

Ranking Policies Under Loss Aversion and Inequality Aversion

Martyna Kobus*, Radosław Kurek†, Thomas Parker‡

Abstract

Strong empirical evidence from laboratory experiments, and more recently from population surveys, shows that individuals, when evaluating their situations, pay attention to whether they experience gains or losses, with losses weighing more heavily than gains. The electorate's loss aversion, in turn, influences politicians' choices. We propose a new framework for welfare analysis of policy outcomes that, in addition to the traditional focus on post-policy incomes, also accounts for individuals' gains and losses resulting from policies. We develop several bivariate stochastic dominance criteria for ranking policy outcomes that are sensitive to features of the joint distribution of individuals' income changes and absolute incomes. The main social objective assumes that individuals are loss averse with respect to income gains and losses, inequality averse with respect to absolute incomes, and hold varying preferences regarding the association between incomes and income changes. We translate these and other preferences into functional inequalities that can be tested using sample data. The concepts and methods are illustrated using data from an income support experiment conducted in Connecticut.

Keywords: Loss aversion, inequality aversion, stochastic dominance, bootstrap inference

JEL class: D04, C14

*Institute of Economics, Polish Academy of Sciences, mkobus@inepan.waw.pl.

†Institute of Economics, Polish Academy of Sciences, radek.kurek@inepan.waw.pl.

‡Department of Economics, University of Waterloo, tmparker@uwaterloo.ca.

1 Introduction

When assessing feasible policy changes, policy makers frequently face the challenge of balancing benefits for those who gain against the setbacks experienced by those who lose. A key insight from the behavioral economics literature (e.g. Tversky and Kahneman, 1991) is that individual welfare is dependent not only on the current state but also on the change in states. Prospect theory, originally developed to explain individual choices in uncertain situations but soon extended to riskless settings (Thaler, 1980) posits that people base their decisions on whether they perceive the choice as leading to a gain or a loss¹, and that losses carry more weight than equivalent gains (i.e. loss aversion) (Kahneman and Tversky, 1979).

Loss aversion is today one of the best-documented concepts in the literature. A global study published in *Nature Human Behaviour* (Ruggeri et al., 2020) repeats Kahneman and Tversky’s original 1979 experiment in 19 countries and 13 languages and confirms that it is broadly replicable. More recently, new evidence has emerged based on large representative population samples rather than laboratory experiments. Blake et al. (2021) confirm that respondents in a UK survey are loss averse. They also confirm another prediction of the theory, namely that individuals are inequality averse for gains but inequality loving (or risk-seeking) for losses.² These observations hold not only for the full sample, but also for every subsample of the data: for both men and women, at any age, income or level of education. Similarly, based on a large demographically representative survey in eight European countries, Meissner et al. (2023) find that respondents are, on average, loss averse and weigh losses about twice as much as gains of the same size, an estimate that is consistent with the result of a meta-analysis of 150 laboratory and field studies by Brown et al. (2024). Higher values are found for France, Italy, Sweden, Germany and the UK, and lower values for Poland, Romania and Spain. Chapman et al. (forthcoming) confirm the asymmetry between gains and losses in the preferences of the US population.³

Given such strong empirical evidence, loss aversion has been invoked to explain a wide range of phenomena such as the endowment effect (Thaler, 1980), the status quo bias (Samuelson and Zeckhauser, 1988), labor supply (Camerer et al., 1997; Dunn, 1996; Kőszegi and Rabin, 2006) and

¹This depends on the reference point and of course stands in contrast to the assumptions of consistency or invariance in preferences made in rational choice theory.

²This is called the *S-shapedness* of the value function and we use this property in Section 3.

³On the other hand, they also find a higher prevalence of loss tolerance in the population than in laboratory experiments.

job search (DellaVigna et al., 2017), the equity premium puzzle (Benartzi and Thaler, 1995), tax evasion (Dhami and al Nowaihi, 2010; Engström et al., 2015; Rees-Jones, 2018), price setting by firms (Ahrens et al., 2017), incumbency biases in elections (Quattrone and Tversky, 1988), consumption decisions (Bowman et al., 1999; Köszegi and Rabin, 2006), the timing of retirement (Seibold, 2021), and many others. It is thus natural that loss aversion, as an inherent feature of human preferences, should be incorporated into the welfare analysis of policy outcomes.

Moreover, loss aversion ought to be considered important for political economy reasons as well. Mounting evidence indicates that loss aversion of the electorate determines policy choices. By incorporating it into a model of trade policy determination, Freund and Özden (2008) are able to explain why industries experiencing losses are more likely to be protected and why existing protectionist policies are persistent. In their framework, as in this paper, society is made up of individuals who exhibit loss aversion and the government takes this into account when maximising social welfare. Consequently, the government is concerned about constituencies who would suffer losses from a policy change. In a similar setting, Tovar (2009) finds that loss aversion of the electorate, if large enough, may be the reason for the anti-trade bias puzzle (Rodrik, 1995). Alesina and Passarelli (2019) show that voters' preference reversals towards policies such as the Affordable Act, the Smoke Free Air Act or carbon taxes can be explained by loss aversion. Rules and institutions that are overly protective and difficult to change may be designed by legislators due to the population's loss aversion (Attanasi et al., 2017).

Beyond the need to account for an important aspect of individual preferences, and aside from political economy considerations, policy makers may also have purely normative reasons for incorporating loss aversion into policy evaluation. As an example of this, the recent pandemic sparked a significant debate on the ethical principles guiding vaccine development (Kahn et al., 2020; Solbakk et al., 2021). Eyal (2020) argues that the Hippocratic maxim of 'first, do no harm' prompted US and EU policy makers to postpone the development of Covid-19 vaccines by refusing to authorize human challenge trials, in which a small group of volunteers would have been deliberately exposed to the virus. In this choice, policy makers prioritized the risk of harm to a few individuals over the potentially enormous gains from speeding up the delivery of vaccines to broad segments of society.

Firpo et al. (forthcoming) develop methods to rank policies in a way that is sensitive to individuals' loss aversion. However, their framework assumes that individuals do not care about incomes at all and focus exclusively on income changes. It seems more realistic to assume that both dimen-

sions matter. Indeed, the behavioral economics literature postulates (see, e.g. Kőszegi and Rabin, 2006) that the extended utility function—commonly referred to as the value function—depends both on outcomes and on gains and losses.⁴ Moreover, a policy that is optimal with respect to gains and losses may redistribute income in a way that perpetuates existing income inequality to socially unacceptable levels. Thus, accounting for inequality aversion with respect to incomes, the traditional concern of welfare analysis, remains indispensable.

We thus significantly extend the framework considered in Firpo et al. (forthcoming). Namely, we consider a bivariate setting in which not only income changes matter, but also final incomes. A policy generates a distribution of incomes and, since it is always preceded by another distribution, it also generates a distribution of gains and losses. The standard approach in the welfare analysis (Atkinson, 1970) considers incomes alone. The approach of Firpo et al. (forthcoming) considers changes only. We combine the two to obtain a more realistic setting for individual and policymaker preferences. This can alter the ranking of distributions obtained from considering only a single dimension. For example, suppose one distribution appears preferable in terms of losses, but primarily because the losses of richer individuals are smaller. Once income inequality aversion is taken into account, that distribution may no longer be preferred; the alternative distribution may be even favored over some range of values (of incomes and of income gains/losses). Conversely, consider a tax reform that reduces inequality—an outcome desirable for any inequality-averse decision maker. What still matters is how this reduction is achieved. If it results from substantial gains accruing to a few winners at the lower end of the income distribution, while many others in the same income range experience small losses, then the inequality reduction effect appears less persuasive. The proportions of winners and losers, and the magnitudes of their respective gains and losses, matter also for the political economy reasons already discussed. The bivariate joint framework makes it possible to capture not only the policy maker’s preferences for equality and loss aversion per se, but also preferences regarding how the two interact. For example, the policy maker may prefer policies for which losses are concentrated among high-income individuals rather than among those with low incomes. This highlights dependence, a distinctive feature of multivariate frameworks as opposed to univariate ones. Overall, adopting a bivariate perspective substantially broadens the scope of policy welfare comparisons.

⁴A common functional form is the sum of a function that depends on outcomes and a function that depends on gains/losses (see O’Donoghue and Sprenger, 2018, for a review). This form is a special case of the classes of value functions considered in this paper (see Corollary 2.1).

We follow the social choice tradition of ranking the welfare resulting from policies by means of a social welfare function. In line with this literature (see e.g. Dalton, 1920; Sen, 1970; Gajdos and Weymark, 2012), the welfare function expresses the preference of a social decision maker who uses a utility function to transform individual outcomes into an interpersonally comparable measure of well-being. To take account of certain regularities in individual preferences, the utility function reflects qualitative properties such as the diminishing marginal utility of income or, in the context of this paper, loss aversion. We are therefore concerned with classes of utility functions and of social welfare functions. To be more precise, we refer to the class of welfare functions as a social value function, because instead of an individual utility function depending on income, there is an extended utility function (as mentioned, called a ‘value function’) depending on both absolute income and income gains/losses. Such extended utility functions, with specific functional forms, are used in models of reference-dependent preferences (e.g. Kőszegi and Rabin, 2006; O’Donoghue and Sprenger, 2018).

In such a setting, we develop bivariate dominance criteria to rank joint distributions of income and income gains and losses. These criteria are called dominance criteria because they hold for whole classes of value functions, thus making the ranking of policies or distributions robust. The theorems below link value functions from a specific class to a functional inequality that can be tested with observable features of the distribution of income and income changes. Such results take us from the realm of the unobservable (utility or value functions) to the realm of implementable conditions.

In standard risk and inequality comparisons, if the utility function is non-decreasing and concave, expected utility or utilitarian welfare is higher in an income distribution that stochastically dominates another distribution at the second order (Rothschild and Stiglitz, 1970; Atkinson, 1970). Similarly, in the main result of the paper (Theorem 4.1), we assume another set of conditions for the value function: that it is loss averse, inequality averse for incomes, and has other properties that can be linked to the so called higher-order risk preferences known from the literature on risk measurement (Eeckhoudt and Schlesinger, 2013). The literatures on the measurement of risk and inequality measurement share a common analytical structure and concepts and results in one literature are used in the other (Rothschild and Stiglitz, 1970; Atkinson, 1970; Gajdos and Weymark, 2012). The first property is an aversion to positive associations between income and income changes, which relates to correlation aversion in risk measurement (Eeckhoudt et al., 2007). It expresses a

preference for less positively correlated incomes and income changes. The bivariate distribution in which there is higher likelihood for low (high) outcomes to be paired with large gains (high losses) is preferred to the distribution when the opposite happens. The next property is a preference for the association aversion to be larger at the bottom of the distribution (for losses and low incomes) than at the top (for high gains and high incomes). This is equivalent to cross-prudence in risk (Eeckhoudt et al., 2007). Here it means that a policy maker's preference for equality in one dimension is stronger the lower the value of the other dimension. He/She prefers more equal outcomes among those who have lost than among those who have gained and a more equal distribution of gains among the poorer than among the richer. Final property is related to cross-temperance in risk preferences (Eeckhoudt et al., 2007). Here it means that a policy maker's preference for equality in one dimension is stronger the lower the degree of equality of the other dimension.

Given this set of qualitative features, expected value is higher in the distribution that dominates in terms of a criterion — a specified set of testable conditions — based on the joint distribution of income and income change induced by a policy. These qualitative features of value functions arise from the joint nature of our setting. This is also reflected in the dominance criteria, which combine Firpo et al.'s criterion for the univariate distribution of gains/losses, second-order stochastic dominance for the univariate distribution of income, and a dominance criterion for the integral of the joint distribution. The opposite result, on the other hand, uses the class of value functions that reverse the sign of the mentioned higher-order preferences and is given in Theorem 4.2.

Apart from the main result, which involves both loss aversion and inequality aversion, we also develop other results. In Theorem 2.1 we combine loss aversion with first-order stochastic dominance so that a better distribution is the one that has lower losses/higher gains and at the same time higher incomes (including higher mean incomes). It also has a lower association between gains/losses and incomes. The opposite result, favoring a higher association, provides a related dominance criterion based on survival dominance (Theorem 2.2). Association is typically not taken into account in models of reference-dependent preferences, e.g. the value function used in Kőszegi and Rabin (2006) is additive. Therefore, the necessary and sufficient conditions for dominance induced by this function are only a better distribution of gains/losses (in the sense of loss-aversion-sensitive dominance as in Firpo et al. (forthcoming)) and a better distribution of income (in the sense of first-order stochastic dominance). Another set of results combines inequality aversion for incomes with inequality aversion for gains but inequality loving for losses. Theorem 3.1 provides a relevant dominance criterion, which

combines second-order stochastic dominance with the criterion developed by Linton et al. (2005) for S-shaped value functions. The class of value functions should also be association-averse, cross-prudent and cross-temperate, or association-loving, cross-imprudent and cross-intemperate if one considers a parallel result that reverses the sign of higher-order derivatives of the value function (Theorem 3.2).

All the criteria described above can be translated into sets of functional inequalities, in a manner analogous to the relationship between stochastic dominance and the functional inequality between distribution functions that is equivalent to dominance. The dominance criteria all take into account more qualitative information on preferences, and as a result are not as simple as the typical stochastic dominance relationship, but tests can be designed that work the same way. Our tests are similar in spirit to Linton et al. (2010) and involve the estimation of “contact sets”, that is, (gain, income) pairs where the outcomes of two policies seem to be very similar. The distribution of test statistics related to the inequalities can be expressed elegantly in the language of directionally-differentiable maps from distribution functions to test statistics, and allows us to borrow a special bootstrap method from Fang and Santos (2019) to conduct inference. As shown in Section 5, tests based on this bootstrap method are consistent against fixed alternatives and control size uniformly in the null region (the collection of probability measures where a dominance hypothesis is satisfied).

The tests are illustrated using data from the experimental evaluation of a well-known welfare policy reform in the US in order to compare two policies that affect income distribution and generate gainers and losers. Specifically, we compare Aid to Families with Dependent Children (AFDC) with Jobs First (JF), the policy that replaced AFDC in Connecticut in the 1990s. Although Jobs First offered more generous income support than AFDC, it had a strict time limit beyond which no support was available. The evaluation randomly assigned households to either AFDC or Jobs First. These data were used by Bitler et al. (2006), who showed that, although the mean impact of Jobs First was positive, the policy created both winners and losers, meaning that its overall evaluation was not as straightforward as the mean impact suggested. Bitler et al. (2006) used final incomes as an outcome. In a richer framework, where individuals care not only about their final income but also about the changes induced by the two policies, Jobs First is not the favored policy. Specifically, according to six dominance criteria developed in the paper, the hypothesis that JF is a dominant policy is always rejected, while the hypothesis that AFDC is a dominant policy is never rejected. More precisely, AFDC almost first-order dominates JF with respect to income

changes, and our dominance criteria for income changes follow from this. We use these data only for illustrative purposes, namely to demonstrate the testing of the six criteria. Since the dominance criteria are linked to social welfare functions, one can conclude that JF is not a favored policy, and this conclusion holds across broad families of welfare functions. The joint evaluation further shows that JF's advantage over AFDC is concentrated primarily among higher-income households with large gains, whereas AFDC provides a more favorable distribution elsewhere and, in particular, entails a lower risk of small losses across the board.

Our framework can be viewed as a framework of stochastic dominance (Levy, 2016), and our results contribute directly to the stochastic dominance literature. The distributions being compared need not result from policy interventions; they may be any distributions, such as lotteries. Since our conditions for value functions can be interpreted as risk preferences, the results yield implementable criteria for evaluating lotteries in terms of these preferences. However, unlike in classic stochastic dominance, the dimensions are treated asymmetrically, with different conditions applying to each of them.

Beyond the literatures on loss aversion in the behavioral economics, risk measurement and stochastic dominance, this paper also relates to the normative evaluation of tax policies. Traditionally, such evaluation is conducted by comparing the post-tax income distribution with the pre-tax distribution. However, this neglects the potential reranking of individuals induced by a tax reform (Aronson et al., 1994). The issue of reranking is closely linked to horizontal equity, which requires that a fair tax system treat equals equally (Musgrave, 1959). According to this principle, a tax reform should preserve the utility ranking of individuals (Feldstein, 1976; King, 1983). Such considerations naturally lead to a bidimensional framework, in which both the final distribution and the initial status quo distribution are taken into account (Auerbach and Hassett, 2002; Bourguignon, 2011; Slesnick, 1989). For example, Bourguignon (2011) focuses on the joint distribution of status quo incomes and income changes, and shows that in this setting sequential stochastic dominance, as developed by Atkinson and Bourguignon (1987), can be applied to compare distributions. In that literature, differently than in this paper, the same dominance criteria applied to incomes are applied to income changes, with no attention to loss aversion or the S-shapedness of the value function that are the focus of this paper. Moreover, the status quo distribution plays a special role in that literature (a point that has been criticized, see e.g. (Kaplow, 1989, 1995)). In our setting it is not necessary that the compared policies start from the same status quo. Here the distribution

of gains and losses is important per se, rather than rerankings.

The paper is organized as follows. The next three sections each develop the results linking qualitative features of utility functions to testable criteria based on joint distribution functions: Section 2 (Theorem 2.1 and 2.2), Section 3 (Theorem 3.1 and 3.2) and Section 4 (Theorem 4.1 and 4.2). We then relate the developed dominance conditions to functional inequalities and tests in Section 6, which also contains an empirical application. Section 7 concludes. The appendix at the end of the paper contains proofs of the theorems.

2 Loss Aversion Sensitive Bivariate dominance

Let X be a random variable that denotes gains and losses with cumulative distribution function F^1 and density function f^1 . Without loss of generality, let $\mathcal{X} = (-a_1, a_2) \subset \mathbb{R}$ denote the support of X .⁵ Further, let Z be a random variable that denotes outcomes in levels, with cumulative distribution function F^2 and density function f^2 . Let $\mathcal{Z} = [0, a_3) \subset \mathbb{R}_+$ denote the support of Z . Finally, let (X, Z) denote a random vector with joint cumulative distribution function F and joint density function f . Let $\mathcal{X} \times \mathcal{Z}$ denote its support and let \mathcal{F} denote the space of all bivariate distributions with support $\mathcal{X} \times \mathcal{Z}$.

We define a bivariate social value function as follows.

Definition 2.1 (Social Value Function (SVF)). *Let $W : \mathcal{F} \rightarrow \mathbb{R}$ denote the social value function*

$$W(F) = \int_{\mathbb{R} \times \mathbb{R}^+} v(x, z) dF(x, z), \quad (1)$$

where $v : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}$ is called a value function.

Throughout the paper we consider various properties of the value function that define classes of SVF. We start with the following ones.

Definition 2.2 (Loss aversion sensitive value function). *The value function $v : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}$ is differentiable and satisfies:*

- *Disutility of losses and utility of gains:* $v(-x, z) \leq 0 \leq v(x, z)$ for all $x > 0, z$;
- *Non-decreasing:* $\frac{\partial}{\partial x} v(x, z) \geq 0, \frac{\partial}{\partial z} v(x, z) \geq 0$ for all x, z ;

⁵For brevity, we often write $\int_{-\infty}^{\infty}$ instead of $\int_{-a_1}^{a_2}$.

- *Loss-averse*: $\frac{\partial}{\partial x}v(-x, z) \geq \frac{\partial}{\partial x}v(x, z)$ for all $x > 0, z$;
- *Association averse (submodular)*: $\frac{\partial^2}{\partial x \partial z}v(x, z) \leq 0$ for all x, z .

The first three conditions regarding gains and losses in the Definition 2.2 are standard requirements in prospect theory (Tversky and Kahneman, 1991): (i) losses hurt (bring negative utility) and gains bring utility, (ii) higher outcomes and higher gains (or smaller losses) are better, (iii) losses hurt more than gains of the same value. The fourth property is an aversion to positive associations between outcomes and gains/losses. That is, it is better to have more individuals with low outcomes but large gains and individuals with large outcomes but high losses than to have more individuals with both high outcomes and high gains or low outcomes and large losses. Association aversion is typically assumed in multivariate welfare and inequality measurement (Atkinson and Bourguignon, 1982), where two dimensions are typically two goods, e.g. income and life expectancy. It is understood as a preference for bringing individuals *multidimensionally* closer together.

Definition 2.3 (Loss Aversion Sensitive Bivariate Dominance). *Let (X_A, Z_A) and (X_B, Z_B) have cumulative distribution functions respectively labeled $F_A, F_B \in \mathcal{F}$. 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 Bivariate Dominance, or LASBD for short, and we write $F_A \succsim_{LASBD} F_B$.*

Theorem 2.1. *Suppose that $F_A, F_B \in \mathcal{F}$. The following are equivalent:*

1. $F_A \succsim_{LASBD} F_B$;
2. For all $x \geq 0, z$, F_A, F_B satisfy

$$F_B^1(-x) - F_A^1(-x) \geq \max\{0, F_A^1(x) - F_B^1(x)\} \quad (2)$$

$$F_A^2(z) \leq F_B^2(z) \quad (3)$$

and for $x \neq a_2$ and $z \neq a_3$

$$F_A(x, z) \leq F_B(x, z); \quad (4)$$

3. For all $x \geq 0, z$, F_A, F_B satisfy (3) and (4) and the following conditions:

$$F_A^1(-x) \leq F_B^1(-x) \quad (5)$$

$$(1 - F_A^1(x)) - F_A^1(-x) \geq (1 - F_B^1(x)) - F_B^1(-x). \quad (6)$$

Theorem 2.1 is a natural extension of Firpo et al. (forthcoming) to a bivariate setting. It states that ranking policy interventions (or distributions in general) over the class of social value functions, as in Definition 2.2, is equivalent to the LASD dominance condition for gains and losses (2) used in Firpo et al. (forthcoming) (which, given their further results, is equivalent to (5) and (6)), first-order stochastic dominance for outcomes (3), and bivariate first-order stochastic dominance for the joint distribution of gains/losses and outcomes (4). LASD is a consequence of loss aversion, and the rest is a consequence of the non-decreasingness of the value function; a well-known result is that for bivariate outcomes non-decreasingness and submodularity of the utility function is equivalent to bivariate first-order stochastic dominance (see for example Levy, 2016). Please note that for gains and losses, the condition does not follow from applying the joint condition to the marginal distribution, as in standard stochastic dominance. In this setting, preferences are asymmetric across the two dimensions, and consequently a weaker property than first-order dominance is required for income changes.

A value function depending not only on the level of consumption but also on gains and losses is considered by Kőszegi and Rabin (2006). They postulate a simple additive form, for which the dominance conditions degenerate to univariate conditions only, as the following corollary shows.

Corollary 2.1. *When the function v is as in Kőszegi and Rabin (2006): $v(x, z) = v_1(x) + v_2(z)$, then conditions (2) and (3) in Theorem 2.1 are necessary and sufficient for $F_A \succsim_{LASBD} F_B$.*

While the first three properties in Definition 2.2 are building blocks of prospect theory, let us check what happens when we reverse the fourth property of submodularity of the value function. In a multivariate welfare and inequality measurement literature there are arguments for favoring higher association in some cases, for example, when goods are complements rather than substitutes (Bourguignon and Chakravarty, 2003). The appropriate definitions and dominance criteria are then the following.

Definition 2.4 (Loss Aversion Sensitive Bivariate Dominance 2). *Let (X_A, Z_A) and (X_B, Z_B) have cumulative distribution functions respectively labeled $F_A, F_B \in \mathcal{F}$. If $W(F_A) \geq W(F_B)$ for all value functions v that satisfy Definition 2.2, except that the last property of v is association loving ($\frac{\partial^2}{\partial x \partial z} v(x, z) \geq 0$, that is, v is supermodular) we say that F_A dominates F_B in terms of Loss Aversion Sensitive Bivariate Dominance 2, or LASBD2 for short, and we write $F_A \succsim_{LASBD2} F_B$.*

Theorem 2.2. *Suppose that $F_A, F_B \in \mathcal{F}$. Let $K(x, z) = F^1(x) + F^2(z) - F(x, z)$. The following are equivalent:*

1. $F_A \succsim_{LASBD2} F_B$;
2. For all $x \geq 0, z$, F_A, F_B satisfy (2), (3) and for $x \neq a_2$ and $z \neq a_3$

$$K_A(x, z) \leq K_B(x, z); \tag{7}$$

3. For all $x \geq 0, z$, F_A, F_B satisfy (5), (6), (3) and (7).

Compared to Theorem 2.1, Theorem 2.2 has a different condition (7). In (4) the integration is performed over rectangles $(-a_1, x) \times (0, z)$ and the condition is that the cumulative distribution function of the dominant distribution is everywhere less than or not greater than that of the dominated distribution. In contrast, (7) is equivalent to the condition that the cumulative distribution of the dominant distribution over the rectangles $(x, a_2) \times (z, a_3)$ is everywhere greater than or equal to that of the dominated distribution. Therefore, dominance using K_A, K_B is equivalent to dominance using a bivariate survival function. That is,

$$K_A(x, z) \leq K_B(x, z) \iff \int_x^{a_2} \int_z^{a_3} dF_A(x, z) \geq \int_x^{a_2} \int_z^{a_3} dF_B(x, z).$$

3 Inequality Aversion Sensitive Dominance

The LASBD dominance condition is of the ‘first order type’. It is similar to standard first-order stochastic dominance except that it accounts for loss aversion. Now we are interested in imposing more structure, which involves the consideration of inequalities induced by policy. Thus, the condition developed is of the ‘second order type’. In particular, the social planner will now be averse to inequality of outcomes and gains, which implies concavity of the value function. For losses, which are negative, the relevant requirement is the opposite, convexity. Overall, the value function is concave for outcomes and S-shaped for gains and losses. The latter is a standard second-order property of the value function in prospect theory (Linton et al., 2005).

Definition 3.1 (Inequality aversion sensitive value function). *The value function $v : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}$ is differentiable and satisfies:*

- *Disutility of losses and utility of gains:* $v(-x, z) \leq 0 \leq v(x, z)$ for all $x > 0, z$;
- *Non-decreasing:* $\frac{\partial}{\partial x}v(x, z) \geq 0, \frac{\partial}{\partial z}v(x, z) \geq 0$ for all x, z ;
- *S-shaped in x :* $\frac{\partial^2 v(x, z)}{\partial x^2} \leq 0$ and $\frac{\partial^2 v(-x, z)}{\partial x^2} \geq 0$ for all $x > 0$;
- *Concave in z :* $\frac{\partial^2 v(x, z)}{\partial z^2} \leq 0$;
- *Association averse (submodular):* $\frac{\partial^2}{\partial x \partial z}v(x, z) \leq 0$ for all x, z ;
- *Decreasingly submodular:* $\frac{\partial^3}{\partial x^2 \partial z}v(x, z) \geq 0$ and $\frac{\partial^3}{\partial x \partial z^2}v(x, z) \geq 0$;
- *Cross-temperate:* $\frac{\partial^4}{\partial x^2 \partial z^2}v(x, z) \leq 0$.

The S-shape of the value function implies a preference for inequality in losses, but an aversion to inequality in gains. It is a preference for equal distribution of gains and unequal distribution of losses, e.g. for all losses to be concentrated in the smallest group of individuals. New properties are decreasing submodularity and cross-temperance. Decreasing submodularity means that not only is a negative association between outcomes and gains/losses preferred (as in association aversion (submodularity)), but it is preferred especially at the bottom of the distribution, that is, among those who have high losses and low outcomes.

Decreasing submodularity concerns the behaviour of the cross derivative of x and z , but another way of looking at it is to consider the behaviour of the second order derivatives $\frac{\partial^2}{\partial x^2}v$ and $\frac{\partial^2}{\partial z^2}v$. Then the equivalent condition is that the value function is decreasingly concave for z and x and increasingly convex for $-x$ ($x > 0$). That is, $\frac{\partial^3}{\partial x \partial z^2}v(x, z) \geq 0$ means that $\frac{\partial^2}{\partial z^2}v$ increases with x . Since we know that $\frac{\partial^2}{\partial z^2}v$ is negative (i.e. the fourth condition), this means that it is less negative at higher x , so the degree of concavity decreases with higher gains. In the same way we analyse $\frac{\partial^2}{\partial x^2}v$ for when $x < 0$ and $x > 0$. Decreasing concavity means that inequality is particularly hurtful at the bottom of the distribution. That is, the social planner prefers more equal outcomes among those who have lost the most than among those who have gained the most, and he/she also prefers more equal distribution of gains among the poorer (i.e. those with low incomes) than among the richer. For losses, on the other hand, the social planner prefers a more unequal distribution among the richer than among the poorer. The decreasing concavity of the value function corresponds to cross-prudence in risk measurement (Eeckhoudt et al., 2007; Jokung, 2011). Cross-prudence is a preference for the disaggregation of two harms: a reduction in any attribute and the addition of zero

mean risk to any attribute. A cross-prudent individual prefers to experience risk in one attribute if the value of the other attribute is higher rather than lower. For example, he prefers monetary risk when his health is better than when both monetary risk and health deterioration occur together. Replacing risk with inequality, more inequality in dimension is tolerated among those who have more of the other dimension. A similar criterion is often invoked in the context of socioeconomic inequalities in health and is said to reduce socioeconomic inequalities in health (see e.g. Makdissi and Yazbeck, 2014). In our context, as mentioned above, the social planner prefers, for example, higher outcome inequality among the winners than among the losers.

The final property, cross-temperance, relates to the fourth-degree derivative. In risk measurement, this means that the decision maker wants to disaggregate risks in both income and gains/losses. In our context, the social planner also prefers to disaggregate inequality in both income and gains/losses. Therefore, his preference for greater equality in one dimension increases with an increasing inequality in the other dimension. With cross-prudence the stronger the preference for equality in one dimension the lower the value of the other dimension; with cross-temperance the stronger the preference for equality in one dimension the less equal the other dimension is.

Definition 3.2 (Inequality Aversion Sensitive Dominance). *Let (X_A, Z_A) and (X_B, Z_B) have cumulative distribution functions respectively labeled $F_A, F_B \in \mathcal{F}$. If $W(F_A) \geq W(F_B)$ for all value functions v that satisfy Definition 3.1, we say that F_A dominates F_B in terms of Inequality Aversion Sensitive Dominance, or IASD for short, and we write $F_A \succsim_{IASD} F_B$*

Theorem 3.1. *Suppose that $F_A, F_B \in \mathcal{F}$. Let $H(x, z) = \int_{-\infty}^x \int_0^z F(t, s) ds dt$, $H^1(x) = \int_{-\infty}^x F^1(t) dt$, $S^1(x) = \int_x^\infty 1 - F^1(t) dt$ and $H^2(z) = \int_0^z F^2(s) ds$. The following conditions are equivalent:*

1. $F_A \succsim_{IASD} F_B$;
2. For all $x > 0 > y, z$, F_A, F_B satisfy

$$S_A^1(x) - S_B^1(x) - (H_A^1(y) - H_B^1(y)) \leq S_A^1(0) - S_B^1(0) - (H_A^1(0) - H_B^1(0)), \quad (8)$$

$$H_A^2(z) \leq H_B^2(z), \quad (9)$$

and for $x \neq a_2$ and $z \neq a_3$

$$H_A(x, z) \leq H_B(x, z); \quad (10)$$

3. For all $x, z > 0 > y$, F_A, F_B satisfy (9), (10) and

$$\int_y^x F_A^1(t)dt \leq \int_y^x F_B^1(t)dt. \quad (11)$$

Theorem 3.1 states that the ranking of distributions induced by the class of social value functions with the properties described in Definition 3.1 is equivalent to second-order stochastic dominance ((9) and (10)), except that for gains and losses we obtain the dominance condition for S-shapedness of the value function (11). Linton et al. (2005) develop a test for this condition. As Levy and Wiener (1998, Theorem 4) point out, the (11) condition follows directly from considering gains (integral over $[0, x]$) and losses (integral over $[y, 0]$) separately, and assuming inequality aversion for the former and inequality loving for the latter. As Theorem 3.1 states it is equivalent to (8) for which we notice the following. Let X^+, X^- denote, respectively, gains and losses (i.e. the positive and negative values of X). Then $S^1(0) = E[X^+]$, that is, $S^1(0)$ is mean gain and $H^1(0) = E[X^-]$ is mean loss. Furthermore $S^1(0) - H^1(0) = E[X]$ is the mean of X . Thus (8) can be rewritten as

$$S_A^1(x) - S_B^1(x) - (H_A^1(y) - H_B^1(y)) \leq E[X_A] - E[X_B]$$

for all $x > 0 > y$. Similarly to $S^1(0)$ and $H^1(0)$, $S^1(x)$ can be interpreted as the average gain *above* x and $H^1(y)$ as the average loss *below* y . Thus, condition (8) states that for F_A to dominate F_B for gains and losses, it has to be that, for all x, y , the difference between the distributions' differences in average gain above x (i.e. $S_A^1(x) - S_B^1(x)$) and in average loss below y (i.e. $H_A^1(y) - H_B^1(y)$), is smaller than the difference in distributions' mean gains and losses (i.e. $E[X_A] - E[X_B]$). Moreover, when $E[X_A^-] = E[X_B^-]$ i.e. mean loss is the same for F_A and F_B , $\int_{-\infty}^y F_A^1(t)dt \geq \int_{-\infty}^y F_B^1(t)dt$ has to hold in the losses region for all y . This is a typical second order stochastic dominance condition but applied to the space of losses. For losses then, the distribution that yields lower welfare (F_B) second-order stochastically dominates the distribution that brings higher welfare (F_A). So the standard condition is reversed, which is not surprising given that S-shapedness means that losses are evaluated using a convex, not a concave function as is typically the case.

We have a parallel class of value functions with some of the preferences changed.

Definition 3.3 (Inequality Aversion Sensitive Dominance 2). *Let (X_A, Z_A) and (X_B, Z_B) have cumulative distribution functions respectively labeled $F_A, F_B \in \mathcal{F}$. If $W(F_A) \geq W(F_B)$ for all value*

functions v that satisfy Definition 3.1, with exception that they are

- Association loving (supermodular): $\frac{\partial^2}{\partial x \partial z} v(x, z) \geq 0$ for all x, z ;
- Decreasingly supermodular: $\frac{\partial^3}{\partial x^2 \partial z} v(x, z) \leq 0$ and $\frac{\partial^3}{\partial x \partial z^2} v(x, z) \leq 0$;
- Cross-intemperate (second-degree supermodular) $\frac{\partial^4}{\partial x^2 \partial z^2} v(x, z) \geq 0$;

we say that F_A dominates F_B in terms of Inequality Aversion Sensitive Dominance 2, or IASD2 for short, and we write $F_A \succsim_{IASD2} F_B$.

Compared to Definition 3.1, value functions in Definition 3.3 favor the association between gains(losses) and outcomes. That is, it is preferable to have more individuals in a society with high (resp. low) incomes and high gains (resp. high losses) than to have those for whom one dimension has higher value and the other has lower value. Decreasing supermodularity is equivalent to increasing concavity for z and x and increasing convexity for $-x$, where $x > 0$. Increasing concavity comes from the second order derivatives of z and $x > 0$, both of which are negative, being even more negative (i.e. increasingly concave) with x and z respectively. Cross-intemperance comes from the fact that the social planner prefers to aggregate inequalities in both dimensions and therefore his preference for equality in one dimension decreases with the degree of inequality in the other dimension.

Theorem 3.2. Suppose that $F_A, F_B \in \mathcal{F}$. Let $L(x, z) = \int_{-\infty}^x \int_0^z K(t, s) ds dt$. The following conditions are equivalent:

1. $F_A \succsim_{IASD2} F_B$;
2. For all $x > 0 > y, z$, F_A, F_B satisfy (8) and (9) and for $x \neq a_2$ and $z \neq a_3$

$$L_A(x, z) \leq L_B(x, z). \quad (12)$$

Since the properties of v for x and z are the same as in Definition 3.1, conditions (8) and (9) from Theorem 3.1 are also necessary and sufficient in Theorem 3.2. The change is in the condition for the joint (12), which is a second-order condition involving K (Theorem 2.2).

4 Loss and Inequality Aversion Sensitive Dominance

Our most interesting condition combines loss aversion in gains and losses with inequality aversion in outcomes, as they are the most often postulated preferences when it comes to gains/losses and outcomes. Furthermore, the value function is averse to the positive association of gains/losses and outcomes, particularly so at the bottom of the distribution. There is also a preference for the disaggregation of inequalities in gains/losses and outcomes.

Definition 4.1 (Loss and inequality aversion sensitive value function). *The value function $v : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}$ is differentiable and satisfies:*

- *Disutility of losses and utility of gains:* $v(-x, z) \leq 0 \leq v(x, z)$ for all $x > 0, z$;
- *Non-decreasing:* $\frac{\partial}{\partial x}v(x, z) \geq 0, \frac{\partial}{\partial z}v(x, z) \geq 0$ for all x, z ;
- *Loss-averse in x :* $\frac{\partial}{\partial x}v(-x, z) \geq \frac{\partial}{\partial x}v(x, z)$ for all $x > 0, z$;
- *Concave in z :* $\frac{\partial^2 v(x, z)}{\partial z^2} \leq 0$;
- *Association averse (submodular):* $\frac{\partial^2}{\partial x \partial z}v(x, z) \leq 0$ for all x, z ;
- *Decreasingly submodular:* $\frac{\partial^3}{\partial^2 x \partial z}v(x, z) \geq 0$ and $\frac{\partial^3}{\partial x \partial^2 z}v(x, z) \geq 0$;
- *Cross-temperate (second-degree submodular):* $\frac{\partial^4}{\partial^2 x \partial^2 z}v(x, z) \leq 0$.

As before, loss and inequality aversion can be formalized as a dominance concept.

Definition 4.2 (Loss and Inequality Aversion Sensitive Dominance). *Let (X_A, Z_A) and (X_B, Z_B) have cumulative distribution functions respectively labeled $F_A, F_B \in \mathcal{F}$. If $W(F_A) \geq W(F_B)$ for all value functions v that satisfy Definition 4.1, we say that F_A dominates F_B in terms of Loss and Inequality Aversion Sensitive Dominance, or LIASD for short, and we write $F_A \succsim_{LIASD} F_B$*

As with the previous classes of value functions, dominance in this class of value functions can be translated to a set of conditions on the distributions of the results of two policies.

Theorem 4.1. *Suppose that $F_A, F_B \in \mathcal{F}$. The following conditions are equivalent:*

1. $F_A \succsim_{LIASD} F_B$;
2. For all x, z , F_A, F_B satisfy (2) and (9) and for $x \neq a_2$ and $z \neq a_3$ they satisfy (10).

According to the previous results, Theorem 4.1 follows quite naturally. The dominance condition prescribed in Theorem 4.1 is the sum of the conditions for loss aversion in gains and losses and second-order stochastic dominance (9) and (10). That is, to rank policy interventions in a way that takes into account loss aversion in gains and losses and inequality aversion in final outcomes, it is necessary to check LASD dominance for gains/losses, second order stochastic dominance for outcomes, and bivariate second order stochastic dominance for the joint distribution.

We have a parallel class of social value function that preserves loss aversion and inequality aversion, but has different conditions for cross derivatives.

Definition 4.3 (Loss and Inequality Aversion Sensitive Dominance 2). *Let (X_A, Z_A) and (X_B, Z_B) have cumulative distribution functions respectively labeled $F_A, F_B \in \mathcal{F}$. If $W(F_A) \geq W(F_B)$ for all value functions v that satisfy Definition 4.1, with exception that it is*

- Association loving (supermodular): $\frac{\partial^2}{\partial x \partial z} v(x, z) \geq 0$ for all x, z ;
- Decreasingly supermodular: $\frac{\partial^3}{\partial x^2 \partial z} v(x, z) \leq 0$ and $\frac{\partial^3}{\partial x \partial z^2} v(x, z) \leq 0$;
- Cross-intemperate (second-degree supermodular) $\frac{\partial^4}{\partial x^2 \partial z^2} v(x, z) \geq 0$;

we say that F_A dominates F_B in terms of Loss and Inequality Aversion Sensitive Dominance 2, or LIASD2 for short, and we write $F_A \succsim_{LIASD2} F_B$.

For this class of functions we have the following result.

Theorem 4.2. *Suppose that $F_A, F_B \in \mathcal{F}$. The following conditions are equivalent:*

1. $F_A \succsim_{LIASD2} F_B$;
2. For all x, z , F_A, F_B satisfy (2) and (9) and for $x \neq a_2$ and $z \neq a_3$ they satisfy (12).

Naturally, Theorem 4.2 is a combination of previously used conditions, (2) as in Theorem 2.1 and (9) and (12) as in Theorem 3.2, as it combines loss aversion in x with inequality aversion in z and favors association in the joint distribution.

5 Statistical inference

We can test the systems of functional inequalities implied by the various dominance criteria. In this section we explain how to conduct these tests.

Each of the dominance concepts defined above is re-expressed using several conditions that describe the relationship between observable features of the dominant and dominated distributions. In particular, all the conditions can be written as functional inequalities, checking for rejection of a dominance hypothesis by checking the sign of the corresponding function.

For example, the first LASBD definition is equivalent to conditions on the joint and marginal distribution functions of distributions A and B as described in Theorem 2.1. To conduct inference about the dominance of distribution A over B , we convert $H_0^1 : F_A \succsim_{LASBD} F_B$ into an equivalent hypothesis on a set of functional inequalities provided by (in the case of the LASBD criterion) displays (3)-(6).

We search for deviations from the null by rearranging the conditions into functions that should be everywhere nonpositive, and search for arguments where that appears to be violated significantly. In the case of the first LASBD condition, we can define the test function

$$g(x, z) = g(F_A, F_B)(x, z) = g^{LASBD}(F_A)(x, z) - g^{LASBD}(F_B)(x, z)$$

where

$$g^{LASBD}(F)(x, z) = \begin{bmatrix} F^2(z) \\ F(-x, z) \\ F(x, z) \\ F^1(-x) \\ F^1(x) + F^1(-x) \end{bmatrix},$$

and the null hypothesis H_0^1 can be rewritten

$$H_0^2 : g(x, z) \leq \mathbf{0}_5 \quad \forall x, z \geq 0.$$

For reference, functions analogous to g^{LASBD} for all the dominance concepts discussed here are collected in the appendix.

We need a way to measure deviations from the hypothesized inequalities, that is, to find (x, z) pairs where it appears that at least one coordinate of the test function is positive. These functions can all be estimated in a straightforward way using plug-in estimates, that is, the empirical (joint and marginal) distribution functions from each observed sample. We define $\hat{g} = g(\hat{F}_A, \hat{F}_B)$, where

(\hat{F}_A, \hat{F}_B) are the empirical distribution functions for (F_A, F_B) . Under the null hypothesis, letting $[x]_+ = \max\{x, 0\}$ (applied coordinate-wise to vectors) and $\|\cdot\|$ be the L_2 norm, $\|[g]_+\| = 0$. Therefore we should have $\|[\hat{g}]_+\| \approx 0$, where the test statistic should only be positive due to random sampling.

The continuous mapping theorem implies that the test statistic $T_n = \|[\hat{g}]_+\|$ converges to zero in probability to zero under the null. However, the asymptotic distribution of T_n is intractable. Usually, one would turn to the bootstrap in this situation, but the pointwise map $x \mapsto [x]_+$ in the definition of the statistic complicates the distribution. We find it convenient to interpret the map $(F_A, F_B) \mapsto \|[g(F_A, F_B)]_+\|$ as a Hadamard directionally differentiable transformation of a pair of distributions into a real-valued statistic (Hadamard directional differentiability of similar maps was discussed extensively in Firpo et al. (2023, forthcoming)). This characterization allows for inference with a modified bootstrap procedure, as described in Fang and Santos (2019). There is other research that might apply to the testing of this problem: Lee et al. (2018) propose a general method for testing functional inequalities, and an alternative bootstrap is described by Hong and Li (2020). The method described below is tailored specifically to these tests.

The tests of all the dominance hypotheses work in the same way and can be described in general. We assume that the null hypothesis has been translated into a multivariate functional inequality in (x, z) , and that the test function g is nonpositive under the null that A dominates B in the chosen sense. We call its plug-in estimate \hat{g} and its estimate using a bootstrapped sample is labeled g^* . Then a hypothesis test is conducted this way:

1. Estimate \hat{g} and $T_n = \|[\hat{g}]_+\|$ using plug-in estimates of the distribution functions.
2. Use \hat{g} to estimate the contact set, that is, the collection of (x, z) such that $g(x, z) = 0$. Call the contact set estimator function $\chi_{c_n}(x, z) = I(|\hat{g}(x, z)| \leq c_n)$. Use $c_n \searrow 0$ such that $c_n \sqrt{n} \rightarrow \infty$ (in our empirical example, we use $c_n = 4 \log \log n / \sqrt{n}$, as suggested by simulation evidence in Linton et al. (2010)).
3. For each iteration $r = 1, \dots, R$ of the bootstrap, resample the data with replacement and calculate g_r^* and $T_r^* = \|[(g_r^* - \hat{g}) \cdot \chi_{c_n}]_+\|$.
4. Use the reference distribution $\{T_r^*\}_{r=1}^R$ for inference: for example, the bootstrap p-value is,

for arbitrarily small $\eta > 0$,

$$p^* = \frac{1}{R} \sum_{r=1}^R I(T_r^* + \eta > T_n).$$

The parts of this algorithm that are not typical of all bootstrap inference techniques are the parts involving contact sets in steps 2 through 4. The reasons behind steps 2 and 3 will be seen in Theorem 5.1 below, and the reason for the η in step 4 will become apparent after the statement of Theorem 5.2.

The formal results stating the consistency of this bootstrap procedure for inference with the loss and inequality averse dominance criteria rely on two minimal regularity assumptions that describe the sample data we assume to be observable.

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

A2 Define $n = n_A + n_B$. n_A and n_B increase such that $n_k/n \rightarrow \lambda_k$ as $n_A, n_B \rightarrow \infty$, where $0 < \lambda_k < 1$ for $k \in \{A, B\}$.

Under the minimal assumptions above, we can show that the bootstrap distribution is a consistent estimator of the limiting distribution of the test statistics for all the dominance concepts. In the following two statements, we use the following notation. We let BL_1 be the space of real-valued Lipschitz functions that are bounded by 1, which is a space of functions typically used to make statements about the weak convergence of a sequence of random variables to its distributional limit. The operators $P\{\cdot\}$ and $E[\cdot]$ denote the probability measure and expected value using the population distribution, while $P^*\{\cdot\}$ and $E^*[\cdot]$ refer to the counterparts using the distribution of the bootstrapped data conditional on the observed sample. The equality $X_n = o_P(1)$ implies that the sequence X_n converges in probability to zero as n diverges. Although many dominance concepts were defined and discussed above, the statistical analysis of tests used for all of the concepts is qualitatively identical, so we refer to all of them in the same way. They all use a sample test function $g(\hat{F}_A, \hat{F}_B)$ to learn about the population test function $g = g(F_A, F_B)$, where the individual coordinates of g may change with each dominance concept, and for each concept there is a set of distributions $\mathcal{F}_0 \subset \mathcal{F}$ that satisfy the null hypothesis, and make it so that $(F_A, F_B) \in \mathcal{F}_0$ implies

$T = \|[g]_+\| = 0$. We will refer to all test functions and all null collections as g and \mathcal{F}_0 in the theorems below, although they change with the particular notion of dominance.

Theorem 5.1. *Let T_n be any of the statistics described above for testing, that is, $T_n = \|[g(\hat{F}_A, \hat{F}_B)]_+\|$ for any of the g described in the appendix. Under conditions **A1** and **A2**, T_n converges weakly to $T = \|[g]_+\|$ in the sense that there exists a random variable \mathcal{T} such that*

$$\sup_{f \in BL_1} |\mathbb{E}[f(\sqrt{n}(T_n - T))] - \mathbb{E}[f(\mathcal{T})]| = o_P(1).$$

Similarly, if $T_n^* = \|[g(F_A^*, F_B^*) - g(\hat{F}_A, \hat{F}_B)]_+ \cdot \chi_{c_n}\|$ denotes the analogous test function evaluated with bootstrap empirical distributions and sample empirical distributions, we have

$$\sup_{f \in BL_1} |\mathbb{E}^*[f(\sqrt{n}T_n^*)] - \mathbb{E}[f(\mathcal{T})]| = o_P(1).$$

The second part of Theorem 5.1 has one unexpected part, which is the form that the bootstrap analog T_n^* takes. In particular, it is not perfectly analogous to the form that weak convergence takes with the sample data because of the presence of the contact set indicator. Because of the positive part map that lies in the definition of the statistics, they are fundamentally less regular than other more conventional statistics and a special bootstrap needs to be devised to ensure bootstrap consistency. This bootstrap technique relies on Fang and Santos (2019) and is justified in another way in Linton et al. (2010).

The previous theorem asserted the consistency of the resampling plan. The next theorem adds to that description. It specifies the size of testing procedures used to infer dominance as described in Sections 2, 3 and 4. To do so, we introduce a sequence of local alternative distributions (F_{A_n}, F_{B_n}) and assume that they satisfy a kind of mean-square convergence condition to the null (F_A, F_B) : assume that for $F_n = F_{A_n}$ or F_{B_n} , and $F = F_A$ or F_B , there is some square integrable h such that

$$\lim_{n \rightarrow \infty} \int \left(\sqrt{n}(\sqrt{dF_n} - \sqrt{dF}) - \frac{h}{2}\sqrt{dF} \right)^2 \rightarrow 0. \quad (13)$$

This form of alternative is used in empirical process theory (van der Vaart and Wellner, 2023, §3.11.1) to discuss distributions that converge to the null at precisely the right rate to find nontrivial limit results.

Theorem 5.2. *Assume conditions **A1** and **A2** are satisfied. Assume that $(F_A, F_B) \in \mathcal{F}_0$, that is, that $g = g(F_A, F_B)$ satisfies $T = \|[g]_+\| = 0$ for one of the g described in the appendix, and let $T_n = \|[g(\hat{F}_A, \hat{F}_B)]_+\|$. Letting $q(1 - \alpha)$ denote the $(1 - \alpha)$ -th quantile of the asymptotic distribution of $\sqrt{n}T_n$, suppose that $q(1 - \alpha) > 0$. Suppose that a local sequence of probability distributions $P_n = (F_{A_n}, F_{B_n})$ satisfies (13) and $\|[g(F_{A_n}, F_{B_n})]_+\| = 0$ for all n . Finally, let $q_n^*(1 - \alpha)$ denote the $(1 - \alpha)$ -th quantile of $T_n^* = \|[g(F_A^*, F_B^*) - g(\hat{F}_A, \hat{F}_B)]_+ \cdot \chi_{c_n}\|$. Then $\limsup_{n \rightarrow \infty} P_n \{\sqrt{n}T_n^* > q_n^*(1 - \alpha)\} \leq \alpha$. This holds with equality when $P_n = P_0$ for all n .*

This result implies that the bootstrap test's size is controlled asymptotically by the intended/nominal test size for the distribution P_0 and all local alternatives that respect the null hypothesis of dominance of A over B . One could in theory compute the power of tests for alternatives that violate the null hypothesis, but this is complicated by the unique features of each testing criterion and no general (yet practical) statements can be made about test power under local alternatives. The regularity condition that the $(1 - \alpha)$ -th quantile of the asymptotic distribution of the test statistic must be positive made in the previous theorem may seem innocuous. However, it has practical implications. If, for example, F_A and F_B are such that we are “far” from rejecting the null hypothesis, it is possible that $T_n = 0$ and $T_r^* = 0$ for all the bootstrap repetitions. In this case, the naive bootstrap p-value $p^\circ = \sum_r I(T_r^* > T_n)/R = 0$. However, the distribution is degenerate here and this seemingly-low p-value does not indicate that the observed test statistic lies in an extreme region of the reference distribution. To address this, we suggest using the modified bootstrap critical value akin to that proposed in Andrews and Shi (2013), namely $p^* = \sum_r I(T_r^* + \eta > T_n)/R$, where $\eta > 0$ is an arbitrarily small constant like $\eta = 10^{-6}$. This has the effect of breaking ties due to degeneracy when they happen, and has no practical effect otherwise.

6 Empirical illustration: Welfare reform in Connecticut

In this section we illustrate the comparisons discussed above using household data from an experiment conducted by the U.S. state of Connecticut in 1996. This data has been discussed at length before (Bitler et al., 2006; Firpo et al., forthcoming) so we only briefly describe the main features of the two policies and the patterns that emerge in the sample.

The treatments in this experiment are both programs that provided income support to low-income families with dependent children. The preexisting Aid to Families with Dependent Children

(AFDC) program was replaced in the 1990s with a different program called Temporary Assistance for Needy Families (TANF). The specific TANF program that was administered by Connecticut was called Jobs First (JF) and had a much different structure than AFDC: it included more generous income support than the AFDC program had, but that support came with a strict time limit. We label the two treatments as AFDC and JF benefit structures. The state was interested in comparing program outcomes and so it experimentally assigned one of AFDC or JF benefit structures to a sample of 4803 households (we observe 2396 households under treatment JF and 2407 households under treatment AFDC). For each household, we observe quarterly income before the treatment was assigned, during the experiment and also after the shorter JF time limit had been reached. Outcomes are measured in the natural logarithm of average household income over post-experiment periods (quarter years), and gains/losses are simply the log post-experiment average minus the log pre-experiment average. Bitler et al. (2006), using average and quantile treatment effects in levels, found that JF made the majority of households better off, but also had significant drawbacks for some households after the time limit had been reached. We will make different comparisons using the loss- and inequality-averse criteria developed above, focusing on household income data.

Let us first take a look at a comparison of marginal and joint income distributions under both policies. They are plotted in Figure 1. We can see that the empirical CDF (ECDF) of income *changes* looks better under the AFDC policy over the entire support of the change distributions. The dominance appears quite strong, namely, the AFDC curve seems to stochastically dominate the JF curve at first order. However, when looking at post-experiment household income *levels*, the policies are not clearly ordered. The levels ECDFs cross, with the marginal AFDC distribution function below that of JF for lower-income households. Above level $z = 7$, the ECDFs cross. This level corresponds to around 2300 US dollars on average quarterly. It is difficult to see in the plots, but above that level of income, the JF and AFDC level distribution functions cross several times. The first-order dominance of AFDC over JF in changes implies that of the loss averse and ‘second order type’ comparisons of the marginal distribution of changes in the test functions, and most likely leads to the pattern of rejections and non-rejections shown in the table of tests below. Therefore, these empirical results should be regarded primarily as illustrations of the methods and their associated testing procedures.

The left two panels of Figure 2 show the joint ECDFs of both benefit distributions. The third

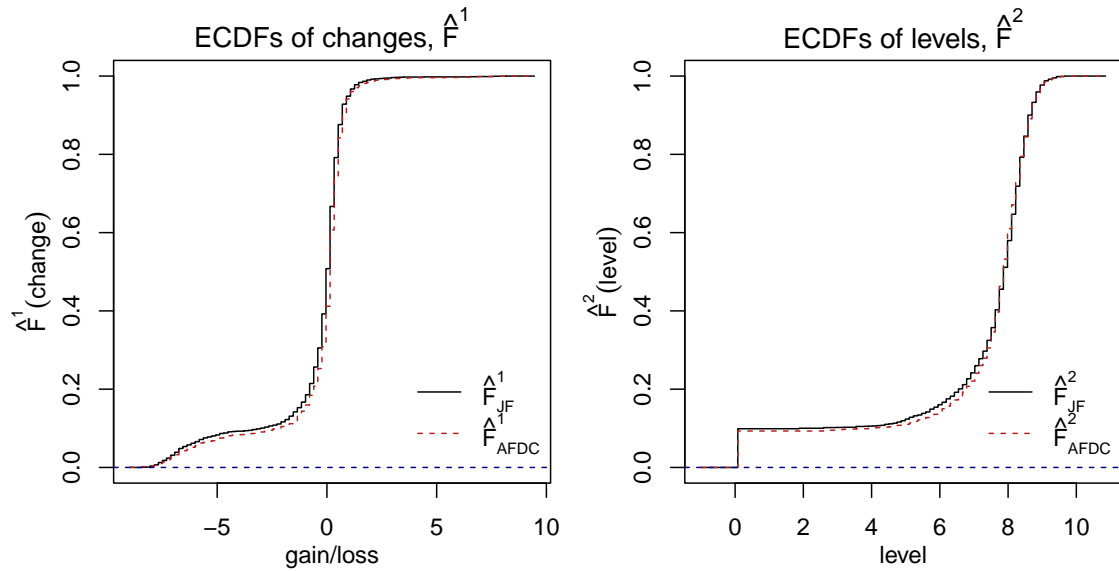


Figure 1: Marginal empirical CDFs of income distributions under JF and $AFDC$ benefit structures, which indicate the distributions of gains/losses and levels respectively. The marginal $AFDC$ gain/loss distribution stochastically dominates the JF gain/loss distribution at first order. In levels, the marginal ECDFs cross.

panel shows the difference between the joint ECDFs, which is not readily apparent in the left two panels. For reference, the “front” corner of all three plots can be used to see where zero is on the vertical axis. The difference in the third panel is calculated such that positive parts indicate that the JF distribution function lies above the $AFDC$ distribution function. The third panel most clearly shows information that cannot be gained by only comparing marginal ECDFs. Broadly speaking, the $AFDC$ structure has a joint distribution function that lies below the JF joint distribution function everywhere except for a dramatic reversal for high-income households. For these households, JF produces a higher likelihood of large gains than $AFDC$ —this is where JF ’s advantages primarily emerge, namely in the high-income, high-gain range. Consequently, households that fare better under JF compared with $AFDC$ are found predominantly among those with higher incomes. Conversely, the positive spike in the region of small gains and losses indicates that high-income households also face a higher risk of losses under JF than under $AFDC$.

We now turn to statistical tests to assess the significance of these observations. We ran tests to check whether JF dominates $AFDC$. These are shown in the left half of Table 1. We also ran

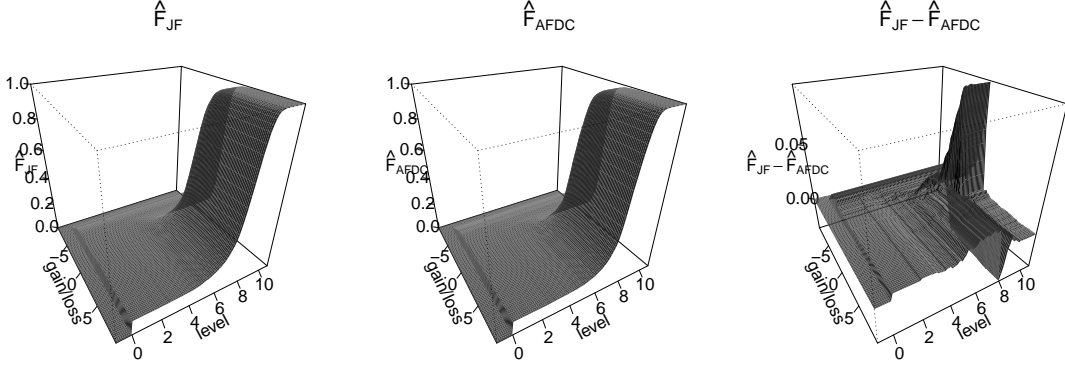


Figure 2: Joint empirical CDFs \hat{F}_{JF} and \hat{F}_{AFDC} and their difference, since the ECDFs look very similar. See the main text for a longer discussion of the spike and dip at the right side of the third panel.

tests to check whether AFDC dominates JF, and those results are presented in the right half of the table. For an exact test statistic, one would need to evaluate the empirical process upon which the test statistics rely at each sample observation, but the combined sample size is prohibitively large. Instead, empirical processes were evaluated on a grid of points in the (x, z) space that used 100 grid points for gains & losses, and 50 points for levels, and test statistics were computed as functionals of that approximated process. Informal experimentation with denser grids resulted in extremely similar results. 999 bootstrap repetitions were used to estimate a reference distribution. Once again, this number could be increased but the results are qualitatively similar.

Several hypothesis testing results are presented in Table 1. Using any of the six dominance concepts developed above, we reject the hypothesis that the JF benefits structure would be preferred by households. On the other hand, we are unable to reject the hypothesis that AFDC benefits would be preferred by households. There are some variations. For example, the first-order-type LASBD test is less decisive than the second-order-type IASD and LIASD tests. However, the basic result is unchanged across all the dominance concepts.

The hypothesis that JF dominates AFDC is rejected using any dominance concept described in this paper. On the other hand, tests of the dominance of AFDC over JF fail to reject in all cases. Assuming that households' behavior is well described by any of the sets of qualitative properties given in the previous sections, there is strong evidence that the AFDC benefit structure is socially preferred comparing to the JF. It appears as though JF carries a greater likelihood of small losses

| | $H_0 : \mathbf{JF} \succsim \mathbf{AFDC}$ | | $H_0 : \mathbf{AFDC} \succsim \mathbf{JF}$ | |
|--------|--|---------|--|---------|
| | statistic | p-value | statistic | p-value |
| LASBD | 0.23 | 0.02 | 0.05 | 0.54 |
| LASBD2 | 0.27 | 0.00 | 0.06 | 0.48 |
| IASD | 7.53 | 0.00 | 0.01 | 0.81 |
| IASD2 | 5.66 | 0.00 | 0.57 | 0.44 |
| LIASD | 7.40 | 0.00 | 0.01 | 0.78 |
| LIASD2 | 5.49 | 0.00 | 0.57 | 0.42 |

Table 1: Tests of the hypotheses that either JF benefits dominate AFDC benefits or that AFDC benefits dominate JF benefits. In all cases, the JF benefits appear to violate the hypothesis that they would be preferred by households. On the other hand, we cannot reject the hypothesis that AFDC benefits dominate JF benefits under any of the dominance concepts. 999 bootstrap repetitions were used to generate reference distributions for all tests, and functions were evaluated on evenly-spaced grids of 100 points for gains/losses and 50 points for incomes in levels.

to household income than AFDC, suggesting that many households had been supplementing their earnings with program support and, when JF assistance ended, their incomes declined.

In the following subsections, we show coordinate processes used in all the tests shown in the left half of Table 1. All the functions are found by rearranging the functions in the corresponding inequalities shown earlier in the text so that the function corresponding to the JF benefit structure has the AFDC function subtracted from it — for example, in Figure 3 below, the left plot shows $\hat{F}_{JF}^2(z) - \hat{F}_{AFDC}^2(z)$ over all levels z , used to test whether the inequality (3) holds (squared values of the positive part of this function are combined in an integral with similar quantities using other coordinate functions to calculate a test statistic).

6.1 Loss aversion sensitive comparison

We illustrate the way that the distributions are compared using the LASBD and LASBD2 concepts. Both the LASBD and LASBD2 concepts use changes in household income from before the program started, when all households were using AFDC benefits, to after the program ended when households subjected to the JF treatment saw their benefits end. This risk, that a household could potentially have a lower income if the JF benefits end and there is no other sizable source of income, is the risk of the JF program that would be of primary concern to a household considering the two policies.

Figure 3 and Figure 4 show all the component functions that go into a test of either of the nulls $F_{JF} \succsim_{LASBD} F_{AFDC}$ or $F_{JF} \succsim_{LASBD2} F_{AFDC}$. In testing, these two concepts have three

common component functions, shown in Figure 3, and differ in one component function, which are contrasted in Figure 4. In the paragraphs to follow, we describe what causes positive values seen in the component functions, which would indicate a violation of either $F_{JF} \succsim_{LASBD} F_{AFDC}$ or $F_{JF} \succsim_{LASBD2} F_{AFDC}$. These positive values are reduced to one-number summaries that are in the top two entries of the leftmost column of Table 1.

All three functions in Figure 3 are positive for some arguments (x, z) . The reasons are different for each panel of the figure. The left panel tracks income levels after the JF time limit. It indicates that $F_{JF}^2(z) \geq F_{AFDC}^2(z)$, roughly speaking for z up to the middle of the income distribution — in other words, the post-experiment incomes are more favorable for low-income households under the AFDC benefit structure. The relationship between the distributions changes at $\exp(7.75) \approx \$2300$, at which point F_{JF} goes below F_{AFDC} and then the functions stay close for larger quarterly incomes. The central panel of Figure 3 reveals that (using (5)) $F_{AFDC}^1 \leq F_{JF}^1$ for almost all levels of loss, but F_{JF} represents a much higher chance of experiencing a small loss as compared to F_{AFDC} , leading to the large values on the right-hand side of the plot. The right panel of Figure 3 is positive for small values of $|x|$ representing absolute gains and losses as in (6), and is positive because F_{AFDC}^1 dominates F_{JF}^1 for small changes especially.

As mentioned above, the LASBD and LASBD2 concepts share three coordinate functions. They differ in one coordinate, which is illustrated in Figure 4. $F_A \succsim_{LASBD} F_B$ requires that A dominates B in the bivariate distribution function, while $F_A \succsim_{LASBD2} F_B$ requires it dominate in K functions (which are the probabilities that a gain/level pair exceed (x, z)). Both functions are positive for some (x, z) , suggesting a rejection of the null hypotheses $F_{JF} \succsim_{LASBD} F_{AFDC}$ or $F_{JF} \succsim_{LASBD2} F_{AFDC}$. For LASBD, this is because F_{JF} does not dominate F_{AFDC} for moderately large income levels and all changes above small losses. That is, the probability of moderately high post-experiment incomes are relatively high and coupled with (usually) some gain in income under AFDC, while for JF that probability is not as high. The LASBD2 concept shows nearly the same information — under F_{JF} , households have a lower probability of having high incomes and small gains.

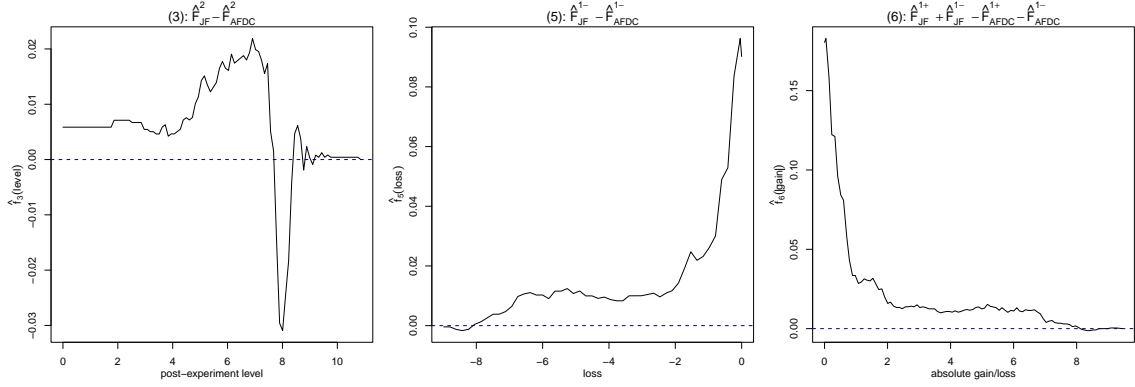


Figure 3: The component functions common to the LASBD and LASBD2 notions used for testing either $H_0 : F_{JF} \lesssim_{LASBD} F_{AFDC}$ or $H_0 : F_{JF} \lesssim_{LASBD2} F_{AFDC}$. The numbers above each panel correspond to the numbered displays in the text. F_k^{1+} is shorthand for the distribution function k for gains/losses evaluated for gains and F_k^{1-} is the same function evaluated for losses.

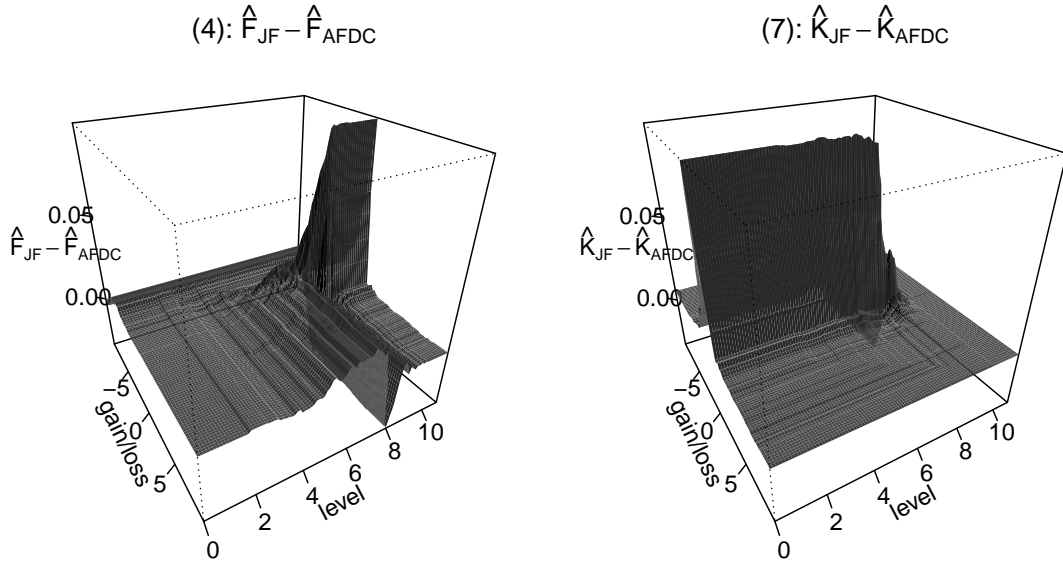


Figure 4: The component functions that differ between the LASBD and LASBD2 concepts. The left panel is analogous to display (4) and shows $\hat{F}_{JF}(x, z) - \hat{F}_{AFDC}(x, z)$ while the right panel is analogous to display (7) and shows $\hat{K}_{JF}(x, z) - \hat{K}_{AFDC}(x, z)$, both evaluated using the empirical distribution estimates of F_{JF} and F_{AFDC} . For reference in these 3-dimensional plots, the functions are zero at the “corners” of the plots, where x and z reach either of their extremes, indicating that the central portions are positive.

6.2 Inequality aversion sensitive comparison

We can make a similar comparison between the functions that go into the IASD and IASD2 criteria. There are two functions that IASD and IASD2 have in common, and are shown in Figure 5.

Both of the functions are exactly zero at the origin, but become positive when evaluating the function anywhere else. Because they are positive, they suggest a violation of the notion that the hypothesis that the JF benefit structure would be preferred by households to the AFDC benefit structure in an IASD sense, that is, they indicate there is evidence against the hypothesis $F_{JF} \succsim_{IASD} F_{AFDC}$.

The difference between the IASD and IASD2 criteria are that IASD uses equation (10) while IASD2 uses (12) for comparison. These two functions are shown in Figure 6. At the left-most corner of each panel of Figure 6, the functions are equal to zero, and they are increasing as gain/loss move away from their lower limits. Once again, these functions are constructed so that significantly positive values suggest a rejection of the hypothesis $F_{JF} \succsim_{IASD} F_{AFDC}$.

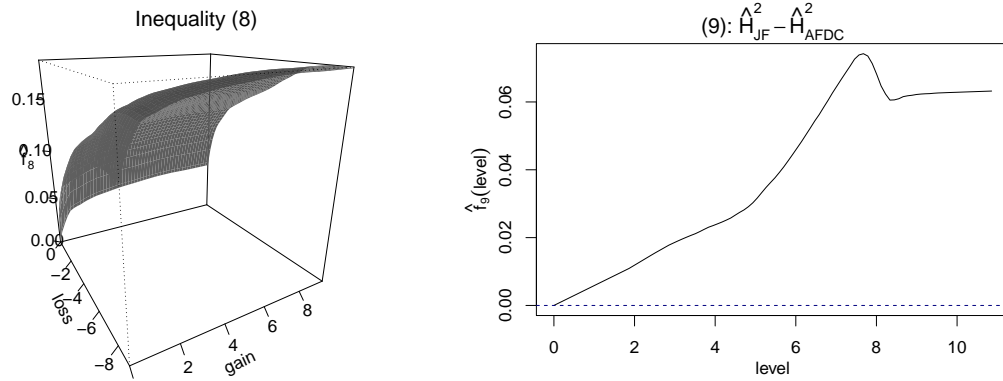


Figure 5: These functions are common to the IASD and IASD2 comparison between JF and AFDC. Because they remain above zero everywhere, they suggest a contradiction of the hypotheses $F_{JF} \succsim_{IASD} F_{AFDC}$ or $F_{JF} \succsim_{IASD2} F_{AFDC}$.

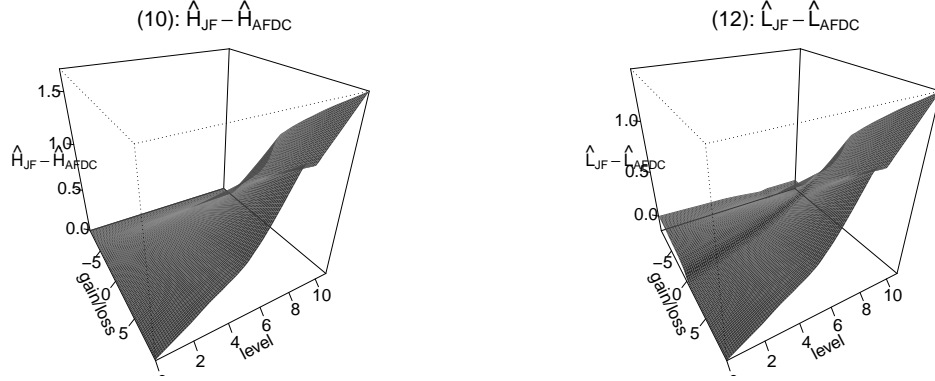


Figure 6: These functions are different between the IASD and IASD2 comparison of JF and AFDC. The IASD concept uses function (10), pictured in the left panel, and the IASD2 concept uses function (12), in the right panel. Both functions remain above zero over their support, with the exception of the right panel (function (12)) for low-income households that experience a large gain over the course of the experiment. The large positive values these functions take suggest a violation of the hypotheses that JF dominates AFDC using either the IASD or IASD2 criterion.

6.3 Loss- and inequality aversion sensitive comparison

The LIASD comparison of these two programs is similar, and indeed, it recombines some of the coordinate functions already pictured in LASBD and IASD comparisons. To test the hypothesis $F_{JF} \succsim_{LIASD} F_{AFDC}$, one would look for positive values of the functions (2), (9) and (10). Equation (2) is easily checked by using equations (5) and (6), which are shown in the center and right panels of Figure 3. The other two coordinate functions were used in the IASD comparison, and their plots can be found in the right panel of Figure 5 and the left panel of Figure 6. The large positive values in all these plots suggest a rejection of the hypothesis, but of course the second-to-last test statistic and its p-value in the left half of Table 1 is used to formally test their joint statistical significance.

The story is similar for a test of the hypothesis $F_{JF} \succsim_{LIASD2} F_{AFDC}$, which also uses functions (2) and (9) and were discussed in the previous paragraph. The way that the LIASD2 comparison differs is in its use of (12), and the difference $L_{JF} - L_{AFDC}$ can be seen in the right panel of Figure 6. These functions also indicate a rejection of the LIASD2 dominance of JF over AFDC, as seen formally in the final row of the left half of Table 1.

7 Conclusions

We propose a variety of bivariate stochastic dominance criteria that may be used to rank prospective policies based on experimental data on both the gains and losses that agents experience and their post-policy outcome in levels. They can also be used to rank any bivariate distributions of absolute outcomes and outcome changes, for example, lotteries. We extend existing univariate approaches so that they incorporate two central empirical regularities that are well-grounded in the literature: individuals dislike losses more than they value equivalent gains, and they are sensitive to the distribution of absolute incomes. The dominance criteria remain non-parametric, which ensures that the rankings of distributions that are obtained are robust to the choice of specific functional forms of value functions. Furthermore, the criteria can be translated into sets of functional inequalities, and testing methods are designed to check whether one policy/distribution is preferred to another.

In this paper we underscore the importance of jointly considering changes in income and the distribution of final incomes. From a policy point of view it is important to know where gains and losses are concentrated along the income dimension. Furthermore, policies that appear attractive when assessed solely on gains and losses may be much less appealing once their distributional consequences are incorporated, and vice versa. The bivariate dominance framework reduces the risk of adopting policies that are “extreme” in a unidimensional sense, instead favoring those that perform robustly well across both key dimensions of individual preferences.

While this is one of the first papers to incorporate loss aversion into welfare analysis, it is important to acknowledge the ongoing debate on this issue in the behavioral economics literature. As already mentioned, we rely on this literature by using an extended utility function whose properties are well established in this literature. It is interesting to note that this literature has almost entirely avoided welfare analysis because it is unclear whether reference dependence represents a bias on the part of decision-makers or non-standard but normative preferences (Reck and Seibold, 2023). Or, as O’Donoghue and Sprenger (2018) put it: “Perhaps first and foremost is the question of whether gain-loss utility should be given normative weight i.e., whether we should assume that the same preferences that rationalize behavior should also be used for welfare analysis.” In other words, while empirical evidence shows that individuals maximize reference-dependent preferences when making decisions, the question is whether the social planner should use these same preferences as an input in the social welfare maximization problem. If the social planner believes that such

preferences represent errors and distort behavior from what would be individually optimal, then paternalistically the social planner should only take into account the “correct” preferences. That question is certainly relevant to our framework.

We agree with the argument put forward in (Tversky and Kahneman, 1991), namely, that decision makers need a good criterion for evaluating policy options, and it is hard to argue that actual experience of the consequences of a policy can be completely discarded as such a criterion. People do evaluate their situations in relation to the reference point, and they do feel an asymmetry between pain and pleasure. Apart from the numerous citations above, this is also well supported by neuroeconomic research (Dhami, 2016) and is known to influence the behavior of humans and primates in general. This leads Rees-Jones (2024) to state “Modern economists have been wary of taking this type of paternalistic stance... who is the researcher to say, confidently, that they know what is best for others?”. Therefore, very recently, a welfare analysis literature is emerging in behavioral economics (Goldin and Reck, 2022; Reck and Seibold, 2023) that includes the loss/gain component and uses an extended utility function. It is, however, different formally than our framework as it includes individual maximizing behavior as a component. Our welfare analysis, as noted, is grounded in the traditional social welfare function approach.

From a broader perspective, this work contributes to the emerging welfare-economic literature that integrates behavioral features—such as loss aversion—into social evaluation without imposing a particular functional specification. It provides a set of implementable tools for researchers and policymakers who wish to respect empirically observed preference patterns while maintaining the robustness and transparency of dominance-based methods. Future research could apply these criteria in other contexts where reference-dependent preferences are relevant, explore their interaction with dynamic considerations such as persistence of losses or intertemporal inequality, and develop social value indices that are characterized by further properties and useful in cases in which the dominance criteria do not produce conclusive results.

Appendix

Functions used in testing

The hypotheses to be tested in the main text can be translated into tests of functional inequalities of one type. For each dominance concept, we write a test function $g = g^k(F_A) - g^k(F_B)$ where k is a stand-in for a concept's label. Recall that the CDF F is bivariate, and the marginal CDFs corresponding to F are labeled F^1 and F^2 (for gains/losses and levels respectively). These functions g can be evaluated over $x, z \geq 0$ — gains and losses are represented by positive or negative values of x respectively and are treated separately, while levels are represented by the argument z (recall it is assumed that $z \in [0, a_3) \subset \mathbb{R}_+$). In the definitions below, we use the following functions, which are also stated in the theorems defining the functions used for testing in the main text:

$$\begin{aligned} H^1(x) &= \int_{-\infty}^x F^1(t)dt, & H^2(z) &= \int_0^z F^2(t)dt & S^1(x) &= \int_x^\infty 1 - F^1(t)dt \\ K(x, z) &= F^1(x) + F^2(z) - F(x, z) & L(x, z) &= \int_{-\infty}^x \int_0^z K(t, s)dsdt. \end{aligned}$$

Then we define the following functions for testing (for IASD/IASD2, we need $x_1, x_2 \geq 0$):

$$\begin{aligned} g^{LASBD}(F)(x, z) &= \begin{bmatrix} F^2(z) \\ F(-x, z) \\ F(x, z) \\ F^1(-x) \\ F^1(x) + F^1(-x) \end{bmatrix} & g^{LASBD2}(F)(x, z) &= \begin{bmatrix} F^2(z) \\ K(-x, z) \\ K(x, z) \\ F^1(-x) \\ F^1(x) + F^1(x) \end{bmatrix}, \\ g^{IASD}(F)(x_1, x_2, z) &= \begin{bmatrix} H^2(z) \\ H(-x_1, z) \\ H(x_1, z) \\ S^1(x_1) - H^1(-x_2) - S^1(0) + H^1(0) \end{bmatrix} & g^{IASD2}(F)(x_1, x_2, z) &= \begin{bmatrix} H^2(z) \\ L(-x_1, z) \\ L(x_1, z) \\ S^1(x_1) - H^1(-x_2) - S^1(0) + H^1(0) \end{bmatrix}, \\ g^{LIASD}(F)(x, z) &= \begin{bmatrix} H^2(z) \\ H(-x, z) \\ H(x, z) \\ F^1(-x) \\ F^1(x) + F^1(-x) \end{bmatrix} & g^{LIASD2}(F)(x, z) &= \begin{bmatrix} H^2(z) \\ L(-x, z) \\ L(x, z) \\ F^1(-x) \\ F^1(x) + F^1(-x) \end{bmatrix}. \end{aligned}$$

Proof of main results

Let us start by formulating the following lemma, which will be useful later in the proofs.

Lemma 1.

$$W(F) = - \int_{-\infty}^0 \frac{\partial}{\partial x} v(x, a_3) F^1(x) dx + \int_0^\infty \frac{\partial}{\partial x} v(x, a_3) (1 - F^1(x)) dx - \int_0^\infty \frac{\partial}{\partial z} v(a_2, z) F^2(z) dz \\ + \int_{-\infty}^\infty \int_0^\infty \frac{\partial^2}{\partial x \partial z} v(x, z) F(x, z) dz dx \quad (14)$$

Proof of Lemma 1.

$$W(F) = \int_{\mathbb{R} \times \mathbb{R}^+} v(x, z) dF(x, z) = \int_{-\infty}^\infty \int_0^\infty v(x, z) f(x, z) dz dx \\ = \int_{-\infty}^\infty \left[v(x, z) \int_0^z f(x, t) dt \Big|_0^{a_3} - \int_0^\infty \frac{\partial}{\partial z} v(x, z) \int_0^z f(x, t) dt dz \right] dx \\ = \int_{-\infty}^\infty v(x, a_3) f^1(x) dx - \int_{-\infty}^\infty \int_0^\infty \frac{\partial}{\partial z} v(x, z) \int_0^z f(x, t) dt dz dx.$$

We will now expand these two terms. We have

$$\int_{-\infty}^\infty v(x, a_3) f^1(x) dx = \int_{-\infty}^0 v(x, a_3) f^1(x) dx + \int_0^\infty v(x, a_3) f^1(x) dx.$$

Starting with the first term and applying integration by parts we get

$$\int_{-\infty}^0 v(x, a_3) f^1(x) dx = v(x, a_3) F^1(x) \Big|_{-\infty}^0 - \int_{-\infty}^0 \frac{\partial}{\partial x} v(x, a_3) F^1(x) dx = - \int_{-\infty}^0 \frac{\partial}{\partial x} v(x, a_3) F^1(x) dx$$

and also

$$\int_0^\infty v(x, a_3) f^1(x) dx = - \int_0^\infty v(x, a_3) d(1 - F^1(x)) = -v(x, a_3) (1 - F^1(x)) \Big|_0^{a_2} + \int_0^\infty \frac{\partial}{\partial x} v(x, a_3) (1 - F^1(x)) dx = \\ = \int_0^\infty \frac{\partial}{\partial x} v(x, a_3) (1 - F^1(x)) dx.$$

Expanding the second term, we have

$$\begin{aligned} \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial}{\partial z} v(x, z) \int_0^t f(x, t) dt dz dx &= \int_0^{\infty} \frac{\partial}{\partial z} v(x, z) F(x, z) dz \Big|_{-a_1}^{a_2} - \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial^2}{\partial x \partial z} v(x, z) F(x, t) dz dx = \\ &= \int_0^{\infty} \frac{\partial}{\partial z} v(a_2, z) F^2(z) dz - \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial^2}{\partial x \partial z} v(x, z) F(x, z) dz dx \end{aligned}$$

Finally, we obtain

$$\begin{aligned} W(F) &= - \int_{-\infty}^0 \frac{\partial}{\partial x} v(x, a_3) F^1(x) dx + \int_0^{\infty} \frac{\partial}{\partial x} v(x, a_3) (1 - F^1(x)) dx \\ &\quad - \int_0^{\infty} \frac{\partial}{\partial z} v(a_2, z) F^2(z) dz + \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial^2}{\partial x \partial z} v(x, z) F(x, z) dz dx \end{aligned}$$

□

Proof of Theorem 2.1. Notice that (2) is equivalent to both (5) and (6) and we will use latter conditions. Let $\Delta W = W(F_A) - W(F_B)$ and similarly for cumulative distribution functions, $\Delta F = F_A - F_B$. Using Lemma 1 we have that

$$\begin{aligned} \Delta W &= - \int_{-\infty}^0 \frac{\partial}{\partial x} v(x, a_3) \Delta F^1(x) dx - \int_0^{\infty} \frac{\partial}{\partial x} v(x, a_3) \Delta F^1(x) dx \\ &\quad - \int_0^{\infty} \frac{\partial}{\partial z} v(a_2, z) \Delta F^2(z) dz + \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial^2}{\partial x \partial z} v(x, z) \Delta F(x, z) dz dx = \\ &= - \int_0^{\infty} \frac{\partial}{\partial x} v(-x, a_3) \Delta F^1(-x) dx - \int_0^{\infty} \frac{\partial}{\partial x} v(x, a_3) \Delta F^1(x) dx \\ &\quad - \int_0^{\infty} \frac{\partial}{\partial z} v(a_2, z) \Delta F^2(z) dz + \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial^2}{\partial x \partial z} v(x, z) \Delta F(x, z) dz dx \geq 0 \end{aligned}$$

Adding and subtracting $\int_0^{\infty} \frac{\partial}{\partial x} v(x, a_3) \Delta F^1(-x) dx$ we get

$$\begin{aligned} \Delta W &= \int_0^{\infty} \frac{\partial}{\partial x} (v(x, a_3) - v(-x, a_3)) \Delta F^1(-x) dx - \int_0^{\infty} \frac{\partial}{\partial x} v(x, a_3) (\Delta F^1(x) + \Delta F^1(-x)) dx \\ &\quad - \int_0^{\infty} \frac{\partial}{\partial z} v(a_2, z) \Delta F^2(z) dz + \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial^2}{\partial x \partial z} v(x, z) \Delta F(x, z) dz dx \geq 0. \quad (15) \end{aligned}$$

Utilizing the assumptions of loss aversion, non-decreasingness and submodularity given in Definition

2.2, (3) and (4), (5) and (6) (or, equivalently (2)) are sufficient for (15) to hold.

We now show that these conditions are also necessary by means of a contradiction. For the first two integrals in (15) the procedure is a modification of Firpo et al. (forthcoming) to two dimensions. Starting with (5), from the fact that the distribution function is right continuous, there is a neighborhood (a, b) , $0 < a < b$, such that for all $x \in (a, b)$, $F_A^1(-x) - F_B^1(-x) > 0$ (i.e. $\Delta F^1(-x) > 0$). Now consider the value function

$$v_1(x, z) = \begin{cases} a - b & x \leq -b \\ x + a & x \in (-b, -a) \\ 0 & x \geq -a \end{cases}$$

Note that $v_1(x, z)$ satisfies Definition 2.2.⁶ In particular, it is non-decreasing with respect to z , because its derivative with respect to z is 0. It is submodular in a trivial way, that is, cross-derivative is 0. It is also loss averse because, for $x \geq 0$, $\frac{\partial v(x, z)}{\partial x} = 0$, while for $x < 0$ the respective derivative is 1 when $x \in (-b, -a)$. Therefore $\int_0^\infty \frac{\partial}{\partial x}(v(x, a_3) - v(-x, a_3))\Delta F^1(-x)dx < 0$ while the rest of integrals in (15) are 0, which contradicts (15). Condition (6) can be proven similarly. Assume that there exists a neighborhood (a, b) , $0 < a < b$ such that for all $x \in (a, b)$ $(1 - F_A^1(x)) - F_A^1(-x) < (1 - F_B^1(x)) - F_B^1(-x)$ (i.e. $\Delta F^1(x) + \Delta F^1(-x) > 0$). Take $v_2(x, z) = -v_1(-x, z)$ for $x > 0$ and $v_2(x, z) = v_1(x, z)$ for $x < 0$, then $-\int_0^\infty \frac{\partial}{\partial x}v(x, a_3)(\Delta F^1(x) + \Delta F^1(-x))dx < 0$ while the rest of integrals in (15) are 0, which contradicts (15).

Now we proceed in a similar fashion with the third integral in (15) and a contradiction to (3). Assume that there exists some $z > 0$ such that $F_A^2(z) - F_B^2(z) > 0$ (i.e. $\Delta F^2(z) > 0$). From the fact that the distribution function is right continuous, it follows that there is a neighborhood (c, d) , $0 < c < d$, such that for all $z \in (c, d)$, $F_A^2(z) - F_B^2(z) > 0$. For $x \geq 0$, consider the value function

$$v_3(x, z) = \begin{cases} 0 & z \leq c \\ z - c & z \in (c, d) \\ d - c & z \geq d \end{cases}$$

⁶The function is not differentiable at the boundaries of the three areas, however, we can always find a differentiable function which is as close to v_1 as one wishes too, e.g. for arbitrarily small $\epsilon > 0$ we consider $\frac{1}{4\epsilon}(x + b + \epsilon)^2 + a - b$ when $x \in (-b - \epsilon, -b + \epsilon)$ that then joins areas $x \leq -b - \epsilon$ and $(-b + \epsilon, -a - \epsilon)$. The same applies to the other value functions in the proofs.

and for $x < 0$ we put $bx, b > 0$. Thus, v_3 fulfills Definition 2.2. In particular $\frac{\partial}{\partial z}v(x, z) > 0$. Then, $-\int_0^\infty \frac{\partial}{\partial z}v(a_2, z)\Delta F^2(z)dz < 0$ while the rest of integrals in (15) are 0, which contradicts (15).

Finally, we prove the necessity of (4). Assume that there exists some x, z such that $F_A(x, z) - F_B(x, z) > 0$. We will first show contradiction for $x < 0$, but we need to define function v that fulfills Definition 2.2 so it is defined on the whole domain of x . Let $x < 0$. From the fact that the distribution function is right continuous, it follows that there is a neighborhood $(-b, -a) \times (c, d)$, $b > a > 0, d > c > 0$, such that for all (x, z) in this neighbourhood, $F_A(x, z) - F_B(x, z) > 0$. Consider the following function

$$v_4(x, z) = \begin{cases} b(c-d) - ac & x \leq -b, z \leq c \\ (d-c)x - ac & x \in (-b, -a), z \leq c \\ dx & 0 > x \geq -a, z \leq c \\ (b-a)z - bd & x \leq -b, z \in (c, d) \\ -xz + dx - az & (x, z) \in (-b, -a) \times (c, d) \\ dx & 0 > x \geq -a, z \in (c, d) \\ -ad & x \leq -b, z \geq d \\ -ad & x \in (-b, -a), z \geq d \\ dx & 0 > x \geq -a, z \geq d \\ 0 & x \geq 0. \end{cases}$$

Let us now check that v_4 fulfills Definition 2.2. Firstly, for $x < 0$, it is negative in each of the nine areas, which follows from $-b < -a < 0$ and $0 < c < d$. It is 0 for $x \geq 0$. Secondly, it is non-decreasing, e.g. the derivative of $(b-a)z - bd$ with respect to z is $b-a > 0$, or the derivative of $-xz + dx - az$ with respect to z is $-x - a > 0$, because $x \in (-b, -a)$. Also, the derivative of $-xz + dx - az$ with respect to x is $-z + d > 0$, because then $z < d$. On the other hand, $\frac{\partial v(x, z)}{\partial x} = 0$ for $x \geq 0$. Thirdly, it is loss averse, as the derivative $\frac{\partial v(-x, z)}{\partial x} > 0 = \frac{\partial v(x, z)}{\partial x}$. Finally, it is submodular given that the cross derivative of $-xz + dx - az$ is -1 when $(x, z) \in (-b, -a) \times (c, d)$ and 0 elsewhere. The values were chosen so that the function is continuous at the boundaries of nine areas.

Finally, for $x > 0$, the following function will lead to a contradiction in the same way as v_4 . “C

large” below means that a constant C is chosen such that its value is sufficient to ensure that the derivative of $\frac{\partial v_5}{\partial x}(-x) \geq \frac{\partial v_5}{\partial x}(x)$.

$$v_5(x, z) = \begin{cases} Cx & x \leq 0, C \text{ large} \\ dx + bz - ac & 0 < x \leq a, z \leq c \\ dx - cx + bz & x \in (a, b), z \leq c \\ dx + bz - cb & x \geq b, z \leq c \\ dx + bz - az & 0 < x \leq a, z \in (c, d) \\ -xz + dx + bz & (x, z) \in (a, b) \times (c, d) \\ dx & x \geq b, z \in (c, d) \\ dx + bz - ad & 0 < x \leq a, z \geq d \\ bz & x \in (a, b), z \geq d \\ dx + bz - bd & x \geq b, z \geq d \end{cases}$$

□

Proof of Corollary 2.1. This is a direct consequence of the fact that in this case

$$W(F) = - \int_{-\infty}^0 v'_1(x) F^1(x) dx + \int_0^\infty v'_1(x) (1 - F^1(x)) dx - \int_0^\infty v'_2(z) F^2(z) dz$$

because $\frac{\partial^2}{\partial x \partial z} v(x, z) = 0$.

□

Proof of Theorem 2.2. Notice that $K^1(x) = K(x, 0)$ and $K^2(z) = K(-a_1, z)$. Proceeding in the same way as in Theorem 2.1 and substituting ΔK in (15) we have that

$$\begin{aligned} \Delta W = & \int_0^\infty \frac{\partial}{\partial x} (v(x, 0) - v(-x, 0)) \Delta K^1(-x) dx - \int_0^\infty \frac{\partial}{\partial x} v(x, 0) (\Delta K^1(x) + \Delta K^1(-x)) dx \\ & - \int_0^\infty \frac{\partial}{\partial z} v(-a_1, z) \Delta K^2(z) dz - \int_{-\infty}^\infty \int_0^\infty \frac{\partial^2}{\partial x \partial z} v(x, z) \Delta K(x, z) dz dx \geq 0. \end{aligned} \quad (16)$$

Given that $K^1(x) = K(x, 0) = F^1(x)$ and $K^2(z) = K(-a_1, z) = F^2(z)$ and utilizing the assumptions of loss aversion, non-decreasingness given in Definition 2.2 and supermodularity, we have that (2), (3) and (7) are sufficient for (16) to hold for any v .

Conditions (2) and (3) are as in Theorem 2.1. We show the necessity of (7) by contradiction. Assume that there exists some x, z such that $K_A(x, z) - K_B(x, z) > 0$. Let $x < 0$. From the fact that the distribution function is right continuous, it follows that there is a neighborhood $(-b, -a) \times (c, d)$, $b > a > 0, d > c > 0$, such that for all (x, z) in this neighbourhood, $K_A(x, z) - K_B(x, z) > 0$. Consider the following function

$$\tilde{v}_4(x, z) = \begin{cases} -bc + bz + dx - bd & x \leq -b, z \leq c \\ (c+d)x + bz - bd & x \in (-b, -a), z \leq c \\ -(c+d)a + bz - bd & 0 > x \geq -a, z \leq c \\ dx - bd & x \leq -b, z \in (c, d) \\ xz + dx + bz - bd & (x, z) \in (-b, -a) \times (c, d) \\ (b-a)z - (a+b)d & 0 > x \geq -a, z \in (c, d) \\ dx - bd & x \leq -b, z \geq d \\ 2xd & x \in (-b, -a), z \geq d \\ -2ad & 0 > x \geq -a, z \geq d \\ 0 & x \geq 0. \end{cases}$$

Let us now check that \tilde{v}_4 fulfills Definition 2.2 but with supermodularity. Firstly, for $x < 0$, it is negative in each of the nine areas, given that $-b < -a < 0$ and $0 < c < d$. It is 0 for $x \geq 0$. Secondly, it is non-decreasing, e.g. the derivative of $(c+d)x + bz - bd$ with respect to x is $c+d > 0$, or the derivative of $xz + dx + bz - bd$ with respect to x is $z+d > 0$, because $z > 0$. Also, the derivative of $xz + dx + bz - bd$ with respect to z is $x+b > 0$ when $x > -b$. Thirdly, it is loss averse, as the derivative $\frac{\partial \tilde{v}_4(-x, z)}{\partial x} > 0 = \frac{\partial \tilde{v}_4(x, z)}{\partial x}$ for $x \geq 0$. Finally, it is supermodular given that the cross derivative of $xz + dx + bz - bd$ is 1 and 0 elsewhere.

Now let $x > 0$ and for all $x, z \in (a, b) \times (c, d)$ we have that $K_A(x, z) - K_B(x, z) > 0$. The following function with \tilde{C} chosen appropriately so that loss aversion is preserved, will lead to a contradiction.

$$\tilde{v}_5(x, z) = \begin{cases} \tilde{C}x & x \leq 0, \tilde{C} \text{ large} \\ ac & 0 < x \leq a, z \leq c \\ cx & x \in (a, b), z \leq c \\ bc & x \geq b, z \leq c \\ az & 0 < x \leq a, z \in (c, d) \\ xz & (x, z) \in (a, b) \times (c, d) \\ bz & x \geq b, z \in (c, d) \\ ad & 0 < x \leq a, z \geq d \\ dx & x \in (a, b), z \geq d \\ bd & x \geq b, z \geq d \end{cases}$$

□

Proof of Theorem 3.1. Let $v_x = \frac{\partial}{\partial x}v$; $v_z = \frac{\partial}{\partial z}v$; further, let v_{xx}, v_{zz} denote respective second order derivatives and $v_{xzz}, v_{xxz}, v_{xxx}$ mixed derivatives. To obtain second order conditions, we need to integrate (14) in Lemma 1 by parts. Let us first concentrate on the first two terms that correspond to, respectively, losses and gains, as the integration here is less standard. Let us denote

$$I := - \int_{-\infty}^0 \frac{\partial}{\partial x}v(x, a_3)F^1(x)dx + \int_0^\infty \frac{\partial}{\partial x}v(x, a_3)(1 - F^1(x))dx$$

in (14). We have that $H^{1'}(t) = F(t)dt$ and $S^{1'}(t) = (F(t) - 1)dt$. Recalling our bounded support assumption, we have

$$\begin{aligned} - \int_{-\infty}^0 \frac{\partial}{\partial x}v(x, a_3)F^1(x)dx &= - \int_{-\infty}^0 \frac{\partial}{\partial x}v(x, a_3)dH^1(x) = - \int_{-\infty}^0 \frac{\partial}{\partial x}v(x, a_3)H^{1'}(x)dx = \\ &= - \frac{\partial}{\partial x}v(x, a_3)H^1(x)|_{-\infty}^0 + \int_{-\infty}^0 \frac{\partial^2}{\partial x^2}v(x, a_3)H^1(x)dx = - \frac{\partial}{\partial x}v(0, a_3)H^1(0) + \int_{-\infty}^0 \frac{\partial^2}{\partial x^2}v(x, a_3)H^1(x)dx = \\ &= - \frac{\partial}{\partial x}v(-a_1, a_3)H^1(0) + \int_{-\infty}^0 \frac{\partial^2}{\partial x^2}v(x, a_3) (H^1(x) - H^1(0)) dx \end{aligned}$$

and

$$\begin{aligned}
& \int_0^\infty \frac{\partial}{\partial x} v(x, a_3) (1 - F^1(x)) dx = \int_0^\infty \frac{\partial}{\partial x} v(x, a_3) (-S^{1'}(x)) dx = \\
& = \frac{\partial}{\partial x} v(x, a_3) (-S^1(x)) \Big|_0^{a_2} - \int_0^\infty \frac{\partial^2}{\partial x^2} v(x, a_3) (-S^1(x)) dx = \frac{\partial}{\partial x} v(0, a_3) S^1(0) + \int_0^\infty \frac{\partial^2}{\partial x^2} v(x, a_3) S^1(x) dx = \\
& = \frac{\partial}{\partial x} v(a_2, a_3) S^1(0) + \int_0^\infty \frac{\partial^2}{\partial x^2} v(x, a_3) (S^1(x) - S^1(0)) dx
\end{aligned}$$

Putting these two pieces together we have

$$\begin{aligned}
I = \frac{\partial}{\partial x} v(a_2, a_3) S^1(0) - \frac{\partial}{\partial x} v(-a_1, a_3) H^1(0) + \int_{-\infty}^0 \frac{\partial^2}{\partial x^2} v(x, a_3) (H^1(x) - H^1(0)) dx + \\
+ \int_0^\infty \frac{\partial^2}{\partial x^2} v(x, a_3) (S^1(x) - S^1(0)) dx
\end{aligned}$$

and

$$\begin{aligned}
\Delta I = \frac{\partial}{\partial x} v(a_2, a_3) (S_A^1(0) - S_B^1(0)) - \frac{\partial}{\partial x} v(-a_1, a_3) (H_A^1(0) - H_B^1(0)) + \\
+ \int_{-\infty}^0 \frac{\partial^2}{\partial x^2} v(x, a_3) ((H_A^1(x) - H_A^1(0)) - (H_B^1(x) - H_B^1(0))) dx + \\
+ \int_0^\infty \frac{\partial^2}{\partial x^2} v(x, a_3) ((S_A^1(x) - S_A^1(0)) - (S_B^1(x) - S_B^1(0))) dx,
\end{aligned}$$

or equivalently

$$\begin{aligned}
\Delta I = \frac{\partial}{\partial x} v(a_2, a_3) (S_A^1(0) - S_B^1(0)) - \frac{\partial}{\partial x} v(-a_1, a_3) (H_A^1(0) - H_B^1(0)) + \\
+ \int_{-\infty}^0 \frac{\partial^2}{\partial x^2} v(x, a_3) ((H_A^1(x) - H_B^1(x)) - (H_A^1(0) - H_B^1(0))) dx + \\
+ \int_0^\infty \frac{\partial^2}{\partial x^2} v(x, a_3) ((S_A^1(x) - S_B^1(x)) - (S_A^1(0) - S_B^1(0))) dx, \quad (17)
\end{aligned}$$

from which we can see that given the S-shapedness of v the conditions $H_A^1(x) - H_B^1(x) \geq H_A^1(0) - H_B^1(0)$ and $S_A^1(x) - S_B^1(x) \leq S_A^1(0) - S_B^1(0)$ for all $x \geq 0$ are sufficient for the last two terms to

be positive. Furthermore, taking $x = -a_1$ in the first condition we get that $0 \geq H_A^1(0) - H_B^1(0)$ and similarly, taking $x = a_2$ in the second condition we get that $0 \leq S_A^1(0) - S_B^1(0)$, so these two conditions are sufficient for $\Delta I \geq 0$ and they are equivalent to (8). Furthermore, there is also equivalence with (11), namely, for all $x > 0 > y$ we have

$$\begin{aligned} S^1(x) - H^1(y) - (S^1(0) - H^1(0)) &= \int_x^\infty 1 - F^1(t) dt - \int_{-\infty}^y F^1(t) dt - \left(\int_0^\infty 1 - F^1(t) dt - \int_{-\infty}^0 F^1(t) dt \right) \\ &= \left(- \int_0^x 1 - F^1(t) dt \right) - \left(\int_{-\infty}^y F^1(t) dt - \int_{-\infty}^0 F^1(t) dt \right) \\ &= - \int_0^x 1 - F^1(t) dt + \int_y^0 F^1(t) dt = - \int_0^x dt + \int_y^x F^1(t) dt. \end{aligned}$$

Coming back to (8) the first term cancels out and we get (11).

Let us now come back to the full expression for ΔW . Rewriting ΔI and further integrating (14) by parts we get

$$\begin{aligned} \Delta W &= v_x(a_2, a_3) \Delta S^1(0) - v_x(-a_1, a_3) \Delta H^1(0) + \int_{-\infty}^0 v_{xx}(x, a_3) (\Delta H^1(x) - \Delta H^1(0)) dx + \\ &+ \int_0^\infty v_{xx}(x, a_3) (\Delta S^1(x) - \Delta S^1(0)) dx - v_z(a_2, a_3) \Delta H^2(a_3) + \int_0^\infty v_{zz}(a_2, z) \Delta H^2(z) dz + v_{xz}(a_2, a_3) \Delta H(a_2, a_3) + \\ &- \int_{-\infty}^\infty v_{xxz}(x, a_3) \Delta H(x, a_3) dx - \int_0^\infty v_{xzz}(a_2, z) \Delta H(a_2, z) dz + \int_{-\infty}^\infty \int_0^\infty v_{xxxz}(x, z) \Delta H(x, z) dz dx \geq 0. \end{aligned} \tag{18}$$

From this we see that, apart from (11) (or (8)), conditions (9) and (10) are sufficient. We now show that these conditions are also necessary by means of a contradiction. In order to violate (11) we will first assume that it is violated in the negative area. Towards a contradiction, assume that there exist intervals $(-b, -a)$, $b > a \geq 0$ such that for all $x \in (-b, -a)$ we have $H_A^1(x) - H_B^1(x) < H_A^1(0) - H_B^1(0)$, or equivalently, $\Delta H^1(x) - \Delta H^1(0) < 0$. The function v_8

$$v_8(x, z) = \begin{cases} a^2 - b^2 & x \leq -b \\ x^2 + 2bx + a^2 & -b < x \leq -a \\ 2(b-a)x & -a < x \leq 0 \\ 0 & x \geq 0. \end{cases}$$

fulfills Definition 3.1. Most importantly, for all $x \in (-b, -a)$ we have $v_{xx} > 0$ and $v_{xx} = 0$ otherwise. Thus using this v_8 we obtain $\int_{-\infty}^{\infty} v_{xx}(x, a_3) (\Delta H^1(x) - \Delta H^1(0)) dx < 0$, and the rest of the terms in (18) is zero, a contradiction. The case of positive area is the same, but the function v should be concave instead of convex.

In a similar fashion, towards a contradiction with (9), assume there exists an interval (c, d) , $d > c \geq 0$, such that for all $z \in (c, d)$ condition (9) is violated: $\Delta H^2(z) > 0$. The function v_9 , for $x \geq 0$

$$v_9(x, z) = \begin{cases} 0 & z = 0 \\ 2(d-c)z & 0 < z \leq c \\ -z^2 + 2dz - c^2 & c < z \leq d \\ d^2 - c^2 & d < z. \end{cases}$$

and $bx, b > 0$ for $x < 0$ fulfills Definition 3.1. Most importantly, for all $z \in (c, d)$ we have $v_{zz} < 0$ and $v_{zz} = 0$ otherwise. Thus using v_9 we obtain $\int_0^{\infty} v_{zz}(a_2, z) \Delta H^2(z) dz < 0$, while the rest of terms in (18) is zero, a contradiction with (18).

The necessity of (10) can be shown by applying the same approach as in Theorem 2.1, but replacing v_4 with respective derivatives. That is, we take $v_4 = \frac{\partial^2}{\partial x \partial z} v_6(x, z)$. Then, $\frac{\partial^4}{\partial x^2 \partial z^2} v_6(x, z) = \frac{\partial^2}{\partial x \partial z} v_4(x, z)$, i.e. $(v_6)_{xxzz} = (v_4)_{xz}$. Similarly, we take $v_3 = \frac{\partial^2}{\partial x \partial z} v_7(x, z)$ and we have that $(v_7)_{xxzz} = (v_3)_z$. A bit more tricky is the case of v_{xxz} where we take $v_1 = \frac{\partial^2}{\partial x \partial z} v_8(x, z)$, and $-v_1(-x, z) = \frac{\partial^2}{\partial x \partial z} v_9(x, z)$. We obtain $(v_8)_{xxz} = (v_1)_x$ on the negative domain of integration and $(v_9)_{xxz}(x, z) = -(v_1)_x(-x, z)$ for $x > 0$ on the positive domain of integration. \square

Proof of Theorem 3.2. Proceeding in the same way as in Theorem 3.1, substituting ΔL into (18) we get

$$\begin{aligned}
\Delta W &= v_x(a_2, 0)\Delta S^1(0) - v_x(-a_1, 0)\Delta H^1(0) + \int_{-\infty}^0 v_{xx}(x, 0) (\Delta H^1(x) - \Delta H^1(0)) dx + \\
&+ \int_0^\infty v_{xx}(x, 0) (\Delta S^1(x) - \Delta S^1(0)) dx - v_z(-a_1, a_3)\Delta H^2(a_3) + \int_0^\infty v_{zz}(-a_1, z)\Delta H^2(z) dz - v_{xz}(a_2, a_3)\Delta L(a_2, a_3) + \\
&+ \int_{-\infty}^\infty v_{xxz}(x, a_3)\Delta L(x, a_3) dx + \int_0^\infty v_{xzz}(a_2, z)\Delta L(a_2, z) dz - \int_{-\infty}^\infty \int_0^\infty v_{xxxz}(x, z)\Delta L(x, z) dz dx \geq 0,
\end{aligned} \tag{19}$$

Conditions (8) and (9) are sufficient and necessary as in Theorem 3.1, and condition (12) is sufficient too.

The necessity of (12) can be shown by noticing that (18) and (19) look similar, except that ΔH is replaced by ΔL and signs of the terms that include ΔL are opposite. In those terms we can consider $-v$ and see that it needs to have opposite sign, that is, if for (18) v needs to be submodular, decreasingly submodular and second-degree submodular and for (19) it needs to be supermodular, decreasingly supermodular and second-degree supermodular. \square

Proof of Theorem 4.1. Using Lemma 1 we have

$$\begin{aligned}
\Delta W &= \int_0^\infty \frac{\partial}{\partial x} (v(x, a_3) - v(-x, a_3)) \Delta F^1(-x) dx - \int_0^\infty \frac{\partial}{\partial x} v(x, a_3) (\Delta F^1(x) + \Delta F^1(-x)) dx \\
&\quad - \int_0^\infty \frac{\partial}{\partial z} v(a_2, z) \Delta F^2(z) dz + \int_{-\infty}^\infty \int_0^\infty \frac{\partial^2}{\partial x \partial z} v(x, z) \Delta F(x, z) dz dx.
\end{aligned}$$

We integrate only the last two terms by parts getting

$$\int_0^\infty \frac{\partial}{\partial z} v(a_2, z) \Delta F^2(z) dz = \frac{\partial}{\partial z} v(a_2, a_3) \Delta H^2(a_3) - \int_0^\infty \frac{\partial^2}{\partial z^2} v(a_2, z) \Delta H^2(z) dz$$

and

$$\begin{aligned}
\int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial^2}{\partial x \partial z} v(x, z) \Delta F(x, t) dz dx = \\
\frac{\partial^2}{\partial x \partial z} v(a_2, a_3) H(a_2, a_3) - \int_{-\infty}^{\infty} \frac{\partial^3}{\partial x^2 \partial z} v(x, a_3) \Delta H(x, a_3) dx \\
- \int_0^{\infty} \frac{\partial^3}{\partial x \partial z^2} v(a_2, z) \Delta H(a_2, z) dz \\
+ \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\partial^4}{\partial x^2 \partial z^2} v(x, z) \Delta H(x, z) dz dx
\end{aligned}$$

Altogether we have

$$\begin{aligned}
\Delta W = \int_0^{\infty} (v_x(x, a_3) - v_x(-x, a_3)) \Delta F^1(-x) dx - \int_0^{\infty} v_x(x, a_3) (\Delta F^1(x) + \Delta F^1(-x)) dx \\
- v_z(a_2, a_3) \Delta H^2(a_3) + \int_0^{\infty} v_{zz}(a_2, z) \Delta H^2(z) dz + v_{xz}(a_2, a_3) H(a_2, a_3) - \int_{-\infty}^{\infty} v_{xxz}(x, a_3) \Delta H(x, a_3) dx \\
- \int_0^{\infty} v_{xzz}(a_2, z) \Delta H(a_2, z) dz + \int_{-\infty}^{\infty} \int_0^{\infty} v_{xxzz}(x, z) \Delta H(x, z) dz dx \geq 0. \quad (20)
\end{aligned}$$

Given the properties of function v as in Definition 4.1 we can see that the conditions (2) (or, equivalently (5) and (6)) in Theorem 2.1 and conditions (9) and (10) in Theorem 3.1 are sufficient for (20) to hold.

Necessity of condition (2) can be shown in the same way as in Theorem 2.1. Similarly, necessity of condition (9) can be shown in the same way as in Theorem 3.1. And finally, necessity of condition (10) can be shown in the same way as in Theorem 3.1 as well with the exception of the case of v_{xxz} , where we take $v_1 = \frac{\partial^2}{\partial x \partial z} v_8(x, z)$, and $v_9 = \begin{cases} \int^z \int^x -v_1(-t, s) dt ds & x > 0 \\ Cx & x \leq 0 \end{cases}$, for some large $C > 0$ to preserve loss aversion in x . We obtain $(v_8)_{xxz} = (v_1)_x$ on the negative domain of integration and $(v_9)_{xxz}(x, z) = -(v_1)_x(-x, z)$ for $x > 0$ on the positive domain of integration. \square

Proof of Theorem 4.2. Proceeding in the same way as in Theorem 4.1 we obtain

$$\begin{aligned}
\Delta W = & \int_0^\infty (v_x(x, 0) - v_x(-x, 0)) \Delta K^1(-x) dx - \int_0^\infty v_x(x, 0) (\Delta K^1(x) + \Delta K^1(-x)) dx \\
& - v_z(-a_1, a_3) \Delta K^2(a_3) + \int_0^\infty v_{zz}(a_2, z) \Delta H^2(z) dz - v_{xz}(a_2, a_3) L(a_2, a_3) + \int_{-\infty}^\infty v_{xxz}(x, a_3) \Delta L(x, a_3) dx \\
& + \int_0^\infty v_{xzz}(a_2, z) \Delta L(a_2, z) dz - \int_{-\infty}^\infty \int_0^\infty v_{xxxz}(x, z) \Delta L(x, z) dz dx \geq 0. \quad (21)
\end{aligned}$$

Given the properties of function v as in Definition 4.1 with some modified as in Definition 4.3 we can see that the conditions (2) in Theorem 2.1, (9) in Theorem 3.1 and (12) in Theorem 3.2 are sufficient and necessary for (21) to hold. \square

Proof of Theorem 5.1. All of the functions g discussed in the main text map a pair of distribution functions to a norm of a (vector) function. There is a common vocabulary of transformations for all the functions. Letting $f = (f_A, f_B) \in (\ell^\infty(\mathbb{R}^k))^2$ (the space of pairs of bounded functions from \mathbb{R}^k to \mathbb{R}), these common elements include the maps

$$f \mapsto f_A \pm f_B, \quad f \mapsto \int_a^b f, \quad f \mapsto \|[f]_+\|. \quad (22)$$

The bootstrap depends on the Hadamard directional derivative of the map $(F_A, F_B) \mapsto \|[g(F_A, F_B)]_+\|$. Hadamard directionally differentiable maps obey a chain rule (Shapiro, 1990) and therefore it is of interest to note the derivatives here. The first two transforms in (22) are linear, and thus fully (therefore also directionally) Hadamard differentiable maps. Finally, for direction $h \in \ell^\infty(\mathbb{R}^k)$, Firpo et al. (forthcoming) show that when $f \leq 0$,

$$\lim_{t \searrow 0} |t^{-1} (\|[f + h]_+\| - \|[f]_+\|) - \|[h \cdot \chi_0]_+\|| = 0, \quad (23)$$

where $\chi_0(x) = 1$ if $f(x) = 0$ and is zero otherwise.

Given the form of the derivatives, all the test statistics have can be composed with some chain of the above maps. Under Assumptions **A1** and **A2**, we have $\sqrt{n}((\hat{F}_A, \hat{F}_B) - (F_A, F_B)) \rightsquigarrow \mathcal{G}_F$, where \mathcal{G}_F is a Gaussian process, and this convergence is uniform over the space \mathcal{F} (apply Theorem 2.8.4 of van der Vaart and Wellner (2023) to the sets $I(X \leq x)$, for $x \in \mathbb{R}^k$). Then Theorem 3.2 of Fang and Santos (2019) implies the result. \square

Proof of Theorem 5.2. The set of distribution functions is convex and the function $g \mapsto \|g \cdot \chi_{c_n}\|$ is convex, while the other mappings of distribution function F to test statistic T are linear, implying the test statistic is a convex map of the distribution functions. Furthermore, Corollary A.2.9 of van der Vaart and Wellner (2023) implies that if $q(1 - \alpha) > 0$, the CDF of the limiting distribution of $\sqrt{n}T_n$ is strictly increasing. Finally, it is assumed that local distributions are in \mathcal{F}_0 . Then Theorem 5.1 and an application of Corollary 3.2 of Fang and Santos (2019) imply the result. \square

References

- Steffen Ahrens, Inske Pirschel, and Dennis J. Snower. A theory of price adjustment under loss aversion. *Journal of Economic Behavior & Organization*, 134:78–95, 2017.
- Alberto Alesina and Francesco Passarelli. Loss aversion, politics and redistribution. *American Journal of Political Science*, 63:936–947, 2019.
- Donald W. K. Andrews and Xiaoxia Shi. Inference based on conditional moment inequalities. *Econometrica*, 81:609–666, 2013.
- J. Richard Aronson, Peter Johnson, and Peter J. Lambert. Redistributive effect and unequal income tax treatment. *The Economic Journal*, 104(423):262–270, 1994.
- Anthony B. Atkinson. On the measurement of inequality. *Journal of Economic Theory*, 2:244–263, 1970.
- Anthony B. Atkinson and François Bourguignon. The comparison of multidimensioned distributions of economic status. *Review of Economic Studies*, 49:183–201, 1982.
- Anthony B. Atkinson and François Bourguignon. Income distribution and differences in needs. In George R. Feiwel, editor, *Arrow and the Foundation of the Theory of Economic Policy*, chapter 12, pages 350–370. Macmillan, 1987.
- Giuseppe Attanasi, Luca Corazzini, and Francesco Passarelli. Voting as a lottery. *Journal of Public Economics*, 146:129–137, 2017.
- Alan J. Auerbach and Kevin A. Hassett. A new measure of horizontal equity. *American Economic Review*, 92(4):1116–1125, 2002.
- Shlomo Benartzi and Richard H. Thaler. Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics*, 110:73–92, 1995.
- Marianne P. Bitler, Jonah B. Gelbach, and Hilary W. Hoynes. What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, 96:988–1012, 2006.
- David Blake, Edmund Cannon, and Douglas Wright. Quantifying loss aversion: Evidence from a UK population survey. *Journal of Risk and Uncertainty*, 63:27–57, 2021.

- François Bourguignon. Status quo in the welfare analysis of tax reforms. *Review of Income and Wealth*, 4:603–621, 2011.
- François Bourguignon and Satya R. Chakravarty. The measurement of multidimensional poverty. *The Journal of Economic Inequality*, 1:25–49, 2003.
- David Bowman, Deborah Minehart, and Matthew Rabin. Loss aversion in a consumption–savings model. *Journal of Economic Behavior & Organization*, 38:155–178, 1999.
- Alexander L. Brown, Taisuke Imai, Ferdinand M. Vieider, and Colin F. Camerer. Meta-analysis of empirical estimates of loss aversion. *Journal of Economic Literature*, 62:485–516, 2024.
- Colin Camerer, Linda Babcock, George Loewenstein, and Richard Thaler. Labor supply of New York City cabdrivers: One day at a time. *The Quarterly Journal of Economics*, 112:407–441, 1997.
- Jonathan Chapman, Erik Snowberg, Stephanie W. Wang, and Colin Camerer. Looming large or seeming small? Attitudes towards losses in a representative sample. *Review of Economic Studies*, forthcoming.
- Hugh Dalton. The measurement of the inequality of incomes. *The Economic Journal*, 30:348–361, 1920.
- Stefano DellaVigna, Attila Lindner, Balázs Reizer, and Johannes F. Schmieder. Reference-dependent job search: Evidence from Hungary. *The Quarterly Journal of Economics*, 132:1969–2018, 2017.
- Sanjit Dhami. *The Foundations of Behavioral Economic Analysis*. Oxford University Press, Oxford, 2016.
- Sanjit Dhami and Ali al Nowaihi. Optimal taxation in the presence of tax evasion: Expected utility versus prospect theory. *Journal of Economic Behavior & Organization*, 75:313–337, 2010.
- Lucia F. Dunn. Loss aversion and adaptation in the labor market: Empirical indifference functions and labor supply. *The Review of Economics and Statistics*, 78:441–450, 1996.
- Louis Eeckhoudt and Harris Schlesinger. Higher-order risk attitudes. In Georges Dionne, editor, *Handbook of Insurance*, chapter 2, pages 41–57. Springer, 2013.
- Louis Eeckhoudt, Béatrice Rey, and Harris Schlesinger. A good sign for multivariate risk taking. *Management Science*, 53:117–124, 2007.
- Per Engström, Katarina Nordblom, Henry Ohlsson, and Annika Persson. Tax compliance and loss aversion. *American Economic Journal: Economic Policy*, 7:132–164, 2015.
- Nir Eyal. Why challenge trials of SARS-CoV-2 vaccines could be ethical despite risk of severe adverse events. *Ethics & Human Research.*, 42:24–34, 2020.
- Zheng Fang and Andres Santos. Inference on directionally differentiable functions. *Review of Economic Studies*, 86:377–412, 2019.
- Martin Feldstein. On the theory of tax reform. *Journal of Public Economics*, 6(1-2):77–104, 1976.

- Sergio Firpo, Antonio F. Galvao, and Thomas Parker. Uniform inference for value functions. *Journal of Econometrics*, 235:1680–1699, 2023.
- Sergio Firpo, Antonio F. Galvao, Martyna Kobus, Thomas Parker, and Pedro Rosa-Dias. Loss aversion and the welfare ranking of policy interventions. *Journal of Econometrics*, forthcoming.
- Caroline Freund and Çağlar Özden. Trade policy and loss aversion. *American Economic Review*, 98:1675–1691, 2008.
- Thibault Gajdos and John A. Weymark. Introduction to inequality and risk. *Journal of Economic Theory*, 147:1313–1330, 2012.
- Jacob Goldin and Daniel Reck. Optimal defaults with normative ambiguity. *The Review of Economics and Statistics*, 104:17–33, 2022.
- Han Hong and Jessie Li. The numerical bootstrap. *The Annals of Statistics*, 48:397–412, 2020.
- Octave Jokung. Risk apportionment via bivariate stochastic dominance. *Journal of Mathematical Economics*, 47:448–452, 2011.
- Jeffrey P. Kahn, Leslie Meltzer Henry, Anna C. Mastroianni, Wilbur H Chen, and Ruth Macklin. For now, it’s unethical to use human challenge studies for SARS-CoV-2 vaccine development. *Proceedings of the National Academy of Sciences USA*, 117:28538–28542, 2020.
- Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47:263–292, 1979.
- Louis Kaplow. Horizontal equity: Measures in search of a principle. *National Tax Journal*, 42(2): 139–154, 1989.
- Louis Kaplow. A fundamental objection to tax equity norms: A call for utilitarianism. *National Tax Journal*, 48(4):497–514, 1995.
- Botond Köszegi and Matthew Rabin. A model of reference-dependent preference. *The Quarterly Journal of Economics*, 121:1133–1165, 2006.
- Mervyn A. King. An index of inequality: With application to horizontal equity and social mobility. *Econometrica*, 51(1):99–115, 1983.
- Sokbae Lee, Kyungchul Song, and Yoon-Jae Whang. Testing for a general class of functional inequalities. *Econometric Theory*, 34:1018–1064, 2018.
- Haim Levy. *Stochastic Dominance: Investment Decision Making Under Uncertainty*. Springer International Publishing, Switzerland, third edition, 2016.
- Haim Levy and Zvi Wiener. Stochastic dominance and prospect dominance with subjective weighting functions. *Journal of Risk and Uncertainty*, 16:147–163, 1998.
- Oliver Linton, Esfandiar Maasoumi, and Yoon-Jae Whang. Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies*, 72:735–765, 2005.
- Oliver Linton, Kyungchul Song, and Yoon-Jae Whang. An improved bootstrap test of stochastic dominance. *Journal of Econometrics*, 154:186–202, 2010.

- Paul Makdissi and Myra Yazbeck. Measuring socioeconomic health inequalities in presence of multiple categorical information. *Journal of Health Economics*, 34:84–95, 2014.
- Thomas Meissner, Xavier Gassmann, Corinne Faure, and Joachim Schleich. Individual characteristics associated with risk and time preferences: A multi country representative survey. *Journal of Risk and Uncertainty*, 66:77–107, 2023.
- Richard A. Musgrave. *The Theory of Public Finance*. McGraw-Hill, New York, 1959.
- Ted O’Donoghue and Charles Sprenger. Reference-dependent preferences. In B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, editors, *Handbook of Behavioral Economics: Applications and Foundations 1*. North-Holland, 2018.
- George A. Quattrone and Amos Tversky. Contrasting rational and psychological analyses of political choice. *American Political Science Review*, 82:719–736, 1988.
- Daniel Reck and Arthur Seibold. The welfare economics of reference dependence. Working Paper 31381, National Bureau of Economic Research, June 2023.
- Alex Rees-Jones. Quantifying loss-averse tax manipulation. *The Review of Economic Studies*, 85:1251–1278, 2018.
- Alex Rees-Jones. Behavioral incentive compatibility and empirically informed welfare analysis: An introductory guide. *Journal of Economic Perspectives*, 38:155–174, 2024.
- Dani Rodrik. Political economy of trade policy. In G.M. Grossman and K. Rogoff, editors, *Handbook of International Economics*, volume 3, chapter 28, pages 1457–1494. Elsevier, 1995.
- Michael Rothschild and Joseph E. Stiglitz. Increasing risk: I. a definition. *Journal of Economic Theory*, 2:225–243, 1970.
- Kai Ruggeri, Sonia Alí, Mari Louise Berge, Giulia Bertoldo, Ludvig D. Bjørndal, Anna Cortijos-Bernabeu, Clair Davison, Emir Dmić, Celia Esteban-Serna, Maja Friedemann, Shannon P. Gibson, Hannes Jarke, Ralitsa Karakasheva, Peggah R. Khorrami, Jakob Kveder, Thomas Lind Andersen, Ingvild S. Lofthus, Lucy McGill, Ana E. Nieto, Jacobo Pérez, Sahana K. Quail, Charlotte Rutherford, Felice L. Tavera, Nastja Tomat, Chiara Van Reyn, Bojana Većkalov, Keying Wang, Aleksandra Yosifova, Francesca Papa, Enrico Rubaltelli, Sander van der Linden, and Tomas Folke. Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, 4:622–633, 2020.
- William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1:7–59, 1988.
- Arthur Seibold. Reference points for retirement behavior: Evidence from German pension discontinuities. *American Economic Review*, 111:1126–1165, 2021.
- Amartya K. Sen. *Collective Choice and Social Welfare*. Elsevier Science, Amsterdam, 1970.
- Alexander Shapiro. On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66:477–487, 1990.

- Daniel T. Slesnick. The measurement of horizontal inequality. *The Review of Economics and Statistics*, 71(3):481–490, 1989.
- Jan Helge Solbakk, Heidi Beate Bentzen, Søren Holm, Anne Kari Tolo Heggstad, Bjørn Hofmann, Annette Robertsen, Anne Hambro Alnæs, Shereen Cox, Reidar Pedersen, and Rose Bernabe. Back to what? the role of research ethics in pandemic times. *Medicine, Health Care and Philosophy*, 24(1):3–20, 2021.
- Richard Thaler. Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, 1:39–60, 1980.
- Patricia Tovar. The effects of loss aversion on trade policy: Theory and evidence. *Journal of International Economics*, 78:154–167, 2009.
- Amos Tversky and Daniel Kahneman. Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106:1039–1061, 1991.
- A. W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, Cham, second edition, 2023.