KU LEUVEN

DEPARTMENT OF ECONOMICS

Revealed social preferences

Arthur DOLGOPOLOV and Mikhail FREER

FACULTY OF ECONOMICS AND BUSINESS

DISCUSSION PAPER SERIES DPS18.06

JULY 2018

REVEALED SOCIAL PREFERENCES

ARTHUR DOLGOPOLOV

Interdisciplinary Center for Economic Science, George Mason University, 4400 University Drive, MSN 1B2, Fairfax, VA 22030, USA

MIKHAIL FREER

Department of Economics, University of Leuven (KU Leuven), Etienne Sabbelaan 53, 8500 Kortrijk, Belgium

Abstract. We use a revealed preference approach to develop tests for the observed behavior to be consistent with theories of social preferences. In particular, we provide revealed preference criteria for the observed set of choices generated by inequality averse preferences and by other-regarding preferences that exhibit increasing benevolence. These tests can be applied to games commonly used to study social preferences, including dictator, ultimatum, investment and carrot-stick games. We further apply these tests to experimental data on dictator and ultimatum games. Finally, we show how to identify the levels of altruism and fair outcomes using the developed revealed preference conditions.

1 Introduction

The revealed preference approach, pioneered by Samuelson (1938), originates in the fact that, although we cannot observe complete preference relation profiles of players, we can observe their choices over some budget sets. This approach was widely used to develop and apply tests for theories of individual behavior. Starting with Richter (1966) and Afriat (1973)

E-mail addresses: `Dolgopolov: adolgopo@gmu.edu, Freer: mikhail.freer@kuleuven.be`.

Date: June 26, 2018.

We would like to thank Oyvind Aas, Christian Basteck, Laurens Cherchye, Bram De Rock, Thomas Demuynck, Daniel Houser and Georg Kirchsteiger for insightful discussions on the earlier versions of the paper.

the approach has been applied to construct tests of individual and collective decision making (see Chambers and Echenique, 2016, for comprehensive overview of the results). Revealed preference theory has been applied to other-regarding preferences starting with Andreoni and Miller (2002). This assumption has testable implications and is supported by the data (see e.g. Andreoni, 1990; Andreoni and Miller, 2002; Charness and Rabin, 2002; Fisman, Kariv and Markovits, 2007; Porter and Adams, 2016; Castillo, Cross and Freer, 2017).

However, other-regarding preferences is a rather weak hypothesis that only requires players to "care" about the payoffs other players receive. In particular, it allows for both altruistic and envious behavior. At the same time there are various theories nested within other-regrading preferences (e.g. inequality aversion, increasing benevolence, fairness preferences, etc.). While these theories attracted a lot of attention in experimental and theoretical literature, the research providing revealed preference characterizations for them is rather limited. Cox, Friedman and Sadiraj (2008) provides necessary conditions for *increasing benevolence* preferences. Deb, Gazzale and Kotchen (2014) applies revealed preference theory to construct criteria for several versions of other-regarding preferences including a special case of *inequality aversion*.[1]

Inequality aversion assumes that players do not like inequality; that is, they encounter some disutility if payoffs are unbalanced. This assumption explains some experimental data (see e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000, 2006; Fehr, Naef and Schmidt, 2006). Inequality aversion is a popular instrument in political economy, and is often applied to studies of income redistribution (Fong, 2001; Tyran and Sausgruber, 2006; Höchtl, Sausgruber and Tyran, 2012; Durante, Putterman and Van der Weele, 2014; Agranov and Palfrey, 2015).

Increasing benevolence assumes that a player's willingness to pay for an additional dollar received by another player increases in the player's own payoff. This is a natural assumption. It has been used to guarantee efficiency in bilateral exchange (Benjamin, 2015). In

[1]In addition, Saito (2013) develops an axiomatization of preferences in risky environments that distinguishes between desire for equality of opportunity and for equality of outcome. Rohde (2010) axiomatizes the Fehr and Schmidt's model of inequity aversion and relates it to rank-dependent utility.

addition, Benjamin (2015) further showed that this assumption is better suited than "altruistic preferences" to explain some common phenomena, e.g., the rotten-kid theorem (Becker, 1974).

This paper constructs revealed preference tests for *inequality averse* and *increasing benevolence* preferences. We start with a standard choice environment over linear budgets, in which a player decides how to allocate money between herself and another player (akin dictator game). Next, we generalize tests for various other games that are used to study social preferences, including ultimatum, trust and carrot-stick games. In addition, we apply the tests to the experimental data on dictator and ultimatum games. We find that different theories of behavior perform best in different environments, and even for different populations for the same environment, providing more evidence for context-dependence of social preferences.

Let us provide closer connection to the previous results on revealed social preferences. Cox, Friedman and Sadiraj (2008) provides a necessary revealed preferences conditions for observed choices to be consistent with increasing benevolence and provides a method of comparing subjects in terms of altruism if the demand functions are completely observed. We show that conditions proposed are also sufficient, and that the comparisons in terms of altruism can be applied even if only the finite set of choices is observed. Deb, Gazzale and Kotchen (2014) constructs revealed preference tests for a special case of inequality aversion (inequality aversion in differences), which can be applied for linear budgets. We offer a more general test (including generalized axiomatization of inequality measures), which can be applied beyond the linear budgets and show that in the context of linear budgets the choice of inequality measure function does not have testable implications. That is, if a subject is consistent with inequality averse preferences given some measure of inequality, she is consistent with inequality aversion preferences given any measure of inequality.

The remainder of this paper is organized as follows. Section 2 presents the general set up and the revealed preference tests as well as extensions of the test for other games. Section 3 provides empirical illustrations. Section 4 presents the partial identification of the level of altruism and notion of fair outcome using the revealed preference conditions. All proofs are collected in Appendix A.

2 Theoretical Framework

We consider a dictator game, which is structured as follows. A player decides how to allocate a given amount of money between herself and the other player, and the chosen allocation is implemented. This game can be written as a decision problem, in which one player chooses a two-dimensional vector allocation: payoff to self and payoff to another player.

Let $X \subseteq \mathbb{R}^2_+$ be the **set of alternatives**. For every $x \in X$ let $x = (x_s, x_o)$, where x_s is the payoff to self and x_o is the payoff to another player. Let $p \in R^2_{++}$ be a price vector.[2] Income is normalized to one at every point, and the budget set is defined as, $B(p) = \{x \in X : px \leq 1\}$. Let $E = (x^t, p^t)_{t=1}^T$ be an **experiment**, which consists of T choices (x^t) at a given price vector (p^t). Moreover, we assume that chosen points x^t are such that $p^t x^t = 1$.[3] A function $u(x) : X \to \mathbb{R}$ **rationalizes** the consumption experiment E if for all $y \in B(p^t)$, $u(x^t) \geq u(y)$ for every $t \in \{1, \ldots, T\}$.

In what follows we present revealed preference tests for each of the theories of social preferences. We start with other-regarding preferences. Next, we present the test for inequality averse preferences and the test for increasing benevolence preferences. Next, we show that the latter two theories are independent. Finally, we show how to apply the tests to ultimatum, trust and carrot-stick games.

2.1 Other-Regarding Preferences (OR). Other-regarding preferences assume that a player cares about her own payoff and the payoff of the recipient. Theory does not make an explicit assumption of whether the player derives utility or disutility from x_o.

Definition 1. *An experiment* $E = (x^t, p^t)_{t=1}^T$ *is* **rationalizable with other-regarding preferences** *if there is a continuous and locally non-satiated utility function* $u(x_s, x_o)$ *that rationalizes* E.

Other-regarding preferences include utility function that is monotone in both payoffs as a special case.

[2]We consider the linear budgets here for the simplicity of exposition however results presented can be extended for the nonlinear budget as in Forges and Minelli (2009).

[3]This technical assumption is dictated by non-satiation of preferences. All the further reasoning can be done without this assumption using more complicated notation.

Definition 2. *An experiment* $E = (x^t, p^t)_{t=1}^T$ *is consistent with* **Generalized Axiom of Revealed Preference (GARP)** *if and only if we have* $p^{t_1}x^{t_n} \leq p^{t_1}x^{t_1}$ *for all sequences* $x^{t_1}, \ldots, x^{t_n}$*, such that* $p^{t_{j+1}}x^{t_j} \leq p^{t_{j+1}}x^{t_{j+1}}$*,* $j \in \{i, \ldots, n-1\}$*.*

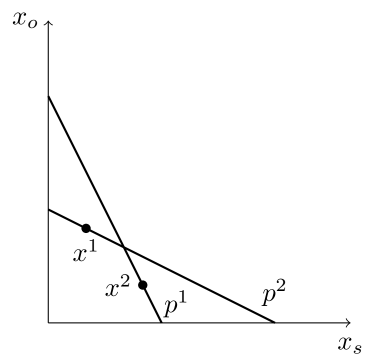


FIGURE 1. Other-Regarding Preferences and GARP

Figure 1 presents the violation of GARP. An allocation x^1 is chosen given prices p^1, therefore, it is better than any allocation which is available at p^1. At the same time x^2 is available at p^1, therefore, x^1 is strictly directly revealed preferred to x^2. Finally x^2 is directly revealed preferred to x^1, since x^1 is available at p^2. Hence, observed choices could not be generated by maximization of utility function.

Proposition 1 (Afriat (1967); Diewert (1973); Varian (1982))**.** *An experiment is rationalizable with other-regarding preferences if and only if it satisfies GARP.*

2.2 Inequality Aversion (IA) Inequality aversion assumes that player gets utility from her own payoff and disutility if payoffs are unbalanced, In order to quantify the "unbalancedness" of the payoffs we use the *inequality measure*. Some examples of commonly used inequality measures are presented below.

– Inequality in differences (e.g. Fehr and Schmidt, 1999; Tyran and Sausgruber, 2006; Agranov and Palfrey, 2015):

$$f(x_s, x_o) = \begin{cases} x_s - x_o & \text{if } x_s \geq x_o \\ \beta(x_s - x_o) & \text{if } x_o > x_s \end{cases}$$

where $\beta \leq 1$.

– Inequality in shares or Lorenz curve[4] (e.g., Bolton and Ockenfels, 2000):

$$f(x_s, x_o) = \begin{cases} \frac{x_s}{x_s+x_o} - \frac{1}{2} & \text{if } x_s \geq x_o \\ \beta\left(\frac{x_s}{x_s+x_o} - \frac{1}{2}\right) & \text{if } x_o > x_s \end{cases}$$

where $\beta \leq 1$.

– Gini Index $f(x_s, x_o) = \frac{|x_s - x_o|}{2(x_s+x_o)}$ (e.g. Durante, Putterman and Van der Weele, 2014)

Further we present an axiomatization of the inequality measure. Closely related axiomatization has been used by Fehr, Kirchsteiger and Riedl (1998) for the gift exchange game. This axiomatization generalizes the examples of inequality measures presented above.

Definition 3. *A continuous function $f(x_s, x_o)$ is an* **inequality measure** *if:*

– $f(x_s, x_o) \geq 0$, *for every x_s, x_o;*

– $f(x_s, x_o) = 0$ *if and only if $x_s = x_o$;*

– *if $x_s > x_o$, then $f(x_s, x_o)$ is decreasing in x_o and increasing in x_s;*

– *if $x_s < x_o$, then $f(x_s, x_o)$ is increasing in x_o and decreasing in x_s;*

– $f(\max\{x_s, x_o\}, \min\{x_s, x_o\}) \leq f(\min\{x_s, x_o\}, \max\{x_s, x_o\})$.

Further we present the definition for rationalization with inequality averse preferences. This rationalization, in general, depends on the inequality measure chosen.

Definition 4. *Let $f(x_s, x_o)$ be an inequality measure. An experiment is* **rationalizable with inequality averse preferences** *if there is a continuous utility function $u(x_s, f(x_s, x_o))$ that is increasing in x_s and decreasing in $f(x_s, x_o)$ that rationalizes it.*

Rationalizability with inequality averse preferences requires every player to choose to allocate to herself at least as much as to the other player. This condition is necessary, because if $x_s < x_o$, then the player could obtain greater utility by increasing x_s at the cost of x_o. Hence, player can set up $x'_s > x_s$ and $x'_o < x_o$ such that $x'_s \leq x'_o$. In this case $f(x'_s, x'_o) < f(x_s, x_o)$ and therefore, (x'_s, x'_o) should be strictly better than (x_s, x_o). That is the player has chosen a strictly dominated outcome and therefore cannot be rationalized with maximization of inequality averse utility function. This condition together with GARP is sufficient for rationalizability with inequality averse preferences.

[4]Intuition for Lorenz curve is the same as for inequality in shares for the two-player case since the shape of the curve is determined by deviation of the lower payoff from equal share split.

Proposition 2. *Let $f(x_s, x_o)$ be an inequality measure. An experiment is rationalizable with inequality averse preferences if and only if it satisfies GARP and $x_s^t \geq x_o^t$ for every $t \in \{1, \ldots, T\}$.*

Condition for rationalization with inequality averse preferences does not depend on the inequality measure. That is, condition is the same for any inequality measure. Hence, the following corollary immediately follows from Proposition 2.

Corollary 1. *An experiment is rationalizable with inequality averse preferences, if and only if it is rationalizable with any inequality measure.*

2.3 Increasing Benevolence (IB) Increasing benevolence means that a player's willingness to pay for an additional dollar given to the other player is increasing in own payoff x_s.[5]

Denote the demand function for x_o by $D_o(p_s, p_o)$ and the demand for x_s by $D_s(p_s, p_o)$. Since we operate in a two-dimensional case, one demand can be immediately derived from another $D_s(p_s, p_o) = \frac{1 - p_o D_o(p_s, p_o)}{p_s}$.

Definition 5. *An experiment is* ***rationalizable with increasing benevolence preferences*** *if there is a rational demand function $D_o(p_s, p_o)$ such that*

- $D_o(p_s^t, p_o^t) = x_o^t$, *and*
- $\frac{p_o}{p_s} \geq \frac{p'_o}{p'_s}$ *and* $\frac{1 - p'_o D_o(p_s, p_o)}{p'_s} \geq D_s(p_s, p_o)$ *implies* $D_o(p_s, p_o) \leq D_o(p'_s, p'_o)$.

Increasing benevolence is equivalent to normality of x_o. A good is said to be normal if its demand is increasing function of income. Necessity of normality for increasing benevolence is quite obvious. Figure 2 illustrates why normality is sufficient for increasing benevolence. Assume that x^1 is a point chosen from the budget defined by p^1; then, the new budget is such that $\frac{p_s}{p_o} \geq \frac{p'_s}{p'_o}$ (x_s is relatively more expensive in the new budget) and the old bundle

[5]This can be defined more formally with the marginal rate of substitution – $WTP = 1/MRS = \frac{u_{x_o}}{u_{x_s}}$ is increasing in x_s. We use the reduced form definition of this, which is necessary but not sufficient. However, it is sufficient to guarantee the empirical implications described by Cox, Friedman and Sadiraj (2008). Moreover, if we define MRS via the ratio of the inverse demand functions (to guarantee the existence of MRS), some sufficiency result can be inferred. Although, one can easily check that if we, for instance, assume that x_s and x_o are substitutes, then the demand conditions would be sufficient for the MRS version.

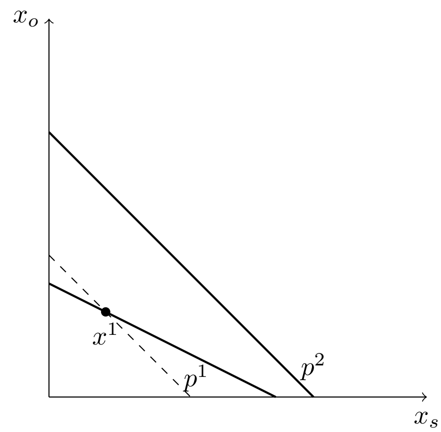


FIGURE 2. Increasing Benevolence and Normality

is attainable. The dashed line shows the parallel downward shift of the budget defined by p_2. Hence, the choice from the dashed budget should be with at least as much x_o as from p^1 (due to the substitution effect). Furthermore, since dashed and p^2 budgets are different only in income, then normality would guarantee that the choice from p^2 would be "above" the x^1.

Definition 6. *An experiment* $E = (x^t, p^t)_{t=1}^T$ *is consistent with* **Normality Axiom of Revealed Preference (NARP)** *if and only if for all observations* $t, v \in \{1, \dots, T\}$ *if* $p_o^t / p_s^t \leq p_o^v / p_s^v$ *and* $x_s^v \leq \frac{1 - p_o^t x_o^v}{p_s^t}$, *then* $x_o^v \leq x_o^t$.

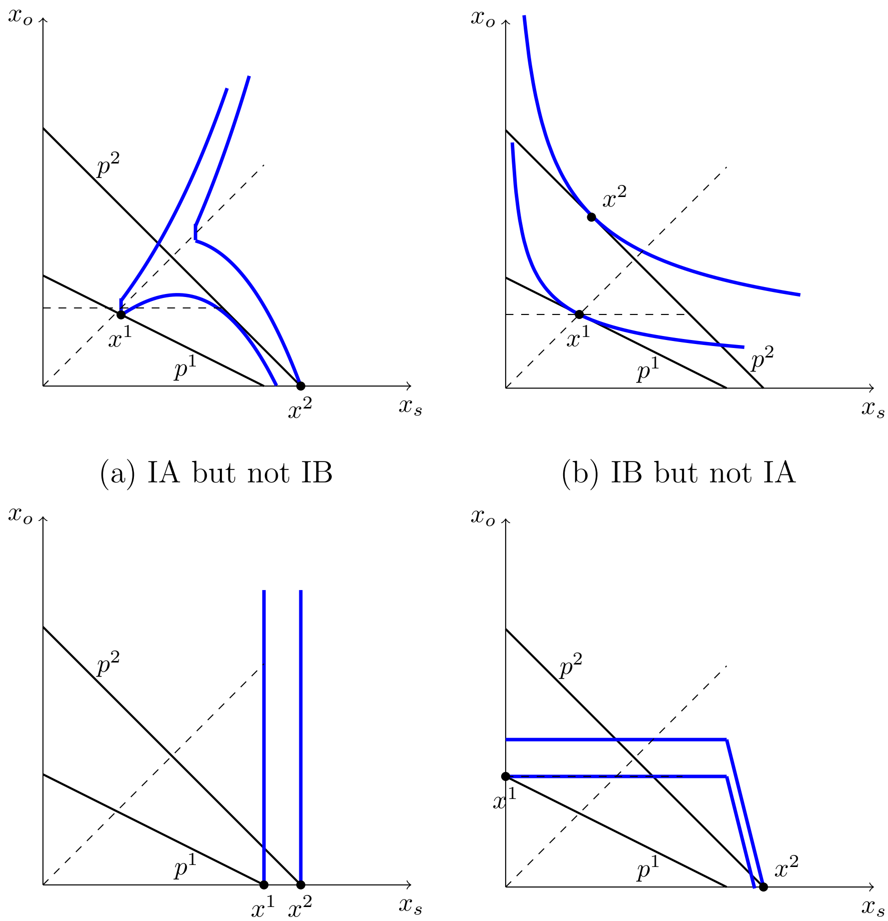
Equivalence between increasing benevolence and normality of demand in x_o allows us to employ the result from Cherchye, Demuynck and De Rock (2018) as the test for increasing benevolence.[6]

Proposition 3 (Cherchye, Demuynck and De Rock (2018))**.** *An experiment is rationalizable with increasing benevolence preferences if and only if it satisfies NARP.*

2.4 Independence of Nested Theories Further we show that nested theories (inequality aversion and increasing benevolence) are independent. Moreover, they are not exhaustive – there can be an other-regarding preference relation, neither inequality averse nor increasing benevolent. Hence, there are four cases: preferences consistent with both nested theories,

[6]If a reader is not convinced by the equivalence argument above, please see p.375 in Cherchye, Demuynck and De Rock (2018) where it is directly proven that the increasing benevolence property is satisfied.

with only one of them or with neither of them. Next we provide examples and intuition for each case.



(a) IA but not IB (b) IB but not IA

(c) IA and IB (d) OR, but neither IA nor IB

Figure 3. Independence of Inequality Aversion and Increasing Benevolence

Consider budgets from Figure 3. Inequality aversion predicts choices to be at or below the 45 degree line from the origin. Increasing benevolence requires the choice from p^2 to have more x_o than the choice from p^1. Figure 3(a) presents the case for the preferences to be consistent with inequality averse preferences, but not increasing benevolent. Indifference curves presented could be generated by the utility function $u(x_s, x_o) = x_s^3 - \max(|x_s - x_o|, 1)$. It guarantees that for p^1 the optimal point is an allocation close to equal split, while for p^2, the optimal choice is to spend all income on x_s. Therefore, this is a violation of NARP and choices are not consistent with increasing benevolence preferences. Figure 3(b) presents the case of preferences that are increasing benevolent, but not inequality averse.

Assume that player maximizes the utility function $u(x_s, x_o) = x_s x_o^2$. Then, for hight enough incomes (and low enough p_o) the choice would lie above 45 degree line ($x_s < x_o$). Hence, such preferences would not be consistent with inequality aversion. Figure 3(c) presents example of preferences consistent with both theories. Examples of such preference relations include selfish preferences ($u(x_s, x_o) = x_s$, presented on the figure) and perfect complements ($u(x_s, x_o) = \min(\alpha x_s, x_o)$ with $\alpha \geq 1$). Finally, Figure 3 presents example of other-regarding preferences not consistent with either of nested theories. Example of such preference is $u(x_s, x_o) = 4\max(x_s, 3) + x_o$. Idea behind, is that for low income, player only cares about x_o and as soon as she gets enough income, she becomes more selfish. Therefore, choice from p^1 would be above 45 degree line at the same time in the budget p^2 player would choose less of x_o then under p^1.

2.5 Revealed Social Preferences Beyond Dictator Games Extending the theory of revealed social preferences to other games is of particular importance, because different motives (that can depend on the game) can trigger different theories to perform better (see for instance Engellman and Strobel, 2004). Further we show how the revealed preference tests of social preferences can be applied for ultimatum, investment and carrot-stick games. Following Cox, Friedman and Sadiraj (2008) we consider second-movers in the two-stage games.

2.5.1 Ultimatum Game. In the first stage a proposer is given an endowment m^t and asked to allocate it between herself and a responder, given that $p_p^t x_p + p_r^t x_r = m^t$, where x_p denotes the proposer's earnings and x_r denotes the responder's earnings. In the second stage, the responder decides whether to accept or reject the allocation. If the allocation is accepted, it is implemented; otherwise, both players get zero.

The case of other-regarding preferences is already considered in Castillo, Cross and Freer (2017). Increasing benevolence is not well-defined for the binary choices. Further, we present the test for inequality aversion. The experiment, on the side of the responder, is a sequence of binary decisions between proposed allocations and zero payoffs. Figure 4 shows the decision problem, as well as acceptance (A^t) and rejection (R^t) regions in $(f(x_r, x_p), x_r)$ coordinates. To be more precise, acceptance and rejection regions can be defined as follows.

$$A^t = \{(f(x_r, x_p), x_r) : x_r \geq x_r^t \text{ and } f(x_r, x_p) \leq f(x_r^t, x_p^t)\}$$

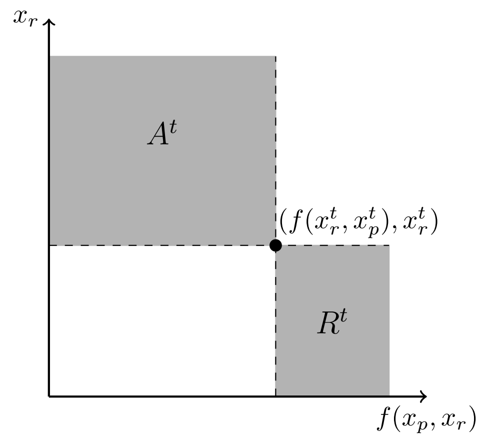


Figure 4. Acceptance and Rejection regions in Ultimatum Game

and

$$R^t = \{(f(x_r, x_p), x_r) : x_r \leq x_r^t \text{ and } f(x_r, x_p) \geq f(x_r^t, x_p^t)\}$$

If point $(f(x_r^t, x_p^t), x_r^t)$ was accepted, then every allocation from its acceptance area should be accepted as well. Otherwise, every allocation from its rejection area should be rejected. Denote the set of all accepted allocations by A_x. Denote the set of all rejected allocations by R_x.

Corollary 2. *Let $f(x_f, x_s)$ be a measure of inequality. An experiment is rationalizable with inequality averse preferences if and only if*

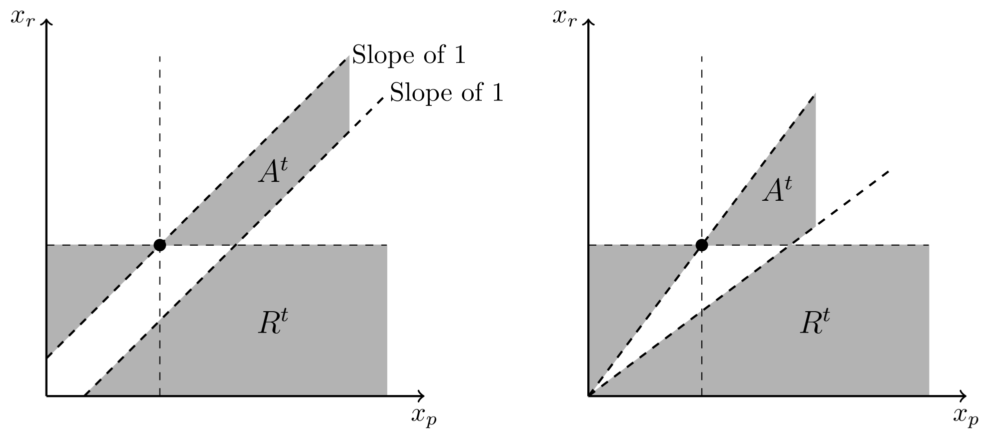
$$R_x \cap \left(\bigcup_{x^t \in A_x} A^t \right) = \emptyset$$

and

$$A_x \cap \left(\bigcup_{x^t \in R_x} R^t \right) = \emptyset$$

Acceptance and rejection regions would translate differently for different measures of inequality. Hence, performance of different measures of inequality can be compared. We consider examples of inequality aversion in differences ($f(x_r, x_p) = |x_r - x_p|$) and inequality aversion in shares ($f(x_r, x_p) = \left|\frac{x_r}{x_r + x_p} - \frac{1}{2}\right|$). We choose this measures for the matter of illustration, since they are among the most widely applied in the literature.

Figure 5(a) shows the acceptance and rejection regions for the inequality aversion in difference. Every point on the line of slope one which goes through x has the same inequality



(a) in Differences (b) in Shares

FIGURE 5. Inequality Aversion in Ultimatum Game for Different Measures of Inequality

level as x. The acceptance and rejection regions differ if different players get higher payoffs, because the direction in which inequality increases differs. Rejection regions (shaded regions below the horizontal line) give the responder lower payoff and increase inequality. Acceptance regions are represented by the stripes, since increased payoff should deliver the measure of inequality, which does not exceed the original level. Figure 5(b) shows acceptance and rejection regions for inequality aversion in shares. Every point on the line that goes through zero and x has the same inequality level as x. Comparing rejection and acceptance regions from Figures 5(a) and 5(b), we can see that they are different. Therefore, these measures of inequality have different testable implications.

2.5.2 Investment Game. Players start with an endowment of I. The first-mover sends an amount $s \in [0, I]$ to the second-mover who receives ks. Then the second-mover returns an amount of $r \in [0, ks]$, and the first-mover receives pr. The final payoffs are $x_f = I - s + pr$ and $x_s = I + ks - r$. Hence, a family of investment games with different p would generate sufficient price variation to apply revealed preference tests.

Figure 6 presents the budget set of the second-mover. Choices on the horizontal segments are not feasible. Therefore, Proposition 1 can be directly used to test consistency with other-regarding preferences. Figure 6 shows possible violation of GARP, hence, there is empirical content even for other-regarding preferences. Denote by p^t price vector, that corresponds to the linear segment of B^t. If all choices are such that $x^t \in B^v$ if and only if $p^v x^t \leq 1$ for every $t, v \in \{1, \ldots, T\}$, then Proposition 3 can be applied to test for increasing benevolence.

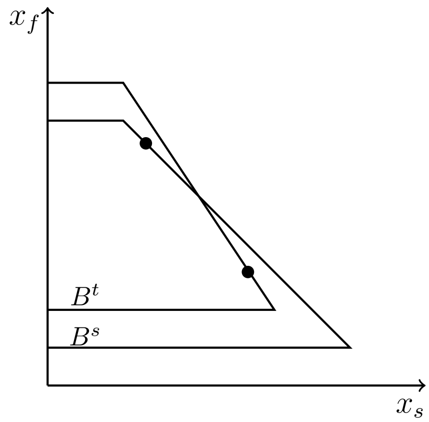
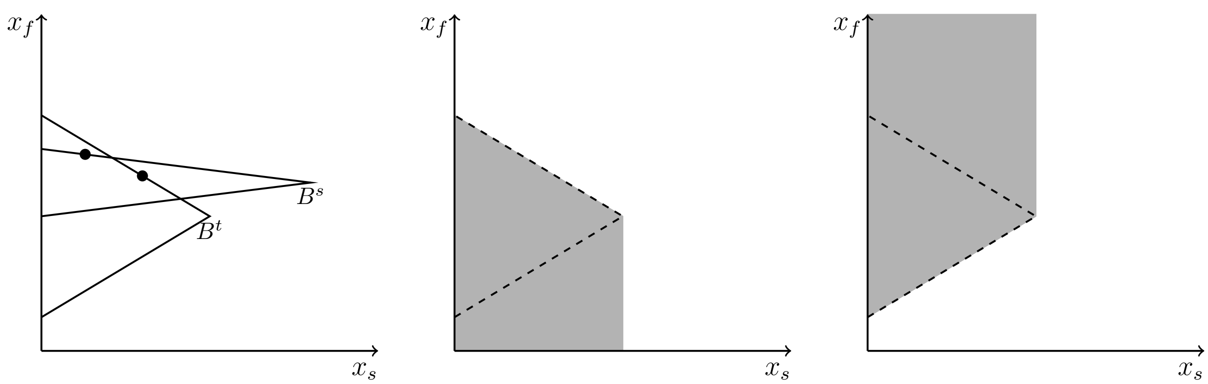


FIGURE 6. Second-mover's budget set in the investment game.

If the parameters are set up such that $x_f = x_s$ outcome is available at every budget, then Proposition 2 can be applied to test for inequality aversion. Therefore, in the investment game (as in dictator), rationalization with one measure of inequality implies rationalization with every measure of inequality.

2.5.3 Carrot-Stick Game. Players have an endowment of I. The first-mover chooses the amount to be sent $s \in [0, I]$. Then, the second-mover can "return" the amount $r \in [-s, s]$ and the first-mover receives pr. The final payoffs are $x_f = I - s + pr$ and $x_s = I + s - |r|$.



(a) B (b) $B^{\downarrow}$ (c) $B^{\uparrow}$

FIGURE 7. Second-mover's budget set in the carrot-stick game.

Figure 7(a) presents the budget sets that the second-mover faces. In this game budgets have both "upper" and "lower" borders. Therefore, it is possible to distinguish whether a player has "altruistic" or "envious" preferences. Preferences are said to be *altruistic* if utility is increasing in both x_s and x_f. Preferences are said to be *envious* if utility is increasing in x_s

and decreasing in x_f. Denote by B^t the budget (as on Figure 7) based on which the player makes a choice. Denote by $B^{\downarrow}[{\uparrow}] = \{(x'_s, x'_f) : (x_s, x_f) \in B$ such that $x_s \geq x'_s$ and $x'_f \geq (\leq) x_f\}$. Denote by $B^{\downarrow\downarrow}[{\uparrow\uparrow}] = \{(x'_s, x'_f) : (x_s, x_f) \in B$ such that $x_s \geq x'_s$ and $x'_f \geq (\leq) x_f\}$ with at least one inequality being strict. Figures 7(b) and 7(c) illustrate the construction of $B^{\downarrow}$ and $B^{\uparrow}$ respectively. The shaded areas show the part of the space added by taking the closure of the budget.

Corollary 3. *An experiment $E = (x^t, B^t)_{t=1}^T$ is rationalizable with altruistic [envious] preferences if and only if there is no sequence $x^{t_1}, \ldots, x^{t_n}$, such that $x^{t_j} \in (B^{t_{j+1}})^{\downarrow}[{\uparrow}]$ for every $j \in \{i, \ldots, n-1\}$ and $x^{t_n} \in (B^{t_1})^{\downarrow}[{\uparrow}]$, implies $x^{t_j} \notin (B^{t_{j+1}})^{\downarrow\downarrow}[{\uparrow\uparrow}]$ and $x^{t_n} \notin (B^{t_1})^{\downarrow\downarrow}[{\uparrow\uparrow}]$.*

Figure 7(a) shows the violation of GARP (equivalent to the condition in Corollary 3). Hence, altruism has empirical content. A symmetric example can be constructed to show that envious preferences also have empirical content. Moreover, if a player is altruistic, she would never use a stick, while an envious player would never use a carrot.

Denote by p^t price vector, that corresponds to the "upper boundary" of B^t. If all choices are such that $x^t \in B^v$ if and only if $p^v x^t \leq 1$ for every $t, v \in \{1, \ldots, T\}$, then Proposition 3 can be applied to test for increasing benevolence. Recall that increasing benevolence is nested within altruistic preferences. Therefore, the stick is not consistent with increasing benevolent preferences.

Further we consider inequality-averse preferences. Similarly to the Corollary 2 we need to map budget to the $(f(x_s, x_f), x_f))$ space. Denote the interior of B by $B^{\downarrow} = \{(f(x'_s, x'_f), x'_f)) : (x_s, x_o) \in B$ such that $x'_f \geq x_f$ and $f(x'_s, x'_f) \leq f(x_s, x_o)\}$ and the strict interior of B by $B^{\downarrow\downarrow}$.

Corollary 4. *Let $f(x_s, x_f)$ be an inequality measure. An experiment $E = (x^t, B^t)_{t=1}^T$ is rationalizable with inequality averse preferences if and only if there is no sequence $x^{t_1}, \ldots, x^{t_n}$, such that $x^{t_j} \in (B^{t_{j+1}})^{\downarrow}$ for every $j \in \{i, \ldots, n-1\}$ and $x^{t_n} \in (B^{t_1})^{\downarrow}$, implies $x^{t_j} \notin (B^{t_{j+1}})^{\downarrow\downarrow}$ and $x^{t_n} \notin (B^{t_1})^{\downarrow\downarrow}$.*

Figure 8(a) presents the case with an inequality averse (in differences) player using the stick. The shaded area (RB) presents the set of points better than the chosen action. None of the points that dominate the chosen one are in the budget. Hence, the choice of stick can

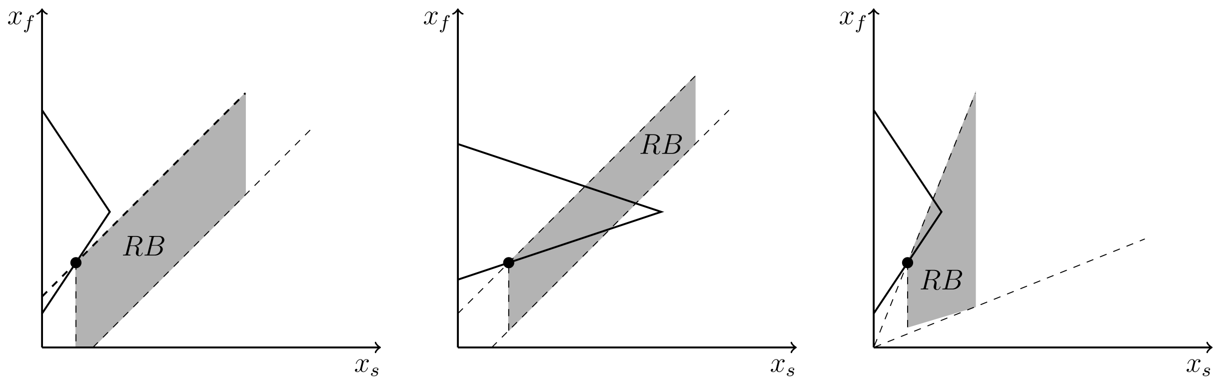


FIGURE 8. Using the stick with inequality averse preferences.

be optimal if a player has inequality averse preferences. Figure 8(b) shows that if p is low enough, then using stick is no longer rational. Figure 8(c) shows that testable implications in the carrot-stick game would depend on the particular measure of inequality, because the budget set contains the points that dominate the chosen one. Hence, the same choice is consistent with inequality aversion in differences (Figure 8(a)), but not with inequality aversion in shares (Figure 8(c)). Finally, let us note that inequality aversion is the only of the above-mentioned theories that can rationalize a player who uses both carrot and stick.

Let us conclude with a remark on possible experimental design in order to apply the tests. Design of experiment directly follows from the construction of the test. In order to guarantee the sufficient price variation one can use discretized first-mover's problem with a strategy method. Hence, the second-mover would have to make decisions over the budgets corresponding to every possible decision (out of the finite set) the first-mover can make for every budget proposer would face (see Castillo and Cross, 2008; Castillo, Cross and Freer, 2017, for the case of ultimatum game). Strategy method allows to avoid possibility of falling short on the power of test due to specific decisions first mover made.

3 Empirical Illustration

We present evidence from dictator and ultimatum games. While dictator game allows for comprehensive test of inequality aversion hypothesis (see Corollary 1), ultimatum game (theoretically) allows to distinguish between different measures of inequality.

3.1 Dictator Game We use data from two studies of dictator games. In both studies, subjects repeatedly played a dictator game with different relative prices and endowments. In every period subjects were asked to allocate tokens between themselves and another person, choosing a point on a linear budget $p_s^t x_s + p_o^t x_o \leq m^t$. The first study (Fisman, Kariv and Markovits, 2007) contains results of experiments with 76 undergraduates from UC Berkeley.[7] In this study, every subject faced 50 different budgets with randomly determined prices. The second study (Porter and Adams, 2016) contains results of experiments with 89 subjects recruited from the general population from the southeast region of the UK. In this study every subject faced 11 different budgets with predetermined prices.

When applied to data, notions of rationality prove to be very strict at least for the first dataset: no more than 16% of subjects can be rationalized with other-regarding preferences and no more than 11% with nested theories. Therefore, it makes sense to relax the notion of rationality and allow for some probability that people make mistakes. For this purpose, we use the **Houtman-Maks index (HMI)**.[8] HMI is the maximum fraction of data that can be rationalized by a given theory. That is, if in a total of T observations, the maximum subset which is consistent with the theory is τ, then $HMI = \tau/T$. For the technical details regarding the implementation of the HMI index, see Appendix B. We report results for the HMI level of .9. The results are robust to other levels of HMI (see Appendix D). The HMI level of .9 allows for deviations from rationality in no more than 10% of budgets.

Next, we want to control for false positives. A false positive is the probability that a random decision making would look consistent with the test. We use two procedures which differ in the assumption about the random behavior. First is the Bronars (1987) power, conducted by generating 1000 random subjects who make decisions uniformly distributed along the budget line. Power of the test is computed as the fraction of random subjects who fail to perform consistently with the test. The second is the bootstrap power (see e.g. Cox,

[7]Experiment contains two other treatments which we do not consider in our analysis. One of the treatments uses step-shaped budgets and another is a dictator game with two recipients.

[8]See Houtman and Maks (1985); Heufer and Hjertstrand (2015); Dean and Martin (2009). We use the HMI because it is the only index that can be applied to test Inequality Aversion. Critical Cost Efficiency Index introduced by Afriat (1973) would not adequately work in the context of inequality aversion. The money pump index introduced by Echenique, Lee and Shum (2011) is defined for GARP only. The swaps index proposed by Apesteguia and Ballester (2015) can be applied only in the context of finite choice sets.

1997; Harbaugh, Krause and Berry, 2001; Andreoni and Miller, 2002). This measure controls for possible behavioral rules that can cause false positive results even if people would take decisions at random. To compute the bootstrap power of the test, we calculate the empirical distribution of the shares of income spent on each commodity – in our case the subject's payoff and the other's payoff – and simulate the pseudo subjects who make their choices at random but distributed according to the empirical distribution function.

Last, to compare pass rates controlling for the power we use the **predictive success index (PSI)** introduced by Selten (1991).[9] The predictive success index is defined as the difference between the share of people that satisfies an axiom at the given level of HMI and the probability that random choices will satisfy the axiom at the same level of HMI. This index ranges between -1 and 1, with -1 meaning no subject passes while all random subjects pass and 1 meaning every subject passes while none of the random subjects do. If PSI is greater than zero, then theory describes the behavior better than random choice, and if PSI is less or equal to zero, then the random choice explains the observed behavior better.

Table 1 shows results of testing other-regarding preferences for both datasets. The second column presents the pass rates (share of subjects who pass the test with HMI at least .9). The third and fourth columns present the power computations according to Bronars and the bootstrap methods.[10] Last two columns present the predictive success index using Bronars and bootstrap powers.

Subjects (in both experiments) are consistent with having other-regarding preferences. In particular 76% of subjects in the first dataset and 91% in the second one are consistent with having other-regarding preferences. Results are robust to controlling for the power of test.

Inequality aversion and increasing benevolence are nested within the other-regarding preferences model. That is, a subject can only be consistent with having inequality averse or

[9]Methodology of using predictive success index in the revealed preference context was introduced by Beatty and Crawford (2011). Statistical interpretation of the index which allows us to construct confidence intervals was proposed by Demuynck (2015).

[10]Power is different for different data sets first of all, because of the different amount of budgets: 50 vs 11. Difference in power is even larger for increasing benevolence. Experiment of Porter and Adams (2016) is aimed at testing GARP, which rather requires price variations while NARP requires rather income variation. Income variation is higher in Fisman, Kariv and Markovits (2007) data set because of the larger amount of budgets.

Theory	Pass Rate	Power of Test Bronars	Power of Test Bootstrap	PSI Bronars	PSI Bootstrap
Fisman, Kariv and Markovits (2007) data					
Other-regarding	58 (76.32%)	100.00%	99.63%	0.76	0.76
95% conf. interval	(65.18% - 85.32%)	(99.99% - 100.00%)	(99.58% - 99.67%)	(0.67 - 0.86)	(0.66 - 0.86)
Porter and Adams (2016) data					
Other-regarding	81 (91.01%)	68.72%	80.01%	0.60	0.71
95% conf. interval	(83.05% - 96.04%)	(68.41% - 69.02%)	(79.75% - 80.28%)	(0.54 - 0.66)	(0.65 - 0.77)

Table 1. Results for Other-Regarding Preferences

increasing benevolence preferences, if she is consistent with having other-regarding preferences. Hence, we report a nested theory analysis; that is, results are presented for the subexperiment, which consists only of subjects who are consistent with other-regarding preferences hypothesis (given HMI=.9).

Theory	Pass Rate	Power of Test Bronars	Power of Test Bootstrap	PSI Bronars	PSI Bootstrap
Fisman, Kariv and Markovits (2007) data					
Inequality Aversion	32 (55.17%)	100.00%	92.58%	0.55	0.48
95% conf. interval	(41.54% - 68.26%)	(100.00% - 100.00%)	(92.39% - 92.76%)	(0.42 - 0.68)	(0.35 - 0.61)
Increasing Benevolence	12 (20.69%)	100.00%	100.00%	0.21	0.21
95% conf. interval	(11.17% - 33.35%)	(100.00% - 100.00%)	(100.00% - 100.00%)	(0.10 - 0.31)	(0.10 - 0.31)
Porter and Adams (2016) data					
Inequality Aversion	51 (62.96%)	98.62%	90.50%	0.62	0.53
95% conf. interval	(51.51% - 73.44%)	(98.55% - 98.70%)	(90.31% - 90.69%)	(0.51 - 0.72)	(0.43 - 0.64)
Increasing Benevolence	69 (85.19%)	85.68%	65.67%	0.71	0.51
95% conf. interval	(75.55% - 92.10%)	(85.44% - 85.90%)	(65.35% - 65.98%)	(0.63 - 0.79)	(0.43 - 0.59)

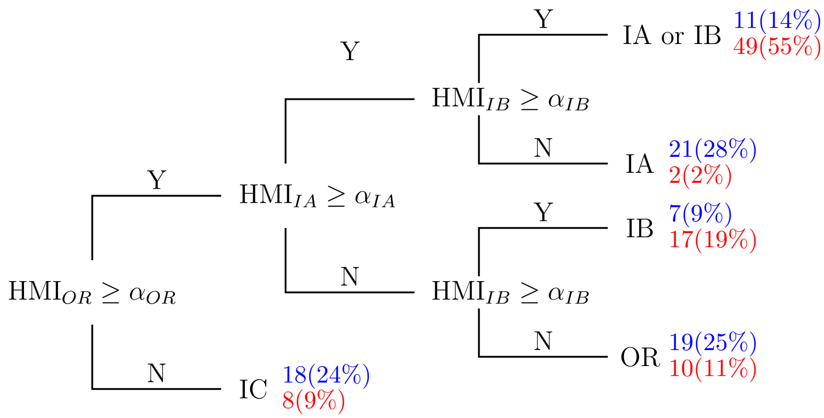
Table 2. Results for Nested Theories

Table 2 presents results for nested theory analysis. Structure of the table is the same to the one of Table 1. Both theories are significantly restrictive – there is significant share of population (15-79%) which is consistent with having other-regarding preferences, but not with increasing benevolence or inequality averse preferences. In addition, while 55% of subjects are consistent with inequality averse preferences in Fisman, Kariv and Markovits (2007) data, only 21% of them is consistent with having increasing benevolence preferences. However, the results are opposite for the Porter and Adams (2016) data: 63% of subjects are consistent with having inequality averse preferences and 85% of them are consistent with having increasing benevolence preferences. To sum up, inequality aversion appears to describe data better in the first dataset; and increasing benevolence performs at least as well as inequality aversion in the second one (difference in PSIs is not statistically significant). This inconsistency between the two datasets provides evidence in line with Fehr, Naef and Schmidt (2006), who showed that social preferences may depend on the demographic characteristics of the population.

3.2 Mixed Types Analysis None of the nested theories can explain the behavior of the entire sample. At the same time, both theories perform well even conditioning on their power. In addition, different theories have quite different empirical implications, and the correlations between pass rates for inequality aversion and increasing benevolence are quite low: .22 (with a confidence interval of $[-.01, .42]$) for Fisman, Kariv and Markovits (2007) data and .51 (with a confidence interval of $[.34, .65]$) for Porter and Adams (2016) data.[11] This shows that there is a non-trivial probability that different subjects can be consistent with different notions of rationality. Therefore, we perform a mixed type analysis. Main goal of this exercise is to find out which theory is the most appropriate to describe behavior at the subject level, while the previous analysis focused at the sample level.

Subjects are assigned to theories according to three sequential binary classification steps presented in Figure 9. First, if a subject is not consistent with other-regarding preferences at threshold α_{OR}, she is classified as inconsistent with other-regarding preferences (IC). Next, we compare whether she is consistent with inequality aversion or increasing benevolence with thresholds α_{IA} and α_{IB} respectively. If the subject is not consistent with either, she is classified as other-regarding (OR). If the subject is consistent with both, she is assigned

[11]Given that the tests are binary, appropriate statistic is ϕ-coefficient. It is a version of correlation coefficient for two binary variables. Both logit and probit regression coefficients are insignificant for Fisman, Kariv and Markovits (2007) data: logit regression coefficient is 1.2 (with 95% confidence interval of $[-0.06, 2.61]$) and probit regression coefficient is 0.75 (with 95% confidence interval of $[-0.04, 1.57]$). That is we can not reject that two nested theories are unrelated for Fisman, Kariv and Markovits (2007) data. For Porter and Adams (2016) data the relationship is much stronger: logit coefficient is 3.09 (with 95% confidence interval of $[1.74, 4.99]$), probit coefficient is 1.84 (with 95% confidence interval of $[1.07, 2.72]$). That is, odds of the subject being consistent with inequality aversion are at least $e^{1.74} = 5.7$ times higher if she is also consistent with increasing benevolence than if she is not.



Top numbers are for Fisman, Kariv and Markovits (2007) data,
bottom numbers are for Porter and Adams (2016) data.

FIGURE 9. Classification Tree

to a separate class of inequality averse or increasing benevolent (IA or IB). If the subject is consistent with only one theory, she is classified as inequality averse (IA) or increasing benevolent (IB).

It is still necessary to determine the thresholds for the classification tree. In order to do this, we modify the unsupervised machine learning methodology from Liu, Xia and Yu (2000). The approach maximizes the information gain from adding a particular cluster. We base this measure on HMI, but the approach is general (for more detailed explanation see Appendix C). The thresholds obtained are as follows: $\alpha_{OR} = 45/50$; $\alpha_{IA} = \alpha_{IB} = 41/50$ for Fisman, Kariv and Markovits (2007) data and $\alpha_{OR} = \alpha_{IA} = \alpha_{IB} = 10/11$ for Porter and Adams (2016) data.

In the first dataset, a large set of subjects can be described by inequality aversion but not by increasing benevolence preferences (28% against 9%). In the second experiment, the distinction is less clear, as the majority of subject (55%) can be described by IA or IB.[12] As a final remark, 11-25% of the population cannot be explained by either inequality aversion

[12]However, as a caveat here, note that power of test for increasing benevolence is significantly lower than the one for inequality aversion. Moreover, in Porter and Adams (2016) experiment, three participants gave more to their counterparts than they kept for themselves in all 11 budgets. Such altruistic behavior is maximally inconsistent with inequality aversion, implying an HMI of zero (theoretically HMI for other theories starts with 1). We exclude these three subjects from the mixed type analysis without significantly affecting the results (see detailed results in Appendix C).

or increasing benevolence, but still has other-regarding preferences. This fact also provides additional evidence for both assumptions being significantly restrictive. The difference between the two datasets in terms of classification for nested theories is consistent with the results in the previous subsection.

Several previous papers (see e.g. Andreoni and Miller, 2002; Porter and Adams, 2016) attempted to classify people into distinct types according to their giving behavior. However, in these papers particular types of preferences were characterized with a parametric specification of the utility function, while the method we offer, is completely non-parametric. Hence, the method proposed is more flexible and allows for more robust classification of subjects, which can further be used for the out-of-sample predictions.

3.3 Ultimatum Game We use data from Castillo, Cross and Freer (2017) who conduct an experiment with total of 123 participants (students from Georgetown and Texas A&M universities). Every subject had to make the accept/reject decision over 13 alternatives drawn from 9 different linear budgets, that adds up to 117 binary choices from every subject. See Castillo, Cross and Freer (2017) for the more detailed description of the design and the data as well as the evidence that subjects are consistent with other-regarding preferences.[13] Next, we present the results on testing inequality aversion in differences and inequality aversion in shares.

Table 3 presents the results of the analysis. About 40-45% of the sample are consistent with either versions of inequality aversion without allowing for any decision making error. If we allow for the decision making error of 5% of choice sets (HMI$\geq$ .95), then we obtain that 87% of subjects are consistent with inequality aversion in differences and 82% of subjects are consistent with inequality aversion of shares. In addition to Bronars and bootstrap power we also conduct the cutoff power analysis. Idea behind is that responder uses cutoff strategies – everything below cutoff is rejected and everything above cutoff is accepted. This cutoff is determined randomly according to uniform distribution (Bronars cutoff) and empirical distribution functions (Bootstrap cutoff). Finally, the power analysis shows that no more than 5% of random subjects are consistent with either theories. Hence, both theories explain

[13]Castillo, Cross and Freer (2017) conducted two sets of sessions, we report the results from them together. Let us note that choices are quite different between the populations, while the consistency with other-regarding preferences is quite uniform for both samples.

	IA in Differences		IA in Shares	
HMI	Num. subj.	% subj.	Num. subj.	% subj.
1.00	55	45%	52	42%
.99	6	5%	8	7%
.98	19	16%	12	10%
.97	10	8%	11	9%
.96	9	5%	12	10%
.95	8	7%	6	5%
$< .95$	16	13%	22	18%

Power Analysis

Test (HMI$\geq$ .95)	Power (IA in Differences)	Power (IA in Shares)
Bronars	1.0000	1.0000
Bronars Cutoff	1.0000	1.0000
Bootstrap	0.9985	0.9990
Bootstrap Cutoff	0.9655	0.9740

TABLE 3. HMI for inequality aversion in differences and shares

data better than the random decision making. Further analysis is aimed on distinguishing which of inequality measures explains the data better.

	Consistent with IA in Differences	Inconsistent with IA in Differences
Consistent with IA in Shares	100 (81 %)	1 (1 %)
Inconsistent with IA in Shares	7 (6 %)	15 (12 %)

TABLE 4. Cross Table (HMI $\geq$ .95)

Table 4 that shows the amounts of subjects who are consistent or not consistent with IA in differences and IA in shares. Most of the subjects (81%) are consistent with both theories, while 12% are consistent with neither theories. Rest 7% of subjects are consistent with only one of the theories. Hence, subjects are either consistent with both theories or with neither.

Therefore, it is necessary to distinguish whether it is the result of subjects' behavior or the design of the tasks.

	Random	Random Cutoff	Bootstrap	Bootstrap Cutoff
Consistent with IA in differences given IA in shares (HMI$\geq$ .95)	0.3835	0.5565	0.4840	0.5125
Consistent with IA in shares given IA in differences (HMI$\geq$ .95)	0.0475	0.0490	0.1825	0.2025

Table 5. Conditional Power

Table 5 presents conditional power analysis for both theories. Power analysis shows that behavioral patterns of subjects make the theories barely distinguishable. Bootstrap cutoff test shows that if a (randomly generated) subject is consistent with IA in differences then she would be consistent with IA in shares with 51% and if a (randomly generated) subject consistent with IA in shares then she would be with 20% consistent with IA in differences. Based on this we can construct an analogue of the PSI, that would allow us to distinguish between the IA in differences and in shares controlling for the power of test. Conditional power would allows us to control for a subjects that are consistent with IA in shares, and just by chance appear to be consistent with the IA in differences. We normalize the amount of subjects who are consistent with both versions of IA by the conditional power. Using the bootstrap cutoff rule we obtain the PSI of $.81(1-.51)+.06=.46$ value for IA in differences and the PSI of $.81(1-.20)+.01=.66$ for IA in shares. That is the figures are larger for IA in shares (despite the lower pass rate), and this holds for all types of random control. However, better performance of IA in shares over IA in differences is rather a weak evidence, since it is mainly driven by the difference in the power of the tests.

4 Revealing Altruism and Fairness

Further we present the revealed preference identification, which allows for comparison of the level of altruism among players and partial identification of the (subjective) notion of fair outcome player may have. If subjects are consistent with increasing benevolence preferences, then we can apply the methodology from Cox, Friedman and Sadiraj (2008) in order to reveal the relative level of altruism of subjects. An alternative is to assume the theory a bit more general than inequality aversion. Inequality aversion assumes that player considers equal split as fair outcome. This issue can be generalized, by allowing players to have subjective notion of fair outcome, which can deviate from equal split. Then, we can generalize conditions from Proposition 2 to identify the fair outcome.

4.1 Revealed Relative Altruism Cox, Friedman and Sadiraj (2008) provide the revealed preference-based approach to classify people in terms of altruism. A demand function $D(p_s, p_o)$ is **more altruistic than** $\tilde{D}(p_s, p_o)$ if $D_o(p_s, p_o) \geq \tilde{D}_o(p_s, p_o)$ for every $p \in \mathbb{R}_{++}$. Original conditions required to observe the whole demand function, while we show that the same method can be applied even if only finite set of choices is observed.

Corollary 5. *Consider consumption experiments E and $\tilde{E}$ in which players faced the same prices. Moreover, assume that both experiments satisfy NARP. Experiments are rationalizable with increasing benevolence preferences such that $D(p_s, p_o)$ is more altruistic than $\tilde{D}(p_s, p_o)$ if and only if $\tilde{x}_o^t \geq x_o^s$ for all $p^s = \tilde{p}^t$.*

The necessity of this condition is quite obvious since we consider experiments with similar prices. Hence, we can reconstruct demand functions such that the first experiment will be more altruistic than the second one.

4.2 Revealed Fair Outcome Further we show how to identify the fair outcome using the revealed preference conditions. For this purpose we need formally define what is the fair outcome. Assume that $\chi^* < 1$ is the *fair ratio of payoffs*, that is, the outcome is fair if and only if $x_s/x_o = \chi^*$. We assume that $\chi^* < 1$, as otherwise every player who is consistent with GARP can be rationalized as inequality averse.

Moreover, we slightly modify the definition of the inequality measure. We assume that $f(x_s, x_o) = 0$ if and only if $x_s/x_o = \chi^*$. Moreover, $f(x_s, x_o)$ is increasing in x_o and decreasing in x_s if $x_s/x_o < \chi^*$; and $f(x_s, x_o)$ is decreasing in x_o and increasing in x_s if $x_s/x_o > \chi^*$. The last property of an inequality measure needs to be restated as follows: if $x_s/x_o = \chi < \chi^*$, then there is $x'_s \leq x_s$, such that $f(x_s, x_o) \geq f(x_o, x'_s)$.

Corollary 6. *Let χ^* determine the notion of fair outcome. An experiment is rationalizable with inequality averse preferences if and only if an experiment satisfies GARP and $x_s^t \geq \chi^* x_o^t$ for every $t \in \{1, \ldots, T\}$.*

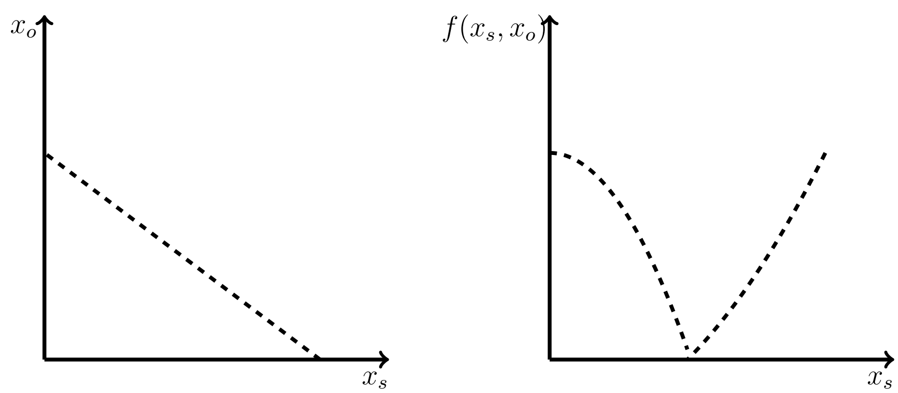
5 Concluding Remarks

The paper constructs the revealed preference tests for increasing benevolence and inequality averse preferences. Moreover, the tests can be applied (beyond the standard revealed preference framework with linear budgets) to other games used to study social preferences. Although all conditions provided in the paper are deterministic, the constraints obtained can be used as the moment inequalities. This would allow one to construct stochastic bounds that take into account decision making and/or measurement errors.[14]

Appendix A: Proofs

Proof of Proposition 2 The reasoning for the further proof does not depend on whether we assume budgets to be linear or not. Idea behind is to show that under any measure of rationality the boundary of the budget set in the allocation space would translate to the boundary of the budget set in the $(x_s, f(x_s, x_o))$ space; and the same for the interior. Therefore, we use linear budgets for the sake of the simplicity of the exposition.

Before we begin the proof, let us introduce some additional notation. Let $IA = (Y, \geq_{IA})$ be a *partially ordered space*, where $Y \subseteq \mathbb{R}^2_+$, with $y = (x_s, f(x_s, x_o))$ for every $y \in Y$ and $y \geq_{IA} y'$ if $x_s \geq x'_s$ and $f(x_s, x_o) \leq f(x'_s, x'_o)$. Denote by $>_{IA}$ the strict part of $\geq_{IA}$. Note that $f(x_s, x_o)$ defines the injective mapping from X to Y.



(a) Budget line in decision space (b) "Budget line" in Y_{IA} space

Figure A.1. Mapping of Budget from $(X, \geq)$ to $(Y, \geq_{IA})$.

[14]See Chernozhukov, Hong and Tamer (2007) for general results on partial identification and Aguiar and Kashaev (2017) for particular applications for revealed preferences with measurement error. Moreover, methodology from the latter paper directly applies to the results we state.

First, let us translate the budget lines to the IA space. Mapped budgets are not necessarily linear. It can be easily seen that the budget set in the new space is compact subset (in natural topology) of IA, since we require the measure of inequality to be a continuous function. Figure A.1 illustrates the mapping of the budget line from X to Y.

Denote by B the budget line in IA obtained from mapping a linear budget $px = 1$. Denote by $B^{\downarrow} = \{y : \text{ there is } y' \in B \text{ such that } y \geq_{IA} y'\}$ the **downward closure** of budget B. Denote by $B^{\downarrow\downarrow} = \{y : \text{ there is } y' \in B \text{ such that } y' >_{IA} y$ the **interior** of budget set B. Denote by $\partial B = B^{\downarrow} \setminus B^{\downarrow\downarrow}$ the **boundary** of $B^{\downarrow}$. Then, an experiment is rationalizable with inequality averse preferences if and only if there is a continuous and monotone (with respect to $\geq_{IA}$) utility function $u(x_s, f(x_s, x_o))$, such that observed choices $x^t \in \underset{x \in (B^t)^{\downarrow} \cap \mathbb{R}^2_+}{\text{argmax}} \{u(x_s, f(x_s, x_o))\}$.

Definition A.1. *A consumption experiment $E = (x^t, B^t)_{t=1}^T$ satisfies* **generalized cyclical consistency (GCC)** *if and only if there is no sequence $x^{t_1}, \ldots, x^{t_n}$, such that $x^{t_j} \in (B^{t_{j+1}})^{\downarrow}$ for every $j \in \{i, \ldots, n-1\}$ and $x^{t_n} \in (B^{t_1})^{\downarrow}$, implies $x^{t_j} \in \partial B^{t_{j+1}}$ and $x^{t_n} \in \partial B^{t_1}$.*

The first lemma can be derived from the results of Nishimura, Ok and Quah (2017).[15]

Lemma A.1. *An experiment satisfies generalized cyclical consistency if and only if there is a continuous, monotone utility function that rationalizes it.*

Figure A.2 illustrates the downward closure of the budget in $(Y, \geq_{IA})$ generated by a linear budget in X. It shows that every point for which $x_s < x_o$ lies in the interior of the budget B.[16] Therefore, such a choice could not be generated by inequality averse utility. Hence, further proof proceeds by showing that if we consider $x_s \geq x_o$, then there is a violation of GARP (in X) if and only if there is a violation of generalized cyclical consistency (in IA). Recall that not all points from $B^{\downarrow}$ can be chosen.

[15]One can easily check that conditions on the space are satisfied. That is $\geq_{IA}$ is a continuous order, since we consider a natural topology of $\geq_{IA}$ and order is always continuous in its natural topology. Finally, IA space is Hausdorff and locally compact. Moreover, the mapped budget is compact since it is a result of continuous mapping of compact set.

[16]Otherwise we can just permute x_s and x_o. Then, the definition of inequality measure would imply that $f(x_o, x_s) \leq f(x_s, x_o)$. Hence, the permuted point is in the interior of the budget.

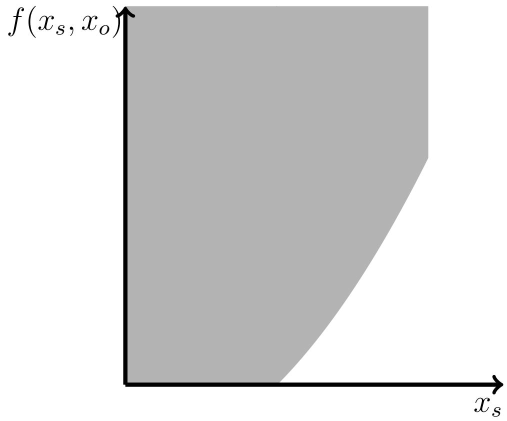


FIGURE A.2. Downward closure of the budget in $(Y, \geq_{IA})$.

To prove that on a half-space such that $x_s \geq x_o$, GARP is equivalent to GCC we show that points are in the linear budget if and only if they are in the $B^{\downarrow}$ and boundary of $B^{\downarrow}$ contains only points such that $px = 1$. We start by characterizing the relation between linear budgets and the budget in IA space.

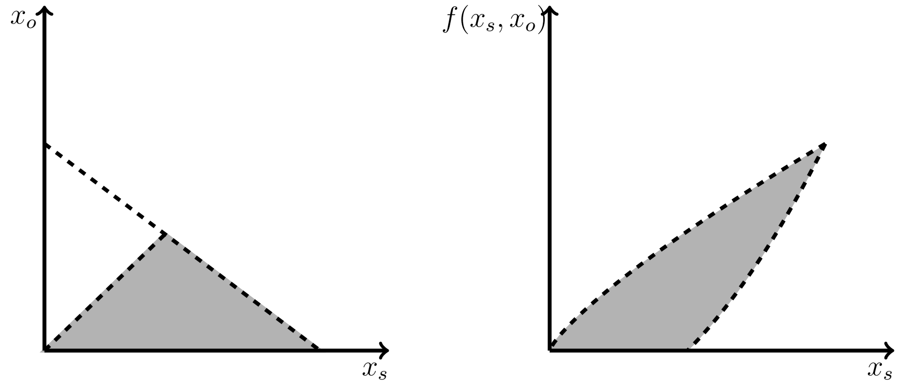
Lemma A.2. *If $x_s \geq x_o$ and $(x'_s, f(x'_s, x'_o)) >_{IA} (\geq_{IA})(x_s, f(x_s, x_o))$, then $(x'_s, x'_o) > (\geq)(x_s, x_o)$.*

Proof. We only prove the case for strict inequalities. Weak inequalities can be proven in a similar fashion. Consider the following cases.

Case 1: $x'_s \geq x'_o$. Then $(x'_s, f(x'_s, x'_o)) >_{IA} (x_s, f(x_s, x_o))$ implies that $x'_s \geq x_s$ and $f(x'_s, x'_o) \leq f(x_s, x_o)$ with at least one inequality being strict. Recall that f is increasing in x_s and decreasing in x_o, hence $x'_o > x_o$. This implies that $(x'_s, x'_o) > (x_s, x_o)$.

Case 2: $x'_s < x'_o$. Then $(x'_s, f(x'_s, x'_o)) >_{IA} (x_s, f(x_s, x_o))$ implies that $x'_s \geq x_s$ and $f(x'_s, x'_o) \leq f(x_s, x_o)$ with at least one inequality being strict. At the same time $x'_o > x'_s \geq x'_s \geq x'_o$, hence $(x'_s, x'_o) > (x_s, x_o)$. □

For the simplicity of the further argument, we allow $(x_s, x_o) \in \mathbb{R}^2$; that is, we relax the assumption of non-negativity of the points. This allows us to establish a one-to-one relation between allocations in linear budgets and those in $B^{\downarrow}$. If this holds for all the space $\mathbb{R}^2$, the same would hold for the corresponding subspace.



(a) Budget line in decision space (b) "Budget line" in Y_{IA} space

FIGURE A.3. Mapping of budgets in half-space $x_s \geq x_o$

Further, we restrict our attention to the half-space of $x_s \geq x_o$. Figure A.3 illustrates the mapping of the linear budget to the $B^{\downarrow}$. The south east boundary of the budget from (b) corresponds to $px = 1$ from (a) and the north west boundary from (b) corresponds to the bisector from (a). Moreover, the shaded area (interior) from (a) corresponds to the shaded area (interior) from (b). We show this in formal terms.

Lemma A.3. *Let $x_s \geq x_o$, then $(x_s, f(x_s, x_o)) \in B^{\downarrow}$ if and only if $px \leq 1$.*

Proof. **(⇒)** Take $(x_s, f(x_s, x_o)) \in B^{\downarrow}$. By construction of $B^{\downarrow}$ there is $x' = (x'_s, x'_o)$ such that $px' = 1$ and $(x'_s, f(x'_s, x'_o) \geq_{IA} (x_s, f(x_s, x_o))$, then $x' \geq x$ (see Lemma A.2). This implies that $px \leq px' = 1$, which is a contradiction.

(⇐) Take $x = (x_s, x_o)$ such that $px \leq 1$. If $x_s > x_o$, then there is $x'_o \geq x_o$ and $x'_s = x_s$, such that $px' = 1$.[17] This implies that $f(x'_s, x'_o) \leq f(x_s, x_o)$. Therefore, $(x_s, f(x_s, x_o)) \leq_{IA}$ $(x'_s, f(x'_s, x'_o)$, and $px' = 1$, hence, $(x_s, f(x_s, x_o)) \in B^{\downarrow}$. If $x_s = x_o$, let $\varepsilon = \frac{1-px}{p_s+p_o} \geq 0$ and $(x'_s, x'_o) = (x_s + \varepsilon, x_o + \varepsilon)$. Then, $(x_s, f(x_s, x_o)) \leq_{IA} (x'_s, f(x'_s, x'_o)$ (since both of them equal to zero), and $px' = 1$, hence, $(x_s, f(x_s, x_o)) \in B^{\downarrow}$. □

Further we show that only points from the boundary of the linear budget can be on the boundary of $B^{\downarrow}$and vice versa.

Lemma A.4. $(x_s, f(x_s, x_o)) \in \partial B$ *if and only if* $x_s \geq x_o$ *and* $px = 1$.

[17]This is due to the fact that $f(x_s, x_o)$ is decreasing in x_o for $x_s \geq x_o$. If for that case $x'_o > x_s$, then set $x'_o = x_s$ and apply construction used for the case of $x_s = x_o$.

Proof. ($\Rightarrow$) If $(x_s, f(x_s, x_o)) \in \partial B$ then $x_s \geq x_o$ is trivially implied. Then, we can just assign $x'_s = x_o$ and $x'_o = x_s$ and obtain a contradiction since $(x_s, f(x_s, x_o)) <_{IA} (x'_s, f(x'_s, x'_o))$. Therefore, we are left to show that $px = 1$. On the contrary, assume that $px < 1$. Then, we can apply the construction as in the proof of Lemma A.3 to get a bundle x' such that $px' = 1$ and $(x_s, f(x_s, x_o)) <_{IA} (x'_s, f(x'_s, x'_o))$. This implies a contradiction

($\Leftarrow$) On the contrary, assume that $(x_s, f(x_s, x_o)) \in B^{\downarrow\downarrow}$. Then, there is $x' = (x'_s, x'_o)$ on the boundary of $B^{\downarrow}$ such that $(x_s, f(x_s, x_o)) <_{IA} (x'_s, f(x'_s, x'_o))$. Hence, Lemma A.2 implies that $1 = px < px'$. This immediately implies a contradiction. □

These two results imply the following corollary.

Corollary A.1. *Let $x_s \geq x_o$, then $(x_s, f(x_s, x_o)) \in B^{\downarrow\downarrow}$ if and only if $px < 1$.*

The proof is omitted, as it is similar to the previous ones. To prove that every point from the interior of $B^{\downarrow}$ is strictly inside of a linear budget we assume the contrary and obtain a contradiction from Lemma A.4. To prove the reverse, we assume that point is in the border and again obtain a contradiction using Lemma A.4.

These preliminary results allow us to finalize the proof.

Corollary A.2. *Let $x^t_s \geq x^t_o$ for all $t \in \{1, \ldots, T\}$. An experiment satisfies GARP if and only if it satisfies generalized cyclical consistency.*

Proof. ($\Rightarrow$) On the contrary, assume that there is a violation of generalized cyclical consistency. Then, there is a sequence $x^{t_1}, \ldots, x^{t_n}$, such that $x^{t_j} \in (B^{t_{j+1}})^{\downarrow}$ for every $j \in \{i, \ldots, n-1\}$ and $x^{t_n} \in (B^{t_1})^{\downarrow\downarrow}$. Lemma A.3 implies that $p^{t_{j+1}} x^{t_j} \leq 1$ and Corollary A.1 implies that $p^{t_1} x^{t_n} < 1$ that is a violation of GARP.

($\Leftarrow$) On the contrary, assume that there is a violation of GARP. Then, there is a sequence $x^{t_1}, \ldots, x^{t_n}$, such that $p^{t_{j+1}} x^{t_j} \leq 1$ for every $j \in \{i, \ldots, n-1\}$ and $p^{t_1} x^{t_n} < 1$. Lemma A.3 implies that $x^{t_n} \in (B^{t_1})^{\downarrow\downarrow}$ Corollary A.1 implies that $x^{t_n} \in (B^{t_1})^{\downarrow\downarrow}$. Hence, there is a violation of generalized cyclical consistency. □

Lemma A.2 together with the observation that choices such as $x_s < x_o$ cannot be rationalized with inequality averse preferences would guarantee that GARP and $x^t_s \geq x^s_o$ for every

$t, s \in \{1, \ldots, T\}$ are necessary and sufficient conditions for rationalization with inequality averse preferences. As Lemma A.1 shows, the generalized cyclical consistency is equivalent to rationalizability with inequality averse preferences.

Proof of Corollary 2 The budget in this case is $B^t = \{(x_s^t, f(x_f^t, x_s^t)), (0, 0)\}$. The comprehensive closure of the budget is $(B^t)^{\downarrow} = B^t \cup R^t$. The interior of the comprehensive closure of the budget is $(B^t)^{\downarrow\downarrow} = R^t$. Moreover, one can easily see that $x^s \in A^t$ if and only if $x^t \in R^t$. Hence, the proof can be concluded by applying Lemma A.1.

Proof of Corollaries 3 and 4 Proofs of both Corollaries directly follow from Lemma A.1 with the corresponding definition of orders.

Proof of Corollary 5 Let us introduce some additional notation. We rely on the result from Cherchye, Demuynck and De Rock (2018). Let $\omega = \frac{p_o}{p_s}$ and $m = \frac{1}{p_s}$. Therefore, we can refer to the demands as functions of only ω and m. Moreover, we reenumerate observations s.t. $\tilde{p}^t = p^t$.

Define $\alpha, \beta > 0$ such that

$$1 + \beta < \min \left\{ \min_{t,s} \left\{ \frac{x_o^t}{x_o^s} : x_o^t > x_o^s \right\}, \min_{t,s} \left\{ \frac{\tilde{x}_o^t}{\tilde{x}_o^s} : \tilde{x}_o^t > \tilde{x}_o^s \right\} \right\}$$

and

$$a(1 + \beta) < \min \left\{ \min_{t,s} \left\{ \frac{x_o^t}{x_o^s} \right\}, \min_{t,s} \left\{ \frac{\tilde{x}_o^t}{\tilde{x}_o^s} \right\} \right\}$$

Let $\delta_{t,v} = \max\{|w^t - w^v|, |m^t - m^v + (w^t - w^v)x_o^t|, |m^t - m^v + (w^t - w^v)\tilde{x}_o^t|\}$ and let $\varepsilon < \min \delta_{t,v}$. Consider the function

$$g(z) = \begin{cases} \alpha & \text{for } z \leq -\varepsilon \\ 1 + \frac{1-\alpha}{\varepsilon} z & \text{for } -\varepsilon \leq z \leq 0 \\ 1 & \text{for } z \geq 0 \end{cases}$$

In addition, consider the function

$$h(z) = \begin{cases} \alpha \frac{1}{z + \varepsilon - 1} & \text{for } z < -\varepsilon \\ 1 + \frac{1-\alpha}{\varepsilon} z & \text{for } -\varepsilon \leq z \leq 0 \\ 1 + \beta \frac{z}{z+1} & \text{for } z \geq 0 \end{cases}$$

Then, according to Cherchye, Demuynck and De Rock (2018), the rationalization of the demand can be obtained as a solution of the following program.

$$
\begin{aligned}
D_o(w,m) = &\max_r r \\
&\text{s.t. } g(w-w^t)h(m^t+(w-w^t)r-m)r \leq x_o^t \ \forall t \in \{1,\ldots,T\} \\
&wr \leq x
\end{aligned}
$$

For both experiments, the functions $g(z)$ and $h(z)$ are the same. Moreover, since both experiments are composed of the same set of prices, every left-hand side of the constraints would be the same for both experiments. In addition, if the left-hand side is decreasing in r, then constraint is not binding, hence, we need to concentrate only on increasing left-hand sides. At the same time every $\tilde{x}_o^t \geq x_o^t$, hence there is larger r and, consequently, a larger demand for x_o at given prices. Therefore, the $\tilde{D}_o(w,m) \geq D_o(w,m)$, i.e. $\tilde{E}$ is more altruistic than E.

Proof of Corollary 6 The proof of this remark is same as the proof of Proposition 2, and is therefore omitted. Note that the last property of the redefined inequality measure still guarantees that if $x_s/x_o < \chi$, then it would be in the strict interior of the budget in IA.

References

Afriat, Sydney N. 1967. “The construction of utility functions from expenditure data.” *International economic review*, 8(1): 67–77.

Afriat, Sydney N. 1973. “On a system of inequalities in demand analysis: an extension of the classical method.” *International economic review*, 460–472.

Agranov, Marina, and Thomas R Palfrey. 2015. “Equilibrium tax rates and income redistribution: A laboratory study.” *Journal of Public Economics*, 130: 45–58.

Aguiar, Victor, and Nail Kashaev. 2017. “Stochastic Revealed Preferences with Measurement Error: Testing for Exponential Discounting in Survey Data.” *Working Paper*.

Andreoni, James. 1990. “Impure altruism and donations to public goods: A theory of warm-glow giving.” *The economic journal*, 100(401): 464–477.

Andreoni, James, and John Miller. 2002. “Giving according to GARP: An experimental test of the consistency of preferences for altruism.” *Econometrica*, 70(2): 737–753.

Apesteguia, Jose, and Miguel A Ballester. 2015. "A measure of rationality and welfare." *Journal of Political Economy*, 123(6): 1278–1310.

Beatty, Timothy KM, and Ian A Crawford. 2011. "How demanding is the revealed preference approach to demand?" *The American Economic Review*, 101(6): 2782–2795.

Becker, Gary S. 1974. "A theory of social interactions." *Journal of political economy*, 82(6): 1063–1093.

Benjamin, Daniel J. 2015. "Distributional preferences, reciprocity-like behavior, and efficiency in bilateral exchange." *American Economic Journal: Microeconomics*, 7(1): 70–98.

Bolton, Gary E, and Axel Ockenfels. 2000. "ERC: A theory of equity, reciprocity, and competition." *American economic review*, 166–193.

Bolton, Gary E, and Axel Ockenfels. 2006. "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment." *The American economic review*, 96(5): 1906–1911.

Bronars, Stephen G. 1987. "The power of nonparametric tests of preference maximization." *Econometrica: Journal of the Econometric Society*, 693–698.

Castillo, Marco E, and Philip J Cross. 2008. "Of mice and men: Within gender variation in strategic behavior." *Games and Economic Behavior*, 64(2): 421–432.

Castillo, Marco E, Philip J Cross, and Mikhail Freer. 2017. "Nonparametric utility theory in strategic settings: Revealing preferences and beliefs from games of proposal and response." *Working Paper.*

Chambers, Christopher P, and Federico Echenique. 2016. *Revealed preference theory.* Vol. 56, Cambridge University Press.

Charness, Gary, and Matthew Rabin. 2002. "Understanding social preferences with simple tests." *The Quarterly Journal of Economics*, 117(3): 817–869.

Cherchye, Laurens, Thomas Demuynck, and Bram De Rock. 2018. "Normality of demand in a two-goods setting." *Journal of Economic Theory*, 173: 361–382.

Chernozhukov, Victor, Han Hong, and Elie Tamer. 2007. "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75(5): 1243–1284.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. "Who is (more) rational?" *The American Economic Review*, 104(6): 1518–1550.

Cox, James C. 1997. "On testing the utility hypothesis." *The Economic Journal*, 107(443): 1054–1078.

Cox, James C, Daniel Friedman, and Vjollca Sadiraj. 2008. “Revealed altruism.” *Econometrica*, 76(1): 31–69.

Dean, Mark, and Daniel Martin. 2009. “How consistent are your choice data?”

Dean, Mark, and Daniel Martin. 2016. “Measuring rationality with the minimum cost of revealed preference violations.” *Review of Economics and Statistics*, 98(3): 524–534.

Deb, Rahul, Robert S Gazzale, and Matthew J Kotchen. 2014. “Testing motives for charitable giving: A revealed-preference methodology with experimental evidence.” *Journal of Public Economics*, 120: 181–192.

Demuynck, Thomas. 2015. “Statistical inference for measures of predictive success.” *Theory and Decision*, 79(4): 689–699.

Diewert, W Erwin. 1973. “Afriat and revealed preference theory.” *The Review of Economic Studies*, 40(3): 419–425.

Durante, Ruben, Louis Putterman, and Joël Van der Weele. 2014. “Preferences for redistribution and perception of fairness: An experimental study.” *Journal of the European Economic Association*, 12(4): 1059–1086.

Echenique, Federico, Sangmok Lee, and Matthew Shum. 2011. “The money pump as a measure of revealed preference violations.” *Journal of Political Economy*, 119(6): 1201–1223.

Engellman, Dirk, and Martin Strobel. 2004. “Inequality aversion, efficiency, and maximin preferences in simple distribution experiments.” *The American Economic Review*, 94(4): 857–869.

Fehr, Ernst, and Klaus M Schmidt. 1999. “A theory of fairness, competition, and cooperation.” *The quarterly journal of economics*, 114(3): 817–868.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. 1998. “Gift exchange and reciprocity in competitive experimental markets.” *European Economic Review*, 42(1): 1–34.

Fehr, Ernst, Michael Naef, and Klaus M Schmidt. 2006. “Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment.” *The American economic review*, 96(5): 1912–1917.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits. 2007. “Individual preferences for giving.” *The American Economic Review*, 97(5): 1858–1876.

Fong, Christina. 2001. "Social preferences, self-interest, and the demand for redistribution." *Journal of Public economics*, 82(2): 225–246.

Forges, Francoise, and Enrico Minelli. 2009. "Afriat's theorem for general budget sets." *Journal of Economic Theory*, 144(1): 135–145.

Harbaugh, William T, Kate Krause, and Timothy R Berry. 2001. "GARP for kids: On the development of rational choice behavior." *The American Economic Review*, 91(5): 1539–1545.

Heufer, Jan. 2014. "Generating random optimising choices." *Computational Economics*, 44(3): 295–305.

Heufer, Jan, and Per Hjertstrand. 2015. "Consistent subsets: Computationally feasible methods to compute the Houtman–Maks-index." *Economics Letters*, 128: 87–89.

Höchtl, Wolfgang, Rupert Sausgruber, and Jean-Robert Tyran. 2012. "Inequality aversion and voting on redistribution." *European economic review*, 56(7): 1406–1421.

Houtman, Martijn, and J Maks. 1985. "Determining all maximal data subsets consistent with revealed preference." *Kwantitatieve methoden*, 19(1): 89–104.

Liu, Bing, Yiyuan Xia, and Philip S Yu. 2000. "Clustering through decision tree construction." 20–29, ACM.

Nishimura, Hiroki, Efe A Ok, and John K-H Quah. 2017. "A Comprehensive Approach to Revealed Preference Theory." *The American Economic Review*, 107(4): 1239–1263.

Porter, Maria, and Abi Adams. 2016. "For love or reward? Characterising preferences for giving to parents in an experimental setting." *The Economic Journal*, 126(598): 2424–2445.

Richter, Marcel K. 1966. "Revealed preference theory." *Econometrica: Journal of the Econometric Society*, 635–645.

Rohde, Kirsten IM. 2010. "A preference foundation for Fehr and Schmidt's model of inequity aversion." *Social Choice and Welfare*, 34(4): 537–547.

Saito, Kota. 2013. "Social preferences under risk: Equality of opportunity versus equality of outcome." *American Economic Review*, 103(7): 3084–3101.

Samuelson, Paul A. 1938. "A note on the pure theory of consumer's behaviour." *Economica*, 5(17): 61–71.

Selten, Reinhard. 1991. "Properties of a measure of predictive success." *Mathematical Social Sciences*, 21(2): 153–167.

Tyran, Jean-Robert, and Rupert Sausgruber. 2006. "A little fairness may induce a lot of redistribution in democracy." *European Economic Review*, 50(2): 469–485.

Varian, Hal R. 1982. "The nonparametric approach to demand analysis." *Econometrica: Journal of the Econometric Society*, 945–973.

Appendix B (Online Only): Computing Houtman-Maks index (HMI)

Multiple approaches to the problem of calculating HMI have been taken in the literature (Choi et al. (2014), Heufer and Hjertstrand (2015)). Choi et al. (2014) and Dean and Martin (2016) used a set cover problem approach to calculate HMI, and we follow in a similar vein. The program below calculates HMI, but does not rely on the linearity of budgets.

We first discuss the calculation of HMI for GARP, which allows us to test for other-regarding preferences and inequality aversion. We will then discuss HMI for NARP, necessary for testing increasing benevolence and other normality-related theories. We require two constants to implement the method, and use the big-M method to selectively activate constraints for a subset of data. Let $M > 1$ be the big-M. Moreover, since strict constraints do not make sense for practical optimization, we introduce an infinitesimal tolerance term $\epsilon < \frac{1}{n}$.

We take as given preference relation $\geq$ with strict part $>$ on space X with $|X| = n$. The (mixed integer) linear program is then to find such $u_i \in [0, 1]$ and $\delta \in \{0, 1\}$ for all elements in X, indexed by $i \leq n$ that minimize $\sum_{i=1}^{n} \delta_i$, so that

$$u_i \geq u_j + \epsilon - M(\delta_i + \delta_j) \text{ for } \{(x_i, x_j) \in X^2 : x_i > x_j\} \tag{1a}$$

$$u_i \geq u_j - M(\delta_i + \delta_j) \text{ for } \{(x_i, x_j) \in X^2 : x_i \geq x_j\} \tag{1b}$$

Binary variables δ in the linear program make constraints active only for the chosen subset of observations, and we thus ensure that this subset is minimal. Then we can simply calculate HMI as $1 - \frac{\sum_{i=1}^{n} \delta_i}{n}$.

To calculate HMI for NARP, we only need to adjust the constraints in the program to Definition 6. We take all pairs of observations $t, v \in \{1, \ldots, T\}$ for which $p_o^t / p_s^t \leq p_o^v / p_s^v$ and $x_s^v \leq \frac{1 - p_o^t x_o^v}{p_s^t}$. These are the observations for which NARP has implications. The (mixed integer) linear program is then to find such $\delta \in \{0, 1\}$ for all t, v above that minimize $\sum_{i=1}^{n} \delta_i$, so that

$$x_o^v \leq x_o^t + M(\delta_t + \delta_v) \tag{2}$$

HMI for JNARP, required for rationalization with both x_s and x_o normal, follows in the same way, but with conditions replaced with those from Definition D.1. We therefore omit it here.

Appendix C (Online Only): Classification

This section discusses our classification methodology and results in more detail. We use the clustering technique from Liu, Xia and Yu (2000), who developed a method to apply decision trees to the problem of organizing unlabeled data.[18] Decision trees are appropriate for our data due to the nested nature of theories.

To select HMI thresholds for the classification rule we use the information gain purity function from Liu, Xia and Yu (2000). Unlike usual distance-based measures (e.g. sum of squares) it can potentially be interpreted regardless of the environment. Information gain is the difference in the expected information needed to identify observed data points against uniform distribution before and after the test. In other words, we select the test that minimizes the weighted entropy for all classes of players. The intuition behind this approach is the following. We apply sequential binary tests for different theories, checking if each data point passes the test at a given threshold level or not. All of these binary tests convey one bit of information about each data point: whether it passes or not. If the performance of the test is indistinguishable from testing uniformly distributed data, then the test conveys no information. If, however, the performance on real data is clearly different from the random data, we would suspect that the test conveys some information about the population, and we would like to maximize this information. Information theory suggests that information in bits conveyed by such tests can be measured as the negative logarithm to the base 2 of the number of possible outcomes described by the test.

To calculate the information gain for a group of n data points we introduce n additional fictional points that have uniformly distributed HMI. We then calculate the expected amount of information needed to classify real points against these uniformly distributed points before and after every binary test. Formally, the information gain from clustering with some threshold α is:

$$1 - \frac{1}{2}\left(\left(\alpha + \frac{N_F}{n}\right)E_F + \left((1-\alpha) + \frac{N_P}{n}\right)E_P\right), \tag{3}$$

[18] Classification trees are widely applied, although mostly as supervised learning technique. That is, it requires having “training” dataset which is already categorized.

where n - number of subjects, N_F (N_P) - number of subjects who fail (pass) the test at $HMI \geq \alpha$, E_F (E_P) - entropy for the class of data that passes (fails) the test. Then $(1-\alpha)$ and α are the fractions of points with uniformly distributed HMI that would respectively pass and fail the test.

Information required to identify a data point against a uniform random draw before clustering is 1 bit. After clustering this information is the weighted sum of entropy in each cluster, which in turn is calculated for the points failing the test as

$$E_F = -D_F^R log_2(D_F^R) - D_F^D log_2(D_F^D),$$

where $D_F^R = \frac{\alpha n}{\alpha n + N_F}$ and $D_F^D = \frac{N_F}{\alpha n + N_F}$. These are the fractions of random and real data points in the cluster that fails the test. The entropy for the cluster that passes the test is calculated in the same manner with fractions $D_P^R = \frac{(1-\alpha)n}{(1-\alpha)n + N_P}$ and $D_P^D = \frac{N_P}{(1-\alpha)n + N_P}$.

By substituting expressions for E_F and E_P in (3), we obtain a simplified expression:

$$1 - \frac{1}{2}\left(\alpha I_P^R + (1-\alpha) I_F^R + \frac{N_P}{n} I_P^D + \frac{N_F}{n} I_F^D\right),$$

where

$$I_P^R = -log_2\left(D_P^R\right),\ I_F^R = -log_2\left(D_F^R\right),$$

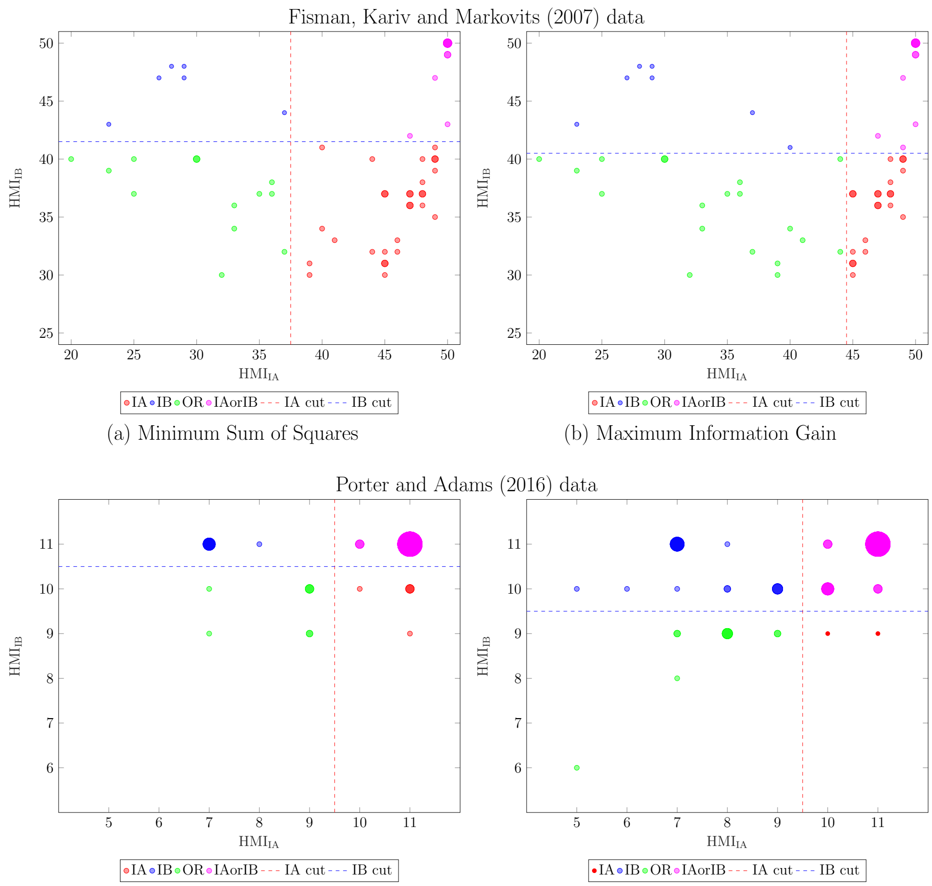
$$I_P^D = -log_2\left(D_P^D\right),\ I_F^D = -log_2\left(D_F^D\right),$$

These four terms represent the information from identifying a point as a data point (D) or as a random point (R) for points that pass the test (P) and fail the test (F). Recall that entropy is the expected amount of information required to decide if some point is an observed data-point or a uniformly-generated random point given the result of a binary test.

We apply this procedure sequentially, first separating the inconsistent cluster by applying the test for other-regarding preferences and then clustering the remaining data according to nested theories. However, we omit the inconsistent points in the figure below.

Classification thresholds for nested theories and classified points are presented in Figure C.1 along with the alternative classification based on minimizing within-cluster sums of square distances from cluster means.

The latter largely agrees with our information gain measure, as can be seen from comparing classification results in Figure 9 and Table C.1. The classification is fairly robust to other



(a) Minimum Sum of Squares
(b) Maximum Information Gain
(c) Minimum Sum of Squares
(d) Maximum Information Gain

Figure C.1. Mixed Types Analysis

Fisman, Kariv and Markovits (2007) data				
IA	IB	IA or IB	OR	IC
29 (38%)	6 (8%)	10 (13%)	13 (17%)	18 (24%)
Porter and Adams (2016) data				
IA	IB	IA or IB	OR	IC
5 (6%)	8 (9%)	40 (45%)	7 (8%)	26 (29%)

Table C.1. Classification of Subjects (Minimal sum of squares)

approaches: 4-means clustering of nested theories agrees with our classification only for half of the data, but qualitatively the results are similar.

Appendix D (Online Only): Detailed Empirical Results

D.1: Additional Theories As an alternative assumption to the increasing benevolence Cox, Friedman and Sadiraj (2008) offered a similar condition, but for one's own payoff. Moreover, it would also be a legitimate assumption to claim normality of both goods. These are two additional theories we are going to test.

Normality of keeping (x_s) may organize the data better, as has been partially shown by Cherchye, Demuynck and De Rock (2018). While the normality of x_s can be checked by applying the permuted version of NARP, the normality in both giving and keeping needs a different test called Joint Normality Axiom of Revealed Preferences.

Definition D.1. *An experiment* $E = (x^t, p^t)_{t=1}^T$ *is consistent with* **Joint Normality Axiom of Revealed Preference (JNARP)** *if and only if for all observations* $t, v \in \{1, \ldots, T\}$ *if* $p_o^t/p_s^t \leq p_o^v/p_s^v$ *and* $x_o^t < x_o^v$*, then* $x_s^t \leq x_s^v$*.*

Both x_s and x_o are normal goods if and only if an experiment satisfies JNARP. Proof of this fact uses the same logic as proof of Proposition 3 and Theorem 2 in Cherchye, Demuynck and De Rock (2018) and is omitted.

D.2: Fisman, Kariv and Markovits (2007) Data.

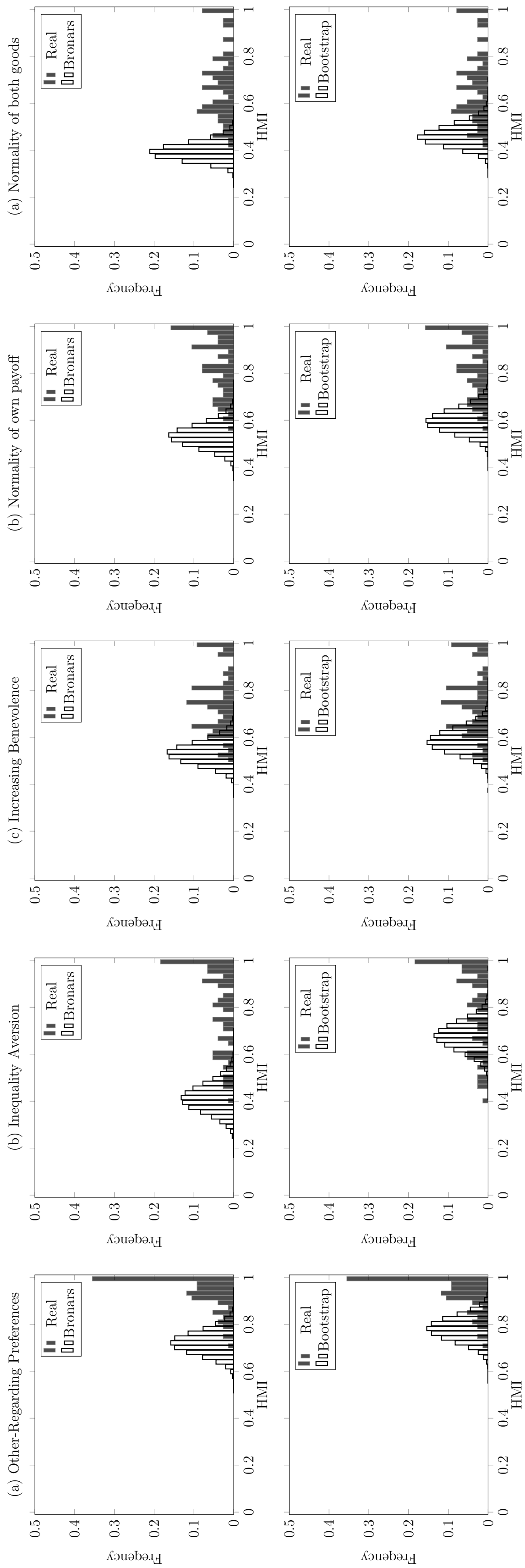


FIGURE D.1. Distributions of HMI for Fisman, Kariv and Markovits (2007) Data

Figure D.1 presents the distributions of HMI indexes for the theories that we are testing. The first row demonstrates the distribution of the HMI for the real subjects in comparison to the distribution of the HMI for the Bronars' power test. The second row presents the distribution of the HMI for the real subjects in comparison to the distribution of the HMI for the bootstrap power test. In order to test the theory, we compare the distribution of its HMI to the distribution of powers of the test. For all five theories, we see that the real subjects pass both the Bronars' and the bootstrap tests; that is, they perform better than random subjects.[19] We confirm that all five theories have empirical support, and at most a quarter of the data needs to be dropped to rationalize an average subject.

Further, we present comparisons of the theories. For this part of the analysis, we restrict our attention to the HMI levels of .8, .9, .95 and 1 (no deviations).

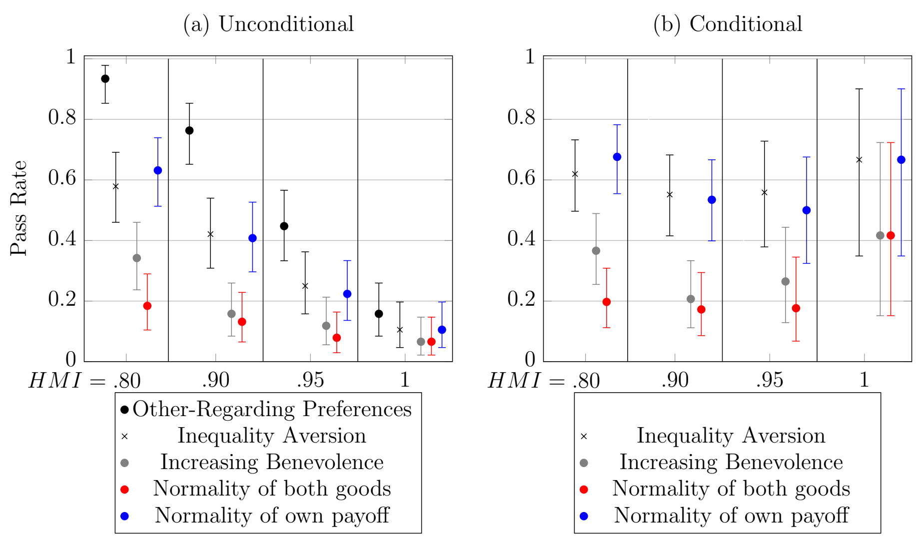


FIGURE D.2. Pass Rates for Fisman, Kariv and Markovits (2007) Data

[19]For other-regarding preferences the mean HMI for real subjects is .92; for Bronars subjects it is .7; for bootstrap subjects it is .76 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests). For inequality aversion the mean HMI for real subjects is .79; for Bronars subjects it is .4; for bootstrap subjects it is .66 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests). For increasing benevolence preferences the mean HMI for real subjects is .74; for Bronars subjects it is .52; for bootstrap subjects it is .57 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests). For normality of x_s the mean HMI for real subjects is .83; for Bronars subjects it is .52; for bootstrap subjects it is .58 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests). For normality of both x_s and x_o the mean HMI for real subjects is .67; for Bronars subjects it is .38; for bootstrap subjects it is .45 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests).

Figure D.2 presents the pass rates with a confidence interval for every theory.[20] We can see that other-regarding preferences demonstrate higher pass rates for all given levels of HMI.

Figure D.2(a) shows that increasing benevolence preferences perform significantly worse than other-regarding preferences. Inequality aversion does better than increasing benevolence (the difference is significant for the HMI levels of $.8, .9$) and worse than other-regarding preferences (the difference is significant for the HMI levels of $.8, .9$). Figure D.2(b) presents the conditional pass rates. In particular, it shows that inequality aversion still outperforms the increasing benevolence preferences (the difference is significant for the HMI levels of $.8, .9$). Moreover, if we assume the inequality aversion preferences, then at least 20% of population who have other-regarding preferences are not behaving as if they have inequality aversion preferences. Assuming increasing benevolence preferences would cost us about 50% of population. This shows that all nested theories are significantly restrictive. Furthermore, the normality of x_s organizes data better than increasing benevolence and joint normality. This result confirms the findings of Cherchye, Demuynck and De Rock (2018) who applied these tests to the Andreoni and Miller (2002) data. Finally, the normality in the own payoff performs as successfully as inequality aversion in this data.

The top row in Figure D.3 shows the predictive success levels with confidence intervals with both Bronars and bootstrap as the control. First of all, we see that the lower bounds of the confidence intervals for the predictive success of all the theories are above zero. Therefore, all of the presented theories predicts the observed behavior better than random (Bronars or bootstrap) decision making. Comparing the predictive success of other-regarding and inequality aversion preferences we see mixed evidence. While the predictive success is always higher for other-regarding preferences, the difference is not always significant. We also see that other-regarding preferences outperform the increasing benevolence preferences at every level of HMI for both Bronars and bootstrap random controls.

In order to compare the predictive success of inequality aversion and increasing benevolence preferences, we take into account the fact that theories are nested. This requires us to not only use the subset of subjects who are consistent with the other-regarding preferences

[20]Confidence intervals are computed using the Clopper-Pearson procedure, since the pass rate can be perceived as a binomial variable.

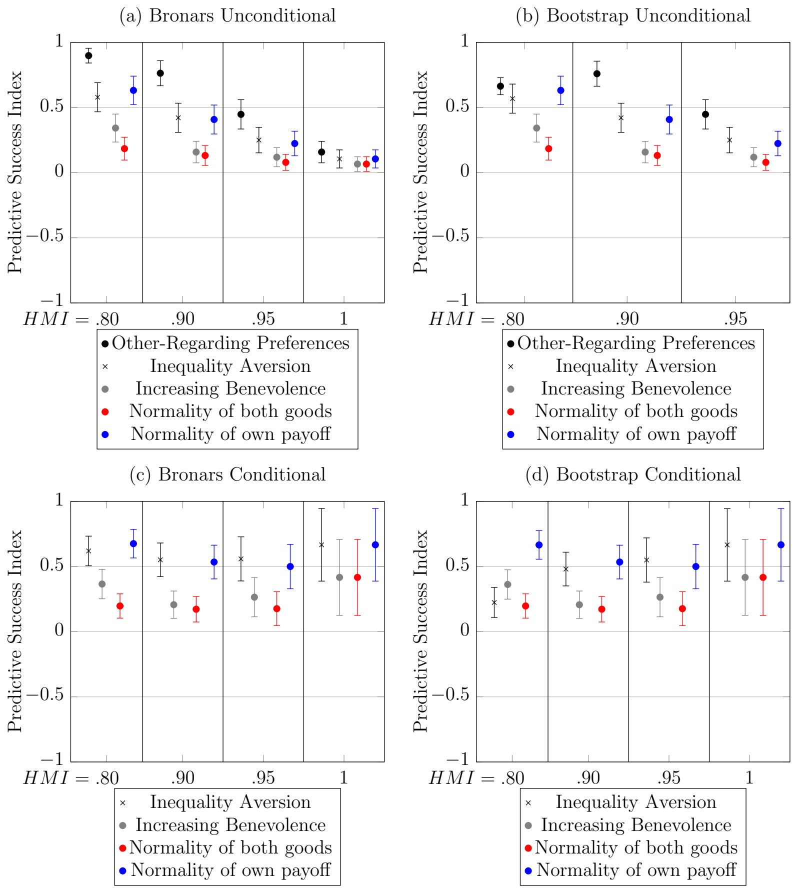


FIGURE D.3. Predictive Success Index for Fisman, Kariv and Markovits (2007) Data

at the given HMI, but to also use random subjects who are also consistent with other-regarding preferences rationalization at the given level of HMI. In particular, if we consider the conditional predictive success for HMI=90%, we take only the subsample of subjects who are consistent with other-regarding preferences rationalization with the HMI of 90%. We then generate for every set of budgets a 1000 random pseudo-subjects who are consistent with other-regarding preferences rationalization with the HMI of 90%. We then compute their HMIs for inequality aversion and increasing benevolence preferences to use those as *random pass rates* for corresponding predictive success indexes. In order to generate random choices that are consistent with other-regarding preferences rationalization at a given HMI, we use the following procedure. We take the random subsample of the experiment that contains 80%, 90% or 95% of budgets and generate the random choices which are consistent with other-regarding preferences rationalization using the Heufer (2014) procedure. We unconditionally place the random choices for the remaining budgets. We calculate this both for the Bronars' and bootstrap tests. The first case follows Heufer (2014) exactly, generating choices that approximate a uniform distribution on each budget, while satisfying GARP. With bootstrap we only need to truncate the empirical distribution at each step to the admissible region, and draw choice points from the resulting distribution. This step is trivial for the Bronars case, since conditional distribution is also a uniform distribution.

The bottom row in Figure D.3 presents the conditional predictive success index. Under the Bronars test, inequality averse preferences theory outperforms increasing benevolence preferences at every level of HMI, and difference is significant at $HMI = .8$ and $HMI = .9$. Under the bootstrap test, we see that for higher levels of HMI, inequality aversion performs better than increasing benevolence, while difference is only significant at $HMI = .9$.

D.3: Porter and Adams (2016) Data. In this experiment every subject has played two sets of budgets in a row. One of them giving to strangers and the other one giving to parents, while the order differs between treatments. We only consider giving to strangers since we want to remain consistent with the analysis we conducted for Fisman, Kariv and Markovits (2007) data.[21] We only consider treatments in which subjects started with a dictator game with strangers to guarantee comparability with the other dataset. Moreover, the second part was a surprise for subjects (it was not announced before the end of the first part), so we can consider the games as comparable.

However, there are some differences in design. The first difference is the population. Fisman, Kariv and Markovits (2007) conducted an experiment with undergraduates from UC Berkley, while Porter and Adams (2016) used a sample of the adults from the southeast region of the UK. Another important difference is that Fisman, Kariv and Markovits (2007) used a constant exchange rate of tokens (experimental currency) to dollars, while Porter and Adams (2016) had a changing exchange rate, through which the price variation was implemented. Let us illustrate this with an example of a decision problem. The subject is given 40 tokens and can decide how much to pass and to hold, while every token she holds converts into 10 pence and every token she passes converts into 30 pence. If one wants to get an equal allocation of tokens (20 pass and 20 hold), then the allocation of real world

[21]Porter and Adams (2016) show that preferences of giving to strangers and parents are significantly different.

currency will not be equal (6 pounds pass and 2 pounds hold). One the other hand, if one wants to keep an equal allocation of real world currency (3 pounds pass and 3 pounds hold), then the allocation of tokens will not be equal (10 tokens pass and 30 tokens hold).

This design feature is important, as the inequality measure for Fisman, Kariv and Markovits (2007) data would be similar whether we think about endowments in tokens or dollars. For Porter and Adams (2016), it would be different. Therefore, we look at both inequality aversion in real currency and in experimental currency. Remark 6 allows us to test for inequality aversion in experimental currency in the same manner as for inequality aversion in real currency. We simply set the equal allocation of tokens as the fair outcome x^*, while the inequality measure is still in terms of real currency payoffs.

We present an analysis similar to one we conducted for Fisman, Kariv and Markovits (2007) data with the only difference that we have two separate versions of inequality aversion. We also report tests for normality of own payoff and normality of both goods. This brings the total number of theories to six. In addition, since this experiment has fewer budgets, we use the following levels of HMI$= 9/11, 10/11, 11/11$.

Figure D.4 presents the distribution of the HMI for all theories. As before, we use the Bronars and bootstrap tests to estimate power. Figure D.4 consists of six panels: (a) for other-regarding preferences; (b) for inequality aversion in real currency; (c) for inequality aversion in experimental currency; (d) for increasing benevolence preferences; (e) for normality of own payoff and (f) for normality of both goods. Most theories outperform random decision making for this data as well.[22] The exception is inequality aversion in experimental

[22]For other-regarding preferences, the mean HMI for real subjects is .96; for Bronars subjects it is .82; for bootstrap subjects it is .78 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests). For inequality aversion in real currency, the mean HMI for real subjects is .84; for Bronars subjects it is .46; for bootstrap subjects it is .59 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests). For increasing benevolence preferences, the mean HMI for real subjects is .93; for Bronars subjects it is .68; for bootstrap subjects it is .7 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests). For normality of x_s, the mean HMI for real subjects is .95; for Bronars subjects it is .68; for bootstrap subjects it is .68 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests). For normality of both x_s and x_o, the mean HMI for real subjects is .92; for Bronars subjects it is .57; for bootstrap subjects it is .6 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests).

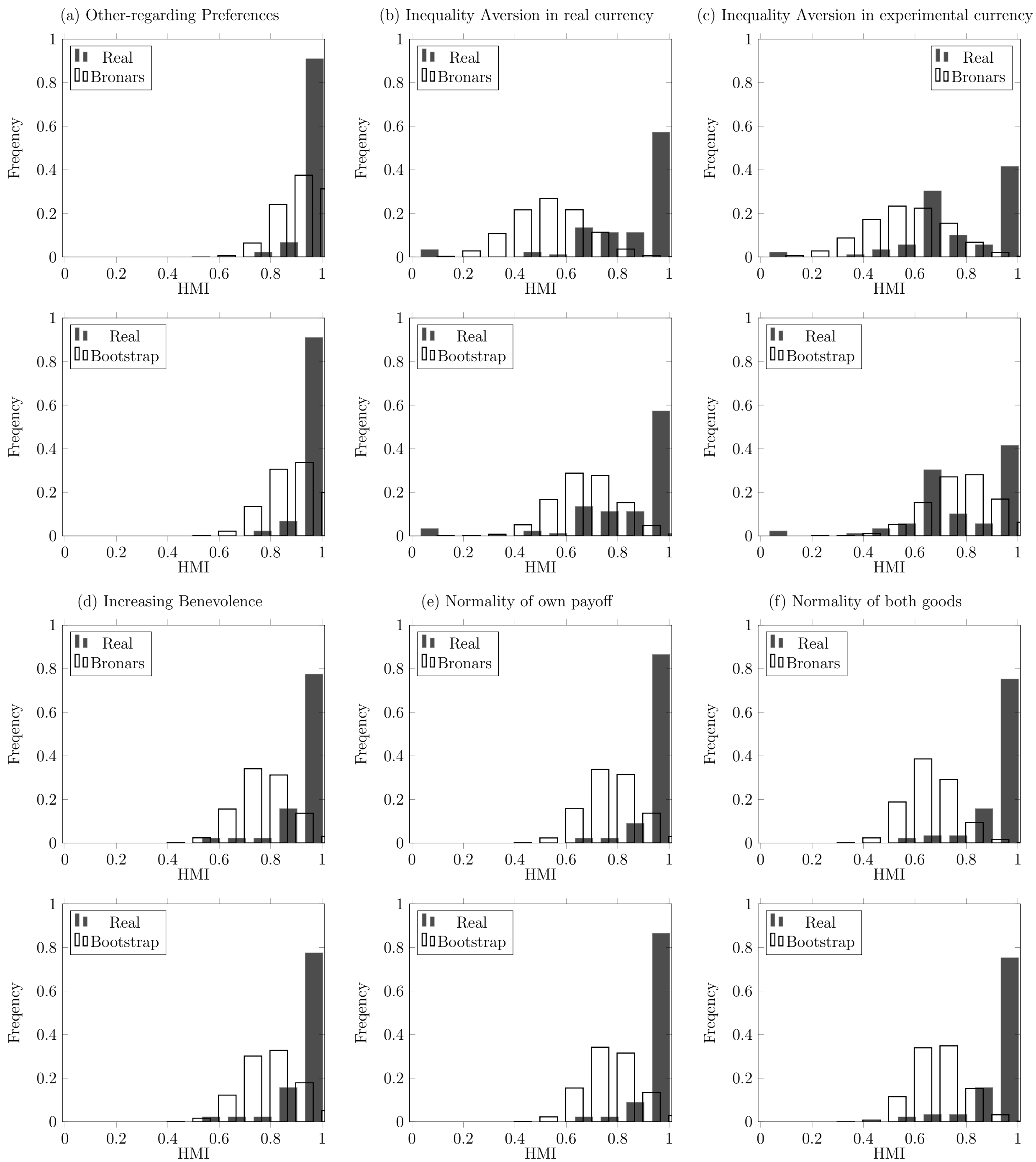


Figure D.4. Distributions of HMI for Porter and Adams (2016) Data

currency. Although it performs better under the Bronars test, it shows almost the same levels of HMI as bootstrap subjects.[23]

[23]For inequality aversion in experimental currency the mean HMI for real subjects is .77; for Bronars subjects it is .49; for bootstrap subjects it is .68 (p-values$<$ .001 for both comparisons using Wilcoxon test and t-tests).

Figure D.1 shows that regardless of framing, subjects are more prone to be inequality averse in real currency than in experimental one. Moreover, power for increasing benevolence is lower in this experiment. Therefore, further comparison should be done based on the predictive success index.

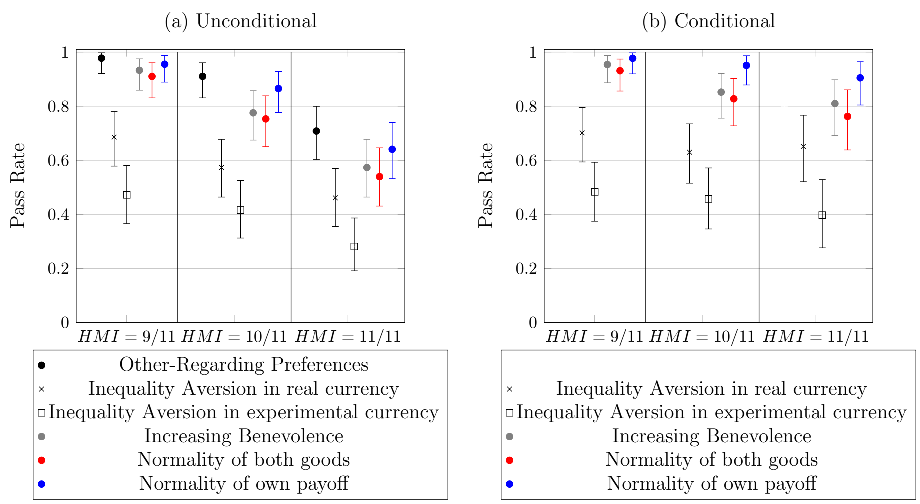


Figure D.5. Pass Rates for Porter and Adams (2016) Data

Figure D.5 presents the pass rates for all theories. Since the latter five theories are nested within the other-regarding preferences, we also present the pass rates conditional on a subject being consistent with other-regarding preferences at a given level of HMI. All nested theories are significantly restrictive, moreover, we observe the ordering of nested theories which is reverse from the one obtained using Fisman, Kariv and Markovits (2007) data. Increasing benevolence preferences tend to be more consistent with the data than inequality averse preferences. Moreover, the difference is statistically significant for $HMI = 9/11$ and $HMI = 10/11$. Let us move on to the predictive success in order to further investigate this, while controlling for the power of the test.

Figure D.6 presents the value of predictive success indexes. As before, we use the Bronars and bootstrap tests to control for both conditional and unconditional predictive success. Ordering of the theories is preserved controlling for power of the test if we restrict HMI for a high enough level ($\geq 10/11$). Observation that the subject's own payoff appears to act as a normal good carries on to this dataset as well. Normality of x_s organizes data better than normality of x_o. Moreover the difference is statistically significant for the bootstrap test and $HMI \leq 10/11$. Additionally, due to the small number of budgets for the low levels of HMI,

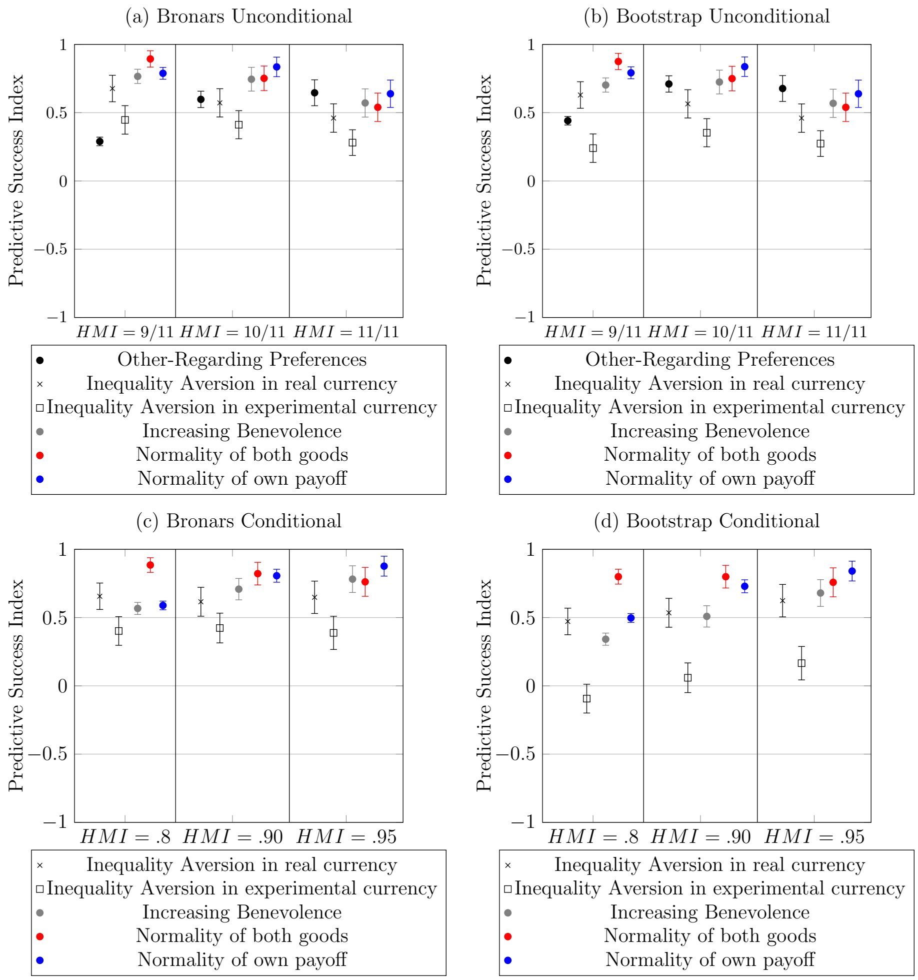


Figure D.6. Predictive Success Index for Porter and Adams (2016) Data

normality in both goods looks rather favorable, since it has the higher power. However, this effect disappears at high enough levels of HMI.