# Preference Recoverability from Inconsistent Choices

Cristián Ugarte*

This version: July 31, 2023
Latest version here

**Abstract**

We study the analysis of choices imperfectly aligned with the preference relation that drives them. First, we develop a measure of decision-making quality that, unlike the existing ones, ensures to asymptotically measure the distance between the subject's choices and her underlying preference (instead of some preference). We then use such a measure to propose a statistically consistent preference estimator. Empirical results suggest consistency is a relevant property when recovering preferences, especially for complex choice environments, compared to estimators based on intuitive motivations.

*Department of Economics; University of California, Berkeley. cugarte@berkeley.edu. I thank Shachar Kariv, Chris Shannon, and Demian Puozo for their invaluable mentorship and advice; Matt Polisson, Yuichiro Kamada, Haluk Ergin, John Quah, and seminar participants at the UC Berkeley Theory lunch, PUC Chile Theory Seminar, UC Berkeley Theory Seminar, UC Berkeley Econometrics Seminar, and 2022 Kansas Workshop in Economic Theory, for helpful discussions and suggestions; María Jesús Ugarte for excellent research assistance; and the Foundations of Data Science Institute for financial support (NSF award #2023505). This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer).

*The econometric theory of demand does study human beings, but only as entities having certain patterns of market behaviour; it makes no claim, no pretence, to be able to see inside their heads.*

Sir John R. Hicks (1956, p. 6)

# 1 Introduction

The problem of recovering preferences from choices is one of the oldest problems in economics, dating back to Antonelli (1886). The focus on finite data starts with Samuelson (1938). It reaches its seminal result in Afriat (1967) theorem, which shows that, in the classical consumer setting, observed choices can be interpreted as driven by a preference relation if, and only if, they satisfy the Generalized Axiom of Revealed Preferences (GARP). However, if a subject's choices are an imperfect implementation of her preferences, she will (with enough observed choices) present violations of GARP. This paper studies how to analyze individual choices that fail GARP.

Starting from the observation that most subjects fail GARP, several alternatives have been proposed to measure *how far* a subject is from satisfying this axiom (Afriat, 1973; Houtman & Maks, 1985; Varian, 1990; Echenique et al., 2011; Dean & Martin, 2016; de Clippel & Rozen, 2021). Such measures are usually interpreted as a proxy for decision making-quality or "economic rationality." However, a measure of decision-making quality should measure the distance between an agent's choices and *her underlying preference*; instead, distance from GARP only measures the distance between an agent's choices and *some* preference. Starting from this observation, we develop a new measure of economic rationality that asymptotically measures the distance between an agent's choices and her underlying preference relation.

The main property to assure that our measure of decision-making quality is adequate is statistical consistency; this is, as the sample size increases, the measured distance is likely close to the distance between agent's the choices and her underlying preference. The proposed measure compares the revealed references (i.e., the preferences we can infer from the observations) with complete and transitive preferences and looks for the preference that minimizes the disagreement between the two. Computationally, we show that our measure can be reduced to solving the Minimum Feedback Arc Set problem.

In order to achieve statistical consistency, we need to impose some structure both on the space of potential preferences and the degree of misalignment between choices and preferences. Concerning the space of preferences, we focus our attention on preferences that are continuous, that follow some objective dominance criteria, for example, the idea that "more is better," and that preferences are Lipschitzian, which is motivated by requiring regularity on how indifference curves fit each other (Mas-Colell, 1977, 1978). For the misalignment between choices and preferences, our main behavioral assumption is that revealed preferences (i.e., the preferences inferred from the data) are more likely to be correct than not, and this has to hold for every subset of the space of bundles.

We use our measure of decision-making quality to propose an estimator of the agent's preferences, which we call the Minimum Mistakes (MM) estimator. The MM estimator is built over two complementary notions. First, it minimizes the number of disagreements between the estimated and revealed preferences, which, given the properties of our rationality measure, assures its statistical consistency. Second, it follows the idea of *partial efficiency* (Afriat, 1973; Varian, 1990), which assures that preferences are inferred not only between observed choices but between an observed choice and every other bundle that was available when the choice was made. To develop our new estimator, we propose a new version of the Afriat Theorem, specifically of its generalization by Nishimura et al. (2017), which includes the possibility of disagreement between the rationalizing preference and some revealed preferences, as well as the possibility of partial efficiency.

We implement our estimator into experimental data from several sources. Our sample comprises 5,345 subjects making choices under risk, choices under uncertainty, or social choices (playing the *dictator game*), all under the experimental design of Choi et al. (2007b). We analyze our estimator regarding how well it predicts choices out-of-sample (this is, choices not used in the original estimation exercise) and compare it with the estimator derived from the Varian (1990) Index (i.e., the Varian estimator). Our results show that the Varian estimator performs better in most sub-samples where subjects choose between two goods (2D environments). In comparison, the MM estimator performs better in all the sub-samples where subjects choose between three goods (3D environments). We interpret these results in light of the Varian Index being motivated by intuitive behavioral descriptions, usually developed in 2D examples, and the MM estimator being motivated by formal properties (statistical consistency). Our results suggest that intuitive explanations of behavior work well in simple environments like the ones usually used for examples that motivate such intuitions. However, as choice problems become more complex and further away from such examples, formal properties become a better tool for developing general estimators.

### 1.1 Related Literature

This paper contributes to the vast literature on revealed preferences. In particular, it addresses two fundamental and closely related questions regarding subjects who fail GARP. The first question is how to measure the misalignment between an agent's choices and her underlying preferences (i.e., to measure decision-making quality). The second is how to recover such preference relation starting from such choices.

The study of revealed preferences in finite data dates back to Samuelson (1938) and Houthakker (1950), who propose the idea of *revealed preferences* to infer elements of a consumer's preference from her observed choices. The Afriat theorem (Afriat, 1967; Varian, 1982) identifies GARP as necessary and sufficient for choices to be consistent with some preference relation. The Afriat theorem has been expanded in several directions, notably by Forges and Minelli (2009) for nonlinear budget sets, Reny (2015) for infinite observed choices, and Nishimura et al. (2017) for general choice spaces and preferences with more restrictive requirements.

The development analysis of subjects whose choices satisfy GARP has focused chiefly on *how far* is a subject from satisfying this property, which is usually interpreted as decision-making quality. The first of such measures is the Critical Cost Efficiency Index (CCEI), proposed by Afriat (1973) based on partial efficiency (see Section 4.1 for a formal definition). The CCEI is defined as one minus the highest partial efficiency level at which we can think of choices as coming from a preference relation. Varian (1990) proposes a generalization in which different choices have different levels of partial efficiency and proposes a measure that aggregates these levels into a single statistic. Although they offer different interpretations, the indices proposed by Houtman and Maks (1985), Echenique et al. (2011), and Dean and Martin (2016) can be considered different aggregations of partial efficiency. de Clippel and Rozen (2021) and Echenique et al. (2022) take a different approach and look at the minimum discrepancy between price rations and the marginal rate of substitution for the data to be consistent.

The problem of identifying an agent's preference from her choices starts with Mas-Colell (1977, 1978). He studies this possibility when choices and preferences are perfectly aligned (i.e., when choices satisfy GARP). He shows that the underlying preference driving the choices can be asymptotically identified. Under imperfect misalignment between choices and preferences, Apesteguia and Ballester (2015) and Chambers et al. (2021) study how to recover the underlying preference from observed choices. Conceptually, these two papers are the closest to this one, although the choice environments they study are different.[1] From the two, Chambers et al. (2021) is the closest regarding the tools used to develop our results.

Technically, our work relies upon several sources. First, we use the large-sample theory of extremum-estimators (Amemiya, 1985; Newey & McFadden, 1994; Jennrich, 1969) and U-statistics (Hoeffding, 1961). Second, we choose the closed-convergence topology as our topology in the space of preference relations and use properties developed for such topology (Kannai, 1970; Hildenbrand, 1974; Grodal, 1974; Redekop, 1993; Border & Segal, 1994). Finally, we rely on the generalizations of the Afriat Theorem by Forges and Minelli (2009) and Nishimura et al. (2017).

The remainder of the paper proceeds as follows. Section 2 describes the choice environment and discusses the main definitions and assumptions. Section 3 proposes our measure of decision-making quality, develops its properties, and presents a computation strategy. Section 4 proposes a preference estimator based on the measure developed in the previous section. Section 5 implements the estimator in experimental data. Finally, Section 6 concludes.

[1] Apesteguia and Ballester (2015) study choice environments with a finite number of alternatives and assumes that all the preferences are strict. Chambers et al. (2021) study general spaces of alternatives and assume that the consumer makes choices from pairwise of alternatives.

## 2 Setup

### 2.1 Choice Environment

Our data set consists of $N$ observations of an agent's consumption decisions. The agent consumes bundles of $K$ nonnegative commodities; the consumption space is $\mathbb{R}^K_+$.[2] In each observation $i \in [N]$, the agent faces a budget set $B^i \subset \mathbb{R}^K_+$ and chooses a bundle $x^i \in B^i$. Since commodities are assumed to be desirable, we define the upper boundary of a budget set as $b(B^i) = \{x \in B^i : y \gg x^i \implies y \notin B^i\}$ and assume $x^i \in b(B^i)$ for all $i$. Following Forges and Minelli (2009), we impose the following restrictions on the budget sets.

**Assumption 1.** Each budget set $B^i$ is compact, comprehensive,[3] and contains at least one element $x \in B^i$ satisfying $x \gg \mathbf{0}$.

Each budget set $B^i$ induces a *gauge* $g^i$, defined as $g^i(x) = \inf\{\lambda > 0 : x \in \lambda B^i\}$, where $\lambda B^i \equiv \{\lambda x : x \in B^i\}$. Given Assumption 1, Lemma 1 in Forges and Minelli (2009) assures that $g^i$ is continuous, increasing, and homogeneous of degree one. Furthermore, the budget set and its upper boundary can be characterized by $B^i = \{x \in \mathbb{R}^K_+ : g^i(x) \leq 1\}$ and $b(B^i) = \{x \in \mathbb{R}^K_+ : g^i(x) = 1\}$.[4] For convenience, we define each budget set in the data set by its gauge $g^i$. Hence, our data set is $\mathcal{D} = (g^i, x^i)_{i\in[N]}$.

We assume an objective dominance relation $\trianglerighteq$ on $\mathbb{R}^K_+$; we interpret $x \trianglerighteq y$ as $x$ being objectively (weakly) better than $y$. Furthermore, we interpret $x \vartriangleright y$ (where $\vartriangleright$ is the asymmetric component of $\trianglerighteq$) as $x$ being objectively strictly better than $y$. Following Nishimura et al. (2017), we impose the following conditions on $\trianglerighteq$.[5]

**Assumption 2.** The dominance relation $\trianglerighteq$ is a continuous preorder and extends $\geq$ (this is, $\geq \subset \trianglerighteq$ and $> \subset \vartriangleright$).

Continuity of $\trianglerighteq$ is assumed because this property is fundamental for inference purposes; intuitively, continuity allows extrapolating from any preference to a neighborhood around it. $\trianglerighteq$ being

[2] We work with the following notation. $\mathbb{N}$ is the set of natural numbers and $\mathbb{R}$ the set of real numbers. $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ is the set of positive numbers including zero, and $\mathbb{R}_{++} = \{x \in \mathbb{R}_+ : x \neq 0\}$ excludes it. For any $M \in \mathbb{N}$ we set $[M] = \{1, 2, \ldots, M\}$. A vector $x \in \mathbb{R}^M$ is $x = (x_1, x_2, \ldots, x_M)$. The vectors $\mathbf{0}$ and $\mathbf{1}$ have all their elements equal to zero and one, respectively (the context implies their dimensionality). For any two vectors $x, y \in \mathbb{R}^M$ we write $x \geq [\gg] y$ if $x_i \geq [>] y_i$ for all $i \in [M]$, and $x > y$ if $x \geq y$ and $x \neq y$ ($\leq$, $\ll$, and $<$ are defined similarly). A function $f : \mathbb{R}^M \to \mathbb{R}$ is increasing if $x > y$ implies $f(x) > f(y)$ and strictly increasing if $x \gg y$ implies $f(x) > f(y)$. For $x, y \in \mathbb{R}^M$, $||x - y||$ denotes the Euclidean distance, and $x \cdot y$ the dot product. A binary relation $R$ on $X$ is a subset of $X \times X$. As usual, we write $xRy$ as equivalent to $(x, y) \in R$ and $x\not{R}y$ as equivalent to $(x, y) \notin R$. The asymmetric component of $R$ is $P$, defined by $xPy$ if $xRy$ and $y\not{R}x$, and its symmetric component is $I$, defined by $xIy$ if $xRy$ and $yRx$. For sets $X$ and $Y$, $|X|$ and $|Y|$ represent their respective cardinality, and $X \setminus Y$ is the set difference of $X$ and $Y$. Finally, $\mathbf{1}\{\cdot\}$ is the indicator function.

[3] A set $B$ is comprehensive if for every $x \in B$ and $y \leq x$ we have $y \in B$.

[4] Forges and Minelli (2009) also include the following assumption (Assumption H) in the requirements for the budget set: "if $x \in b(B)$ then $kx \in B \setminus b(B)$ for all $k \in [0, 1)$". However, this assumption is redundant: as $B$ is comprehensive and $k < 1$, then $x \in B$ implies $kx \in B$, and as $x \gg kx$ we have $kx \notin b(B)$.

[5] A binary relation $R$ on $X$ is a preorder if it is reflexive ($xRx$ for all $x \in X$) and transitive ($xRy$ and $yRz$ imply $xRz$); it is continuous if, for all $x \in \mathbb{R}^K_+$, the upper and lower contour sets, $\{y : xRy\}$ and $\{y : yRx\}$, are closed.

a preorder is a natural assumption, as $x$ is always weakly better to itself, and transitivity is a reasonable requirement for every criterion of desirability. Finally, that $\trianglerighteq$ extends $\geq$ implies that our dominance relation respects the "more is better" criterion.

The existence of an objective dominance relation $\trianglerighteq$ leads to the following definition of revealed preferences.

**Definition 1.** For $i, j \in [N]$, $x^i$ is

- *directly revealed preferred* to $x^j$ (denoted $x^i \succsim^D x^j$) if there is $y \in \mathbb{R}^K_+$ satisfying $g^i(y) = 1$ and $y \trianglerighteq x^j$;
- *directly revealed strictly preferred* to $x^j$ ($x^i \succ^D x^j$) if there is $y \in \mathbb{R}^K_+$ satisfying $g^i(y) = 1$ and $y \triangleright x^j$;
- *revealed preferred* to $x^j$, ($x^i \succsim^R x^j$) if there is a sequence of observations $(m_\ell)_{\ell \in [L]}$, such that $x^i \succsim^D x^{m_1} \succsim^D x^{m_2} \succsim^D \ldots \succsim^D x^{m_L} \succsim^D x^j$; and
- *revealed strictly preferred* to $x^j$ ($x^i \succ^R x^j$) if there are observations $m, m' \in [N]$ such that $x^i \succsim^R x^m \succ^D x^{m'} \succsim^R x^j$.

If $\trianglerighteq = \geq$ and every $g^i$ is of the form $g^i(x) = p^i \cdot x$ for some $p^i \gg \mathbf{0}$, we refer to the choice environment as the *classical consumer environment*. In such a setting, Definition 1 reduces to the classical definition of revealed preferences (see Section 1 in Varian, 1982).

The intuition for the above definition is that if the agent chooses $x^i$ when $y$ is also available, then $x^i$ is revealed preferred to $y$, and if $y \trianglerighteq x^j$ then $y$ is (objectively) preferred to $x^j$; hence transitivity implies that $x^i$ is preferred to $x^j$. The same logic follows for the definition of strict preferences, with the change that $y \triangleright x$ allows us to conclude that $x^i$ is strictly preferred to $x^j$. The definitions of (indirectly) revealed preferences follow from the direct revealed preferences and transitivity. The following remark shows that an equivalent definition to the directly revealed strict preferences could follow the following logic: if $x^i$ is chosen over $y$ and something else and $y \trianglerighteq x^j$, then $x^i$ is strictly preferred to $x^j$. The proofs of all the remarks are in Section 3.1 of the Online Appendix.

*Remark* 1. $x^i \succ^D x^j$ if, and only if, there is a bundle $y$ such that $g^i(y) < 1$ and $y \trianglerighteq x^j$.

Since Afriat (1967), we know that, in the classical consumer environment, observed choices are consistent with a preference relation if and only if they satisfy the *Generalized Axiom of Revealed Preferences* (GARP). GARP states that for any two $i, j \in [N]$, if $x^i \succsim^R x^j$ then $x^j \succ^D x^i$ has to be false. Nishimura et al. (2017) expand Afriat Theorem to a more general choice environment that includes the one studied in this paper.[6]

We interpret failures of GARP as an indication of imperfect implementation of the agent's preferences. However, we do not specify the source of such limitation. Instead of dropping the assumption that the agent has a preference relation, we assume that such preference exists, but

[6]As in the original result by Afriat (1967), Nishimura et al. (2017) do not state their result in terms of GARP but in terms of *cyclical consistency*. Both approaches are equivalent.

the agent's choices are imperfectly aligned with it. In our view, this is the natural interpretation: to interpret a failure of GARP as a sign that the agent has no unique preference relation leaves us with no path to meaningfully analyze any finite number of observations without adopting ad-hoc models (see Afriat, 1973; Bernheim & Rangel, 2009, for further discussions on this problem).

### 2.2 Preferences

The primary motivation of our problem is that choices are an imperfect implementation of the agent's underlying preference relation. We denote this preference by $\succsim^\star$. In the classical consumer environment, and assuming perfect alignment between choices and preferences,[7] Mas-Colell (1977, 1978) studies the problem of asymptotically recovering preferences. He shows that the space of preferences is, in general, too complex to pin down the agent's preferences uniquely. However, he shows that such exercise can be done under certain (non-testable) regularity conditions. The first required property is continuity. Intuitively, continuity allows us to extend each strict preference that we learn to a neighborhood of the bundles being compared.[8]

A crucial property that Mas-Colell (1977) imposes to recover preferences is for preferences to be Lipschitzian, a property he describes as capturing the idea of "indifference curves fitting together not too wildly" (p. 1411). Given a preference $\succsim$, denote by $P^{\succsim}(x) = \{y : x \succsim y\}$ the upper contour set of $x$. Also, for any sets $X, Y \subset \mathbb{R}^K_+$ define $\delta(X, Y) = \inf_{x\in X, y\in Y} ||x - y||$; for simplicity we write $\delta(\{x\}, Y) = \delta(x, Y)$. Finally, for any $r \geq 0$ let $K_r = \{x \in \mathbb{R}^K_+ : (1 + r)^{-1}\mathbf{1} \leq x \leq (1 + r)\mathbf{1}\}$.

**Definition 2.** A preference relation $\succsim$ is *Lipschitzian* if for every $r > 0$ there are reals $H > 0$ and $\varepsilon > 0$ such that if $x, x', y \in K_r$, $x \sim x'$, and $||x - y|| < \varepsilon$, then $\delta(x, P^{\succsim}(y)) \leq H\delta(x', P^{\succsim}(y))$. A set of preferences $\mathcal{P}_0$ is *uniformly Lipschitzian* if, for every $r > 0$, the preferences $\succsim \in \mathcal{P}_0$ admit common Lipschitz constants $H$ and $\varepsilon$.

Paraphrasing Rader (1972, p. 171), Lipschitzian preferences can be explained as follows.[9] Suppose that a change from $x$ to the preference level of $y$ can be achieved with a certain minimum magnitude (in distance) $\delta(x, P^{\succsim}(y))$. Then a change from any other $x'$ indifferent to $x$ can be achieved only with movement at least some fraction $H$ of the minimum distance necessary from $x$. The important point is that the fraction does not depend on $x$ or on how close $y$ is to $x$. However, the sensitivity of preferences to the magnitude of a change in position is always limited by the sensitivity of preferences at any given point on the indifference set of $x$.

We also impose for any "reasonable" preference to agree with the objective dominance relation $\trianglerighteq$.

**Definition 3.** A preference $\succsim$ is $\trianglerighteq$-*monotone* if $x \trianglerighteq y$ implies $x \succsim y$ and $x \triangleright y$ implies $x \succ y$.

[7]This is, if $x^i \succsim^\star x$ whenever $g^i(x) \leq 1$, for all $i \in [N]$.

[8]Since he focuses on linear prices, Mas-Colell (1977, 1978) also imposes for the preferences to be convex; the reason for this is that under linear prices, an agent choosing optimally will never choose in the non-convex portion of her preferences; hence this property cannot be tested.

[9]Rader (1972, Chapter 8) studies Lipschitzian preferences under the name *uniformly sensitive preferences*.

We assume that the agent's underlying preference satisfies all the previous properties.

**Assumption 3.** $\succsim^\star \in \mathcal{P}$, where $\mathcal{P}$ is a uniformly Lipschitzian set of continuous and $\rhd$-monotone preferences.

We endow the space of continuous preferences with the *closed convergence topology*. Intuitively, this topology tells us that if two preferences are close to each other, then comparisons of bundles do not change abruptly between the two. Specifically, if we have bundles $x$ and $y$ and a preference relation $\succ$ such that $x \succ y$, then for any sequences of bundles $x^n \to x$ and $y^n \to y$ and sequence of preferences $\succsim^n \to \succsim$ (in the topology of the space of preferences) for $n$ large enough we have $x^n \succ^n y^n$.[10] The closed convergence topology is the standard choice to study preferences (see Chambers et al., 2021, for a review of the literature). Under this topology, the set of continuous binary relations is compact and metrizable (see Theorem 2 in Chapter B of Hildenbrand, 1974).

**Theorem 1.** *$\mathcal{P}$ is compact.*

The previous result, whose proof is in Appendix A, gives us known bounds on the space of preferences. In particular, it assures that any convergent sequence of preferences in $\mathcal{P}$ will converge to an element of $\mathcal{P}$. For more details about the relevance of this property to identify the true parameter (in this case $\succsim^\star$), see Newey and McFadden (1994).

### 2.3 Data Generating Process

In order to learn about the agent's underlying preference $\succsim^\star$, we need to impose some regularity between such preferences and the data; this is, we need to specify a data generating process. First, we assume independence between observations. As mentioned before, we also assume that the agent consumes in the upper boundary of the consumption space, i.e., that $g^i(x^i) = 1$. Let $\mathcal{G}$ be the set of gauges induced by budget sets satisfying Assumption 1. We denote by $\mathcal{O}$ the space of possible observations, i.e., $\mathcal{O} = \{(g, x) \in \mathcal{G} \times \mathbb{R}^K_+ : g(x) = 1\}$, and by $\mu$ the probability measure that generates observations.

Revealed preferences, specifically directly revealed preferences, are our primary sources of information. As directly revealed preferences are constructed by comparing pairs of observations, we must focus on how pairs of observations are constructed. For this, we denote by $\mu^2$ the probability measure induced by $\mu$ in $\mathcal{O} \times \mathcal{O}$. Our main behavioral assumption is the following.

[10] A complete characterization also requires that if $x \succsim y$ and $\succsim^n \to \succsim$, then there have to be sequences $x^n \to x$ and $y^n \to y$ such that for $n$ large enough $x^n \succsim^n y^n$. In terms of open sets, the closed convergence topology is the smallest topology for which sets of the form $\{x, y, \succsim : x \succ y\}$ are open in the induced product topology. This characterization is due to Kannai (1970), and Redekop (1993) shows that in most economic environments, specifically ours, it is equivalent to the classical characterization of the topology of closed convergence. He also proposes a third equivalent topology named the *questionnaire topology*. A detailed description can be found in Section 2 of Hildenbrand (1970) and Section B.II of Hildenbrand (1974).

**Assumption 4.** If $A, B \subset \mathbb{R}^K_+$ are nonempty and open such that $A \succ^\star B$ (this is, $x \succ^\star y$ for all $x \in A$ and $y \in B$) then

$$\mu^2\left(\left\{((g,x),(f,y)) \in \mathcal{O} \times \mathcal{O} : x \in A,\, y \in B, \text{ and } x \succ^D y\right\}\right) > \mu^2\left(\left\{((g,x),(f,y)) \in \mathcal{O} \times \mathcal{O} : x \in A,\, y \in B, \text{ and } y \succsim^D x\right\}\right) \tag{1}$$

For expositional purposes, in the rest of the paper we will simplify the notation to write the value $\mu^2$ of a set of pairs of observations. For example, we write (1) as

$$\mu^2\left(x \in A,\, y \in B, \text{ and } x \succ^D y\right) > \mu^2\left(x \in A,\, y \in B, \text{ and } y \succsim^D x\right) .$$

Assumption 4 is a joint restriction on the distribution of prices, the agent's choice rule, and her underlying preference relation. It implies that it is more likely to observe correct revealed preferences than incorrect ones. Moreover, that has to be true even when we focus only on subsets of bundles; if not, then we would not be able to use revealed preferences to differentiate between $\succsim^\star$ and other preferences that disagree only on how to compare these subsets of bundles.

The following result (whose proof is in Appendix B) facilitates the interpretation of Assumption 4.

**Proposition 1.** *Let $\mu_X$ be the marginal probability on choices.*[11] *Assumption 4 holds if, and only if*

1. *for each nonempty open set $A \subset \mathbb{R}^K_+$, $\mu_X(A) > 0$; and*
2. *for nonempty open sets $A, B \subset \mathbb{R}^K_+$ such that $A \succ^\star B$*

$$\mu^2\left(x \succ^D y \middle| x \in A,\, y \in B\right) > \mu^2\left(y \succsim^D x \middle| x \in A,\, y \in B\right) .$$

The first part of Proposition 1 tells us that as the number of choices increases, these choices become dense in the space of bundles. This requirement is necessary to learn about an agent's preferences over all the space of alternatives through revealed preferences, as it allows us to compare bundles in all the consumption space. As many preferences agree on how to compare everywhere except in a subset of bundles, without this assumption, it would be impossible to differentiate between such preferences.

The second part of Proposition 1 tells us that whenever we observe directly revealed preferences, they are more likely to be correct than not; this is, that revealed preferences are trustworthy. The idea of this property is that if we want to use revealed preferences as our source of information, then we should, in expected value, trust the revealed preferences we infer. To see this, suppose the second condition in Proposition 1 fails; then some revealed preferences give (in expected value) incorrect information. Unless we have an *a priori* idea of how to differentiate between revealed

[11] This is, for any $Y \subset \mathbb{R}^K_+$, $\mu_X(Y) = \int_{\mathcal{G}} \mu((g, Y))dg$.

preferences being likely correct and those being likely incorrect (which is unclear how to obtain), revealed preferences are not an adequate tool to infer preferences.

We also assume that observing choices that are indifferent to each other is unlikely. This assumption ensures that the revealed preferences we infer are maintained at the limit (almost surely); if not, two relations may be close, and yet the revealed preferences will behave differently concerning them.

**Assumption 5.** For any $\succsim \in \mathcal{P}$, observing two different choices that are indifferent is a zero probability event, i.e., $\mu^2 \left( x \neq y,\, x \sim y \right) = 0$

Assumptions 4 and 5 are common in the literature studying how to learn about an agent's preferences from her choices. For the case when preferences and choices are perfectly aligned, Mas-Colell (1978) requires the first statement of Proposition 1 under a "boundary condition" property, and Assumption 5 is implied by his definition focus on the Strong Axiom of Revealed Preferences instead of GARP (Definition 2 in his paper); as choices and preferences are perfectly aligned, the second condition in Proposition 1 holds as well. For the case with imperfect implementation, both Apesteguia and Ballester (2015) and Chambers et al. (2021) require properties analogous to the ones we require here.[12]

## 3 Measure of Rationality

We propose the following measure of distance between the observed choices in $\mathcal{D}$ and a preference relation $\succsim$

$$d(\mathcal{D}, \succsim) = \frac{|\succ^D \setminus \succ|}{|\succ^D|}.$$

The function $d(\mathcal{D}, \succsim)$ takes the set of directly revealed strict preferences $\succ^D$ and measures the share of its elements that disagree with the preference relation $\succsim$. This function is motivated by the Swaps Index developed by Apesteguia and Ballester (2015) and the objective function for the estimator in Chambers et al. (2021). The function $d(\cdot, \cdot)$ is a modified version of traditional U-statistics (Hoeffding, 1961). To see this, note that

$$d(\mathcal{D}, \succsim) = \left( \sum_{i,j} \mathbf{1}\left\{ x^i \succ^D x^j \right\} \right)^{-1} \sum_{i,j} \mathbf{1}\left\{ x^i \succ^D x^j \right\} \mathbf{1}\left\{ x^j \succsim x^i \right\}$$

where $\sum_{i \neq j}$ is short notation for $\sum_{i \in [N]} \sum_{j \in [N] \setminus \{i\}}$. While traditional U-statistics divide by the number of possible combinations (in our case $N^2$), our distance measure normalizes by the number

[12] Apesteguia and Ballester (2015) require $\sum_{a \in A} f(A, a) > 0$ for every menu $A$ (stated in their Theorem 1), analogous to the first property in Proposition 1, and their $P$-monotonicity, analogous to the second one; Assumption 5 is vacuously satisfied as in their paper all preferences are strict. Chambers et al. (2021) require the first property of Proposition 1 and Assumption 5 in their Assumption 3', and the second property of Proposition 1 by requiring $x \succ y \implies q(\succsim; x, y) > 1/2$ for their statistical choice function $q$.

of revealed preferences $|\succ^D|$.

We propose the following measure of decision-making quality for a given data set

$$\Delta(\mathcal{D}) = \min_{\succsim \in \mathcal{P}} d(\mathcal{D}, \succsim)\,. \tag{2}$$

Our main motivation for proposing a new measure of economic rationality is that departures from rationality should be measured by the misalignment between an agent's observed choices and her underlying preference relation, not between her choices and some preference relation. The following result, which is the main result of our paper, tells us that, as the sample size increases, $\Delta(\mathcal{D})$ is likely to measure the desired misalignment.[13]

**Theorem 2.** *Let $\succsim^N$ be a minimizer of (2) when the data set has $N$ observations. Then, as $N \to \infty$,*

$$\succsim^N \xrightarrow{p} \succsim^\star \,.$$

The proof of this result is in Appendix C.

### 3.1 Computation

Our proposed computation method relies on the following assumption.

**Assumption 6.** Let $\mathcal{C}$ be the set of all continuous and $\rhd$-monotone preferences. For every $\succsim \in \operatorname{argmin}_{\succsim \in \mathcal{C}} d(\mathcal{D}, \succsim)$ there is $\succsim' \in \mathcal{P}$ such that $\succ^D \setminus \succ = \succ^D \setminus \succ'$.

Assumption 6 tells us that obtaining a lower distance between the data set and a preference is impossible if we drop the Lipschitzian condition on the preferences. Since the Lipschitzian property is a property on how indifference curves that are close to each other behave with finite data, which is always sparse, it is impossible to test this assumption.

Assumption 6 implies that our measure of distance from rationality results from a minimization of the objective function over *all* continuous and $\trianglerighteq$-monotone preference relations. The fact that we do not exclude any of such relations beforehand is crucial, as the space of candidates includes all the binary relations that are transitive and excludes all the ones that are not. Hence, any estimator's strict component $\succ$ will be $\rhd$-monotone and acyclic. This is, for any sequence $\{x^{m_\ell}\}_{\ell=1}^L$ such that $x^{m_1} \succ x^{m_2} \succ \ldots \succ x^{m_L}$ we have $x^{m_L} \not\succ x^{m_1}$. Acyclicality allows us to reduce the computation of the MM Index to a vastly studied problem in computer science: the *minimum feedback arc set* (MFAS) problem.

[13] As usual, the notation $\succsim^N \xrightarrow{p} \succsim$ means that for every $\varepsilon > 0$

$$\lim_{N\to\infty} \Pr\left(\rho\left(\succsim^N, \succsim\right) \geq \varepsilon\right) = 0\,,$$

where $\rho$ is any metric compatible with the chosen topology.

We start by reducing our data to a directed graph (or digraph). A digraph $G = (V, E)$ is composed of a set of vertices $V$ and a set of edges $E \subset V \times V$, where if $x, y \in V$ are two vertices and $(x, y) \in E$ is an edge. A *cycle* is a sequence of vertices $x_1, x_2, \ldots, x_M$ such that $(x_m, x_{m+1}) \in E$ for all $m \in [M-1]$, and $(x_M, x_1) \in E$. A digraph is *acyclic* if it has no cycles. After associating a positive cost to each edge, the MFAS problem is to find the minimum cost of removing edges to make the resulting digraph acyclic.

**Definition 4.** Take a digraph $G = (V, E)$ and a set of weights $\Omega = (\omega_e)_{e\in E}$, where $\omega_e > 0$ for all $e \in E$. The *minimum feedback arc set* (MFAS) problem of $[G, \Omega]$ is

$$\min_{E'\subset E} \sum_{e\in E'} \omega_e, \text{ subject to } (V, E \setminus E') \text{ is acyclic.} \tag{3}$$

To reduce the computation of $\Delta(\mathcal{D})$ to an MFAS problem, we first reduce the choice data to a digraph. In order to do so, we need to take into account that the dominance relation $\rhd$ imposes some order in the observed data: first, $\rhd$ informs us about strict preferences; second, it is possible to infer indifferences between two bundles $x, y$ if $x \unrhd y$ and $y \unrhd x$. We denote such relation as $\sim_\rhd$. In order to include the potential indifferences in our data, we reduce the set of observations to one per indifference set (according to $\unrhd$). Then, we assure to respect $\rhd$ by giving the revealed preferences that agree with $\rhd$ enough weight to ensure that they will not be removed when solving the MFAS problem.

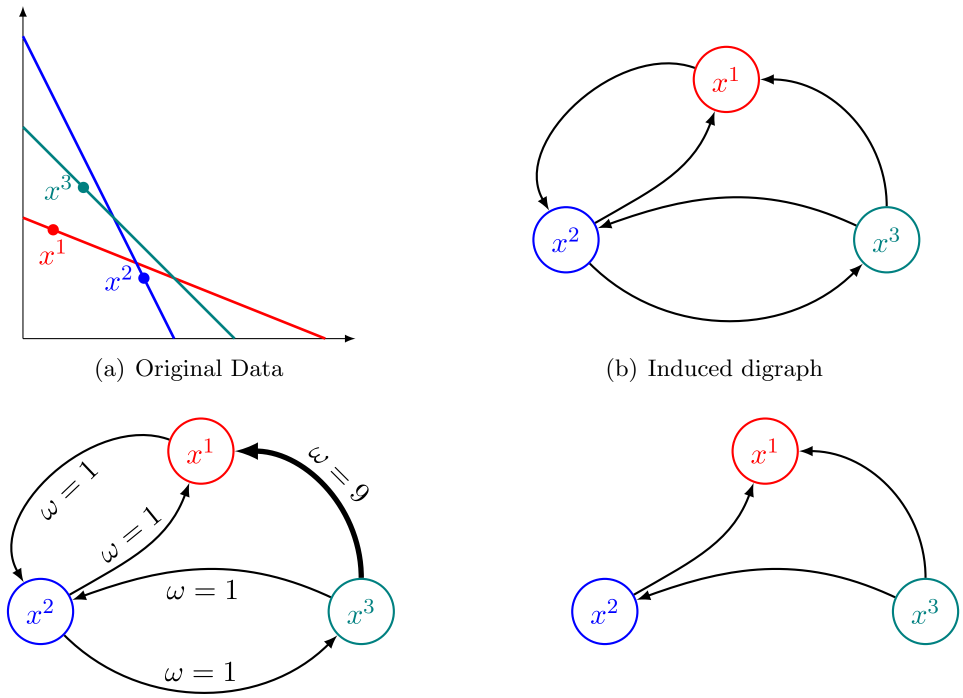
**Proposition 2.** *Let $\mathbf{x} = \{x^i : \nexists\, j < i$ such that $x^i \sim_\rhd x^j\}$, and*

$$\succ^D_\rhd = \{(x, y) \in \mathbf{x} \times \mathbf{x} : \exists\, i, j \in [N] \text{ such that } x^i \sim_\rhd x,\, x^j \sim_\rhd y,\, x^i \succ^D x^j, \text{ and } x^j \not\unrhd x^i\}.$$

*Define the graph $G = (\mathbf{x}, \succ^D_\rhd)$ and the weights $\Omega = (\omega_v)_{v\in\succ^D_\rhd}$ by $\omega_{(x,y)} = N^2$ if $x \rhd y$, and $\omega_{(x,y)} = \left|\{(w, z) \in \mathbf{x}_0 \times \mathbf{x}_0 : w \sim_\rhd x,\, z \sim_\rhd y, \text{ and } w \succ^D z\}\right|$ otherwise, where $\mathbf{x}_0 = \{x^i\}_{i\in[N]}$ is the set of observed choices. The preference $\succsim \in \mathcal{P}$ solves (2) if, and only if, $\succ^D_\rhd \setminus \succ$ solves the MFAS problem of $[G, \Omega]$.*

The proof of the previous result, in Appendix D combines $\rhd$-monotonicity and acyclicality of strict preferences with the rationalizability characterization in Nishimura et al. (2017).

Figure 1 presents an example of how to compute (2) for a data set of three observations ($N = 3$) in the classical consumer environment. Panel (a) shows the original data set. Panel (b) converts the data set to a digraph, where choices are nodes revealed preferences are vertices (an arrow going from $x$ to $y$ represents that $x$ is directly revealed strictly preferred to $y$). Panel (c) introduces the weights in the digraph; as $x^3 > x^1$ is the only relation according to $\geq$, the arrow from $x^3$ to $x^1$ has a weight of $N^2 = 9$; all the others have weight equal to 1. Finally, panel (d) shows the graph resulting from eliminating the cycles via the MFAS problem. By Proposition 2, we have that any preference satisfying $x^3 \succ x^2 \succ x^1$ solves (2). Finally, as two of the five vertices are eliminated by solving the MFAS problem, we have $\Delta(\mathcal{D}) = 2/5$.




(a) Original Data (b) Induced digraph

(c) Digraph with weights (d) Solution to MFAS problem

Figure 1: Computation of $\Delta(\mathcal{D})$ by solving the MFAS problem (with $\trianglerighteq \, = \geq$). Panel (a) shows original data; (b) shows digraph induced by data; (c) add weights to the digraph, where the edge $(x^3, x^1)$ has weight $N^2 = 9$ as $x^3 > x^1$; finally (d) shows solution to MFAS problem. Proposition 2 implies that any preference $\succsim$ satisfying $x^3 \succ x^2 \succ x^1$ solves (2), and $\Delta(\mathcal{D}) = 2/5$.

# 4 Preference Estimator

In this section, we propose an estimator of preferences based on our measure of decision-making quality.

## 4.1 Rationalization

If the data satisfies GARP, a preference *rationalizes* the data if every observed choice is optimal according to such preference. If that is the case, then the preference rationalizing the data will agree with all the revealed preferences. If choices and preferences are imperfectly aligned, GARP will not hold, and no preference relation agrees with all the revealed preferences. The starting point of our estimator is to identify which revealed preferences we interpret as incorrect.

In our distance between a data set and a preference, $d(\mathcal{D}, \succsim)$, the set $\succ^D \setminus \succ$ contains the directly revealed strict preferences in $\succ^D$ that $\succsim$ disagrees with. If $\succsim$ is interpreted as an estimator of the underlying preference, then the revealed preferences in $\succ^D \setminus \succ$ must be interpreted as incorrect. For lack of a better word, we refer to these discarded revealed preferences as *mistakes*.

Mistakes are revealed preferences in both $\succsim^D$ or $\succ^D$ with which a preference relation might disagree. The following definition specifies the requirements for these mistakes.

**Definition 5.** A *tuple of mistakes* $\mathcal{M}$ is a pair $(\mathcal{M}^w, \mathcal{M}^s)$, where $\mathcal{M}^w \subset \succsim^D$ and $\mathcal{M}^s \subset \succ^D$, satisfying the following characteristics

1. $x^i \trianglerighteq x^j$ implies $(x^i, x^j) \notin \mathcal{M}^w$ and $x^i \triangleright x^j$ implies $(x^i, x^j) \notin \mathcal{M}^s$;
2. if $x^i \succsim^D x^j$ and $x^j \triangleright x^i$ then $(x^i, x^j) \in \mathcal{M}^w$, and if $x^i \succ^D x^j$ and $x^j \trianglerighteq x^i$, then $(x^i, x^j) \in \mathcal{M}^s$; and
3. if $x^i \succ^D x^j$ and $(x^i, x^j) \in \mathcal{M}^w$, then $(x^i, x^j) \in \mathcal{M}^s$.

The tuple of mistakes is a collection of mistakes $\mathcal{M}^w$ taken from $\succsim^D$, and another one, $\mathcal{M}^s$, taken from $\succ^D$. The first requirement rules out the possibility of interpreting the relations derived directly from $\trianglerighteq$-monotonicity as mistakes. Similarly, the second requirement tells us that if a revealed preference contradicts $\trianglerighteq$-monotonicity, it must be interpreted as a mistake. Finally, the third requirement states that if we interpret a $x \succsim^D y$ as a mistake, then by definition of strict preferences, we should also interpret $x \succ^D y$ as a mistake (since $x \succ y$ implies $x \succsim y$); in other words, if we interpret a strict revealed preference as correct, then we should also interpret the corresponding weak preference as correct.

Starting from the directly revealed preferences, we want to recover a preference relation that agrees with all the revealed preferences that are not mistakes. This is our first rationalization requirement, which we denote as "discarding only" the mistakes $\mathcal{M}$.

**Definition 6.** Given choice data $\mathcal{D}$ and a tuple of mistakes $\mathcal{M}$, the preference relation $\succsim$ *discards only* $\mathcal{M}$ if $x \succsim y$ for all $(x, y) \in (\succsim^D \setminus \mathcal{M}^w)$, and $x \succ y$ for all $(x, y) \in (\succ^D \setminus \mathcal{M}^s)$.

Given a tuple of mistakes $\mathcal{M}$, for a preference to discard only $\mathcal{M}$ is a natural requirement. However, it is insufficient as a rationalization requirement as revealed preferences are only defined between observed choices.[14] Figure 2 presents a simple example of this insufficiency; even though the preference agrees with the only revealed preference $x^1 \succ^D x^2$, it does not rationalize the data as none of the choices are optimal from the budget sets.

To address the limitation of discarding only $\mathcal{M}$, we also require a preference estimator to satisfy *partial efficiency*. The idea of partial efficiency is proposed by Afriat (1973), who proposes to set a threshold $e$ (where $0 \leq e \leq 1$) and, with linear prices, to interpret a choice $x$ as preferred to another bundle $y$ only if $y$'s cost is less than $e$ times the cost of $x$ (at the prices at which $x$ was chosen). A data set is $e$-rationalizable by a preference if such preference satisfies this new comparison of bundles. Varian (1990) expands the previous requirement by using different levels of partial efficiency in different observations. Specifically, he proposes to take a vector $v \in [0, 1]^N$

[14] Usually, revealed preferences are defined between an observed choice and *any* bundle in the budget set from which it is chosen. However, in Definition 1, we define them only between chosen bundles as it allows us to develop our measure of decision-making quality (2). Furthermore, as GARP is a test that involves only observed choices, modifying the definition does not present any practical limitation.

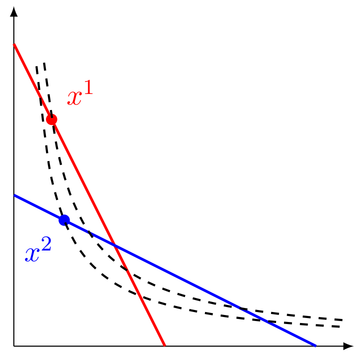



Figure 2: A preference that agrees with the revealed preferences but does not rationalize the data.

and to interpret a choice $x^i$ as preferred to a bundle $y$ only if its cost is less than $v_i$ (instead of one) times the cost of $x^i$. Then, a preference $v$-rationalizes the data if it agrees with all the revealed preferences given the vector $v$. We extend the $v$-rationalization requirement to our setting, where budget sets are not necessarily linear.

**Definition 7.** Given $v \in [0,1]^N$, $\mathcal{D}$ is *$v$-rationalizable* if there is a preference relation $\succsim$ such that $x^i \succsim x$ whenever $g^i(x) \leq v_i$. Such preference *$v$-rationalizes* the data.

In order to include both rationalization requirements, the vector $v$ must be small enough such that none of the revealed preferences interpreted as mistakes are included in the sets $\{x : g^i(x) \leq v_i\}$. This requirement is achieved with the following family of vectors.

**Definition 8.** Given a tuple of mistakes $\mathcal{M}$, an *$\mathcal{M}$-vector* $v^{\mathcal{M}}$ is any vector in $[0,1]^N$ satisfying the following properties

$$v_i^{\mathcal{M}} < g^i(x^j) \qquad \text{for all } (x^i, x^j) \in \mathcal{M}^w, \text{ and} \tag{4}$$

$$v_i^{\mathcal{M}} \leq g^i(x^j) \qquad \text{for all } (x^i, x^j) \in \mathcal{M}^s. \tag{5}$$

Condition (4) is motivated by the fact that if $x^i$ being preferred to $x^j$ is interpreted as a mistake then, from the vector $v$, we should not infer that $x^i$ is preferred to $x^j$; therefore we require $g^i(x^j) > v_i$. The same logic applies to (5).

A suitable preference estimator will discard only $\mathcal{M}$ and $v^{\mathcal{M}}$-rationalize the data. To develop a test for these conditions, we start by defining revealed preferences that consider the existence of mistakes.

**Definition 9.** Take a tuple of mistakes $\mathcal{M}$ and two choices $x^i, x^j$; $x^i$ is

- $\mathcal{M}$-*directly revealed preferred* to $x^j$ (denoted $x^i \succsim_{\mathcal{M}}^D x^j$) if $(x^i, x^j) \in (\succsim^D \setminus \mathcal{M}^w)$;
- $\mathcal{M}$-*directly revealed strictly preferred* to $x^j$ ($x^i \succ_{\mathcal{M}}^D x^j$) if $(x^i, x^j) \in (\succ^D \setminus \mathcal{M}^s)$;
- $\mathcal{M}$-*revealed preferred* to $x^j$ ($x^i \succsim_{\mathcal{M}}^R x^j$) if there is a sequence of choices $(x^{m_\ell})_{\ell=1}^L$ such that $x^i \succsim_{\mathcal{M}}^D x^{m_1} \succsim_{\mathcal{M}}^D x^{m_2} \succsim_{\mathcal{M}}^D \ldots \succsim_{\mathcal{M}}^D x^{m_L} \succsim_{\mathcal{M}}^D x^j$; and

- $\mathcal{M}$*-revealed strictly preferred* to $x^j$ ($x^i \succ^R_{\mathcal{M}} x^j$) if there are choices $x^m$, $x^{m'}$ such that $x^i \succsim^R_{\mathcal{M}} x^m \succ^D_{\mathcal{M}} x^{m'} \succsim^R_{\mathcal{M}} x^j$.

The previous definition removes the mistakes from the classical definition of revealed preferences; this is, it discards the revealed preferences in $\mathcal{M}$.

From Afriat Theorem and its generalization by Nishimura et al. (2017), we know that rationalization by a continuous and $\trianglerighteq$-monotone preference is equivalent to GARP. Here we extend the previous definition to include the existence of mistakes.

**Definition 10.** Given a tuple of mistakes $\mathcal{M}$, we say that choices satisfy the *Generalized Axiom of Revealed Preferences with Mistakes $\mathcal{M}$ (GARP$_{\mathcal{M}}$)* if for any $i, j \in [N]$

$$x^i \succsim^R_{\mathcal{M}} x^j \implies x^j \not\succ^D_{\mathcal{M}} x^i\,.$$

The definitions of $\mathcal{M}$-revealed preferences and GARP$_{\mathcal{M}}$ are the natural extensions to generate a (relaxed) test of consistency starting from the idea of mistakes. It is clear that the more elements we add to the tuple of mistakes, the less restrictive the test becomes.[15]

The following result shows that GARP$_{\mathcal{M}}$ is a necessary and sufficient test to find a preference relation that discards only $\mathcal{M}$ and $v^{\mathcal{M}}$-rationalizes the data.

**Theorem 3** (Afriat's Theorem with Mistakes)**.** *There is a continuous and $\triangleright$-monotone preference relation discarding only $\mathcal{M}$ and $v^{\mathcal{M}}$-rationalizing the data if, and only if, the data satisfies GARP$_{\mathcal{M}}$.*

The proof of the previous theorem is in Appendix E. It is based on the generalization of the Afriat Theorem by Nishimura et al. (2017). As usual in rationalization results, it implies that after satisfying the comparison between observed choices, which in our case is discarding only $\mathcal{M}$, the $v$-rationalization requirement has no empirical content.

### 4.2 Preference Estimator

A basic property of any estimator is to be consistent; this is, to asymptotically recover the true value of the underlying parameter. In the words of Newey and McFadden (1994, p. 2114),

> *an estimator that is not even consistent is usually considered inadequate.*

From Theorem 2, we know that any estimator that solves (2) will be consistent. We propose an estimator combining this idea with the rationalization requirements derived in the previous section.

Intuitively, we would like for an estimator to discard only a minimum tuple of mistakes $\mathcal{M}$, in the sense that if we remove elements from $\mathcal{M}$ then GARP$_{\mathcal{M}}$ fails. Moreover, we would like for the

[15]This is, starting from two tuple of mistakes $\mathcal{M} = (\mathcal{M}^w, \mathcal{M}^s)$ and $\mathcal{M}_0 = (\mathcal{M}^w_0, \mathcal{M}^s_0)$ such that $\mathcal{M}^w \supset \mathcal{M}^w_0$ and $\mathcal{M}^s \supset \mathcal{M}^s_0$, if the data satisfies GARP$_{\mathcal{M}_0}$ then it satisfies GARP$_{\mathcal{M}}$.

$v^{\mathcal{M}}$ vector to be "as big as possible." We propose to obtain a suitable tuple of mistakes through the following algorithm.

**Algorithm 1.** *Let $\widehat{\mathcal{M}} = (\widehat{\mathcal{M}}^w, \widehat{\mathcal{M}}^s)$ be the tuple of mistakes obtained from the following algorithm:*

1. *Let $E_0$ be a solution to the MFAS problem defined in Proposition 2. Set*

$$\widehat{\mathcal{M}}^s = \left\{(x,y) \in \succ^D \colon \exists\ (w,z) \in E_0 \ s.t.\ x \sim_{\trianglerighteq} w \text{ and } y \sim_{\trianglerighteq} z\right\} \bigcup \left\{(x,y) \in \succ^D \colon y \trianglerighteq x\right\} .$$

2. *Let $C = (\succ^D \setminus \widehat{\mathcal{M}}^s) \bigcup \{(x,y) \in \succsim^D \colon x \trianglerighteq y\}$, and $\widehat{\mathcal{M}}^w_0 = \succsim^D \setminus C$. Enumerate the elements in $\widehat{\mathcal{M}}^w_0$ from 1 to $T = |\widehat{\mathcal{M}}^w_0|$, i.e., $\widehat{\mathcal{M}}^w_0 = (m_t)_{t \in [T]}$. Starting at $t = 1$ and increasing $t$ by one until $t = T$ perform the following computation:*
   *2.1 Define $\widetilde{M}_t = \left(\widehat{\mathcal{M}}^w_{t-1} \setminus \{m_t\}, \widehat{\mathcal{M}}^s\right)$.*
   *2.2 If $\mathcal{D}$ satisfies $\mathrm{GARP}_{\widetilde{M}_t}$ set $\widehat{\mathcal{M}}^w_t = \widehat{\mathcal{M}}^w_{t-1} \setminus \{m_t\}$; if not, set $\widehat{\mathcal{M}}^w_t = \widehat{\mathcal{M}}^w_{t-1}$.*
   *Finally, set $\widehat{\mathcal{M}}^w = \widehat{\mathcal{M}}^w_T$.*

The first step of the previous definition assures that any estimator arising from $\widehat{\mathcal{M}}$ will be a solution to (2), and therefore is consistent. Step 2 identifies which elements from $\succsim^D$ should be interpreted as mistakes. First, the set $C$ takes all the elements from $\succsim^D$ that we have to interpret as correct (i.e., not as mistakes); these come from either $\trianglerighteq$-dominance or from such elements being interpreted as correct strong preferences (which imply weak preferences).[16] The set $\widehat{\mathcal{M}}^w_0$ includes all the elements of $\succsim^D$ that could potentially be interpreted as mistakes; of the elements in this set, the ones that violate $\trianglerighteq$ monotonicity are necessarily mistakes, but the rest (the ones in $E$) are not. Hence, we evaluate the elements in $E$ individually regarding whether they can be removed from the set of mistakes without violating GARP.

The next results show properties of $\widehat{\mathcal{M}}$ that suggest it is an adequate instrument to construct a preference estimator. Propositions 3 and 4 show that $\widehat{\mathcal{M}}$ satisfies minimum requirements for rationalization. Proposition 5 shows that, among the tuples satisfying those requirements, the one obtained from Algorithm 1 is minimum; this is, no element can be removed without creating violations of GARP. The proofs these results are presented in Appendices F, G, and H, respectively.

**Proposition 3.** *The tuple $\widehat{\mathcal{M}}$ obtained from Algorithm 1 is a tuple of mistakes, i.e., it satisfies the properties of Definition 5.*

**Proposition 4.** *Let $\widehat{\mathcal{M}}$ be the tuple of mistakes obtained from Algorithm 1. $\mathcal{D}$ satisfies $\mathrm{GARP}_{\widehat{\mathcal{M}}}$.*

**Proposition 5.** *Let $\widehat{\mathcal{M}} = (\widehat{\mathcal{M}}^w, \widehat{\mathcal{M}}^s)$ be the tuple of mistakes obtained from Algorithm 1. For any $\mathcal{M} = (\mathcal{M}^w, \mathcal{M}^s)$ such that $\mathcal{M}^w \subset \widehat{\mathcal{M}}^w$ and $\mathcal{M}^s \subset \widehat{\mathcal{M}}^s$, with at least one inclusion being proper, $\mathcal{D}$ fails $\mathrm{GARP}_{\mathcal{M}}$.*

We propose an estimator based on the minimum mistakes tuple. Although we would also like to set a specific $v^{\widehat{\mathcal{M}}}$ vector for our estimator, the strict inequality in (4) implies that the biggest $v^{\widehat{\mathcal{M}}}$

[16]Note that conditions 3 in Definition 5 implies that for any tuple of mistakes $\mathcal{M} = (\mathcal{M}^w, \mathcal{M}^s)$, if $x^i \succ^D x^j$ and $(x^i, x^j) \notin \mathcal{M}^s$ then $(x^i, x^j) \notin \mathcal{M}^w$.

vector might not exist. Hence, we leave the choice of a specific vector open. In the Section 5, we propose a method to choose a suitable $v^{\widehat{\mathcal{M}}}$ vector, which we use for our empirical implementation.

**Definition 11.** Take the minimum mistakes $\widehat{\mathcal{M}}$ from Algorithm 1, and a $v^{\widehat{\mathcal{M}}}$ an $\widehat{\mathcal{M}}$-vector from Definition 8. A *minimum mistakes (MM) estimator* is a preference relation $\succsim^{\widehat{\mathcal{M}}} \in \mathcal{P}$ that $v^{\widehat{\mathcal{M}}}$-rationalizes the data and discards only $\widehat{\mathcal{M}}$.

Proposition 4 and Theorem 3 imply that our estimator discards only $\widehat{\mathcal{M}}$ and $v^{\widehat{\mathcal{M}}}$-rationalizes $\mathcal{D}$, and Proposition 5 implies that the estimator presents a minimum number of disagreements with the revealed preferences. The final result of this section (proof in Appendix I) shows that the estimator is also consistent.

**Theorem 4.** *Let $\succsim^{\widehat{\mathcal{M}},N}$ be a MM estimator when $\mathcal{D}$ has $N$ observations. Then as $N \to \infty$,*

$$\succsim^{\widehat{\mathcal{M}},N} \xrightarrow{p} \succsim^{\star} .$$

# 5 Empirical Implementation

In this section, we implement the MM estimator into laboratory data from several sources. We specifically analyze how well our estimator predicts out-of-sample choices, a method popular in the machine learning literature and proposed by Fudenberg et al. (2022) to analyze economic models. We use the estimator derived from the Varian (1990) Index as a benchmark for comparison (Section 1 of the Online Appendix presents a formal definition of the Varian Index), which we refer to as the Varian estimator. The Varian Index is, within the existing measures of decision-making quality, the better suited to recover preferences (see Section III in Halevy et al., 2018, for a detailed discussion of this issue). We focus our analysis on the classical consumer environment with linear budget sets (a characteristic of the data we analyze) and with the "more is better" dominance relation ($\trianglerighteq = \geq$).

The main conceptual difference between the MM and the Varian estimators is their motivation. On the one hand, the Varian estimator is motivated by partial efficiency, which, although intuitively compelling, lacks the statistical properties usually required in the econometrics literature. On the other hand, the main property of the MM estimator is its statistical consistency (Theorem 4).

## 5.1 Data Description

We analyze the choices of 5,345 subjects from previous experimental studies. They all follow the design of Choi et al. (2007b); subjects graphically choose bundles of different goods from randomly generated linear budget sets. We include subjects making choices in settings with two goods (which we refer to as 2D environments) and three goods (3D environments). Our sample includes subjects making choices in different environments: choices under risk, choices under ambiguity, and social choices. In choices under risk, the different goods are Arrow securities for different states of the world, and the probability of each state is known; choices under ambiguity differ from choices under risk only in that not all the states have known probabilities; finally, in social choices subjects play the *dictator game*, choosing between own consumption and consumption of one (in 2D) or two (in 3D) anonymous players. When possible, we also split our samples between sub-samples representative of the general population and sub-samples including only undergraduate or graduate students. Our sub-samples are:

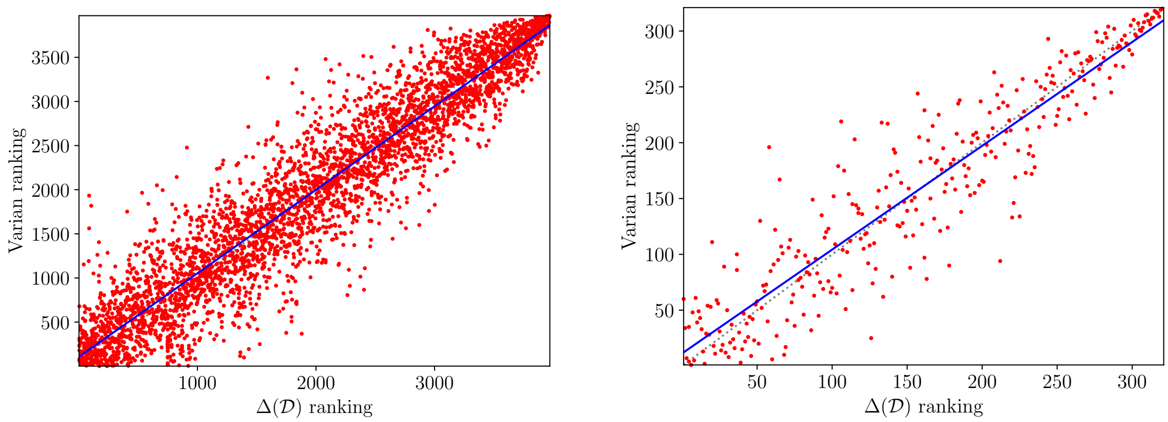
- 2D-Risk-General: 1,182 subjects, taken from Choi et al. (2014). The sample is representative of the Dutch-speaking population in the Netherlands. All subjects face a symmetric environment, where both states have probability 1/2.
- 2D-Risk-Students: 1,020 undergraduate students. 974 subjects make choices in a symmetric environment, and the remaining 46 subjects face an asymmetric one with probabilities 2/3 and 1/3. The symmetric data is taken from Choi et al. (2007a), Zame et al. (2020), Cappelen et al. (2021), and Dembo et al. (2021), and the asymmetric one from Choi et al. (2007a).
- 2D-Social-General: 1,698 subjects, taken from Fisman et al. (2017), Fisman et al. (2022). Both experiments are embedded in the American Life Panel, an internet survey administered by the RAND Corporation to adult Americans.
- 2D-Social-Students: 1,058 students, taken from Fisman et al. (2007), Fisman, Jakiela, Kariv, and Markovits (2015), Fisman, Jakiela, and Kariv (2015), and Li et al. (2017). Subjects are undergraduate students (Fisman et al., 2007; Fisman, Jakiela, & Kariv, 2015), Law students (Fisman, Jakiela, Kariv, & Markovits, 2015), and medical students (Li et al., 2017).
- 3D-Risk: 168 undergraduate students, taken from Dembo et al. (2021). All three states of the world have the same probability 1/3.
- 3D-Ambiguity: 154 undergraduate students, taken from Ahn et al. (2014). Subjects knew that one state of the world is 1/3 but did not know the probabilities of the other two (except that they are positive and add up to 2/3).
- 3D-Social: 65 undergraduate students, taken from Fisman et al. (2007).

All subjects make 50 choices each except the ones in the 2D-Risk-General sub-sample, who make 25 choices each.

Since the $\Delta(\mathcal{D})$ and the Varian Index measure different objects, they cannot be directly compared. Instead, Figure 3 compares how they rank subjects who fail GARP, splitting between the 2D and the 3D samples.[17] Although both measures rank most of the subjects similarly, there are more subjects with a (relative) high value of the Varian Index and a (relative) low value of $\Delta(\mathcal{D})$ than the opposite (this is, there are more subjects near the southeast end of the plot than near the northwest one): the slope of the blue line (which represents a linear regression) is less than one.[18] The reason is that while $\Delta(\mathcal{D})$ only penalizes subjects by the number of violations, i.e., the number of revealed preferences that need to be removed, the Varian Index measures its intensity

[17]Figure 3 of the Online Appendix presents comparisons for each sub-sample; all present the same pattern.
[18]The 99% confidence intervals are [.933, .960] for the 2D data and [.876, .983] for the 3D data.

(by how much does the income needs to be reduced in order to eliminate the revealed preference). Thus there are subjects for whom few revealed preferences need to be removed. Hence their $\Delta(\mathcal{D})$ ranking is low, but the Varian Index associates a high cost to remove them.




(a) 2D

(b) 3D

Figure 3: Ordering of subjects according to $\Delta(\mathcal{D})$ and Varian Index. Linear regression in blue; gray dotted line is 45° line.

### 5.2 Out-of-sample prediction

For any finite data set, any tuple of mistakes $\mathcal{M}$ (such that $\text{GARP}_{\mathcal{M}}$ holds) is consistent not with one but with a set of estimators, and the same is true for the Varian estimator. Hence, we measure the quality of the recovered preferences by their out-of-sample accuracy, i.e., by how consistent the set of estimators is with choices that were not included in the original estimation. For this purpose, we split each subject observations into two groups, using the first (the train data) to estimate preferences and the second (the test data) to check whether the observed choices are consistent with some preference derived from the estimators. This approach is based on Fudenberg et al. (2022), who motivate it in standard machine-learning practices. We perform our analysis using ten observations for testing and the rest for training (i.e., 40 choices for training, except for the 2D-Risk-General sub-sample, which uses 15 choices for training). In each case, choices were chosen randomly. We start by measuring accuracy, denoted $a_{\text{test}}$, as the share of choices in the test data consistent with some preference estimator.

Since both estimators provide a set set of possible preferences, not only one, one may generate a "tighter bound" on the set of preferences than the other. For example, suppose the set of preferences from one estimator is always included in the set of preferences of the other. Then, the second one will automatically have a higher accuracy than the first (in the limit case, an estimator that never discards any preference will always have a 100% accuracy). Following Fudenberg et al. (2022), we compare each estimator's completeness, which is a measure that corrects accuracy by how tight the bounds of the estimator are. To measure completeness, we first generate 1,000 uniform random choices from each budget set in the test data, then compute the accuracy of the random choices, denoted $a_{\text{random}}$. Completeness is measured as

$$\frac{a_{\text{test}} - a_{\text{random}}}{1 - a_{\text{random}}} .$$

As a robustness exercise, Table 2 in the Online Appendix includes the same analysis using five observations for testing. The results are qualitatively similar for both sizes of test data.

### 5.2.1 Characterization

Starting from a tuple of mistakes $\mathcal{M}$ satisfying GARP$_{\mathcal{M}}$, we want to know whether a new observation is consistent with some preference that discards only $\mathcal{M}$ and $v^{\mathcal{M}}$ rationalizes the data. For this purpose, we denote the set of all such preferences by

$$\mathcal{R}(\mathcal{M}) = \left\{ \succsim \in \mathcal{P} : \succsim \ v^{\mathcal{M}}\text{-rationalizes the data and discards only } \mathcal{M} \right\} .$$

Suppose we observe a new budget set $\{x : g(x) \leq 1\}$ that satisfies Assumption 1 (this is, $g$ is continuous, increasing, and homogeneous of degree one). We are interested in testing whether the choice from this budget set is consistent with *some* preference in $\mathcal{R}(\mathcal{M})$. We denote the set of all such choices by $C_{\mathcal{M}}(B)$:

$$C_{\mathcal{M}}(g) = \{x \in g(x) \leq 1 : \text{ there is } \succsim \in \mathcal{R}(\mathcal{M}) \text{ for which } x \succsim y \text{ whenever } g(y) \leq 1\} .$$

*Remark* 2. If $\mathcal{D}$ satisfies GARP$_{\mathcal{M}}$ then $C_{\mathcal{M}}(g) \neq \emptyset$ for any budget set $\{x : g(x) \leq 1\}$ satisfying Assumption 1.

To characterize the set $C_{\mathcal{M}}(g)$, we follow the approach in Varian (1982) and define a partial order between the new budget set and the ones observed in the data.

**Definition 12.** Given choice data $\mathcal{D}$, a tuple of mistakes $\mathcal{M}$ and a budget set generated by $g$, we say that $g$ is

1. $\mathcal{M}$-*directly revealed preferred* to $g^i$, denoted $g \gtrsim^D_{\mathcal{M}} g^i$, if there is $x$ such that $g(x) = 1$ and $x \trianglerighteq x^i$;
2. $\mathcal{M}$-*directly revealed strictly preferred* to $g^i$, denoted $g >^D_{\mathcal{M}} g^i$, if there is $x$ such that $g(x) = 1$ and $x \triangleright x^i$;
3. $\mathcal{M}$-*revealed preferred* to $g^i$, denoted $g >^R_{\mathcal{M}} g^i$, if there exist an observation $x^j \in [N]$ such that $g \gtrsim^D g^j$ and $x^j \succsim^R_{\mathcal{M}} x^i$; and
4. $\mathcal{M}$-*revealed strictly preferred* to $g^i$, denoted $g >^R_{\mathcal{M}} g^i$, if there exist an observation $x^j \in [N]$ such that either $g >^D_{\mathcal{M}} g^j$ and $x^j \succsim^R_{\mathcal{M}} x^i$, or $g \gtrsim^D_{\mathcal{M}} g^j$ and $x^j \succ^R_{\mathcal{M}} x^i$.

The idea of directly revealed preferences in the previous definition is as follows. If $x^i$ is chosen from $g^i$ and $g$ contains a bundle that is at least as good as $x^i$, then the agent (weakly) prefers to choose from $g$ rather than from $g^i$. Similarly, if $g$ contains a bundle that is strictly better than $x^i$, the agent prefers to choose from $g$ over $g^i$. Finally, if there is a bundle in $g$ that is strictly better than another choice $x^j$ and we have inferred (through $\mathcal{M}$ revealed preferences) that $x^j$ is preferred to $x^i$, then we also infer that the agent prefers to choose from $g$ than from $g^i$. Following Fact 8 in Varian (1982), we obtain the following characterization of $C_{\mathcal{M}}$.

**Proposition 6.** *Suppose $\mathcal{D}$ satisfies GARP$_{\mathcal{M}}$. For any $\mathcal{M}$-vector $v^{\mathcal{M}}$ and budget set generated by $g$, a bundle $x$ is in the set $C_{\mathcal{M}}(g)$ if, and only if, the following conditions hold:*

$$\{y : g(y) = 1 \text{ and } y \rhd x\} = \emptyset \tag{6}$$

$$x^i \not\rhd x \text{ and } \{y : g^i(y) = v_i^{\mathcal{M}} \text{ and } y \rhd x\} = \emptyset \qquad \text{whenever } g \succsim_{\mathcal{M}}^{R} g^i, \text{ and} \tag{7}$$

$$x^i \not\trianglerighteq x \text{ and } \{y : g^i(y) = v_i^{\mathcal{M}} \text{ and } y \trianglerighteq x\} = \emptyset \qquad \text{whenever } g \succ_{\mathcal{M}}^{R} g^i\,. \tag{8}$$

Intuitively, this characterization requires the new choice not to add more violations of GARP than the ones that $\mathcal{M}$ already removed. The proof of this result is in Section 3.2 of the Supplemental Material. Additionally, Section 1 of the Online Appendix describes how to apply the previous result to compute out-of-sample predictions for the Varian estimator.

In the classical consumer environment (where $\trianglerighteq\, =\geq$ and $g^i(y) = p^i \cdot y$) the characterization in Proposition 6 gets a simpler form. If $g(y) = p \cdot y$ and $\trianglerighteq\, =\geq$, then the requirement in (6) can be restated as $p \cdot x = 1$. Similarly, the requirement $\{y : g^i(y) = v_i^{\mathcal{M}}$ and $y \rhd x\} = \emptyset$ in (7) is equivalent to $p^i \cdot x \geq v_i^{\mathcal{M}}$, and the requirement $\{y : g^i(y) = v_i^{\mathcal{M}}$ and $y \trianglerighteq x\} = \emptyset$ in (8) is equivalent to $p^i \cdot x > v^i$ .

#### 5.2.2 Results

Before checking whether choices in the test data are consistent with the estimator obtained from the train data, we need to define the vector $v$ for which the preference relation $v$-rationalizes the test data. In the case of the MM Index, Definition 8 implies that for any tuple of mistakes $\mathcal{M}$ there is a continuous of possible $\mathcal{M}$-vectors. Specifically, as (5) is a strict inequality, there is no "biggest" $\mathcal{M}$-vector. We overcome this by finding the biggest vector that satisfies (4) (which exists as this condition is a weak inequality) and then multiplying each component of the vector that is less than one by a factor of .999. In the case of the Varian Index, if $\mathcal{D}$ fails GARP, then the data is not $v$-rationalizable using the vector recovered by the Varian Index (see Ugarte, 2023). We overcome this by multiplying each component less than one of such vector by the same factor .999.

Table 1 presents the average accuracy and completeness for the MM and the Varian estimators and the average difference between them. Accuracy results are mixed for the 2D data. There are no significant differences between indices for the Risk - General and Social - Students sub-samples (the $t$-statistics of the differences are .50 and -1.67, respectively), the MM estimator has a higher accuracy for the Risk - Students sub-sample ($t = 7.02$), and the Varian estimator has higher accuracy for the Social - General sub-sample ($t = -3.04$). Furthermore, the average difference for the complete 2D sample is not statistically significant ($t = 1.64$). For the 3D data, the MM estimator presents a higher accuracy than the Varian one for all sub-samples ($t$-statistics are 5.55, 7.08, and 5.27 for the Risk, Ambiguity, and Social sub-samples, respectively, and 10.10 for the complete sample).

Table 1: Out-of-sample prediction (Test: 10 choices)

| Sample | | # Sub | Accuracy MM | Accuracy Varian | Accuracy Diff | Completeness MM | Completeness Varian | Completeness Diff |
|---|---|---|---|---|---|---|---|---|
| 2D | Risk - General | 951 | 0.766<br>(0.006) | 0.765<br>(0.006) | 0.001<br>(0.002) | 0.344<br>(0.015) | 0.352<br>(0.014) | -0.008<br>(0.004) |
| | Risk - Students | 762 | 0.814<br>(0.006) | 0.790<br>(0.008) | 0.025<br>(0.004) | 0.619<br>(0.012) | 0.608<br>(0.012) | 0.011<br>(0.005) |
| | Social - General | 1549 | 0.704<br>(0.005) | 0.710<br>(0.005) | -0.006<br>(0.002) | 0.438<br>(0.010) | 0.457<br>(0.009) | -0.019<br>(0.004) |
| | Social - Students | 709 | 0.794<br>(0.007) | 0.798<br>(0.007) | -0.004<br>(0.002) | 0.597<br>(0.014) | 0.612<br>(0.013) | -0.015<br>(0.005) |
| | ALL | 3971 | 0.756<br>(0.003) | 0.754<br>(0.003) | 0.002<br>(0.001) | 0.478<br>(0.006) | 0.489<br>(0.006) | -0.010<br>(0.002) |
| 3D | Risk | 141 | 0.809<br>(0.012) | 0.760<br>(0.017) | 0.050<br>(0.009) | 0.592<br>(0.024) | 0.550<br>(0.025) | 0.042<br>(0.012) |
| | Ambiguity | 134 | 0.805<br>(0.014) | 0.750<br>(0.017) | 0.055<br>(0.008) | 0.631<br>(0.024) | 0.562<br>(0.025) | 0.069<br>(0.013) |
| | Social | 45 | 0.704<br>(0.025) | 0.620<br>(0.032) | 0.084<br>(0.016) | 0.536<br>(0.040) | 0.456<br>(0.041) | 0.079<br>(0.018) |
| | ALL | 320 | 0.793<br>(0.009) | 0.736<br>(0.012) | 0.057<br>(0.006) | 0.600<br>(0.016) | 0.542<br>(0.016) | 0.059<br>(0.008) |

Out-of-sample average accuracy and completeness for each sub-sample, including only subjects who fail GARP. standard errors in parenthesis. Diff is difference between MM Varian estimators, D is number of dimensions, and # Sub. is number of subjects. Test data is 10 observations, and the remaining are used for training.

Completeness results are different from Accuracy ones for the 2D data: the Varian Index estimator performs better in the Risk - General, Social - General, and Social - Students sub-samples ($t$-statistics are -1.87, -5.43, and -2.87, respectively), while the MM estimator does better in the Risk - Students one ($t = 2.06$). Overall, for the 2D data, the Varian estimator presents a higher average completeness ($t = -4.59$). For the 3D data, the completeness results are similar to the ones for accuracy: the MM estimator presents higher average completeness in all sub-samples ($t$-statistics are 3.50, 5.29, and 4.33 for the Risk, Ambiguity, and Social sub-samples, respectively, and 7.25 for the complete sample).

The fact that the Varian estimator has higher average completeness for 2D data and the MM estimator performs better in the 3D data paints a compelling picture. One possible explanation for this pattern is that the Varian estimator is motivated by intuitive rather than formal arguments, and most intuitive explanations and examples are constructed on 2D data. Hence, intuitive arguments might be more likely to work in choice environments closer to those used in such examples. On the other hand, formal properties such as the consistency of the MM estimator (Theorem 4) do not rely on the simplicity or intuitiveness of the choice environment. Hence, as the environment becomes more complex, we expect formal properties to generate more suitable estimators.

## 6 Concluding Remarks

If a subject's choices are an imperfect implementation of her preference, any measure of decision-making quality should measure the distance between the subject's choices and her preference. Hence, measuring the distance between the observed choices and GARP might be insufficient, as the distance is measured according to some preference but not necessarily the subject's. In this paper, we propose a new measure of decision-making quality that approaches the distance between the subject's choices and underlying preference relation as the sample size increases.

Since preferences are not directly observable, the question about measuring the distance between an agent's choices and her underlying preference is inevitably related to learning about the agent's preference from her choices. We use our measure of decision-making quality to propose a new preference estimator, which, to the best of our knowledge, is the first statistically consistent estimator for the choice environment we study (which includes the traditional consumer environment as a particular case).

We implement our preference estimator on experimental data by analyzing the quality of its out-of-sample prediction. We compare our results to the estimator derived from the Varian (1990) Index. Our results suggest that intuitive arguments may be well-suited to recover preferences in the simple choice environments where those arguments are motivated. However, formal properties, such as the statistical consistency of the proposed estimator, become more relevant as the complexity of the choice environment increases.

## References

Afriat, S. N. (1967). The Construction of Utility Functions from Expenditure Data. *International Economic Review*, *8*(1), 67.

Afriat, S. N. (1973). On a System of Inequalities in Demand Analysis: An Extension of the Classical Method. *International Economic Review*, *14*(2), 460–472.

Ahn, D., Choi, S., Gale, D., & Kariv, S. (2014). Estimating ambiguity aversion in a portfolio choice experiment. *Quantitative Economics*, *5*(2), 195–223.

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press.

Antonelli, G. B. (1886). Sulla teoria matematica della economia politica. *Pisa: Nella tipografia del Folchetto*.

Apesteguia, J., & Ballester, M. A. (2015). A Measure of Rationality and Welfare. *Journal of Political Economy*, *123*(6), 1278–1310.

Bernheim, B. D., & Rangel, A. (2009). Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *Quarterly Journal of Economics*, *124*(1), 51–104.

Border, K. C., & Segal, U. (1994). Dynamic Consistency Implies Approximately Expected Utility Preferences. *Journal of Economic Theory*, *63*(2), 170–188.

Cappelen, A. W., Kariv, S., Sørensen, E. Ø., & Tungodden, B. (2021). *The Development Gap in Economic Rationality of Future Elites*, Working Paper.

Chambers, C. P., Echenique, F., & Lambert, N. (2021). Recovering Preferences from Finite Data. *Econometrica*, *89*(4), 1633–1664.

Choi, S., Fisman, R., Gale, D., & Kariv, S. (2007a). Consistency and Heterogeneity of Individual Behavior under Uncertainty. *American Economic Review*, *97*(5), 1921–1938.

Choi, S., Fisman, R., Gale, D., & Kariv, S. (2007b). Revealing Preferences Graphically: An Old Method Gets a New Tool Kit. *American Economic Review: papers and proceeding*, *97*(2), 153–158.

Choi, S., Kariv, S., Müller, W., & Silverman, D. (2014). Who Is (More) Rational? *American Economic Review*, *104*(6), 1518–1550.

Dean, M., & Martin, D. (2016). Measuring Rationality with the Minimum Cost of Revealed Preference Violations. *Review of Economics and Statistics*, *98*(3), 524–534.

de Clippel, G., & Rozen, K. (2021). Relaxed Optimization: How Close is a Consumer to Satisfying First-Order Conditions? *Review of Economics and Statistics*.

Dembo, A., Kariv, S., Polisson, M., & Quah, J. (2021). *Ever Since Allais*, Working Paper.

Echenique, F., Imai, T., & Saito, K. (2022). Approximate Expected Utility Rationalization. *Journal of the European Economic Association*.

Echenique, F., Lee, S., & Shum, M. (2011). The Money Pump as a Measure of Revealed Preference Violations. *Journal of Political Economy*, *119*(6), 1201–1223.

Fisman, R., Jakiela, P., & Kariv, S. (2015). How did distributional preferences change during the Great Recession? *Journal of Public Economics*, *128*.

Fisman, R., Jakiela, P., & Kariv, S. (2017). Distributional preferences and political behavior. *Journal of Public Economics*, *155*.

Fisman, R., Jakiela, P., Kariv, S., & Markovits, D. (2015). The distributional preferences of an elite. *Science*, *349*(6254).

Fisman, R., Jakiela, P., Kariv, S., & Vannutelli, S. (2022). The Distributional Preferences of Americans, 2013-2016. *Experimental Economics (forthcoming)*.

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, *97*(5).

Forges, F., & Minelli, E. (2009). Afriat's theorem for general budget sets. *Journal of Economic Theory*, *144*(1), 135–145.

Fudenberg, D., Kleinberg, J., Liang, A., & Mullainathan, S. (2022). Measuring the Completeness of Economic Models. *Journal of Political Economy*, *130*(4).

Grodal, B. (1974). A note on the space of preference relations. *Journal of Mathematical Economics*, *1*(3), 279–294.

Halevy, Y., Persitz, D., & Zrill, L. (2018). Parametric Recoverability of Preferences. *Journal of Political Economy*, *126*(4), 1558–1593.

Hicks, J. R. (1956). *A Revision of Demand Theory.* Claredon Press.

Hildenbrand, W. (1970). On economies with many agents. *Journal of Economic Theory*, *2*, 161–188.

Hildenbrand, W. (1974). *Core and Equilibria of a Large Economy. (PSME-5).* Princeton University Press.

Hoeffding, W. (1961). The strong law of large numbers for U-statistics. *Institute of Statistics Mimeo Series*, *302*(302).

Houthakker, H. S. (1950). Revealed Preference and the Utility Function. *Economica*, *17*(66), 159.

Houtman, M., & Maks, J. (1985). Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve methoden*, *19*, 89–104.

Jennrich, R. I. (1969). Asymptotic Properties of Non-Linear Least Squares Estimators. *The Annals of Mathematical Statistics*.

Kannai, Y. (1970). Continuity Properties of the Core of a Market. *Econometrica*, *38*(6), 791–815.

Li, J., Dow, W. H., & Kariv, S. (2017). Social preferences of future physicians. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48).

Mas-Colell, A. (1977). The Recoverability of Consumers' Preferences from Market Demand Behavior. *Econometrica*, *45*(6), 1409.

Mas-Colell, A. (1978). On Revealed Preference Analysis. *The Review of Economic Studies*, *45*(1), 121–131.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics* (pp. 2111–2245). Elsevier.

Nishimura, H., Ok, E. A., & Quah, J. (2017). A Comprehensive Approach to Revealed Preference Theory. *American Economic Review*, *107*(4), 1239–1263.

Rader, T. (1972). *Theory of Microeconomics* (1st). Academic Press.

Redekop, J. (1993). The questionnaire topology on some spaces of economic preferences. *Journal of Mathematical Economics*, *22*(5), 479–494.

Reny, P. J. (2015). A Characterization of Rationalizable Consumer Behavior. *Econometrica*, *83*(1), 175–192.

Samuelson, P. A. (1938). The Empirical Implications of Utility Analysis. *Econometrica*, *6*(4).

Ugarte, C. (2023). *The generality of the Strong Axiom*, Working Paper.

Varian, H. R. (1982). The Nonparametric Approach to Demand Analysis. *Econometrica*, *50*(4), 945.

Varian, H. R. (1990). Goodness-of-fit in optimizing models. *Journal of Econometrics*, *46*(1-2), 125–140.

Zame, W. R., Tungodden, B., Sørensen, E. Ø., Kariv, S., & Cappelen, A. W. (2020). *Linking Social and Personal Preferences: Theory and Experiment*, Working Paper.

# A Proof of Theorem 1

In order to proof Theorem 1 we adapt results from Mas-Colell (1977).

**Definition 13.** A function $u : \mathbb{R}^K_+ \to \mathbb{R}$ is *Lipschitzian* if, for every $r > 0$ there is a number $M > 0$ such that $|f(x) - f(y)| \leq M\, ||x - y||$ for every $x, y \in K_r$. It is $\trianglerighteq$*-regular* if for every $r > 0$ there is $\delta > 0$ such that $u(x) - u(y) \geq \delta ||x - y||$ whenever $x, y \in K_r$ and $x \rhd y$.

A set of function $\mathcal{U}$ is *uniformly Lipschitzian* if for every $r > 0$ the functions $u \in \mathcal{U}$ admit the same Lipschitz constant $M$. It is *uniformly regular* if for every if for every $r > 0$ the functions $u \in \mathcal{U}$ admit the same constant $\delta$.

Mas-Colell (1977) focuses on functions that are strictly $\geq$-increasing. Here we adapt his definition of regularity for a general dominance relation, and call it $\trianglerighteq$-regular. In Theorems 1 and 1', Mas-Colell (1977) shows the equivalence between Lipschitzian preferences and Lipschitzian and regular utility functions. Although his focus is on $\geq$-increasing preferences, there is nothing in his proof that uses the specific structure of $\geq$ instead of a generic dominance relation $\trianglerighteq$.[19] Regarding the assumption of convexity of preferences, Mas-Colell (1977) uses it only in the proof of his Lemma 4; below we provide a proof (which relies on two auxiliary results) that does not rely on the preferences being convex. For any preference $\succsim$ and bundle $x$, let $I^{\succsim}(x)$ be the indifference set of $x$.

**Lemma 1.** *If $\succsim$ is Lipschitzian and $x \succ y$ then $\delta(I^{\succsim}(x), I^{\succsim}(y)) > 0$.*

*Proof.* Towards a contradiction suppose $\delta(I^{\succsim}(x), I^{\succsim}(y)) = 0$. Take $x' \in I^{\succsim}(x)$ and $y' \in I^{\succsim}(y)$ close enough such that $||x' - y'|| < \varepsilon$. As $x \sim x' \succ y \sim y'$ we have $\delta(y', P(x)) = d(y', I(x)) > 0$. Hence $H||x'' - y''|| \geq \delta(x', I(y)) > 0$ for all $x'' \in I(x)$ and $y'' \in I(y)$. Therefore $\delta(I(x), I(y)) \geq H\delta(x', I(y)) > 0$, a contradiction. □

**Lemma 2.** *If $\succsim$ is continuous and $x \succsim y \succsim z$ then $\delta(I^{\succsim}(x), I^{\succsim}(z)) \geq \delta(I^{\succsim}(x), I^{\succsim}(y)) + \delta(I^{\succsim}(y), I^{\succsim}(z))$.*

*Proof.* Take $\eta > 0$, $x' \in I(x)$ and $z' \in I(z)$ such that $||x' - z'|| < d(I^{\succsim}(x), I^{\succsim}(z)) + \eta$, and $y' \in \langle x', z' \rangle$ such that $y' \in I(y)$, which exists by continuity of $\succsim$. Then

$$d(I^{\succsim}(x), I^{\succsim}(y)) + d(I^{\succsim}(y), I^{\succsim}(z)) \leq ||x' - y'|| + ||y' - z'||$$

[19] In particular, the only change we need to make in his proof is to replace "$x > y$" for $x \rhd y$ in the statement of Lemma 5.

$$
\begin{aligned}
&= ||x' - z'|| \\
&< d(I^{\succsim}(x), I^{\succsim}(z)) + \eta \,.
\end{aligned}
$$

The result follows as $\eta$ is arbitrary. $\square$

**Lemma 3** (Mas-Colell, 1977, Lemma 4)**.** *Let $\succsim$ be Lipschitzian and strictly increasing, $r > 0$, and $x^n, y^n \in K_r$ for all $n$. If $x^n \sim y^n$ and $x^n \neq y^n$ for all $n$, then the sequence $z^n = ||x^n - y^n||^{-1}(x^n - y^n)$ does not have an accumulation point in $\mathbb{R}^K_+$.*

*Proof.* Towards a contradiction suppose $z^n$ has an accumulation point $z$ in $\mathbb{R}^K_+$. By construction we have $||z^n|| = 1$ for all $n$, therefore $||z|| = 1$. As $z \in \mathbb{R}^K_+$, this implies $z > \mathbf{0}$. Since $K_r$ is compact, and up to subsequences if necessary, let $x^n \to x$ and $y^n \to y$. Since $z > \mathbf{0}$ we have $x > y$, and therefore $x \succ y$; let $C = \delta(I^{\succsim}(x), I^{\succsim}(y)) > 0$ (Lemma 1).

We analyze three different exhaustive cases, showing that all of them are impossible.

- If $x \succ x^n \succ y$, let $w^n$ be the unique element in the segment $\langle x, y \rangle$ satisfying $w^n \sim x^n$. Up to a sub sequence if necessary, let $w^n \to w$. There are two possible subcases:
  - $w = x$, let $z = {}^{w+y}/_2$; then for $n$ large enough $w^n \succ z \succ y$. Lemma 2 implies $\delta(I(w^n), I(y)) \geq \delta(I(w^n), I(z)) + \delta(I(z), I(y)) > 0$. Also, for all $\eta > 0$ and $n$ large enough we have $\delta(I(w^n), I(y)) \leq \delta(y^n, I(y)) < \eta$, a contradiction.
  - $w \neq x$, let $z = {}^{x+w}/_2$; then for $n$ large enough $x \succ z \succ w^n$ and, by Lemma 2, $\delta(I(x), I(w^n)) \geq \delta(I(x), I(z)) + \delta(I(z), I(w^n)) > 0$. Also, for all $\eta > 0$ and $n$ large enough we have $\delta(I(x), I(w^n)) \leq \delta(I(x), I(x^n)) < \eta$, a contradiction.
- If $x^n \succsim x \succ y$, then, by Lemma 2, $\delta(I(x^n), I(y)) \geq \delta(I(x^n), I(x)) + d(I(x), Y(y)) \geq C > 0$. For all $n$ sufficiently large we have $\delta(I(x^n), I(y)) \leq ||y^n - y|| < C$, a contradiction.
- If $x \succ y \succsim x^n$, then, by Lemma 2, $\delta(I(x), I(x^n)) \geq \delta(I(x), I(y)) + \delta(I(y), I(x^n)) \geq C > 0$. For all $n$ sufficiently large we have $\delta(I(x^n), I(x)) \leq ||x - x^n|| < C$, a contradiction.

$\square$

Now we can directly adapt Theorem 1' in Mas-Colell (1977).

**Theorem 5** (Generalization of Theorem 1' in Mas-Colell, 1977)**.** *Let $\Phi$ be the mapping that assigns to every utility function the preference relation it represents. A set $\mathcal{P}_0$ of $\rhd$-increasing preferences is uniformly Lipschitzian if, and only if, there exists a set of uniformly Lipschitzian and uniformly $\unrhd$-regular utility functions $\mathcal{U}_0$ such that $\Phi(\mathcal{U}_0) = \mathcal{P}_0$.*

*Proof of Theorem 1.* Let $\mathcal{P}$ be a uniformly Lipschitzian set of continuous and $\unrhd$-increasing preferences. For every preference $\succsim \in \mathcal{P}_0$ let $u_{\succsim}$ be the unique utility that represents $\succsim$ and satisfies $u_{\succsim}(x) = \alpha$ for every $x$ of the form $x = \alpha\mathbf{1}$, and let $\mathcal{U} = \{u_{\succsim} : \succsim \in \mathcal{P}\}$. By Theorem 8 in Border and

Segal (1994) we have $\succsim^n \to \succsim$ if $u_{\succsim^n} \to u_{\succsim}$ in the topology of compact convergence.[20] By Theorem 5 the set $\mathcal{U}$ is uniformly Lipschitzian and uniformly $\trianglerighteq$-regular, and by the Arzela-Ascoli theorem, $\mathcal{U}$ is compact (see Section 6.4 in Chambers et al., 2021). Hence, by Proposition 1 in Chambers et al. (2021), $\mathcal{P}$ is compact. □

# B Proof of Proposition 1

*Proof of Proposition 1.* For sufficiency suppose Assumption 4 holds.

1. Take $A \subset \mathbb{R}^K_+$ open and $x \in A$. Then the sets $B = \{y \in A : y \gg x\}$ and $C = \{y \in A : x \gg y\}$ are nonempty, open, and by $\trianglerighteq$-monotonicity of $\succsim^\star$ we have $B \succ^\star C$. Then

$$\mu_X(A) \geq \mu^2(x \in B,\, y \in C,\, x \succ^D y) > \mu^2(x \in B,\, y \in C,\, y \succsim^D x) \geq 0\,.$$

   The first inequality follows from $B, C \subset A$ and the independence of observations, and the strict inequality from Assumption 4.

2. Take nonempty open sets $A, B \subset \mathbb{R}^K_+$ such that $A \succ^\star B$. From the previous condition we have $\mu_X(A), \mu_X(B) > 0$, and since choices are independent $\mu^2(x \in A,\, y \in B) = \mu_X(A)\mu_X(B) > 0$. Hence

$$\begin{aligned}
\mu^2(x \succ^D y | x \in A,\, y \in B) &= \frac{\mu^2(x \in A,\, y \in B, \text{ and } x \succ^D y)}{\mu^2(x \in A,\, y \in B)} \\
&> \frac{\mu^2(x \in A,\, y \in B, \text{ and } y \succsim^D x)}{\mu^2(x \in A,\, y \in B)} \\
&= \mu^2(y \succsim^D x | x \in A,\, y \in B)\,.
\end{aligned}$$

   The two equalities follow from the definition of conditional probability and the inequality from Assumption 4.

For necessity suppose both conditions in Lemma 1 hold, and take nonempty open sets $A, B \subset \mathbb{R}^K_+$ such that $A \succ^\star B$. From condition 1 we have $\mu_X(A), \mu_X(B) > 0$ and from independence $\mu^2(x \in A,\, y \in B) = \mu_X(A)\mu_X(B) > 0$. Hence

$$\begin{aligned}
\mu^2(x \in A,\, y \in B, \text{ and } x \succ^D y) &= \mu^2(x \succ^D y | x \in A,\, y \in B)\, \mu^2(x \in A,\, y \in B) \\
&> \mu^2(y \succsim^D x | x \in A,\, y \in B)\, \mu^2(x \in A,\, y \in B) \\
&= \mu^2(x \in A,\, y \in B, \text{ and } y \succsim^D x)\,.
\end{aligned}$$

The two equalities follow from the definition of conditional probability and the inequality from condition 2. □

[20] A sequence of functions $f^n : \mathbb{R}^K_+ \to \mathbb{R}$ converges to $f$ in the topology of compact convergence if it converges uniformly to $f$ in every compact $X \subset \mathbb{R}^K_+$.

# C Proof of Theorem 2

For the proofs in this section we use the function

$$\kappa(\succsim) = \frac{\mu^2(x \succ^D x' \text{ and } x \succ x')}{\mu^2(x \succ^D x')}\,.$$

**Lemma 4.** *For every data set $\mathcal{D}$ and preference $\succsim \in \mathcal{P}$ the function $d(\mathcal{D}, \succsim)$ is continuous with probability one.*

*Proof.* Take choice data $\mathcal{D}$, a preference $\succsim \in \mathcal{P}$, and a sequence $\succsim^m \to \succsim$. Take $i, j \in [N]$. From Assumption 5 having two observations such that $x^i \sim x^j$ is a zero probability event. If $x^i \succ x^j$ then for $m_0$ large enough we have that $x^j \not\succsim^m x^i$, and hence $x^i \succ^m x^j$, for all $m \geq m_0$. Similarly if $x^j \succ x^i$ for $m_1$ large enough we have $x^j \succ^m x^i$ for all $m \geq m_1$. Hence $\lim_{m\to\infty} d(\mathcal{D}, \succsim^m) = d(\mathcal{D}, \succsim)$ with probability one. ◻

**Lemma 5.** $1 - d(\mathcal{D}, \succsim) \xrightarrow{p} \kappa(\succsim)$.

*Proof.* There are $N^2$ pairs of observations over which we could potentially observe strict preference relations (in particular, $x \succ^D x$ is possible). Since observations are i.i.d., it follows from Hoeffding's (1961) Strong Law of Large Numbers for U-statistics that $|\succ^D|/N^2 \xrightarrow{a.s.} \mu^2(x \succ^D y) + \mu^2(y \succ^D x)$, and $|\succ^D \cap \succ|/N^2 \xrightarrow{a.s.} \mu^2(x \succ^D y \text{ and } x \succ y) + \mu^2(y \succ^D x \text{ and } y \succ x)$. Assumption 4 implies $\mu^2(x \succ^D y) + \mu^2(y \succ^D x) > 0$. Since $1 - d(\mathcal{D}, \succsim) = \left(|\succ^D|/N^2\right)^{-1} \left(|\succ^D \cap \succ|/N^2\right)$, the continuous mapping theorem implies

$$1 - d(\mathcal{D}, \succsim) \xrightarrow{a.s.} \frac{\mu^2(x \succ^D y \text{ and } x \succ y) + \mu^2(y \succ^D x \text{ and } y \succ x)}{\mu^2(x \succ^D y) + \mu^2(y \succ^D x)} = \kappa(\succsim)\,.$$

The last equality is given by the independence between observations. Almost sure convergence implies convergence in probability. ◻

**Lemma 6.** *If $\succsim \in \mathcal{P}$ and $\succsim \neq \succsim^\star$, then $\kappa(\succsim^\star) > \kappa(\succsim)$.*

*Proof.* Take $\succsim \neq \succsim^\star$ ($\succsim \in \mathcal{P}$). The denominator $\mu^2(x \succ^D y)$ of $\kappa$ is independent of $\succsim$, and hence constant. Thus to maximize $\kappa(\succsim)$ we only need to maximize the numerator. Let $\mathbf{I}_{\succsim}(x, y) \equiv \mathbf{1}\{x \succsim y\}$. The numerator of $\kappa(\succsim)$ is

$$\mu^2(x \succ^D y \text{ and } x \succ y) = \int_{\mathcal{O}\times\mathcal{O}} \mathbf{1}\left\{x \succ^D y\right\} \mathbf{I}_{\succ}(x, y) d\mu^2((g, x), (f, y))\,.$$

The difference in numerators is

$$\begin{aligned}
&\mu^2(x \succ^D y \text{ and } x \succ^\star y) - \mu^2(x \succ^D y \text{ and } x \succ y) = \\
&\qquad = \int_{\mathcal{O}\times\mathcal{O}} \mathbf{1}\left\{x \succ^D y\right\} \left(\mathbf{I}_{\succ^\star}(x, y) - \mathbf{I}_{\succ}(x, y)\right) d\mu^2((g, x), (f, y))
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathcal{O}\times\mathcal{O}} \mathbf{1}\left\{x \succ^D y\right\} \left(\mathbf{I}_{\succ^\star\setminus\succ}(x,y) - \mathbf{I}_{\succ\setminus\succ^\star}(x,y)\right) d\mu^2((g,x),(f,y)) \\
&\geq \int_{\mathcal{O}\times\mathcal{O}} \mathbf{1}\left\{x \succ^D y\right\} \left(\mathbf{I}_{\succ^\star\setminus\succsim}(x,y) - \mathbf{I}_{\succ\setminus\succ^\star}(x,y)\right) d\mu^2((g,x),(f,y)) \\
&= \int_{\mathcal{O}\times\mathcal{O}} \mathbf{1}\left\{x \succ^D y\right\} \left(\mathbf{I}_{\succ^\star\setminus\succsim}(x,y) - \mathbf{I}_{\succ\setminus\succsim^\star}(x,y)\right) d\mu^2((g,x),(f,y)) \qquad (9)
\end{aligned}
$$

The last equality follows from Assumption 5, as $\mathbf{I}_{\succ\setminus\succ^\star}$ and $\mathbf{I}_{\succ\setminus\succsim^\star}$ are equal $\mu^2$-almost surely. From the second term within the parenthesis in (9) we have

$$
\begin{aligned}
\int_{\mathcal{O}\times\mathcal{O}} \mathbf{1}\left\{x \succ^D y\right\} \mathbf{I}_{\succ\setminus\succsim^\star}(x,y) d\mu^2((g,x),(f,y)) &= \int_{\mathcal{O}\times\mathcal{O}} \mathbf{1}\left\{y \succ^D x\right\} \mathbf{I}_{\succ\setminus\succsim^\star}(y,x) d\mu^2((p,x),(q,y)) \\
&= \int_{\mathcal{O}\times\mathcal{O}} \mathbf{1}\left\{y \succ^D x\right\} \mathbf{I}_{\succ^\star\setminus\succsim}(x,y) d\mu^2((g,x),(f,y)) \,.
\end{aligned}
$$

The first equality interchanges the names of the observations, which is irrelevant as they are independent. The second one uses the fact that $(y,x) \in \succ \setminus \succsim^\star$ if and only if $(x,y) \in \succ^\star \setminus \succsim$. Replacing this last expression in (9) we get

$$
\begin{aligned}
&\mu^2(x \succ^D y \text{ and } x \succ^\star y) - \mu^2(x \succ^D y \text{ and } x \succ y) = \\
&\qquad = \int_{\mathcal{O}\times\mathcal{O}} \mathbf{I}_{\succ^\star\setminus\succsim}(x,y) \left(\mathbf{1}\left\{x \succ^D y\right\} - \mathbf{1}\left\{y \succ^D x\right\}\right) d\mu^2((g,x),(f,y)) \\
&\qquad \geq \int_{\mathcal{O}\times\mathcal{O}} \mathbf{I}_{\succ^\star\setminus\succsim}(x,y) \left(\mathbf{1}\left\{x \succ^D y\right\} - \mathbf{1}\left\{y \succsim^D x\right\}\right) d\mu^2((g,x),(f,y)) \qquad (10)
\end{aligned}
$$

Since $\succsim$ and $\succsim'$ are continuous and strictly increasing the set $\succ^\star \setminus \succsim$ is open in $\mathbb{R}^K_+ \times \mathbb{R}^K_+$ (see the proof of Lemma 7 in Chambers et al., 2021). Then there are sets of sets $\{U_u\}_{u\in\mathcal{U}}$ and $\{V_v\}_{v\in\mathcal{V}}$, with $U_u, V_v \subset \mathbb{R}^K_+$ nonempty and open, such that $\succ^\star \setminus \succsim = \bigcup_{(u,v)\in\mathcal{U}\times\mathcal{V}} U_u \times V_v$. Hence, it follows from Assumption 4 that the right hand side of (10) is strictly positive. Therefore $\kappa(\succsim^\star) > \kappa(\succsim)$. □

**Lemma 7.** *The function $\kappa(\succsim)$ is continuous on $\mathcal{P}$.*

*Proof.* As the denominator of $\kappa(\succsim)$ is independent of $\succsim$ we only focus on the numerator. Take $\succsim \in \mathcal{P}$ and a sequence $\succsim^n \to \succsim$ and $(x,y)$ such that $x \not\sim y$. If $x \succ y$ then for $n$ large enough we have $x \succ^n y$, and if $y \succ x$ for $n$ large enough we have $y \succ^n x$. By Assumption 5 to have $x \sim y$ is a zero probability event. Since observations are independent

$$
\begin{aligned}
&\lim_{n\to\infty} \int_{\mathcal{O}} \int_{\mathcal{O}} \mathbf{1}\left\{x \succ^D y\right\} \mathbf{1}\left\{x \succ^n y\right\} d\mu((g,x)) d\mu((f,y)) \\
&\quad = \lim_{n\to\infty} \int_{\mathcal{O}} \int_{\mathcal{O}\setminus\{(f',y'):y'\neq x,\, y'\sim x\}} \mathbf{1}\left\{x \succ^D y\right\} \mathbf{1}\left\{x \succ^n y\right\} d\mu((g,x)) d\mu((f,y)) \\
&\quad = \int_{\mathcal{O}} \int_{\mathcal{O}\setminus\{(f',y')\in\mathcal{O}:y'\neq x,\, y'\sim x\}} \mathbf{1}\left\{x \succ^D y\right\} \lim_{n\to\infty} \mathbf{1}\left\{x \succ^n y\right\} d\mu((g,x)) d\mu((f,y))
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{\mathcal{O}} \int_{\mathcal{O} \setminus \{(f',y') \in \mathcal{O} : y' \neq x,\, y' \sim x\}} \mathbf{1}\left\{x \succ^D y\right\} \mathbf{1}\left\{x \succ y\right\} d\mu((g,x)) d\mu((f,y)) \\
&= \int_{\mathcal{O}} \int_{\mathcal{O}} \mathbf{1}\left\{x \succ^D y\right\} \mathbf{1}\left\{x \succ y\right\} d\mu((g,x)) d\mu((f,y))
\end{aligned}
$$

which is the numerator of $\kappa$. The interchange between the limit and the integration follows from Lebesgue dominated convergence. $\square$

*Proof of Theorem 2.* Lemma 6 implies that $1-\kappa$ is uniquely maximized by $\succsim^\star$. Furthermore, sine $\mathcal{P}$ is compact (Theorem 1) and $d \to 1-\kappa$ (Lemma 5), it follows from a slight variation of Theorem 1 in Jennrich (1969) that $d$ converges uniformly in probability to $1-\kappa$. Since $\kappa$ is continuous (Lemma 7) the result follows from Theorem 2.1 in Newey and McFadden (1994). $\square$

# D Proof of Proposition 2

**Lemma 8.** *Let $E^\star$ be a solution to the MFAS problem of $[G,\Omega]$ defined in Proposition 2. If $x \rhd y$ then $(x,y) \notin E^\star$.*

*Proof.* Towards a contradiction suppose there are $x,y \in \mathbf{x}$ such that $x \rhd y$ and $(x,y) \in E^\star$. Define $\widetilde{E} = \{(x',y') \in \succ^D_{\rhd} : x' \not\rhd y'\}$. Since there are at most $N^2$ elements in $\succ^D$, $(x,y) \in \succ^D_{\trianglerighteq}$, and $(x,y) \notin \widetilde{E}$ we have $\sum_{e \in E^\star} \omega_e \geq \omega_{(x,y)} = N^2 > |\widetilde{E}| = \sum_{e \in \widetilde{E}} \omega_e$. As $\rhd$ is acyclic the graph $(\mathbf{x}, \succ^D_{\rhd} \setminus \widetilde{E}) = (\mathbf{x}, \succ^D_{\rhd} \cap \rhd)$ is acyclic; this contradicts $E^\star$ being a solution to the MFAS problem of $[G,\Omega]$. $\square$

**Lemma 9.** *Let $\trianglelefteq = \{(x,y) : y \trianglerighteq x\}$. For every continuous and $\rhd$-monotone preference $\succsim$ we have $|\succsim^D \setminus \succ| = |\succsim^D \cap \trianglelefteq| + \sum_{e \in (\succ^D_{\rhd} \setminus \succ)} \omega_e$.*

*Proof.* As $\succsim$ is $\trianglerighteq$-monotone, $\succ \cap \trianglelefteq = \emptyset$; hence $\succ^D \setminus \succ = (\succ^D \cap \trianglelefteq) \cup ((\succ^D \setminus \trianglelefteq) \setminus \succ)$, where $\succ^D \cap \trianglelefteq$ and $(\succ^D \setminus \trianglelefteq) \setminus \succ$ have an empty intersection. For any $(x,y) \in \succ^D \setminus \succ$ for which $y \not\trianglelefteq x$ there are $w,z \in \mathbf{x}_0$ such that $x \sim_{\rhd} w$, $y \sim_{\rhd} z$, and $(w,z) \in \succ^D_{\rhd} \setminus \succ$. Furthermore, since $\succsim$ is $\trianglerighteq$-monotone we have $\left|(\succ^D \setminus \trianglelefteq) \setminus \succ\right| = \sum_{e \in \succ^D_{\rhd} \setminus \succ} \omega_e$. As $\succ^D \cap \trianglelefteq$ and $(\succ^D \setminus \trianglelefteq) \setminus \succ$ have an empty intersection, the previous equality implies the desired result. $\square$

*Proof of Proposition 2.* For sufficiency, towards a contradiction suppose $\succsim \in \mathcal{P}$ is such that $\succ^D_{\rhd} \setminus \succ$ solves the MFAS problem of $[G,\Omega]$ but is not a solution to (2). Then there is $\succsim' \in \mathcal{P}$ such that $|\succ^D \setminus \succ'| < |\succ^D \setminus \succ|$ which. As both $\succsim$ and $\succsim'$ are continuous and $\trianglerighteq$-monotone, Lemma 9 implies $\sum_{e \in (\succ^D_{\rhd} \setminus \succ')} \omega_e < \sum_{e \in (\succ^D_{\rhd} \setminus \succ)} \omega_e$. Since $\succsim'$ is a preference the graph $(\mathbf{x}, \succ^D_{\rhd} \setminus (\succ^D_{\rhd} \setminus \succ')) = (\mathbf{x}, \succ^D_{\rhd} \cap \succ')$ is acyclic. Therefore $\succ^D_{\rhd} \setminus \succ$ does not solve the MFAS problem of $[G,\Omega]$, a contradiction.

For necessity suppose $\succsim$ is a solution of (2), and towards a contradiction assume $E \subset \succ_{\triangleright}^{D}$ is a solution to the MFAS problem of $[G, \Omega]$ and satisfies

$$\sum_{e \in E} \omega_e < \sum_{e \in (\succ_{\triangleright}^{D} \setminus \succ)} \omega_e \,. \tag{11}$$

We now show that there is a continuous and $\trianglerighteq$-monotone preference $\succsim'$ for which $\succ_{\triangleright}^{D} \cap \succ' = \succ_{\triangleright}^{D} \setminus E$.

As $\trianglerighteq$ is continuous, for every $x$ the set $\{y : x \trianglerighteq y\}$ is closed, and hence its complement $\{y : x \not\trianglerighteq y\}$ is open. As $\mathbf{x}_0$ is finite and $\{y : x \not\trianglerighteq y\}$ is open there is $\varepsilon > 0$ small enough such that for any two $x, y \in \mathbf{x}_0$ such that $x \not\trianglerighteq y$ we have $x + \varepsilon\mathbf{1} \not\trianglerighteq y$. Take such $\varepsilon$ and for each $x \in \mathbf{x}_0$ define the set $A_x = \{x\} \bigcup \{z + \varepsilon\mathbf{1} : (x, z) \in (\succ_{\triangleright}^{D} \setminus E)\}$, and let $\mathcal{A} = (A_x)_{x \in \mathbf{x}}$. In terms of Nishimura et al. (2017, Chapter II.C), $((\mathbb{R}_{+}^{K}, \trianglerighteq), \mathcal{A})$ is a *continuous choice environment*.[21] Let $A_x^{\downarrow} = \{z : w \trianglerighteq z \text{ for some } w \in A_x\}$ be the decreasing closure of $A_x$ with respect to $\trianglerighteq$. We now show that for any two $x, y \in \mathbf{x}_0$ if $y \neq x$ and $y \in A_x^{\downarrow}$ then $x \operatorname{tran}(\succ_{\triangleright}^{D} \setminus E) y$. If $x \neq y$ and $y \in A_x^{\downarrow}$ then one of the following is true: (i) $x \triangleright y$, (ii) $(x, y) \in (\succ_{\triangleright}^{D} \setminus E)$ (since $(x, y + \varepsilon\mathbf{1}) \in A_y$ and $y + \varepsilon\mathbf{1} \trianglerighteq y$), or (iii) there is $z \in \mathbf{x}$ such that $(x, z) \in (\succ_{\triangleright}^{D} \setminus E)$ and $z + \varepsilon\mathbf{1} \trianglerighteq y$. In (i) and (ii) is clear that $(x, y) \in (\succ_{\triangleright}^{D} \setminus E)$. In (iii), our choice of $\varepsilon$ ensures $z \trianglerighteq y$; if $z \triangleright y$ we have $(z, y) \in (\succ_{\triangleright}^{D} \setminus E)$ and hence $(x, y) \in \operatorname{tran}(\succ_{\triangleright}^{D} \setminus E)$, and if $y \trianglerighteq z$ by the definition on $\mathbf{x}$ we have $z = y$ and go back to case (ii). In all cases we conclude that $(x, y) \in \operatorname{tran}(\succ_{\triangleright}^{D} \setminus E)$.

Let $A_x^{\downarrow\downarrow} = \{z : w \triangleright z \text{ for some } w \in A_x\}$ be the strictly decreasing closure of $A_x$ with respect to $\trianglerighteq$. We now show that for every $x \in \mathbf{x}$ we have $x \notin A_x^{\downarrow\downarrow}$. If $x \in A_x^{\downarrow\downarrow}$ then there is $y \in A_x$ such that $y \triangleright x$, which implies that $y - \varepsilon\mathbf{1} \in \mathbf{x}$, $(x, y - \varepsilon\mathbf{1}) \in (\succ_{\triangleright}^{D} \setminus E)$, and (by the choice of $\varepsilon$) $y - \varepsilon\mathbf{1} \triangleright x$. By Lemma 8 we have that $(y - \varepsilon\mathbf{1}, x) \in (\succ^{D} \setminus E)$, which violates the fact that $(\mathbf{x}_0, \succ_{\triangleright}^{D} \setminus E)$ is acyclic.

Finally, let the choice correspondence $c : \mathcal{A} \rightrightarrows \mathbb{R}_{+}^{K}$ be defined by $c(A_x) = x$ for every $A_x \in \mathcal{A}$. Take a sequence of chosen bundles $(z^{\ell})_{\ell \in [L]}$ ($z^{\ell} \in \mathbf{x}$) satisfying $z^2 \in A_{z^1}^{\downarrow}$, $z^3 \in A_{z^2}^{\downarrow}$, …, $z^L \in A_{z^{L-1}}^{\downarrow}$, and $z^1 \in A_{z^L}^{\downarrow}$. The previous relations imply

$$z^1 \operatorname{tran}(\succ_{\triangleright}^{D} \setminus E)\, z^2 \operatorname{tran}(\succ_{\triangleright}^{D} \setminus E) \ldots \operatorname{tran}(\succ_{\triangleright}^{D} \setminus E)\, z^L \operatorname{tran}(\succ_{\triangleright}^{D} \setminus E)\, z^1 \,.$$

If there were different elements $z^{\ell'} \neq z^{\ell''}$ in the previous equation, then $z^{\ell'} \operatorname{tran}(\succ_{\triangleright}^{D} \setminus E) z^{\ell''} \operatorname{tran}(\succ_{\triangleright}^{D} \setminus E) z^{\ell'}$ would be a cycle of $(\mathbf{x}, \succ_{\triangleright}^{D} \setminus E)$, which is impossible. Hence all the elements in $(z^{\ell})_{\ell \in [L]}$ have to be equal. Since $x \notin A_x^{\downarrow\downarrow}$ for every $x \in \mathbf{x}$, we conclude that, following again the terminology of Nishimura et al. (2017), $c$ satisfies cyclical $\trianglerighteq$-consistency.[22]

As $c$ satisfies cyclical $\trianglerighteq$-consistency, by Theorem 2 in Nishimura et al. (2017) there is a continuous and strictly $\trianglerighteq$-increasing utility function $u$ that rationalizes $c$. As for every $x, y$ for which

[21] Nishimura et al. (2017) define a continuous choice environment as $((X, \trianglerighteq), \mathcal{A})$, where $X$ is a locally compact and separable metric space; $\trianglerighteq$ is a continuous preorder; and $\mathcal{A}$ a is a collection of nonempty compact subsets of $X$.

[22] A choice correspondence $c$ satisfies *cyclical consistency* if, for every sequence of observations $(x^{\ell})_{\ell \in [L]}$ ($x_{\ell} \in \mathbf{x}$), if whenever $x^1 \in c(A_{x^1}) \cap A_{x^2}^{\downarrow}$, $x^2 \in c(A_{x^2}) \cap A_{x^3}^{\downarrow}$, …, $x^{L-1} \in c(A_{x^{L-1}}) \cap A_{x^L}^{\downarrow}$, and $x^L \in c(A_{x^L}) \cap A_{x^1}^{\downarrow}$, we also have $x^1 \in A_{x^2}^{\downarrow} \setminus A_{x^2}^{\downarrow\downarrow}$, $x^2 \in A_{x^3}^{\downarrow} \setminus A_{x^3}^{\downarrow\downarrow}$, …, $x^{L-1} \in A_{x^L}^{\downarrow} \setminus A_{x^L}^{\downarrow\downarrow}$, and $x^L \in A_{x^1}^{\downarrow} \setminus A_{x^1}^{\downarrow\downarrow}$.

$x \left(\succ^D_{\rhd} \setminus E\right) y$ we have $(x, y+\varepsilon\mathbf{1}) \in A_x$, then $u(x) \geq u(y+\varepsilon\mathbf{1})$, which as $u$ is strictly $\trianglerighteq$-increasing implies $u(x) > u(y)$. Take the preference relation $\succsim'$ represented by $u$, which is continuous and $\trianglerighteq$-monotone, as $u$ is continuous and strictly $\trianglerighteq$-increasing. Then $(\succ^D_{\trianglerighteq} \setminus E) \subset (\succ^D \cap \succ')$. Moreover, as for any proper subset $E'$ of $E$ the graph $(\mathbf{x}_0, \succ^D \setminus E')$ has a cycle and $\succ$ does not have cycles we have $(\succ^D \setminus E) = (\succ^D \cap \succ')$. As $E \subset \succ^D_{\rhd}$, from (11) we have $\sum_{e\in(\succ^D_{\rhd}\setminus\succ')} \omega_e < \sum_{e\in(\succ^D_{\rhd}\setminus\succ)} \omega_e$, which by Lemma 9 implies $| \succ^D \setminus \succ' | < | \succ^D \setminus \succ |$. Finally, by Assumption 6 there is $\succsim'' \in \mathcal{P}$ such that $| \succ^D \setminus \succ'' | = | \succ^D \setminus \succ' |$, and therefore $\succsim$ is not a solution to (2). □

# E Proof of Theorem 3

*Proof of Theorem 3.* First suppose there is a continuous and $\trianglerighteq$-monotone preference relation $\succsim$ that $v^{\mathcal{M}}$-rationalizes the data and discards only $\mathcal{M}$. take any two $i, j \in [N]$ such that $x^i \succsim^D_{\mathcal{M}} x^j$, and towards a contradiction suppose $x^j \succ^D_{\mathcal{M}} x^i$. As $\succsim$ discards only $\mathcal{M}$ we have both $x^i \succsim x^j$ and $x^j \succ x^i$, a contradiction.

Now suppose that the data satisfies GARP$_{\mathcal{M}}$. As $\trianglerighteq$ is continuous, the set $\{y : x \trianglerighteq y\}$ is closed, and hence its complement $\{y : x \not\trianglerighteq y\}$ is open. As the data is finite and $g^i$ is continuous for all $i \in [N]$ there is $\varepsilon > 0$ small enough such that for any two observed bundles $x^i, x^j$ we have that (1) $x^i \not\trianglerighteq x^j$ implies $x^i + \varepsilon\mathbf{1} \not\trianglerighteq x^j$ and (2) $g^i(x^j) < 1$ implies $g^i(x^j + \varepsilon\mathbf{1}) < 1$. Take such $\varepsilon$ and for each $i \in [N]$ define the set $A_i = \{x^i\} \bigcup \{x^j : x^i \succsim^D_{\mathcal{M}} x^j\} \bigcup \{x^j + \varepsilon\mathbf{1} : x^i \succ^D_{\mathcal{M}} x^j\} \bigcup \{y : g^i(y) \leq v^{\mathcal{M}}_i\}$, and let $\mathcal{A} = (A_i)_{i\in[N]}$. Then $((R^K_+, \trianglerighteq), \mathcal{A})$ is, following Nishimura et al. (2017), a continuous choice environment.

We have the following two properties:

1. For any two observations $i, j \in [N]$ if $x^j \in A^{\downarrow}_i$ then $x^i \succsim^R_{\mathcal{M}} x^j$: If $x^j \in A^{\downarrow}_i$ then one of the following is true: (i) $x^i \trianglerighteq x^j$, (ii) there is $x^m$ such that $x^i \succsim^D_{\mathcal{M}} x^m$ and $x^m \trianglerighteq x^j$, (iii) there is $x^m$ such that $x^i \succ^D_{\mathcal{M}} x^m$ and $x^m + \varepsilon\mathbf{1} \trianglerighteq x^j$, or (iv) there is $x$ such that $g^i(x) \leq v^{\mathcal{M}}_i$ and $x \trianglerighteq x^j$. In (i) that $x^i \succsim^D_{\mathcal{M}} x^j$ follows from the definition of $\mathcal{M}$-revealed preferences; in (ii) we have $x^m \succsim^D_{\mathcal{M}} x^j$, hence $x^i \succsim^R_{\mathcal{M}} x^j$; in (iii) by the definition of $\varepsilon$ we have that $x^m \trianglerighteq x^j$, hence $x^m \succsim^D_{\mathcal{M}} x^j$ and $x^i \succsim^R_{\mathcal{M}} x^j$; finally in (iv) the definition of $v^{\mathcal{M}}$ implies that $x^i \succsim^D_{\mathcal{M}} x^j$. In all cases we conclude that $x^i \succsim^R_{\mathcal{M}} x^j$.
2. For any two observations $i, j \in [N]$ if $x^j \in A^{\downarrow\downarrow}_i$ then $x^i \succ^R_{\mathcal{M}} x^j$: If $x^j \in A^{\downarrow\downarrow}_i$ then one of the following is true: (i) $x^i \rhd x^j$, (ii) there is $x^m$ such that $x^i \succsim^D_{\mathcal{M}} x^m$ and $x^m \rhd x^j$, (iii) there is $x^m$ such that $x^i \succ^D_{\mathcal{M}} x^m$ and $x^m + \varepsilon\mathbf{1} \rhd x^j$, or (iv) there is $x$ such that $g^i(x) \leq v^{\mathcal{M}}_i$ and $x \rhd x^j$. In (i) $x^i \succ^D_{\mathcal{M}} x^j$ by definition; in (ii) by $x^m \rhd x^j$ implies $x^m \succ^D_{\mathcal{M}} x^j$ and hence $x^i \succ^R_{\mathcal{M}} x^j$; in (iii) the definition of $\varepsilon$ implies $x^m \trianglerighteq x^j$, hence $x^m \succsim^D_{\mathcal{M}} x^j$ and $x^i \succ^R_{\mathcal{M}} x^j$; (iv) the definition of $v^{\mathcal{M}}$ implies that $x^i \succ^D_{\mathcal{M}} x^j$. In all cases we conclude $x^i \succ^R_{\mathcal{M}} x^j$.

Define the choice correspondence $c : \mathcal{A} \rightrightarrows \mathbb{R}^K_+$ by $c(A_i) = x^i$ for every $i \in [N]$, hence $c(\mathcal{A})$ is the set of chosen bundles. To show that $c$ satisfies cyclical $\trianglerighteq$-consistency suppose there is a chain of chosen bundles $(x^{m_\ell})_{\ell\in[L]}$ such that $x^{m_2} \in A^{\downarrow}_{m_1}$, $x^{m_3} \in A^{\downarrow}_{m_2}$, …, $x^{m_L} \in A^{\downarrow}_{m_{L-1}}$, and

$x^{m_1} \in A^{\downarrow}_{m_L}$. Hence $x^{m_1} \succsim^R_{\mathcal{M}} x^{m_2} \succsim^R_{\mathcal{M}} \dots \succsim^R_{\mathcal{M}} x^{m_L} \succsim^R_{\mathcal{M}} x^{m_1}$. Towards a contradiction (and without loss of generality) suppose $x^{m_1} \in A^{\downarrow\downarrow}_{m_L}$, which implies that $x^{m_L} \succ^R_{\mathcal{M}} x^{m_1}$. Then there are chosen bundles $x^s, x^{s'}$ such that $x^{m^L} \succsim^R_{\mathcal{M}} x^s \succ^D_{\mathcal{M}} x^{s'} \succsim^R_{\mathcal{M}} x^{m_1}$. Furthermore, the previous relation implies $x^{s'} \succsim^R_{\mathcal{M}} x^{m_1} \succsim^R_{\mathcal{M}} x^{m_L} \succsim^R_{\mathcal{M}} x^s$, and hence $x^{s'} \succsim^R_{\mathcal{M}} x^s$. This relation along with $x^s \succ^D_{\mathcal{M}} x^{s'}$ imply that choices fails $\text{GARP}_{\mathcal{M}}$. Hence $c$ satisfies cyclical $\trianglerighteq$-consistency.

As $c$ satisfies cyclical $\trianglerighteq$-consistency, by Theorem 2 in Nishimura et al. (2017) there is a continuous and strictly $\trianglerighteq$-increasing utility function $u$ such that rationalizes $c$. Let $\succsim$ be the preference relation represented by $u$. As $u$ rationalizes $c$ we have that $x^i \succsim x$ whenever $p^i x \leq v_i^{\mathcal{M}}$, hence $\succsim$ $v^{\mathcal{M}}$-rationalizes the data. We also have that $x^i \succsim x^j$ whenever $x^i \succsim^D_{\mathcal{M}} x^j$, so $\succsim \supset \succsim^D_{\mathcal{M}}$. Finally if $x^i \succ^D_{\mathcal{M}} x^j$ as $x^j + \varepsilon \mathbf{1} \in A_i$ we have $x^i \succsim x^j + \varepsilon \mathbf{1}$, which as $\succsim$ is $\trianglerighteq$-monotone implies $x^i \succ x^j$. Hence $\succ \supset \succ^D_{\mathcal{M}}$. Therefore $\succsim$ $v$-rationalizes the data and discards only $\mathcal{M}$. □

# F Proof of Proposition 3

We follow the notation in Algorithm 1 for $C$, $\widehat{\mathcal{M}}^s$, $\widehat{\mathcal{M}}^w$, $\widehat{\mathcal{M}}^w_0$, $\widetilde{M}_t$ and $\widehat{\mathcal{M}}$, and refer to properties 1, 2, and 3 following Definition 5.

*Proof of Proposition 3.* Suppose $x^i \triangleright x^j$, which implies $x^i \succ^D x^j$. By Lemma 8 we have $(x, y) \notin E_0$ and therefore $(x, y) \notin \widehat{\mathcal{M}}^s$. Similarly, if $x \trianglerighteq y$ then $(x, y) \in C$ and $(x, y) \notin \widehat{\mathcal{M}}^w_0$. As $\widehat{\mathcal{M}}^w \subset \widehat{\mathcal{M}}^w_0$, then $(x, y) \notin \widehat{\mathcal{M}}^w$, and $\widehat{\mathcal{M}}$ satisfies property 1.

If $x^i \succ^D x^j$ and $x^j \trianglerighteq x^i$ then $(x^i, x^j) \in \widehat{\mathcal{M}}^s$. If $x^i \succsim^D x^j$ and $x^j \triangleright x^i$, then, $x^j (\succ^D \setminus \widehat{\mathcal{M}}^s) x^i$ by property 1. Suppose $(x^i, x^j)$ is the $t^{\text{th}}$ element in $\widehat{\mathcal{M}}^w_0$ according to the enumeration in Algorithm 1, so $\widetilde{M}_t = \left( \widehat{\mathcal{M}}^w_{t-1} \setminus \{(x^i, x^j)\}, \widehat{\mathcal{M}}^s \right)$. Since $(x^j, x^j) \in C$, we have $x^i \succsim^D_{\widetilde{M}_t} x^j \succsim^D_{\widetilde{M}_t} x^j \succ^D_{\widetilde{M}_t} x^i$. Therefore $\text{GARP}_{\widetilde{M}_t}$ fails and, given the construction of $\widehat{\mathcal{M}}^w_T$, $(x^i, x^j) \in \widehat{\mathcal{M}}^w_T = \widehat{\mathcal{M}}^w$. We conclude that $\widehat{\mathcal{M}}$ satisfies property 2.

Finally, let $x^i \succ^D x^j$ and $(x^i, x^j) \in \widehat{\mathcal{M}}^w$, and towards a contradiction suppose $(x^i, x^j) \notin \widehat{\mathcal{M}}^s$. Since $(x^i, x^j) \in \succ^D \setminus \widehat{\mathcal{M}}^s \subset C$, then $(x^i, x^j) \notin \widehat{\mathcal{M}}^w_0$. As $\widehat{\mathcal{M}}^w \subset \widehat{\mathcal{M}}^w_0$, then $(x, y) \notin \widehat{\mathcal{M}}^w$, a contradiction. Therefore $\widehat{\mathcal{M}}$ satisfies property 3. □

# G Proof of Proposition 4

We follow Algorithm 1 for the definitions of $E_0$, $C$, $\widehat{\mathcal{M}}^s$, $\widehat{\mathcal{M}}^w$, $\widehat{\mathcal{M}}^w_t$, and $\widehat{\mathcal{M}}$.

**Lemma 10.** $\succsim^D \setminus \widehat{\mathcal{M}}^w_0 = C = (\succ^D \setminus \widehat{\mathcal{M}}^s) \bigcup \{(x, y) \in \succsim^D : x \sim_{\trianglerighteq} y\}$.

*Proof.* Since $\succ^D \subset \succsim^D$ and $x^i \sim_{\trianglerighteq} x^j$ implies $x^i \succsim^D x^j$, we have $C \subset \succsim^D$. This implies $\succsim^D \setminus \widehat{\mathcal{M}}^w_0 = \succsim^D \setminus (\succsim^D \setminus C) = \succsim^D \cap C = C$. Now take $(x, y) \in \succsim^D$ such that $x \triangleright y$; Proposition 3 implies $(x, y) \in \succ^D \setminus \widehat{\mathcal{M}}^s$, which implies the desires result. □

**Lemma 11.** *Let $G = (\mathbf{x}, \succ^D_{\triangleright})$ be the digraph defined in Proposition 2, $\Omega$ the respective weights, and $E^\star$ the solution to the MFAS problem of $[G, \Omega]$. If $x^i(\succ^D \setminus \widehat{\mathcal{M}}^s)x^j \sim_{\trianglerighteq} x^m$, then there are $y, z \in \mathbf{x}$ such that $y \sim_{\triangleright} x^i$, $z \sim_{\triangleright} x^m$, and $y(\succ^D_{\triangleright} \setminus E^\star)z$.*

*Proof.* Follows from transitivity of $\sim_{\trianglerighteq}$ and the definitions of $\mathbf{x}$ and $\widehat{\mathcal{M}}^s$. $\square$

*Proof of Proposition 4.* For $t \in [T]$ Let $\widehat{\mathcal{M}}_t = \left(\widehat{\mathcal{M}}^w_t, \widehat{\mathcal{M}}^s\right)$. We show that $\mathcal{D}$ satisfies $\text{GARP}_{\widehat{\mathcal{M}}_T}$, which is the same as $\text{GARP}_{\widehat{\mathcal{M}}}$, by induction on $t$.

- Let $t = 0$ and suppose $x^i \succsim^R_{\widehat{\mathcal{M}}_0} x^j$; then there is a sequence $(x^{m_\ell})_{\ell \in [L]}$ satisfying

$$x^i \left(\succsim^D \setminus \widehat{\mathcal{M}}^w_0\right) x^{m_1} \left(\succsim^D \setminus \widehat{\mathcal{M}}^w_0\right) \dots \left(\succsim^D \setminus \widehat{\mathcal{M}}^w_0\right) x^{m_L} \left(\succsim^D \setminus \widehat{\mathcal{M}}^w_0\right) x^j \,. \tag{12}$$

  By Lemma 10, $y(\succsim^D \setminus \widehat{\mathcal{M}}^w_0)z$ implies either $y(\succ^D \setminus \widehat{\mathcal{M}}^s)z$ or $y \sim_{\trianglerighteq} z$, and by transitivity of $\sim_{\trianglerighteq}$ it is without loss of generality to assume that there are no two consecutive $\sim_{\trianglerighteq}$ relations in (12). Take $\mathbf{x}$ as defined in Proposition 2; if $y(\succ^D \setminus \widehat{\mathcal{M}}^s)z$, then there are $w, w' \in \mathbf{x}$ such that $w \sim_{\trianglerighteq} y$, $w' \sim_{\trianglerighteq} z$, and $w(\succ^D_{\triangleright} \setminus E_0)w'$. Hence, by Lemma 11 there is a sequence $(w^s)_{s \in [S]}$, $w^s \in \mathbf{x}$, such that $w^1 \sim_{\trianglerighteq} x^i$, $w^S \sim_{\trianglerighteq} x^j$, and

$$w^1 \left(\succ^D_{\triangleright} \setminus E_0\right) w^2 \left(\succ^D_{\triangleright} \setminus E_0\right) \dots \left(\succ^D_{\triangleright} \setminus E_0\right) w^{S-1} \left(\succ^D_{\triangleright} \setminus E_0\right) w^S \,. \tag{13}$$

  Towards a contradiction, suppose $x^j \succ^D_{\widehat{\mathcal{M}}_0} x^i$. Then there are $y, y' \in \mathbf{x}$ such that $y \sim_{\trianglerighteq} xj$, $y' \sim_{\trianglerighteq} x^i$, and $y(\succ^D_{\triangleright} \setminus E_0)y'$. Since there are no two elements in $\mathbf{x}$ related by $\sim_{\trianglerighteq}$, it follows that $y = w^S$ and $y' = w^1$; therefore $w^S(\succ^D_{\triangleright} \setminus E_0)w^1$. Along with (13), the previous relation implies that $\succ^D_{\triangleright} \setminus E_0$ has a cycle, a contradiction with $E_0$ solving the MFAS problem of Proposition 2. We conclude that $\mathcal{D}$ satisfies $\text{GARP}_{\widehat{\mathcal{M}}_0}$.
- Suppose $\text{GARP}_{\widehat{\mathcal{M}}_{t-1}}$ holds. That $\text{GARP}_{\widehat{\mathcal{M}}_t}$ holds follows directly from its definition.

$\square$

# H Proof of Proposition 5

In this section we use the notation from Algorithm 1 for the binary relations $E$, $\widehat{\mathcal{M}}^s$, $\widehat{\mathcal{M}}^w$, and $\widehat{\mathcal{M}}^w_t$, and for the tuple of mistakes $\widehat{\mathcal{M}}$, and from Proposition 2 for $\mathbf{x}$ and $\sim^D_{\triangleright}$.

*Proof of Proposition 5.* First, suppose there is $(x^i, x^j) \in \widehat{\mathcal{M}}^s \setminus \mathcal{M}^s$. There are two possible cases:

- There are $w, y \in \mathbf{x}$ such that $w \sim_{\trianglerighteq} x^i$, $y \sim_{\trianglerighteq} x^j$, and $(w, z) \in E_0$: Let $E_1 = E_0 \setminus \{(w, z)\}$. Since $E_0$ solves the MFAS problem in Proposition 2, $\succ^D_{\triangleright} \setminus E_1$ has a cycle; this is, there is a sequence $(w^s)_{s \in [S]}$, such that

$$w^1 \left(\succ^D_{\triangleright} \setminus E_1\right) w^2 \left(\succ^D_{\triangleright} \setminus E_1\right) \dots \left(\succ^D_{\triangleright} \setminus E_1\right) w^S \left(\succ^D_{\triangleright} \setminus E_1\right) w^1 \,.$$

Since $\mathcal{M}^s \subset \widehat{\mathcal{M}}^s$, and following the definitions of $\mathbf{x}$ and $\succ_{\triangleright}^{D}$, there is a sequence of observed choices $(x^{m_s})_{s\in[S]}$, where $x^{m_s} \sim_{\trianglerighteq} w^s$ for every $s \in [S]$, such that $x^{m_1} \succ_{\mathcal{M}}^{D} x^{m_2} \succ_{\mathcal{M}}^{D} \dots \succ_{\mathcal{M}}^{D} x^{m_S} \succ^{D} x^{m_1}$. By Definition 5, $y \succ_{\mathcal{M}}^{D} z$ implies $y \succsim_{\mathcal{M}}^{D} z$; hence $x^{m_1} \succsim_{\mathcal{M}}^{R} x^{m_S}$. Therefore $\mathcal{D}$ fails $\text{GARP}_{\mathcal{M}}$.

- $x^j \trianglerighteq x^i$: This implies $x^j \succsim^{R} x^i$ and, by Definition 5, $x^j \succsim_{\mathcal{M}}^{D} x^i$. Since $x^i \succ_{\mathcal{M}}^{D} x^j$, $\mathcal{D}$ fails $\text{GARP}_{\mathcal{M}}$.

Now suppose $\mathcal{M}^s = \widehat{\mathcal{M}}^s$, and there is $(x^i, x^j) \in \widehat{\mathcal{M}}^w \setminus \mathcal{M}^w$. Let $(x^i, x^j)$ be the $t^{\text{th}}$ element of $\widehat{\mathcal{M}}_0^w$. As $(x^i, x^j) \in \widehat{\mathcal{M}}^w$ then $\mathcal{D}$ fails $\widetilde{M}_t$, so there are $\ell, m \in [N]$ such that $x^\ell \succsim_{\widetilde{M}_t}^{R} x^m$ and $x^m \succ_{\widetilde{M}_t}^{D} x^\ell$. Since $\mathcal{M}^w \subset \widehat{\mathcal{M}}^w \subset \widehat{\mathcal{M}}_{t-1}^w$ and $(x^i, x^j) \notin \mathcal{M}^w$, then $\mathcal{M}^w \subset \widehat{\mathcal{M}}_{t-1}^w \setminus \{(x^i, x^j)\}$, which implies $x^\ell \succsim_{\mathcal{M}}^{R} x^m$. As $\mathcal{M}^s = \widehat{\mathcal{M}}^s$ we have $x^m \succ_{\mathcal{M}}^{D} x^\ell$, and $\text{GARP}_{\mathcal{M}}$ fails. □

# I Proof of Theorem 4

We take the notation for $[G, \Omega]$ and $\succ_{\triangleright}^{D}$from Proposition 2.

*Proof of Theorem 4.* By definition of $\widehat{\mathcal{M}}^s$ (Algorithm 1), $\succ_{\triangleright}^{D} \cap \widehat{\mathcal{M}}^s = E_0$. Since $\succ_{\triangleright}^{D} \subset \succ^{D}$ and $\succ^{\widehat{\mathcal{M}}}$ extends $\succ^{D} \setminus \widehat{\mathcal{M}}^s$ we have $\succ^{D} \setminus \widehat{\mathcal{M}}^s = \succ_{\triangleright}^{D} \cap \widehat{\mathcal{M}}^s = E_0$. Therefore $\succ^{D} \setminus \widehat{\mathcal{M}}^s$ solves the MFAS problem of $[G, \Omega]$. By Proposition 2 $\succsim^{\widehat{\mathcal{M}},N}$ solves (2) and, by Theorem 2, $\succsim^{\widehat{\mathcal{M}},N} \xrightarrow{p} \succsim^{\star}$. □