing Treatment Assignment under Budget Constraints," *Journal of Econometrics*, 167, 168–196.

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2006): "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, 96, 988–1012.

Boyd, S., and L. Vandenberghe (2004): *Convex Optimization*. Cambridge University Press, Cambridge.

Cárcamo, J., A. Cuevas, and L.-A. Rodríguez (2019): "Directional Differentiability for Supremum-Type Functionals: Statistical Applications," *arXiv e-prints*, p. arXiv:1902.01136.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2001): "Removing the Veil of Ignorance in Assessing the Distributional Impacts of Social Policies," *Swedish Economic Policy Review*, 8, 273–301.

Cattaneo, M. D., M. Jansson, and K. Nagasawa (2017): "Bootstrap-Based Inference for Cube Root Consistent Estimators," Working paper.

Chetverikov, D., A. Santos, and A. M. Shaikh (2018): "The Econometrics of Shape Restrictions," *Annual Review of Economics*, 10, 31–63.

Chew, S. H. (1983): "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox," *Econometrica*, 51, 1065–1092.

Cho, J. S., and H. White (2018): "Directionally Differentiable Econometric Models," *Econometric Theory*, 34, 1101–1131.

Christensen, T., and B. Connault (2019): "Counterfactual Sensitivity and Robustness," Working paper.

Dehejia, R. (2005): "Program Evaluation as a Decision Problem," *Journal of Econometrics*, 125, 141–173.

Dümbgen, L. (1993): "On Nondifferentiable Functions and the Bootstrap," *Probability Theory and Related Fields*, 95, 125–140.

Eeckhoudt, L., and H. Schlesinger (2006): "Putting Risk in Its Proper Place," *American Economic Review*, 96, 280–289.

Fang, Z., and A. Santos (2019): "Inference on Directionally Differentiable Functions," *Review of Economic Studies*, 86, 377–412.

Fishburn, P. C. (1980): "Continua of Stochastic Dominance Relations for Unbounded Probability Distributions," *Journal of Mathematical Economics*, 7, 271–285.

Frank, M. J., R. B. Nelsen, and B. Schweizer (1987): "Best-Possible Bounds for the Distribution of a Sum — A Problem of Kolmogorov," *Probability Theory and Related Fields*, 74, 199–211.

Freund, C., and c. Özden (2008): "Trade Policy and Loss Aversion," *American Economic Review*, 98, 1675–1691.

Gajdos, T., and J. A. Weymark (2012): "Introduction to Inequality and Risk," *Journal of Economic Theory*, 147, 1313–1330.

Heckman, J. J., J. Smith, and N. Clements (1997): "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487–535.

Heckman, J. J., and J. A. Smith (1998): "Evaluating the Welfare State," in *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, ed. by S. Strom. Cambridge University Press, New York.

Hirano, K., and J. Porter (2009): "Asymptotics for Statistical Treatment Rules," *Econometrica*, 77, 1683–1701.

Hong, H., and J. Li (2018): "The Numerical Delta Method," *Journal of Econometrics*, 206, 379–394.

Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–292.

Kasy, M. (2016): "Partial Identification, Distributional Preferences, and the Welfare Ranking of Policies," *Review of Economics and Statistics*, 98, 111–131.

Kőszegi, B., and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, CXXI, 1133–1165.

Kitagawa, T., and A. Tetenov (2018): "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616.

——— (2019): "Equality-Minded Treatment Choice," *Journal of Business and Economic Statistics*, forthcoming.

Levy, H. (2016): *Stochastic Dominance: Investment Decision Making Under Uncertainty, 3rd edition*. Springer International Publishing, Switzerland.

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

Linton, O., K. Song, and Y.-J. Whang (2010): "An Improved Bootstrap Test of Stochastic Dominance," *Journal of Econometrics*, 154, 186–202.

Makarov, G. (1982): "Estimates for the Distribution Function of a Sum of Two Random Variables when the Marginal Distributions are Fixed," *Theory of Probability and its Applications*, 26(4), 803–806.

Manski, C. F. (2004): "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 1221–1246.

Masten, M. A., and A. Poirier (2020): "Inference on Breakdown Frontiers," *Quantitative Economics*, 11, 41–111.

Rabin, M., and R. H. Thaler (2001): "Anomalies: Risk Aversion," *Journal of Economic Perspectives*, 15, 219–232.

Rick, S. (2011): "Losses, Gains, and Brains: Neuroeconomics Can Help to Answer Open Questions about Loss Aversion," *Journal of Consumer Psychology*, 21, 453–463.

Roemer, J. E. (1998): *Theories of Distributive Justice*. Harvard University Press, Cambridge.

Rüschendorf, L. (1982): "Random Variables with Maximum Sums," *Advances in Applied Probability*, 14, 623–632.

Samuelson, W., and R. Zeckhauser (1988): "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, 1, 7–59.

Sen, A. K. (2000): *Freedom, Rationality and Social Choice: The Arrow Lectures and Other Essays*. Oxford University Press, Oxford.

Shaked, M., and G. J. Shanthikumar (1994): *Stochastic Orders and Their Applications*. Academic Press, San Diego, CA.

Shapiro, A. (1990): "On Concepts of Directional Differentiability," *Journal of Optimization Theory and Applications*, 66, 477–487.

Stoye, J. (2009): "Minimax Regret Treatment Choice With Finite Samples," *Journal of Econometrics*, 151, 70–81.

Tetenov, A. (2012): "Statistical Treatment Choice Based on Asymmetric Minimax Regret Criteria," *Journal of Econometrics*, 166, 157–165.

Tversky, A., and D. Kahneman (1991): "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, 106, 1039–1061.

——— (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.

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

Wellner, J. A. (1992): "Empirical Processes in Action: A Review," *International Statistical Review*, 60, 247–269.

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

Williamson, R. C., and T. Downs (1990): "Probabilistic Arithmetic I. Numerical Methods for Calculating Convolutions and Dependency Bounds," *International Journal of Approximate Reasoning*, 4, 89–158.

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

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

Online supplemental appendix to "Loss aversion and the welfare ranking of policy interventions"

Sergio Firpo* Antonio F. Galvao† Martyna Kobus‡ Thomas Parker§
Pedro Rosa-Dias¶

April 17, 2020

This supplement appendix contains numerical Monte Carlo simulations studying the empirical size and power of the statistical methods proposed in the main text and additional results for the empirical application in Section 5 of the main text.

1 Monte Carlo simulations

In this section, we compare the finite sample performances tests proposed in the text for testing the LASD null hypothesis. We describe the results of simulation experiments used to investigate the size and power properties of the tests described in the main text. There are three simulation settings: a normal location model and a triangular model under point identification, and a normal location model under partial identification.

1.1 Normal model, identified case

In this experiment there are two independent, Gaussian random variables that represent point-identified outcomes. The scale of both distributions is set to unity, the location of distribution A is set to zero and the location of distribution B is allowed to vary. Letting μ_B denote the

*Insper, Sao Paulo, Brazil. E-mail: firpo@insper.edu.br
†Department of Economics, University of Arizona, Tucson, USA. E-mail: agalvao@email.arizona.edu
‡Institute of Economics, Polish Academy of Sciences, Warsaw, Poland. E-mail: mkobus@inepan.waw.pl
§Department of Economics, University of Waterloo, Waterloo, Canada. E-mail: tmparker@uwaterloo.ca
¶Department of Economics and Public Policy, Imperial College Business School, Imperial College London, UK. E-mail: p.rosa-dias@imperial.ac.uk

location of distribution B, tests should not reject the null $H_0 : F_A \succeq_{LASD} F_B$ when $\mu_B \leq 0$ and should reject the null when $\mu_B > 0$. This is a case where $\mathcal{P}_{00}$ is a singleton, which is when $\mu_B = 0$.

We select constant sequences in the following way. Let $n = n_A + n_B$. The estimated contact sets $\hat{\mathcal{X}}_0^k = \{x \in \mathcal{X} : |\hat{m}_{kn}(x)| \leq a_n\}$ worked well using $a_n = 4\log(\log(n))/\sqrt{n}$. For estimated ϵ-maximizer sets $\hat{\mathcal{M}}^k = \{x \in \mathcal{X} : \hat{m}_{kn}(x) > \sup \hat{m}_{kn}(x) - b_n\}$ we used $b_n = \sqrt{\log(\log(n))/n}$. For deciding on which coordinate appeared significantly larger than the other, or whether both coordinates reached approximately the same supremum, that is, when estimating $|\max \hat{m}_{1n}(x) - \max \hat{m}_{2n}(x)| \leq c_n$, we used the same constant sequence as b_n, that is, $c_n = \sqrt{\log(\log(n))/n}$. These sequences were used after preliminary simulations with the normal model, and were used in the other two simulations as well (with $n = n_0 + n_A + n_B$ in the partially-identified setting).

The size and power of the tests is good in this example, as can be seen in Figure 1. The mean of distribution B ran from $-2/\sqrt{n}$ to $4/\sqrt{n}$ so the alternatives are local to the boundary of the null region. Sample sizes were identical for both samples and set equal to 100, 500 or 1,000. When resampling, the number of bootstrap repetitions was set equal to 499 (for samples of size 100), 999 (for samples of size 500) or 1,999 (for samples of size 1,000). Figure 1 plots empirical rejection probabilities from 1,000 simulation runs.

From Figure 1 it can be seen that the empirical rejection probabilities are relatively close to the nominal 5% rejection probability at the boundary of the null region when $\mu_B = 0$. The behavior of supremum norm tests was identical so only V_{1n} test results are shown. The W_{1n} and W_{2n} results are close and the differences are due to numerical integration that occurs over one or two dimensions depending on the statistic.

1.2 Triangular model, identified case

In this experiment we use two independent triangular random variables, where we let $\theta = (\alpha, \beta, \gamma)$ denote the lower endpoint of the support, the mode of the distribution and the upper endpoint of the support. Distribution A uses $\theta_A = (-1, 0, 1)$, while the shape of distribution B is allowed to vary. For a parameter $\epsilon \in [-1/2, 1/2]$ we let $\theta_B = (-1-\epsilon/\sqrt{n}, -\epsilon/\sqrt{n}, 1+\epsilon/\sqrt{n})$, so that all the distributions are local to the boundary of the null region represented by $\epsilon = 0$. Two distributions are depicted in Figure 2, in which $\epsilon = 1/4$. This implies that $F_A \succeq_{LASD} F_B$. From the right panel of the plot it can be seen that these distributions satisfy an LASD ordering, but they would not be ordered by FOSD.

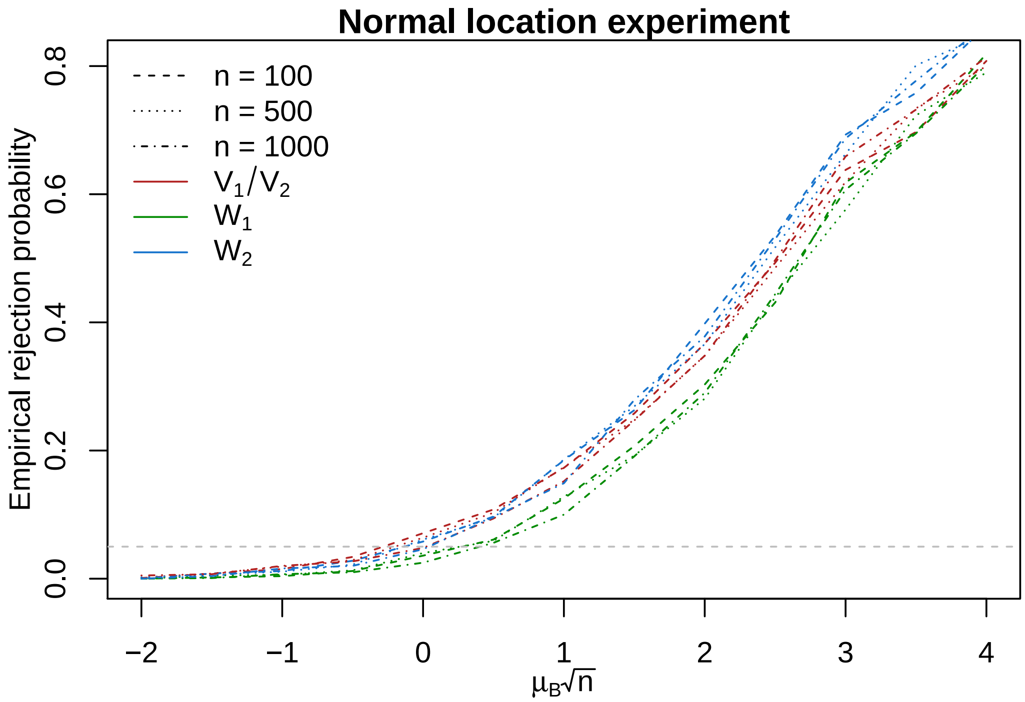

Figure 1: Empirical rejection probabilities of the LASD tests in the point identified normal location model experiment. The tests are of nominal 5% size, should have exactly 5% rejection probability when $\mu_B = 0$ and should reject when $\mu_B > 0$. V_{1n} and V_{2n} tests have identical behavior so only V_{1n} results are shown. Samples of sizes 100, 500 and 1000 correspond respectively to 499, 999 and 1999 bootstrap repetitions. Distributions are local to the boundary of the null region, which is where $\mu_B = 0$. 1000 simulation repetitions.

Figure 3 shows the empirical rejection results from the triangular model experiment. We allow ϵ, which controls the shape of distribution B, to vary between $-1/2$ and $1/2$. The tests in this experiment should reject the null when $\epsilon < 0$, should equal the nominal size at $\epsilon = 0$ and should not reject when $\epsilon > 0$. Because of the restricted supports of the distributions and the relatively small region for ϵ, the horizontal axis for the power curves shown in Figure 3 is the value of the alternative parameters in absolute scale and not local alternatives. Therefore the power curves show a noticeable change over different values of the sample sizes used.

1.3 Normal model, partially identified case

In this experiment we use three independent normal random variables (Z_0, Z_A, Z_B) with scales set to unity and location parameters $\mu = (0, 0, \mu_B)$, where μ_B is allowed to vary. We denote this triple of marginal normal CDFs by $G(\mu_B)$. Rounding to one decimal place, the null $H_0 : F_A \succeq_{LASD} F_B$ should be rejected when $\mu_B > 2.8$. We let μ_B vary locally around this

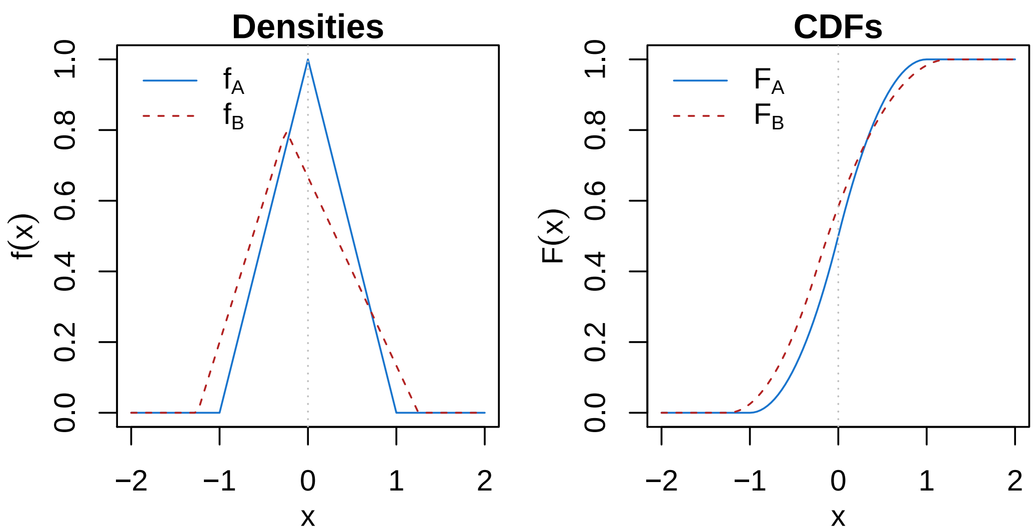

Figure 2: Triangular model densities and distribution functions. In this example distribution $F_A \succeq_{LASD} F_B$ (in terms of the description in the text, $\epsilon = 1/4$ for distribution B). Heuristically, the higher gains under policy B are outweighed by the probability of larger losses so that distribution A dominates distribution B in the LASD sense, but $F_A \not\succeq_{FOSD} F_B$.

approximate boundary point. Figure 4 depicts the $T_3(G(\mu_B))$ function for $\mu_B = 2.7, 2.8$ or 2.9. Tests are designed to detect the positive deviation in the right-most panel of the figure, when $T_3(G)(x) > 0$ for some $x \geq 0$.

Figure 5 shows empirical rejection probabilities for tests with three independent normal distributions. The tests are not conducted under any assumptions about the independence of the samples. The rejection probabilities are different than those in the point-identified experiments — more evidence is needed to detect deviations from the null region than in the identified case, because the bound U_B combines observations from the control and sample B. Although more information is necessary, it is important to note that these alternatives (like in the other experiments) are local to the boundary of the $\mathcal{P}_0^{nec}$ set.

As can be seen in Figure 5, the tests in the partially identified case do not reject the null with as high a probability as in the point identified case, which is a direct result of the lack of knowledge about inter-sample correlations that dictates the form of the T_3 function defined in the main text. Also, it appears as though these deviations from the null are not very well detected by the Cramér-von Mises tests in relation to the Kolmogorov-Smirnov tests. However, it is important to note that in this example, alternatives are local alternatives, and represent smaller and smaller deviations from the null region as sample sizes increase.

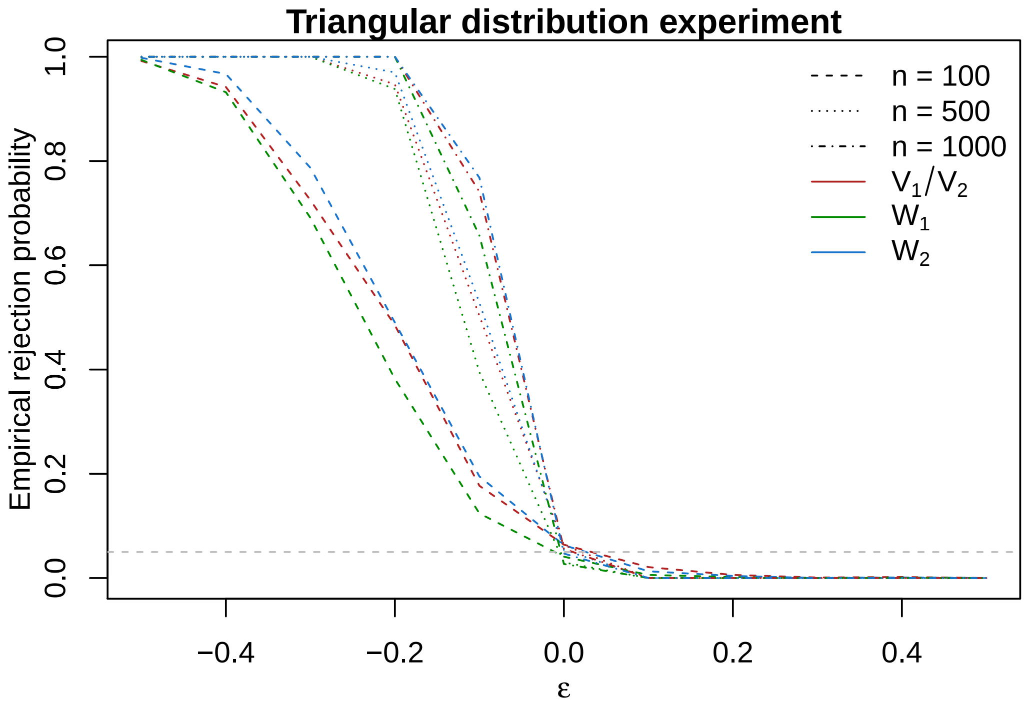

Figure 3: Empirical rejection probabilities of the LASD tests in the point identified triangular model experiment. The tests are of nominal 5% size, should have exactly 5% rejection probability when $\epsilon = 0$ and should reject when $\epsilon < 0$. Samples of sizes 100, 500 and 1000 correspond respectively to 499, 999 and 1999 bootstrap repetitions. Distributions are around the boundary of the null region, which is where $\epsilon = 0$, but plotted on an absolute, not local, scale. 1000 simulation repetitions.

2 Application

In this section we present the additional test results for the empirical application discussed in Section 5 of the main paper. Table 1 includes results of V_{1n} statistics, the second Table 2 contains V_{2n} statistics, Table 3 the third contains W_{1n} statistics and finally Table 4 reproduces the table of W_{2n} results used in the main text. The tables reveal that all the tests have very similar qualitative conclusions. Some of the entries are exactly the same across tables and are indeed repetitions of the same tests, but the tables are shown in entirety to facilitate comparison.

Finally, we note that the example could be used to conduct tests under partial identification, as if we had no knowledge of the longitudinal structure of the data. However, tests using V_{3n} or W_{3n} statistics were all identically zero and had p-values equal to 1, and the table of corresponding results is omitted.

	LASD in changes			FOSD in levels		
	$F_{AFDC} \succeq F_{JF}$	$F_{JF} \succeq F_{AFDC}$	equality	$G_{AFDC} \succeq G_{JF}$	$G_{JF} \succeq G_{AFDC}$	equality
avg-RA	4.3714	0.3283		4.1198	0.7006	
p-value	0.0180	0.8954		0.0005	0.7599	
lastQ-RA	1.5022	1.3864	1.7449			
p-value	0.3757	0.4012	0.3832			
avg-TL	0.1150	12.9315		2.5374	1.5446	2.5374
p-value	0.8914	0.0000		0.0390	0.2896	0.0655
lastQ-TL	0.7333	7.7585				
p-value	0.4682	0.0005				

Table 1: Table of sup-norm tests. LASD tests use the T1 process.

	LASD in changes			FOSD in levels		
	$F_{AFDC} \succeq F_{JF}$	$F_{JF} \succeq F_{AFDC}$	equality	$G_{AFDC} \succeq G_{JF}$	$G_{JF} \succeq G_{AFDC}$	equality
avg-RA	4.3714	0.3283		4.1198	0.7006	
p-value	0.0180	0.8954		0.0005	0.7599	
lastQ-RA	1.5022	1.3864	1.7449			
p-value	0.3757	0.4012	0.3832			
avg-TL	0.1150	12.9315		2.5374	1.5446	2.5374
p-value	0.8914	0.0000		0.0390	0.2896	0.0655
lastQ-TL	0.7333	7.7585				
p-value	0.4682	0.0005				

Table 2: Table of sup-norm tests. LASD tests use the T2 process.

	LASD in changes			FOSD in levels		
	$F_{AFDC} \succeq F_{JF}$	$F_{JF} \succeq F_{AFDC}$	equality	$G_{AFDC} \succeq G_{JF}$	$G_{JF} \succeq G_{AFDC}$	equality
avg-RA	3.4014	0.1789		2.8028	0.2240	
p-value	0.0510	0.9060		0.0070	0.8609	
lastQ-RA	0.7836	1.7883	2.1269			
p-value	0.6333	0.4237	0.6413			
avg-TL	0.0709	7.7449		1.2789	2.1285	2.4831
p-value	0.9290	0.0000		0.3387	0.1446	0.2081
lastQ-TL	0.6751	5.1723				
p-value	0.6453	0.0315				

Table 3: Table of L2-norm tests. LASD tests use the T1 process.

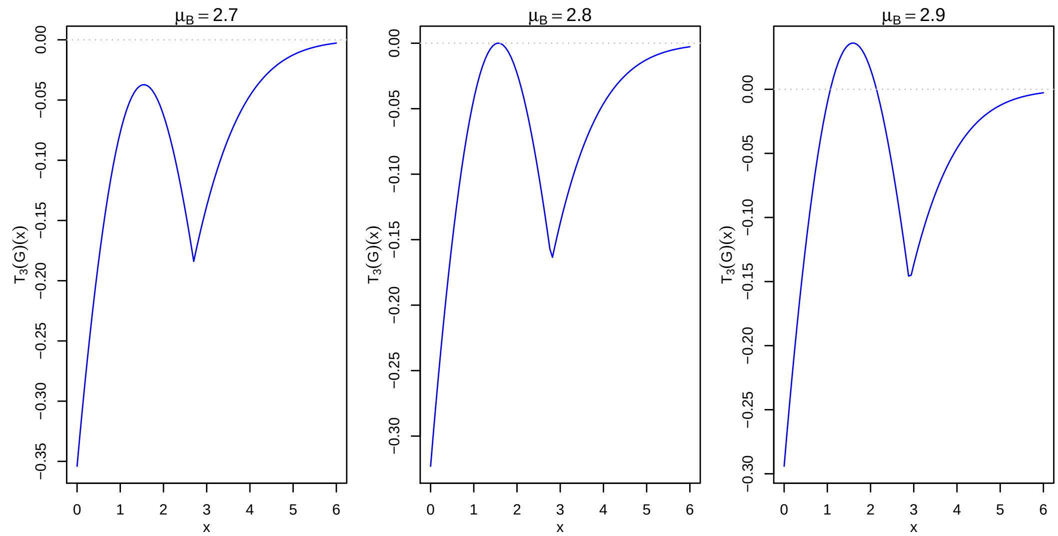

Figure 4: The $T_3(G(\mu_B))$ function for different values of the location of the marginal distribution function G_B. Tests should reject the null hypothesis when $T_3(G)(x) > 0$ for some x as in the right panel.

	LASD in changes			**FOSD in levels**		
	$F_{AFDC} \succeq F_{JF}$	$F_{JF} \succeq F_{AFDC}$	equality	$G_{AFDC} \succeq G_{JF}$	$G_{JF} \succeq G_{AFDC}$	equality
avg-RA	3.4335	0.1790		2.8028	0.2240	
p-value	0.0500	0.9055		0.0070	0.8609	
lastQ-RA	0.8000	2.1583	2.1269			
p-value	0.6273	0.3637	0.6413			
avg-TL	0.0858	9.1380		1.2789	2.1285	2.4831
p-value	0.9150	0.0000		0.3387	0.1446	0.2081
lastQ-TL	0.8238	5.8269				
p-value	0.5963	0.0175				

Table 4: Table of L2-norm tests. LASD tests use the T2 process.

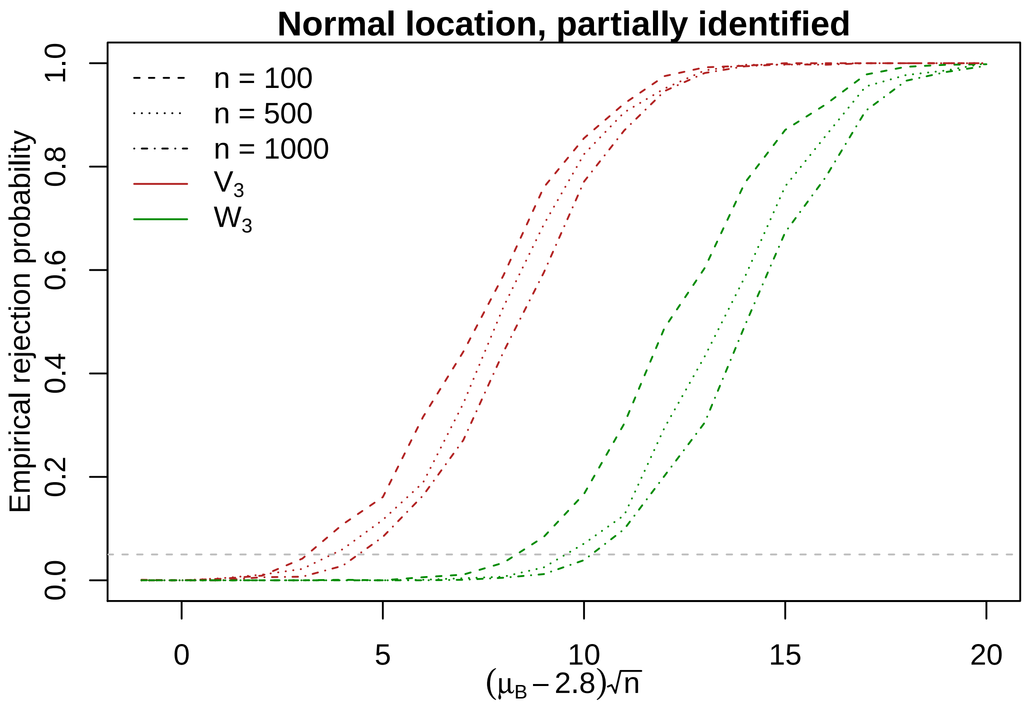

Figure 5: Empirical rejection probabilities of the LASD tests in the partially identified normal location model experiment. The control and policy A distributions have means set to zero, while the location of policy B is allowed to vary. The tests are of nominal 5% size, should have exactly 5% rejection probability when $(\mu_B - 2.8)\sqrt{n} = 0$ and should reject when $(\mu_B - 2.8)\sqrt{n} > 0$ (alternatives are local to the boundary of the set $\mathcal{P}_0^{nec}$ described in the text). Samples of sizes 100, 500 and 1000 correspond respectively to 499, 999 and 1999 bootstrap repetitions. 1000 simulation repetitions.