

Parametric Recoverability of Preferences*

Yoram Halevy[†] Dotan Persitz[‡] Lanny Zrill[§]

September 25, 2016

Abstract

Revealed preference theory is brought to bear on the problem of recovering approximate parametric preferences from consistent and inconsistent consumer choices. We propose measures of the incompatibility between the revealed preference ranking implied by choices and the ranking induced by the considered parametric preferences. These incompatibility measures are proven to characterize well-known inconsistency indices. We advocate a recovery approach that is based on such incompatibility measures, and demonstrate its applicability for misspecification measurement and model selection. Using an innovative experimental design we empirically substantiate that the proposed revealed-preference-based method predicts choices significantly better than a standard distance-based method.

*We are grateful to the editor Ali Hortaçsu and anonymous referees for insightful comments and suggestions that significantly improved this work. We thank Jose Apesteguia, Mark Dean, Eddie Dekel, Erwin Diewert, Tim Feddersen, Dave Freeman, Faruk Gul, Shachar Kariv, Hiro Kasahara, Yuichi Kitamura, Charles Manski, Vadim Marmer, Yusufcan Masatlioğlu, Hiroki Nishimura, Wolfgang Pesendorfer, Phil Reny, Ariel Rubinstein, Tomasz Strzalecki, Andy Skrzypacz, Elie Tamer, Peter Wakker, Yaniv Yedid-Levi and participants in various seminars, workshops and conferences. Shervin Mohammadi-Tari programmed the experimental interface and improved the recovery algorithm. Hagit Labiod and Ori Nofekh provided excellent research assistance in proofreading Sections 2, 3 and 4, the related appendices and the Matlab code. Yoram Halevy and Lanny Zrill gratefully acknowledge financial support from SSHRC. Dotan Persitz gratefully acknowledges the support of the ISF (grant number 1390/14) and the Henry Crown Institute for business research in Israel. Work on this project started when Zrill was a PhD student at UBC and Persitz was a postdoctoral fellow at UBC. The experimental work was done under UBC's BREB approval H14-00510.

[†]Vancouver School of Economics, University of British Columbia. Email: yoram.halevy@ubc.ca.

[‡]The Coller School of Management, Tel Aviv University. Email: persitzd@post.tau.ac.il.

[§]Department of Economics, Langara College. Email: lzzrill@langara.bc.ca.

1 Introduction

This paper studies the problem of recovering stable preferences from individual choices. The renewed interest in this problem emerges from the recent availability of relatively large data sets composed of *individual* choices made from linear budget sets. These rich data sets allow researchers to recover approximate *individual* stable utility functions and report the magnitude and distribution of behavioral characteristics in the population. We bring revealed preference theory to bear on this problem of recovering approximate parametric preferences from both consistent and inconsistent consumer choices.

Classical revealed preference theory studies the conditions on observables (choices) that are equivalent to the maximization of some utility function. If a data set is constructed from consumer choice problems in an environment with linear budget sets, Afriat (1967) proves that no revealed preferences cycles among observed choices, a condition known as the Generalized Axiom of Revealed Preference (henceforth GARP), is equivalent to assuming that the consumer behaves as if she maximizes some locally non satiated utility function. In his proof, Afriat constructs a well behaved piecewise linear utility function that is consistent with the consumer choices. Theorem 1 shows that similar reasoning may be applied for approximate preferences when GARP is not satisfied, by adjusting the revealed preference information to exclude cycles.

The method above requires recovering twice the number of parameters as there are observations and therefore the behavioral implications of the constructed functional forms may be difficult to interpret and apply to economic problems. In many cases researchers assume simple functional forms with few parameters that lend themselves naturally to behavioral interpretations. The drawback of this approach is that simple functional forms are often too structured to capture every nuance of individual decision making. Thus, preferences recovered in this way are almost always misspecified. That is, the ranking implied by the recovered preferences may be incompatible with the ranking information implied by the decision maker's choices (summarized through the

revealed preference relation¹). We argue that given a parametric utility specification, one should seek a measure to quantify the extent of misspecification and minimize it as a criterion for selecting from the functional family.

Our proposed measures of misspecification rely on insights gained from the literature that quantifies internal inconsistencies inherent in a data set. The [Houtman and Maks \(1985\)](#) Inconsistency Index searches for the minimal subset of observations that should be removed from a data set in order to eliminate cycles in the revealed preference relation. Similarly, the [Varian \(1990\)](#) Inconsistency Index is calculated by aggregating the minimal budget adjustments required to remove revealed preference relations that cause the data set to fail GARP. A special case of the Varian Inconsistency Index is the Critical Cost Inefficiency Index ([Afriat 1972; 1973](#)) in which adjustments are restricted to be identical across all observations.

[Theorem 2](#) provides the following novel theoretical characterization of these indices: for every utility function a loss can be calculated that aggregates budget adjustments required to remove incompatibilities between the ranking information induced by the utility function and the revealed preference information contained in the observed choices. The loss function corresponding to the Houtman-Maks Inconsistency Index is the Binary Incompatibility Index (henceforth BII), which counts the observations that are not rationalized by a given utility function. Similarly, the loss function corresponding to the Varian Inconsistency Index is the Money Metric Index (proposed by [Varian, 1990](#), henceforth MMI), which aggregates the *minimal* budget adjustments required to remove all incompatibilities. We prove that the inconsistency indices equal the infimum of their corresponding loss functions taken over *all* continuous, acceptable² and locally non-satiated utility functions. Hence, the inconsistency indices lend themselves naturally as benchmarks for minimizing incompatibilities between the data set and all considered utility functions.

We argue that parametric recovery should generalize the principle intro-

¹If choices are inconsistent the “revealed preference relation” refers to the ranking remaining after excluding cycles in some “minimal” way (see [Definition 1](#) below).

²A utility function is acceptable if the zero bundle is weakly worse than every other non-negative bundle. See also [Definition 13](#) in [Appendix A.4](#).

duced in characterizing the inconsistency indices, by calculating the infimum of the loss function over a *restricted* subset of utility functions. If a data set is consistent (satisfies GARP), the incompatibility measures we propose quantify the extent of misspecification that arises solely from considering a specific family of utility functions, rather than all continuous, acceptable and locally non-satiated utility functions. If the data set does not satisfy GARP, each measure can be additively decomposed into the respective inconsistency index and a misspecification index. Since for a given data set the inconsistency index is constant, the incompatibility measures can be minimized to recover parametric preferences within some parametric family.

This discussion continues a line of thought proposed by [Varian \(1990\)](#), who was unsatisfied with the standard approach which relies on parametric specification when testing for optimizing behavior. Varian suggested separating the analysis into two parts. The first part, which does not rely on a parametric specification, tests for consistency and quantifies how close choices are to being consistent using an inconsistency index. The second part uses the money metric as a “natural measure of how close the observed consumer choices come to maximizing a particular utility function” (page 133) and employs it as a criterion for recovering preferences. Varian argued that measuring differences in utility space has a more natural economic interpretation than measuring distances between bundles in commodity space.

We augment Varian’s intuition by providing theoretical and practical substance for the use of loss functions as measures of misspecification. First, we relate the budget adjustments implied by the proposed loss functions to the Houtman-Maks, Varian and Afriat inconsistency indices. Second, we advocate recovery methods that utilize as much ranking information encoded in observed choices rather than distance-based methods, since making a choice from a menu reveals that the chosen alternative is preferred to *every* other feasible alternative, not only to the predicted one. Therefore, our rationale for using the MMI is different from Varian’s, and could be equally applied to other loss functions, as the BII. Third, since we show that the goodness of fit can be decomposed into an inconsistency index and a misspecification mea-

sure, it lends itself naturally to several novel applications including evaluating parametric restrictions and model selection. Thus, ultimately we show that the two parts proposed by [Varian \(1990\)](#) are closely related, as the difference between them can be attributed to the sets of utility functions considered. Finally, while Varian takes the theory to a representative agent data, we use individual level data gathered in the laboratory to provide evidence for the predictive superiority of the MMI.

As an illustration of a practical application, we use the MMI to recover parameters for the data set collected by [Choi et al. \(2007\)](#) in which subjects choose a portfolio of Arrow securities. Using the Disappointment Aversion model of [Gul \(1991\)](#) with the CRRA functional form, we recover parameters using Non-Linear Least Squares (henceforth NLLS) and MMI. We find substantial numerical differences with respect to the recovered parameters that in some cases imply significant quantitative and qualitative differences in preferences.

However, the data collected by [Choi et al. \(2007\)](#) was not designed to compare the accuracy in which different recovery methods represent the decision maker's preferences. Therefore, we propose a general empirical-experimental methodology whereby recovery methods are evaluated based on their predictive success and apply it in an experimental setting similar to [Choi et al. \(2007\)](#). The experiment utilizes a unique two-part design. In the first part of the experiment we collect choice data from linear budget sets and instantaneously recover individual parameters from this data using the two different parametric recovery methods (MMI and NLLS). We use the individually recovered parameters to construct a sequence of *pairs* of portfolios (per individual) such that one of the portfolios in each pair is preferred according to the parametric preferences recovered by the MMI and the other is preferred by the parametric preferences recovered by the NLLS. Then, in the second part of the experiment, the subject is presented with these individually constructed pairs of portfolios and their choices are used to evaluate the predictive success of each recovery method.

This methodology enables us not only to compare the relative predictive

success of the recovery method but also to observe subject’s choices in regions that may otherwise be unobservable. In particular, when subjects choose from linear budget sets, non-convex preferences imply the existence of bundles that are never chosen if the subject chooses optimally. This may make it difficult to identify different sets of parameters that may nevertheless imply substantially different behavior (e.g. the extent of local risk seeking). By offering the subjects pairwise choices located in the region of non-convexity we can directly observe their true preferences in this region and identify which set of recovered parameters more accurately represents their underlying preferences.

For our sample of 203 subjects, we find that the MMI recovery method predicted subjects’ choices significantly more accurately than the NLLS recovery method. At the aggregate level, approximately 54% of pairwise choices are predicted by the MMI recovery method. At the individual level, consider those subjects for whom one of the methods correctly predicted more than two thirds of the pairwise choices. The choices of almost 60% of those subjects were more accurately predicted by the MMI recovery method. Moreover, when we focus our attention to only those subjects for which the recovered parameters imply non-convex preferences (i.e. local risk-seeking behavior), the MMI recovery method predicted more accurately in 62.5% of pairwise choices and for 75% of subjects for which more than two thirds of the choices are correctly predicted. We interpret these results as suggesting that our proposed MMI recovery method is more reliable than measures based on the distance between observed and predicted choices in commodity space, especially in decision making environments where closeness does not necessarily imply similarity.

We use the data from the experiment and the data collected by [Choi et al. \(2007\)](#) to show that the preferences of approximately 40% of the subjects are well approximated by expected utility compared to the general Disappointment Aversion functional form. In addition, we demonstrate non-nested model selection, by providing evidence that the choices of most subjects are better approximated by the Disappointment Aversion model with the *CRRA* utility index than by the Disappointment Aversion model with the *CARA* utility index.

In the next section we generalize the standard definitions of revealed preference relations and provide a proof of an extension of Afriat (1967) Theorem for inconsistent data sets (Theorem 1). In Section 3 we introduce the main inconsistency indices discussed in the paper and in Section 4 we introduce the Money Metric and the Binary Incompatibility measures and use them to characterize the inconsistency indices (Theorem 2). In Section 5 we analyze the data gathered by Choi et al. (2007) and point out the need for an external criterion to decide between the recovery methods. The experimental design is described in Section 6 while the results are reported in Section 7. Section 8 demonstrates the use of our theoretical results for hypothesis testing and model selection. Section 9 concludes.

2 Preliminaries

Consider a decision maker (henceforth DM) who chooses bundles $x^i \in \mathfrak{R}_+^K$ ($i \in 1, \dots, n$) from budget menus $\{x : p^i x \leq p^i x^i, p^i \in \mathfrak{R}_{++}^K\}$. Let $D = \{(p^i, x^i)_{i=1}^n\}$ be a finite data set, where x^i is the chosen bundle at prices p^i . The following definitions generalize the standard definitions of revealed preference (for similar concepts see Afriat, 1972, 1987; Varian, 1990, 1993; Cox, 1997).

Definition 1. Let D be a finite data set. Let $\mathbf{v} \in [0, 1]^n$.³ An observed bundle $x^i \in \mathfrak{R}_+^K$ is

1. \mathbf{v} -*directly revealed preferred* to a bundle $x \in \mathfrak{R}_+^K$, denoted $x^i R_{D, \mathbf{v}}^0 x$, if $v^i p^i x^i \geq p^i x$ or $x = x^i$.
2. \mathbf{v} -*strictly directly revealed preferred* to a bundle $x \in \mathfrak{R}_+^K$, denoted $x^i P_{D, \mathbf{v}}^0 x$, if $v^i p^i x^i > p^i x$.

³Throughout the paper we use bold fonts (as \mathbf{v} or $\mathbf{1}$) to denote vectors of scalars in \mathfrak{R}^n . We continue to use regular fonts to denote vectors of prices and goods. For $\mathbf{v}, \mathbf{v}' \in \mathfrak{R}^n$ $\mathbf{v} = \mathbf{v}'$ if $\forall i : v_i = v'_i$, $\mathbf{v} \geq \mathbf{v}'$ if $\forall i : v_i \geq v'_i$, $\mathbf{v} \geq \mathbf{v}'$ if $\mathbf{v} \geq \mathbf{v}'$ and $\mathbf{v} \neq \mathbf{v}'$ and $\mathbf{v} > \mathbf{v}'$ if $\forall i : v_i > v'_i$.

3. \mathbf{v} -revealed preferred to a bundle $x \in \mathfrak{R}_+^K$, denoted $x^i R_{D,\mathbf{v}} x$, if there exists a sequence of observed bundles (x^j, x^k, \dots, x^m) such that $x^i R_{D,\mathbf{v}}^0 x^j, x^j R_{D,\mathbf{v}}^0 x^k, \dots, x^m R_{D,\mathbf{v}}^0 x$.

When $\mathbf{v} = \mathbf{1}$ Definition 1 reduces to the standard definition of revealed preference relation. When \mathbf{v} decreases, more revealed preference information is being relaxed as summarized in the following observation (for a proof see Appendix A.1).

Fact 1. Let $\mathbf{v}' \leq \mathbf{v}$. Then: $R_{D,\mathbf{v}'}^0 \subseteq R_{D,\mathbf{v}}^0$, $P_{D,\mathbf{v}'}^0 \subseteq P_{D,\mathbf{v}}^0$ and $R_{D,\mathbf{v}'} \subseteq R_{D,\mathbf{v}}$.

Consider the following notion of consistency for data sets (Varian, 1990):

Definition 2. Let $\mathbf{v} \in [0, 1]^n$. D satisfies the *General Axiom of Revealed Preference Given \mathbf{v}* ($GARP_{\mathbf{v}}$) if for every pair of observed bundles, $x^i R_{D,\mathbf{v}} x^j$ implies not $x^j P_{D,\mathbf{v}}^0 x^i$.

When $\mathbf{v} = \mathbf{1}$ Definition 2 is equivalent to Afriat's (1967) cyclical consistency (GARP, see Varian (1982)). Practically, the vector \mathbf{v} is used to generate an adjusted relation $R_{D,\mathbf{v}}^0$ that contains no strict cycles while $R_{D,\mathbf{1}}^0$ may contain such cycles. Obviously, usually there are many vectors such that D satisfies $GARP_{\mathbf{v}}$. Following are two useful and trivial properties of $GARP_{\mathbf{v}}$ (proofs in appendices A.2 and A.3, respectively):

Fact 2. Every D satisfies $GARP_{\mathbf{0}}$.

Fact 3. Let $\mathbf{v}, \mathbf{v}' \in [0, 1]^n$ and $\mathbf{v} \geq \mathbf{v}'$. If D satisfies $GARP_{\mathbf{v}}$ then D satisfies $GARP_{\mathbf{v}'}$.

The following definition of \mathbf{v} -rationalizability relates the revealed preference information implied by observed choices to the ranking induced by a utility function.

Definition 3. Let $\mathbf{v} \in [0, 1]^n$. A utility function $u(x)$ \mathbf{v} -rationalizes D , if for every observed bundle $x^i \in \mathfrak{R}_+^K$, $x^i R_{D,\mathbf{v}}^0 x$ implies that $u(x^i) \geq u(x)$. We say that D is \mathbf{v} -rationalizable if such $u(\cdot)$ exists.

That is, the intersection between the set of bundles which are ranked strictly higher than an observed bundle x^i according to u , and the set of bundles to which x^i is revealed preferred when the budget constraint is adjusted by v^i , is empty. $\mathbf{1}$ -rationalizability reduces to the standard definition of Rationalizability (Afriat, 1967).⁴

\mathbf{v} -Rationalizability does not imply uniqueness. There could be different utility functions (not related through monotonic transformation) that \mathbf{v} -rationalize the same data set. Afriat's (1967) celebrated theorem provides tight conditions for the rationalizability of a data set.⁵ Afriat's (1967) theorem was generalized in many directions. For example, Reny (2015) extended to infinite data sets, Forges and Minelli (2009) to general budget sets and Fujishige et al. (2012) to indivisible goods. The following Theorem generalizes Afriat's result to inconsistent data sets.

Theorem 1. *The following conditions are equivalent:*

1. There exists a non-satiated utility function that \mathbf{v} -rationalizes the data.
2. The data satisfies $GARP_{\mathbf{v}}$.
3. There exists a continuous, monotone and concave utility function that \mathbf{v} -rationalizes the data.

Proof. See Appendix A.4.⁶ □

⁴Throughout the paper *rationalizability* means $\mathbf{1}$ -*rationalizability*, D is *rationalizable* if it is $\mathbf{1}$ -*rationalizable* and D satisfies $GARP$ if it satisfies $GARP_{\mathbf{1}}$.

⁵For discussion and alternative proofs of the original theorem see Diewert (1973); Varian (1982); Teo and Vohra (2003); Fostel et al. (2004); Geanakoplos (2013).

⁶Afriat (1973) uses the Theorem of the Alternative to provide a non-constructive proof for the uniform case. Afriat (1987) states Theorem 1 without a proof (Theorem 6.3.I on page 179). In his unpublished PhD dissertation Houtman (1995, Theorem 2.5) considers non-linear pricing and monotone adjustments. While the proof in Afriat (1973) can be easily generalized to our case, we preferred to adapt the construction suggested in Houtman (1995) for the case of scale adjustments of linear budget sets. In addition, while Afriat (1973) does not require the chosen bundle to remain feasible following an adjustment, our proof (as the one in Houtman (1995)) respects this requirement.

3 Inconsistency Indices

For some of the following inconsistency measures we make use of a general aggregator function across observations.⁷

Definition 4. $f_n : [0, 1]^n \rightarrow [0, M]$, where M is finite, is an *Aggregator Function* if $f_n(\mathbf{1}) = 0$, $f_n(\mathbf{0}) = M$ and $f_n(\cdot)$ is continuous and weakly decreasing.⁸

Varian (1990) proposed an inconsistency index that measures the minimal adjustments of the budget sets that remove cycles implied by choices. While Varian suggests to aggregate the adjustments using the sum of squares, we define this index with respect to an arbitrary aggregator function.⁹

Definition 5. Let $f : [0, 1]^n \rightarrow [0, M]$ be an aggregator function. *Varian's Inconsistency Index* is,¹⁰

$$I_V(D, f) = \inf_{\mathbf{v} \in [0, 1]^n : D \text{ satisfies } GARP_{\mathbf{v}}} f(\mathbf{v})$$

Varian (1990) suggested this index as a non-parametric measure for the extent of utility maximizing behavior implied by a data set of consumer choices. Varian's Inconsistency Index is a generalization of the Critical Cost Efficiency Index (suggested earlier by Afriat (1972; 1973)) that is restricted to uniform adjustments. Denote the set of vectors with equal coordinates by

⁷In most of this paper we assume a fixed data set of size n , therefore we will abuse notation by omitting the subscript, unless required for clarity.

⁸An aggregator function f_n is weakly decreasing if for every $\mathbf{v}, \mathbf{v}' \in [0, 1]^n$:
 $\mathbf{v} \geq \mathbf{v}' \implies f_n(\mathbf{v}) \leq f_n(\mathbf{v}')$
 $\mathbf{v} > \mathbf{v}' \implies f_n(\mathbf{v}) < f_n(\mathbf{v}')$. One may wish to restrict the set of potential aggregator functions to include only separable functions that satisfy the cancellation axiom. The results do not require the richness of possible aggregator functions.

⁹Alcantud et al. (2010) follow Varian (1990) to suggest the Euclidean norm of the adjustments vector. Tsur (1989) uses $\frac{\sum_{i=1}^n (\log v_i)^2}{n}$ while Varian (1993) and Cox (1997) mention the maximal adjustment and Smeulders et al. (2014) consider the generalized mean $\sum_{i=1}^n (1 - v_i)^\rho$ where $\rho \geq 1$.

¹⁰Consider a data set of two points $D = \{(p^1, x^1); (p^2, x^2)\}$ such that $p^1 x^2 = p^1 x^1$ but $p^2 x^1 < p^2 x^2$. D is inconsistent with GARP (since $x^1 R_{D, \mathbf{1}} x^2$ and $x^2 P_{D, \mathbf{1}}^0 x^1$), but consider the sequence $\mathbf{v}_l = (1 - \frac{1}{l}, 1)$ where $l \in \mathbb{N}_{>0}$. It is easy to verify that for every $l \in \mathbb{N}_{>0}$, D satisfies $GARP_{\mathbf{v}_l}$. Therefore $I_V(D, f) = 0$.

$\mathcal{I} = \left\{ \mathbf{v} \in [0, 1]^n : \mathbf{v} = v\mathbf{1}, \forall v \in [0, 1] \right\}$ and a coordinate of a typical vector $\mathbf{v} \in \mathcal{I}$ by v .

Definition 6. *Afriat's Inconsistency Index* is,

$$I_A(D) = \inf_{\mathbf{v} \in \mathcal{I}: D \text{ satisfies } GARP_{\mathbf{v}}} 1 - v$$

Houtman and Maks (1985) proposed an inconsistency index based on the maximal subset of observations that satisfies *GARP*. This is identical to restricting the adjustments vector to belong to $\{0, 1\}^n$ (see also Smeulders et al. (2014); Heufer and Hjertstrand (2015)) and to aggregate using the sum ($n - \sum_{i=1}^n v_i$). Again, we define this index with respect to an arbitrary aggregator function.

Definition 7. Let $f : [0, 1]^n \rightarrow [0, M]$ be an aggregator function. *Houtman-Maks Inconsistency Index* is,

$$I_{HM}(D, f) = \inf_{\mathbf{v} \in \{0, 1\}^n: D \text{ satisfies } GARP_{\mathbf{v}}} f(\mathbf{v})$$

Fact 4. $I_V(D, f)$, $I_A(D)$ and $I_{HM}(D, f)$ always exist.

Proof. See Appendix A.5. □

Afriat's and Houtman-Maks inconsistency indices are considerably more prevalent in the empirical-experimental literature than Varian's Inconsistency Index, mainly due to computational considerations (discussed in Appendix B.1).¹¹ However, definitions 5, 6 and 7 demonstrate that Afriat's and Houtman-Maks inconsistency indices are merely reductions of Varian's Inconsistency Index to subsets of adjustment vectors (and a specific functional form in the case of Afriat's Inconsistency Index). Moreover, in Appendix B.1 we

¹¹The Money Pump Inconsistency Index proposed by Echenique et al. (2011), the Minimum Cost Inconsistency Index suggested by Dean and Martin (2015) and the Area Inconsistency Index mentioned in Heufer (2008, 2009) and Apesteguia and Ballester (2015) are discussed and compared to the Varian Inconsistency Index in Appendix B.2. In Appendix B.3 we discuss an inconsistency index based on Euclidean distance rather than on revealed preference related to an index mentioned in Beatty and Crawford (2011).

claim that, practically, for most individual level data sets, the Varian Inconsistency Index can be computed exactly or with an excellent approximation.

In the consistency literature, Afriat (1973) and Varian (1990; 1993) view the extent of the adjustment of the budget line as the amount of income wasted by a DM relative to a fully consistent one (hence the term “Inefficiency Index”). An alternative interpretation (due to Manzini and Mariotti, 2007, 2012; Masatlioglu et al., 2012; Cherepanov et al., 2013), views the adjusted budget set as a consideration set which includes only the alternatives from the original budget menu that the DM compares to the chosen alternative. By construction, those bundles not included in the attention set are irrelevant for revealed preference consideration. Houtman (1995), for example, holds that the DM overestimates prices and hence does not consider all feasible alternatives. Another line of interpretation for inconsistent choice data, is measurement error (Varian, 1985; Tsur, 1989; Cox, 1997). These errors could be the result of various circumstances as (literally) trembling hand, indivisibility, omitted variables, etc. All the interpretations above take literally the existence of an underlying “welfare” preferences that generate the data (Bernheim and Rangel, 2009). In addition there exist other plausible data generating processes that may result in approximately (and even exactly) consistent choices (Simon, 1976; Rubinstein and Salant, 2012).

We do not find a clear reason to favor one interpretation over the other, and would rather remain agnostic about the nature of the adjustments required to measure inconsistency. Moreover, this paper takes the data set as the primitive and the utility function as an approximation. As such, the adjustments serve as a measurement tool (“ruler”) for quantifying the extent of misspecification.

4 Parametric Recoverability

The proof of Theorem 1 is constructive: if a data set D of size n satisfies $GARP_{\mathbf{v}}$ then finding a utility function that \mathbf{v} -rationalizes the data reduces to finding $2n$ real numbers that satisfy a set of n^2 inequalities (see the proofs

of Lemma 4 and Theorem 1 in Appendix A.4).¹² Although the constructed utility function does not rely on any parametric assumptions, the large number of parameters makes it difficult to directly learn from it about behavioral characteristics of the DM, which are typically summarized by few parameters (e.g. attitudes towards risk, ambiguity and time).¹³ Moreover, generically, a data set can be \mathbf{v} -rationalized by more than a single utility function. Hence, if one can find a “simpler” (parametric) utility function that rationalizes the data set, it will have an equal standing in representing the ranking information implied by the data set. If one accepts that “simple” may be superior, then one should consider the tradeoff between simplicity and misspecification. We pursue this line of reasoning by considering the minimal misspecification implied by certain parametric specifications.

The problem of *parametric* recoverability is to *approximately* rationalize observed choice data by a parametric utility function. We approach this problem by acknowledging that in the case where the data set is consistent (satisfies GARP) the representation of choice data by utility function almost always entails some tension between two rankings over alternatives. The first is the ranking implied by choices, which is captured by the revealed preference (partial) relation, and the other is the complete ranking induced by the parametric utility function. If the utility function rationalizes the data then the two rankings are compatible. Otherwise, the two rankings are incompatible and we say that the utility function is misspecified with respect to the data. The incompatibility is manifested by the existence of a pair of alternatives on which the two rankings disagree.

In Section 4.1 we propose two loss functions that measure the incompati-

¹²Varian (1982) builds on the celebrated Afriat (1967) theorem to construct non parametric bounds that partially identify the utility function, assuming that preferences are convex (see Halevy et al. (2016)). His approach has been extended and developed in Blundell et al. (2003, 2008) (see also Section 3.2 in Cherchye et al. (2009)). However, to the best of our knowledge, it has not been expanded to include treatment of inconsistent data sets. The parametric approach developed in the current paper extends naturally to inconsistent data sets and easily accommodates non-convex preferences.

¹³We thank a referee for pointing out that this problem resembles the known issue of “overfitting” in statistical estimation.

bility between the two rankings. Obviously, there are loss functions that are not based on the incompatibility between the suggested utility function and the revealed preference relation. For example, Non Linear Least Squares is a loss function that is based on the distance between the choice predicted by the suggested utility function and the observed choice. In sections 5 and 7 we demonstrate empirically the difference between these two types of loss functions.

The main theoretical contribution of the paper is presented in Section 4.2. This result establishes that the loss functions we propose do not depend on the choice data being consistent. In the case of inconsistent choices, the loss functions capture both the extent of inconsistency and the misspecification of the parametric utility function with respect to the data. We prove that the loss functions can then be additively decomposed into a corresponding inconsistency index and a misspecification measure. Section 8 demonstrates the empirical implications of this decomposition to model selection.

4.1 Incompatibility Indices

4.1.1 The Money Metric Index

Consider a bundle x^i that is chosen at prices p^i and a utility function $u(\cdot)$. While x^i is revealed preferred to all feasible bundles, u may rank some of these bundles above x^i . The first loss measure for the incompatibility between a data set D and a utility function u is based on the Money Metric Utility Function (Samuelson, 1974) and was suggested by Varian (1990) (see also Gross (1995)). It measures the minimal budget adjustment that makes bundles that u ranks above x^i infeasible, thus eliminating the incompatibility between the two rankings.

Definition 8. The *normalized money metric vector* for a utility function $u(\cdot)$, $\mathbf{v}^*(D, u)$, is such that $v^{*i}(D, u) = \frac{m(x^i, p^i, u)}{p^i x^i}$ where $m(x^i, p^i, u) = \min_{\{y \in \mathbb{R}_+^K : u(y) \geq u(x^i)\}} p^i y$. Let $f : [0, 1]^n \rightarrow [0, M]$ be an aggregator function. The *Money Metric Index* for a utility function $u(\cdot)$ is $f(\mathbf{v}^*(D, u))$.

Let \mathcal{U}^c denote the set of all locally non-satiated, acceptable and continuous utility functions on \mathfrak{R}_+^K .

Proposition 1. *Let $D = \{(p^i, x^i)_{i=1}^n\}$, $u \in \mathcal{U}^c$ and $\mathbf{v} \in [0, 1]^n$. $u(\cdot)$ \mathbf{v} -rationalizes D if and only if $\mathbf{v} \leq \mathbf{v}^*(D, u)$.*

Proof. See Appendix A.6. □

Proposition 1 establishes that $f(\mathbf{v}^*(D, u))$ may be viewed as a function that measures the loss incurred by using a specific utility function to describe a data set. $\mathbf{v}^*(D, u)$ measures the minimal adjustments to the budget sets required to remove incompatibilities between the revealed preference information contained in D and the ranking information induced by u . It also implies that each coordinate of $\mathbf{v}^*(D, u)$ is calculated independently of the other observations in the data set.^{14,15}

If $\mathbf{v}^*(D, u) = \mathbf{1}$ then Proposition 1 is merely a restatement of the familiar definition of rationalizability using the money metric as a criterion. A utility function $u \in \mathcal{U}^c$ rationalizes the observed choices if and only if there is no observation such that there exists an affordable bundle that u ranks above the observed choice. In this case we would say that the utility function is correctly specified.

Recall that given an aggregator function $f(\cdot)$, $f(\mathbf{v}^*(D, u))$ measures the incompatibility between a data set D and a specific preference relation represented by the utility function u . Given a set of utility functions $\mathcal{U} \subseteq \mathcal{U}^c$, the Money Metric Index measures the incompatibility between \mathcal{U} and D .

¹⁴One may intuitively believe that such independent calculation uses only the directly revealed preference information and may fail to rationalize the data based on the indirect revealed preference information. However, since $R_{D, \mathbf{v}}$ is the transitive closure of $R_{D, \mathbf{v}}^0$, it follows that a utility function is compatible with the directly revealed preference information if and only if it is compatible with all the indirectly revealed preference information.

¹⁵An additional implication of this property is that given m data sets D_i of n_i observations, and utility function $u(\cdot)$, since u $\mathbf{v}^*(D_i, u)$ -rationalizes D_i for every i , then u $\mathbf{v}^*(\bigcup_{i=1}^m D_i, u)$ -rationalizes $\bigcup_{i=1}^m D_i$ where $\mathbf{v}^*(\bigcup_{i=1}^m D_i, u) = (\mathbf{v}^*(D_1, u)', \dots, \mathbf{v}^*(D_m, u)')$. Moreover, if $f_n(\cdot)$ is additive separable for every n then $f_{\sum_{i=1}^m n_i}(\mathbf{v}^*(\bigcup_{i=1}^m D_i, u)) = \sum_{i=1}^m f_{n_i}(\mathbf{v}^*(D_i, u))$.

Definition 9. For a data set D and an aggregator function $f(\cdot)$ let $\mathcal{U} \subseteq \mathcal{U}^c$. The *Money Metric Index* of \mathcal{U} is

$$I_M(D, f, \mathcal{U}) = \inf_{u \in \mathcal{U}} f(\mathbf{v}^*(D, u))$$

4.1.2 The Binary Incompatibility Index

In this subsection we introduce a new loss measure that treats all incompatibilities similarly, by assigning them a maximal loss value.

Definition 10. The *Binary Incompatibility vector* for a utility function $u(\cdot)$, $\mathbf{b}^*(D, u)$, is such that $b^{*i}(D, u) = 1$ when $\nexists x$ such that $p^i x^i \geq p^i x$ and $u(x) > u(x^i)$, and $b^{*i}(D, u) = 0$ otherwise. Let $f : [0, 1]^n \rightarrow [0, M]$ be an aggregator function. The *Binary Incompatibility Index* for a utility function $u(\cdot)$ is $f(\mathbf{b}^*(D, u))$.

Consider a datum set that includes only the i -th observation from D . Then, the i -th element of the Binary Incompatibility vector tests whether the utility function rationalizes this datum set. While the Money Metric Index is restricted to the classical environment of choice from linear budget sets, the Binary Incompatibility Index may be easily applied to more general settings of choice from menus. The following proposition is the counterpart of Proposition 1 for the Binary Incompatibility Index.

Proposition 2. Let $D = \{(p^i, x^i)_{i=1}^n\}$, $u \in \mathcal{U}^c$ and $\mathbf{b} \in \{0, 1\}^n$. $u(\cdot)$ \mathbf{b} -rationalizes D if and only if $\mathbf{b} \leq \mathbf{b}^*(D, u)$.

Proof. See Appendix A.7. □

Definition 11. For a data set D and an aggregator function $f(\cdot)$, let $\mathcal{U} \subseteq \mathcal{U}^c$. The *Binary Incompatibility Index* of \mathcal{U} is

$$I_B(D, f, \mathcal{U}) = \inf_{u \in \mathcal{U}} f(\mathbf{b}^*(D, u))$$

4.1.3 Monotonicity of the Incompatibility Indices

The following observation follows directly from the definitions of $I_M(D, f, \mathcal{U})$ and $I_B(D, f, \mathcal{U})$ and concerns their monotonicity with respect to \mathcal{U} (see proof in Appendix A.8).

Fact 5. *For every $\mathcal{U}' \subseteq \mathcal{U} : I_M(D, f, \mathcal{U}) \leq I_M(D, f, \mathcal{U}')$ and $I_B(D, f, \mathcal{U}) \leq I_B(D, f, \mathcal{U}')$.*

In particular, Fact 5 implies that for every $\mathcal{U} \subseteq \mathcal{U}^c : I_M(D, f, \mathcal{U}^c) \leq I_M(D, f, \mathcal{U})$ and $I_B(D, f, \mathcal{U}^c) \leq I_B(D, f, \mathcal{U})$. That is, the value of the loss measures calculated for *all* continuous, acceptable and locally non-satiated utility functions is a lower bound on the incompatibility indices for every subset of utility functions.

4.2 Decomposing the Incompatibility Indices

The methods we propose to construct $\mathbf{v}^*(D, u)$ and $\mathbf{b}^*(D, u)$ do not depend on the consistency of the data set D . Therefore, even if a DM does not satisfy GARP,¹⁶ we can recover preferences (within the parametric family \mathcal{U}) that *approximate* the consistent revealed preference information encoded in choices. The difficulty with this approach arises from the fact that the loss indices include both the inconsistency with respect to GARP and the misspecification implied by the chosen parametric family.

We show that the suggested incompatibility indices can be decomposed into these two components. Our strategy in developing the decomposition is to use an inconsistency index as a measure of internal inconsistency, which is independent of the parametric family under consideration. We prove that the incompatibility indices calculated for *all* locally non-satiated, acceptable and continuous utility functions coincide with the respective inconsistency indices. That is, $I_M(D, f, \mathcal{U}^c)$ equals Varian’s Inconsistency Index (in particular, using

¹⁶Andreoni and Miller (2002); Porter and Adams (2015) find that a great majority of the subjects satisfy GARP. However, other experimental studies (Ahn et al., 2014; Choi et al., 2007, 2014; Fisman et al., 2007) report that more than 75 percent of the subjects did not satisfy GARP.

the minimum aggregator, $I_M(D, f, \mathcal{U}^c)$ equals Afriat Inconsistency Index), and $I_B(D, f, \mathcal{U}^c)$ coincides with the Houtman-Maks Inconsistency Index. The proof of the Theorem invokes Theorem 1 and is provided in Appendix A.9.

Theorem 2. *For every finite data set D and aggregator function f :*

1. $I_V(D, f) = I_M(D, f, \mathcal{U}^c)$.
2. $I_{HM}(D, f) = I_B(D, f, \mathcal{U}^c)$.
3. *If $f(\mathbf{v}) = 1 - \min_{i \in \{1, \dots, n\}} v^i$, then $I_A(D) = I_M(D, f, \mathcal{U}^c)$.*

Theorem 2 enables us to decompose the loss indices into familiar measures of inconsistency and natural measures of misspecification that quantify the cost of restricting preferences to a subset of utility functions (possibly through a parametric form). By the monotonicity of I_M and I_B (Fact 5), for every $\mathcal{U} \subseteq \mathcal{U}^c$ we can write the loss indices of \mathcal{U} in the following way:

$$\begin{aligned} I_M(D, f, \mathcal{U}) &= I_V(D, f) + (I_M(D, f, \mathcal{U}) - I_M(D, f, \mathcal{U}^c)) \\ I_B(D, f, \mathcal{U}) &= I_{HM}(D, f) + (I_B(D, f, \mathcal{U}) - I_B(D, f, \mathcal{U}^c)) \end{aligned}$$

In each decomposition, the first addend is a measure of the cost associated with inconsistent choices that is *independent* of any parametric restriction and depends only on the DM's choices, while the second addend measures the cost of restricting the preferences to a specific parametric form by the researcher who tries to recover the DM's preferences. A graphical demonstration of this decomposition appears in Appendix C.

Two reasons lead us to believe that such decomposition is essential for any method of recovering preferences of a DM who is inconsistent. First, since for a given data set the inconsistency index is constant (zero if GARP is satisfied), the decomposition implies that minimizing $I_M(D, f, \mathcal{U})$ or $I_B(D, f, \mathcal{U})$ is equivalent to minimizing the misspecification within some parametric family \mathcal{U} . Second, only when the incompatibility measure can be decomposed, one can truly evaluate the cost of restricting preferences to some parametric family compared to the cost incurred by the inconsistency in the choices. The

following sections demonstrate the importance of these theoretical insights in analyzing experimental data.

5 Application to Choice under Risk

The goal of this section is to demonstrate the empirical applicability of the Money Metric Index (MMI) as a criterion for recovering parametric preferences.¹⁷ We show that the suggested method can be used to recover approximate preferences for both consistent and inconsistent decision makers. For the inconsistent subjects, we use Theorem 2 to assess the degree to which these recovered preferences encode the revealed preference information contained in the choices. We compare the parameters resulting from employing the MMI and a recovery method that minimizes a loss function that is based on the Euclidean distance between observed and predicted choices in the commodity space (Non-Linear Least Squares, NLLS) and show that important qualitative differences arise.

As a starting point, we analyze in this section a data set of portfolio choice problems collected by Choi et al. (2007). In their experiment, subjects were asked to choose the optimal portfolio of Arrow securities from linear budget sets with varying prices. We focus our analysis only on the treatment where the two states are equally probable. For each subject, the authors collect 50 observations and proceed to test these choices for consistency (i.e. GARP). Then, they estimate a parametric utility function in order to determine the magnitude and distribution of risk attitudes in the population. Choi et al. (2007) estimate a Disappointment Aversion (DA) functional form introduced by Gul (1991) (for more details see Appendix D).

$$u(x_1^i, x_2^i) = \gamma w(\max\{x_1^i, x_2^i\}) + (1 - \gamma)w(\min\{x_1^i, x_2^i\}) \quad (5.1)$$

¹⁷In analyzing choices from budget menus, recovery based on Money Metric Index retains more ranking information from the data than recovery based on the Binary Incompatibility Index.

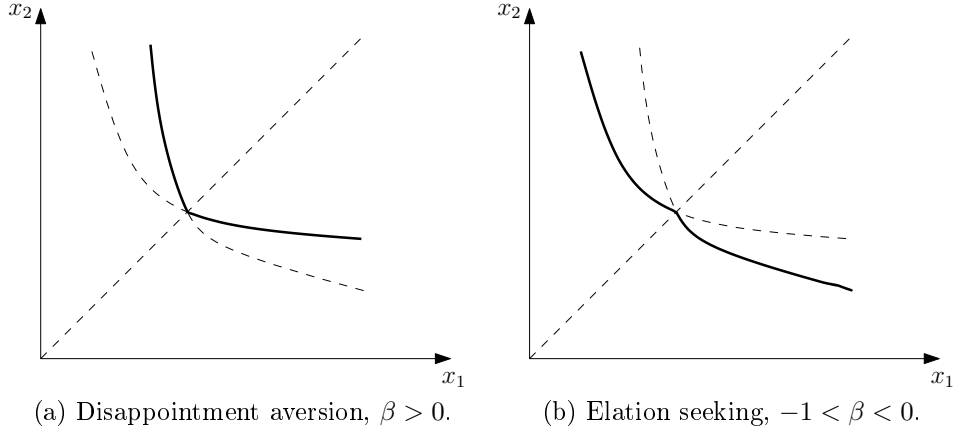


Figure 5.1: Typical indifference curves induced by Gul (1991) Disappointment Aversion function with $\beta \neq 0$.

where

$$\gamma = \frac{1}{2 + \beta} \quad \beta > -1 \quad w(z) = \begin{cases} \frac{z^{1-\rho}}{1-\rho} & \rho \geq 0 \quad (\rho \neq 1) \\ \ln(z) & \rho = 1 \end{cases}$$

The parameter γ is the weight placed on the better outcome. For $\beta > 0$, the better outcome is under-weighted relative to the objective probability (of 0.5) and the decision maker is *disappointment averse*. For $\beta < 0$, the better outcome is over-weighted relative to the objective probability (of 0.5) and the decision maker is *elation seeking*. In the knife-edge case, where $\beta = 0$, Expression (5.1) reduces to expected utility.

The parameter β has an important economic implication: if $\beta > (=) 0$ the decision maker exhibits *first-order (second order) risk aversion* (Segal and Spivak, 1990). That is, the risk premium for small fair gambles is proportional to the standard deviation (variance) of the gamble. First-order risk aversion can account for important empirical regularities that expected utility (with its implied second-order risk aversion) cannot, such as in portfolio choice problems (Segal and Spivak, 1990), calibration of risk aversion in the small and large, and disentangling inter-temporal substitution from risk aversion (see Epstein,

1992 for a survey). A negative value of β corresponds to a DM who is locally risk-seeking. Figure 5.1 illustrates characteristic indifference curves for disappointment averse and elation seeking (locally non-convex) subjects, respectively. Additionally, $w(x)$ is a standard utility function and is represented here by the *CRRA* functional form (we also report results where the utility for wealth function is *CARA*, i.e. $w(z) = -e^{-Az}$ where $A \geq 0$).

We recover parameters using two different methods. The first is the NLLS which is based on the Euclidean distance between the predicted and the observed choices,

$$\min_{\beta, \rho} \sum_{i=1}^n \left\| x^i - \arg \max_{x: p^i x \leq p^i x^i} (u(x; \beta, \rho)) \right\| \quad (5.2)$$

where $\|\cdot\|$ is the Euclidean norm. The second is the MMI, $I_M(D, f, \mathcal{U})$, using the normalized average sum-of-squares (henceforth, SSQ) aggregator, $f(\mathbf{v}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (1 - v^i)^2}$. For both methods, we use an optimization algorithm that allows us to recover individual parameters from observed choices for each subject.¹⁸

5.1 Recovering Preferences for Inconsistent Subjects

In Section 4.2 we prove the decomposition of the Money Metric Index into the Varian Inconsistency Index - which serves as a measure of inconsistency, and a remainder - which measures misspecification. As such, by using the MMI, we recover parameters that are closest to approximate preferences for those

¹⁸The recovery code implements an individual level data analysis and includes four modules. The first module implements the GARP test and calculates various inconsistency indices (see Appendix B.1). The other three modules implement the NLLS, MMI (with various aggregators) and BI recovery methods. Each of these three modules can recover preferences in the Disappointment Aversion (CRRA and CARA) functional family for portfolio choice data and in the CES functional family for other-regarding preferences data. The MATLAB code package is available online and user instructions are included in the package. The disaggregated results (using NLLS, MMI-SSQ and MMI-MEAN) of the Choi et al. (2007) data are available in a separate Excel file named “Choi et al. (2007) - Results”.

Subject	I_V	β	ρ	I_M
320	0	-0.509	0.968	0.1322
209	0.0288	0.164	0.352	0.0563

Table 1: Comparing consistent and inconsistent subjects.

subjects who fail GARP.¹⁹ Throughout the analysis, we exclude subjects with an unreliable Varian Inconsistency Index (9 out of 47 subjects).²⁰

To illustrate, consider Table 1 that compares the recovered parameters using the Money Metric Index with the SSQ aggregator for two subjects taken from Choi et al. (2007). Subject 320’s choices are consistent with GARP while Subject 209’s choices are inconsistent. In spite of the fact that Subject 320 is consistent, the parametric preferences considered do not accurately encode the ranking implied by her choices, as it requires 13.22% wasted income on average. On the other hand, the revealed preference information implied by Subject 209’s choices are nicely captured by the parametric family, since it implies incompatibility of only 5.63%, in spite of the fact that her choices are inconsistent (114 violations of GARP). Additionally, since $I_V = 0.0288$, the decomposed misspecification for Subject 209 amounts to only 2.75% ($I_M - I_V$) wasted income on average with respect to her approximate preferences. The lesson from this example is that although Subject 320 is consistent with GARP, the choices of Subject 209 are better approximated using the Disappointment

¹⁹Approximate preferences are defined by the set $\tilde{\mathcal{U}} = \{u \in \mathcal{U}^c : I_V(D, f) = I_M(D, f, \{u\})\}$. In general, this set is not a singleton as the vector of budget adjustments, \mathbf{v} , required by the calculation of the Varian Inconsistency Index, is not unique nor is the utility function that rationalizes a given revealed preference relation, $R_{D, \mathbf{v}}$, for a particular vector of adjustments.

²⁰Computing the Varian Inconsistency Index is a hard computational problem (see the discussion in Appendix B.1.2). The data of Choi et al. (2007) includes 47 subjects, 12 are consistent (pass GARP) and 35 are inconsistent. We take advantage of the sample size and calculate the exact index for 22 of the 35 inconsistent subjects (63%) and for 4 additional subjects we are able to provide a very good approximation. For the other 9 subjects we report a weak approximation computed using an algorithm that over-estimates the real index. The implication of overestimation is that the decomposition of the MMI overestimates the inconsistency component and underestimates the misspecification component. That said, while the extent of misspecification with respect to the approximate preferences may be underestimated, the recovered parameters are independent of the calculation of the Varian Inconsistency Index.

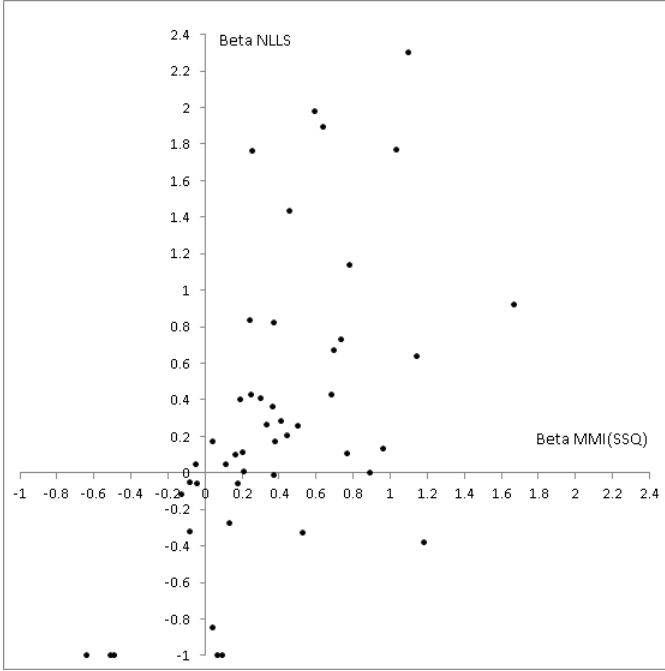


Figure 5.2: Disappointment Aversion Parameter: NLLS vs. MMI (SSQ).

Aversion with CRRA functional form. As such, the MMI can be applied uniformly to all data sets, and the appropriateness of a certain functional form can be evaluated ex-post (as will be further demonstrated in Section 8).

5.2 Comparison of Recovered Parameters by Method

Figure 5.2 demonstrates graphically the difference between the recovered parameters by comparing the disappointment aversion parameter (β) as recovered by the NLLS and MMI (SSQ) recovery methods. When NLLS recovers convex preferences ($\beta > 0$) then usually MMI recovers convex preferences as well, although there may be considerable quantitative differences between the recovered parameters. However, when the preferences recovered by NLLS are non-convex ($\beta < 0$), there seem to be no qualitative relation between the recovered parameters by the two methods.

Moreover, the parameters recovered by NLLS in some of the non-convex cases imply extreme elation seeking. This property can also be seen clearly

from the distribution of the disappointment aversion parameter (β) and the curvature of the utility function (ρ) across subjects which is reported in Appendix E.²¹

In light of the considerable differences between the recovered parameters, an essential next step is to compare these two recovery methods based on an out of sample criterion that is independent of the objective function of the candidate methods.

6 Experimental Design and Procedures

In this section we propose and describe a controlled experiment designed to perform a comparison between NLLS and MMI based on predictive power. Specifically, in the first part of the experiment we used a design inspired by Choi et al. (2007), to collect individual level portfolio choices from linear budget sets. From each subject’s choices we instantaneously recovered approximate parametric preferences by each of the two recovery methods. Using this information, we constructed pairs of portfolios such that the ranking induced by each set of approximate preferences on these portfolios disagree. Therefore, each recovery method implied opposite prediction on the subject’s choice from each pair of constructed portfolios. In the second and final part of the experiment, the subject chose a portfolio from each of the constructed pairs of portfolios, thus providing an out of sample direct criterion for the relative predictive success of each method.

6.1 Procedures and Details

For the experiment we recruited 203 subjects using the ORSEE system (Greiner, 2015) which is operated by the Vancouver School of Economics (VSE)

²¹Note that the recovered parameters for NLLS may differ from those reported in Choi et al. (2007) for several reasons: we allow for elation seeking ($-1 < \beta < 0$); we permit boundary observations ($x^i = 0$); we use Euclidean norm (instead of the geometric mean) and we use multiple initial points (including random) in the optimization routine (instead of a single predetermined point). We were able to replicate the results reported by Choi et al. (2007).

at the University of British Columbia. Subjects participated voluntarily and were primarily undergraduate students representing many disciplines within the university. Before subjects began the experiment, the instructions were read aloud as subjects followed along by viewing a dialog box on-screen (see Appendix F.1 for the instructions). The experiments were conducted over several sessions in October 2014 and February 2015 at the Experimental Lab at the Vancouver School of Economics (ELVSE). Each experimental session lasted approximately 45 minutes.

In the first part of the experiment, the subjects selected portfolios of contingent assets from a series of 22 linear budget sets with differing price ratios and/or relative wealth levels. These choices were used instantaneously to recover individual preferences using the two recovery methods introduced above. From these two sets of recovered parameters we constructed, uniquely for each subject, a sequence of 9 pairs of portfolios from which subjects chose during the second part of the experiment. Each pair included one *risky* portfolio, where outcomes differed across states, and one *safe* portfolio, where the subject obtained a certain payoff regardless of the state. Note that the subjects were unaware of the background calculation and the relation between the two parts of the experiment.

In total, each subject made 31 choices across the two parts of the experiment. After both rounds were completed, one of these rounds was selected randomly to be paid according to the subject's choice. For whichever round was selected, subjects were asked to flip a coin in order to determine for which state they would be paid. The choices were made over quantities of tokens which were converted at a 2 to 1 exchange rate to CAD. Subjects were paid privately upon completion of the experiment and their earnings averaged about 19.53 CAD in addition to a fixed fee of 10 CAD for showing up to the experiment on time.

6.2 Part 1: Linear Budget Sets

In this part of the experiment subjects chose portfolios of contingent assets from linear budget sets. Each portfolio, $x^i = (x_1^i, x_2^i)$, consisted of quantities of tokens such that subjects received x_1^i tokens if state 1 occurred and x_2^i tokens if state 2 occurred, with each state equally likely to occur. Portfolios were selected from a linear budget set, defined by normalized prices, p^i , and displayed graphically via a computer interface. All participants faced the same budget sets and in the same order, however, this was not known to the subjects.

The interface was a two-dimensional graph that ranged from 0 to 100 tokens on each axis. Subjects were able to adjust their choices in increments of 0.2 tokens with respect to the x-axis. Additionally, token allocations are rounded to one decimal place. Screen shots of the graphical interface are included in Appendix F.1. Subjects chose a particular portfolio by left-clicking on their desired choice on the budget line, and were asked to confirm their choice before moving on to the next round. Subjects were restricted to choose only those points which lie on the boundary of the budget set to eliminate potential violations of monotonicity.²²

The budget sets, and associated prices, were specifically chosen to address two issues. First, a sufficient overlap between budget sets is required so that GARP test will have sufficient power.²³ Second, an emphasis on moderate price ratios was required to identify the role of First-order Risk Aversion/Seeking (represented by β) in the subject's preferences. For further details on the budget lines selection see Appendix F.2.

²²Two special cases were treated slightly differently by the interface. First, when subjects chose a point close to the certainty line, a dialog box appeared that asked them if they meant to choose the allocation where the value in both accounts is equal, guaranteeing themselves a sure payoff, or if they prefer to stick with the point they chose. Second, when subjects chose a point that is close to either axis, a dialog box appeared that asked them if they meant to choose a corner choice or if they prefer to stick with the point they chose. This is done to overcome mechanical aspects of precision in the interface at points that have specific qualitative significance.

²³For a detailed analysis of a test that demonstrates that this set of budget sets is sufficiently powerful, see Appendix F.2.

6.3 Part 2: Pairwise Choices

Upon completion of the tasks in Part 1, the subject’s choices were used to recover structural parameters for the Disappointment Aversion functional form with CRRA using both NLLS and MMI (SSQ). These two sets of parameters were used to construct a sequence of 9 pairwise choice problems. In each pairwise comparison, subjects chose one of two portfolios - one *risky* portfolio (where payoffs differ across states) and one *safe* portfolio (where the payoff is certain) - represented as points in the coordinate system.²⁴

As preferences are a binary relation over bundles, pairwise choices allow us to directly observe the subject’s preferences in their most fundamental form. Therefore, we employed pairwise choice procedure to adjudicate between the two sets of recovered parameters, $\hat{\theta}_{NLLS} = \{\hat{\beta}_{NLLS}, \hat{\rho}_{NLLS}\}$ and $\hat{\theta}_{MMI} = \{\hat{\beta}_{MMI}, \hat{\rho}_{MMI}\}$. Given a risky portfolio, x^R , we calculated the certainty equivalent, $CE_i(CE_j)$, for both sets of parameters, $\hat{\theta}_i(\hat{\theta}_j)$ where $i, j \in \{NLLS, MMI\}$. In the case where both $\hat{\beta}_{NLLS} > 0$ and $\hat{\beta}_{MMI} > 0$ (both recovered preferences are convex) we selected the safe portfolio to be the mid-point between the two certainty equivalents, $x^S = (CE_i + CE_j) / 2$. Then, if $CE_i > CE_j$, in ranking the risky portfolio x^R and the safe portfolio x^S , $\hat{\theta}_i$ induces a preference for the risky portfolio while $\hat{\theta}_j$ induces a preference for the safe one. Since pairwise choices reveal the DM’s underlying preferences, choice of the risky portfolio reveals that the set of parameters $\hat{\theta}_i$ better approximates the DM’s preferences, while choosing the safe portfolio reveals the opposite.

In the case where at least one recovery method resulted in an elation seeking preferences ($\hat{\beta}_{NLLS} < 0$ or $\hat{\beta}_{MMI} < 0$), Part 2 of the experiment enabled us to identify the extent of non-convexity of the underlying preferences, in addition to driving a wedge between the two sets of parameters. To achieve this

²⁴A fundamental design requirement was that subjects would view the two related but distinct tasks in the same frame. Hence, the interface was designed so that the pairwise choice problems were presented in the same two-dimensional coordinate system as the budget lines task. Moreover, as most subjects view the pairwise choice as a more primitive task, the instructions were written so that the presentation of Part 1’s interface was through a natural extension of a pairwise choice task. See the instructions in Appendix F.1.

additional goal we note that for locally non-convex preferences the certainty equivalent may exceed the expected value for some risky portfolios. Therefore, the pairwise choice procedure searched for a risky portfolio x^R , such that $CE_j(x^R) < E[x^R] < CE_i(x^R)$ and chose the safe portfolio, x^S such that $x^S = E[x^R]$.²⁵ Similarly to the mid-point design, choice of the risky portfolio reveals that the set of parameters $\hat{\theta}_i$ better approximates the DM's preferences, while choosing the safe portfolio reveals the opposite. In addition, the choice of the safe (risky) portfolio reveals local risk aversion (seeking) in the neighborhood of the portfolio x^R , providing a direct evidence to the extent of non convexity of the underlying DM's preferences.²⁶

To investigate the nature of *local* risk attitudes across subjects, the pairwise choice problems were constructed so that in 6 of them the risky portfolio was of *low variability* while in the other 3 problems, the risky portfolio was of *high variability*. For a detailed description of the algorithm that constructs the pairwise choices see Appendix F.3.

6.4 Incentive Compatibility

Finally, two comments regarding the incentive compatibility of this design. First, since this is a chained experimental design, had subjects been aware that parts of the experiment are connected and understood the precise structure of the pairwise choice procedure, they may have been able to manipulate their choices in order to maximize their expected gains. We are confident that this is not the case since the instructions and the experimental procedure were designed carefully not to reveal that the portfolios offered in Part 2 were calculated based on the choices in Part 1. Moreover, an extremely detailed knowledge of the experimental design and the recovery procedures is essential in order to manipulate the choices successfully.

²⁵Since risk attitude depends on both β and ρ it is possible to have $\beta < 0$ and have the associated utility function exhibit risk aversion with respect to some risky portfolio. However, $\beta < 0$ is sufficient for a utility function to display, at least locally, risk seeking behavior with respect to portfolios with small variance.

²⁶The safe portfolio was the preferred alternative by the MMI recovery method in 927 of the 1827 pairwise choices in our sample (50.7%).

Second, subjects were paid according to their decision in a randomly selected problem. If subjects isolate their decisions in different problems this payment system is incentive compatible. If they had integrated their decisions (by reducing the compound lottery induced by the random incentive system and their decisions), their choices would have been biased towards expected utility behavior ($\beta = 0$), a pattern observed for only about 40% of the subjects, as will be shown in Section 8.2.

7 Results: Pairwise Choice

The results of Part 1 of the experiment exhibit patterns broadly similar to those reported in Section 5 for the data sets gathered by Choi et al. (2007) (see Appendix G).²⁷ We use these results extensively (together with the results of Choi et al. (2007)) in Section 8 to demonstrate several important implications of Theorem 2.

The current section, however, is devoted to the results from Part 2 of the experiment. This part was designed so that in each pairwise comparison, one of the portfolios is preferred according to the recovered parameters of the MMI(SSQ) and the other is preferred according to the recovered parameters of the NLLS. Hence, in this section we analyze the choices of the subjects to infer on the relative predictive accuracy of the two recovery methods.

The results provided here are based on the full sample. As the complete sample includes subjects and choices that arguably should not be included in such a comparison (as the choices in Part 1 are too inconsistent or the algorithm could not meaningfully separate the recovery methods), Appendix H reports similar results for a refined sample.

In the following, statistical significance is defined with respect to the null hypothesis that MMI predictions are not better than random predictions, which entails a one-sided binomial test. The p -values should be interpreted as

²⁷The data gathered in the experiment are available in a separate Excel file named “Halevy et al (2016) - Data”. The disaggregated results of Part 1 are available in a separate Excel file named “Halevy et al (2016) Part 1 - Results”.

	# of Observations	Correct Predictions by MMI (%)	p -value
Complete Sample	1827	986 (54.0%)	0.0004
Low-variability	1218	652 (53.5%)	0.0074
High-variability	609	334 (54.8%)	0.0093

Table 2: Aggregate Results

the likelihood that the MMI correctly predicts x or more out of n choices correctly by chance alone. Results are reported at the aggregate and individual levels.

7.1 Results

7.1.1 Aggregate Results

In the aggregate analysis we treat all observations as a single data set. The first row of Table 2 reports the predictive success of the MMI recovery method over all 1827 observations (203 subjects times 9 observations per subject). The next two rows report similar results for the low-variability and high-variability portfolios separately. These results suggest that the MMI is a significantly (p -value smaller than 1%) better predictor of subjects' choices both overall and for the two sub-classes of portfolios separately (at an odds ratio of approximately 1.17).

7.1.2 Individual Results

For the individual level analysis each subject is treated as a single data point. Denote the number of correct MMI predictions by X . With only 9 choices per subject it may be difficult to declare one of the two methods as decisively better for moderate values ($X \in \{3, 4, 5, 6\}$), as the probability to get each one of these values at random is greater than 15%. Hence, Table 3 reports the number of subjects for whom one method was decisively better - able to predict more than two thirds of the choices correctly ($X \in \{0, 1, 2, 7, 8, 9\}$).

There are 103 subjects for which one recovery method was decisively better. The probability that one recovery method would be decisively better by

Subjects in Complete Sample (203)		
$X \geq 7$	$X \leq 2$	p -value
61	42	0.0378

Table 3: Individual Level Results

random prediction alone for a single subject is approximately 18%, so the probability of having 103 decisive predictions out of 203 subjects is close to zero. One preliminary conclusion is that our design and algorithm was able to separate the predictions made by NLLS and MMI effectively.

The empirical distribution of correct MMI predictions is significantly different from a null-hypothesis of random prediction.²⁸ As is evident from Table 3, MMI is a significantly better predictor at the individual level as well (one-sided p -value²⁹ 0.038), as it is decisively better predictor for 45% more subjects than NLLS.

7.2 Disappointment Aversion

7.2.1 Definite vs. Indefinite Disappointment Aversion

To further our understanding of the results we divide the sample into two classes according to the recovered parameters. The Definite Disappointment Averse (DDA) group is composed of those subjects for which both methods recover $\beta \geq 0$, whereas the Indefinite Disappointment Averse (IDA) group is composed of those subjects for which β is negative for one or both recovery methods. The DDA group includes 150 subjects while the other 53 subjects belong to the IDA group.

²⁸The statistic for the multinomial likelihood ratio test is $-2 \ln(L/R) = -2 \sum_{i=1}^k x_i \ln(\pi_i/p_i)$ where the categories are the number of correct predictions by the MMI, π_i is the theoretical probability of category i if the prediction is random while p_i is the frequency of category i in the data. This statistic for the complete sample equals 85.523 which, by a chi-squared distribution with 9 degrees of freedom has a p -value of approximately zero. Pearson’s chi-squared test provides similar results.

²⁹The p -value in the third column is calculated for the group of 103 subjects for whom one recovery method was decisively better than the other, under the null hypothesis that each recovery method has an equal chance to be decisive.

	# Observations	# Correct Predictions by MMI	% Correct Predictions by MMI	<i>p</i> -value
DDA	1350	706	52.3%	0.0484
IDA	477	280	58.7%	< 0.0001

Table 4: Aggregate Results by Group.

DDA (150)			IDA (53)		
$X \geq 7$	$X \leq 2$	<i>p</i> -value	$X \geq 7$	$X \leq 2$	<i>p</i> -value
38	30	0.1981	23	12	0.0448

Table 5: Individual Level Results by Group

In the aggregate analysis we treat the whole set of observations as a single data set with 1350 observations for the DDA group and 477 for the IDA group. Table 4 demonstrates that the MMI recovery method remains a better predictor in both groups. When the sample includes only the DDA group the advantage of the MMI is significant at 5% level (but the advantage is not significance in the refined sample, see Table 12 in Appendix H.3). However, when the sample includes only the IDA group, the advantage of the MMI recovery method is highly significant in spite of the smaller sample size (and is robust to the refinement).

At the individual level Table 5 shows that although the MMI recovery method predicts decisively better than NLLS in both DDA and IDA, the difference in predictive accuracy within the DDA group is insignificant. However, the difference within the IDA group is substantial and statistically significant as MMI predicts decisively for almost twice as many subjects for which NLLS predicts decisively.

7.2.2 Definite Elation Seeking

Further, we focus on a subset of IDA group, referred to as the Definite Elation Seeking (DES) group, that includes the 29 subjects for whom both recovery methods recover $\beta < 0$. The MMI recovery method predicted correctly 163 of the 261 choice problems these subjects encountered, which amount to 62.5% of the observations. Hence, the difference between the recovery methods within

the DES group is even more substantial than in the whole IDA group and it is highly significant (p -value smaller than 0.0001).

The individual results are similar: for 20 out of the 29 subjects in the DES group, one recovery method predicted decisively better (more than two thirds of pairwise choices) than the other, and for 75% of them (15 out of 20) the MMI produced the better prediction (p -value equals 0.0207). These results suggest that the difference in predictive success between the MMI and NLLS recovery methods can be attributed mostly (but not only) to subjects for which the recovery methods resulted in apparent non-convex preferences.

7.2.3 MMI vs. NLLS when Preferences are Non-convex

The pairwise comparisons in Part 2 of the experiment allow us to directly observe the subject's preferences in these non-convex regions of their indifference curves. Our results imply that the MMI recovers a significantly more accurate representation of subject preferences when the underlying preferences are non-convex.

Specifically, for 21 of the 29 subjects in the DES group (72.4%) the disappointment aversion parameter recovered by the NLLS is more negative than the one recovered by the MMI.³⁰ While we cannot conclude that NLLS *systematically* overstates the extent of elation seeking, this pattern of differences does correspond to particular patterns of choices observed in Part 1 of the experiment. Figure 7.1 illustrates the choices from Part 1 of the experiment for three characteristic subjects as well as their corresponding parameter estimates. Generally, as the subject's choices drift farther from the certainty line the greater is the difference between the parameter recovered by the NLLS and the MMI recovery methods.

³⁰For 19 of these 21 subjects the difference is more than 0.1. For 6 of the 8 subjects where the parameter recovered by the NLLS is less negative than the one recovered by the MMI, the difference is less than 0.1.

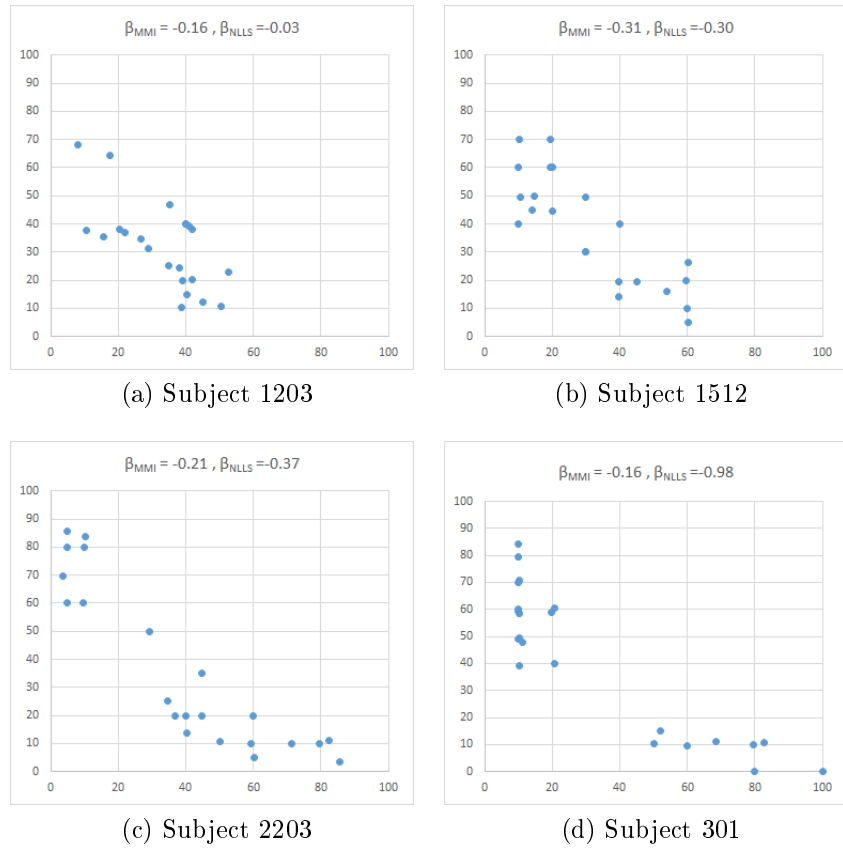


Figure 7.1: Patterns of Choice - Non-convex Preferences

7.3 Illustrative Discussion

To conclude this section we wish to suggest an informal explanation for our finding. Briefly, when choices exhibit non-convex preferences (in our context, elation seeking behavior), many parametric utility functions can provide an equally good approximation of the underlying preferences. In these cases, the NLLS recovery method will most probably pick a set of parameters that imply greater non-convexity than implied by the set of parameters recovered by the MMI method. The results of Part 2 of the experiment suggest that the parameters recovered by the MMI are considerably better in predicting the subjects' choices in the non-convex region.

To demonstrate the multiplicity of approximated preferences given the

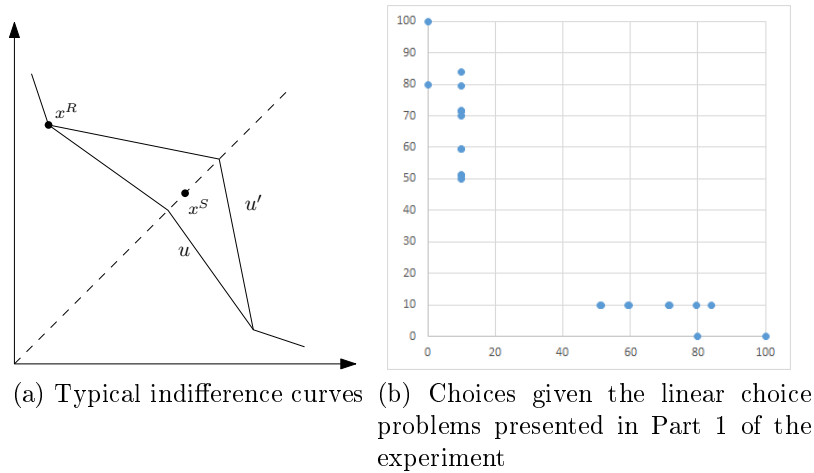


Figure 7.2: Two Simulated Subjects

same data set, consider two simulated subjects with preferences represented by the utility functions u and u' with the characteristic indifference curves shown in Figure 7.2a. Faced with the same sequence of linear budget sets as our subjects in Part 1 of the experiment, the implied optimal choices for these simulated subjects are *exactly the same* and are illustrated in Figure 7.2b.³¹ This pattern of choices is highly structured and may result from a reasonable heuristic according to which the subject wants to guarantee a payment of 10 tokens, but is willing to bet with the remainder of her income on the cheaper asset (unless the relative prices are extreme). In order to accommodate this behavior, NLLS resorts to substantial non-convexity while the MMI can rationalize these choices within the DA model without making strong claims on behavior that is unobservable using linear budget lines. For an informal demonstration, see Appendix I.

³¹Notice that the pattern of choice for these simulated subjects is very similar to Subject 301 in Figure 7.1d. Not surprisingly, the recovered parameters for our simulated subject are also very similar to Subject 301, $\beta_{MMI} = -0.24$, $\rho_{MMI} = 0.40$, $\beta_{NLLS} = -0.91$, $\rho_{NLLS} = 1.55$.

8 Results: Choice from Budget Lines

The usage of the MMI as a recovery method relies on the observation that it can be decomposed into an inconsistency index, which is independent of the specific utility function evaluated, and a misspecification index – which depends on the subset of utility functions considered. Given two parametric families \mathcal{U} and \mathcal{U}' , a researcher will calculate the value of the MMI loss index for each family ($I_M(D, f, \mathcal{U}')$ and $I_M(D, f, \mathcal{U})$), and since both incorporate the same inconsistency measure - $I_V(D, f)$, the data set D may be better approximated by \mathcal{U} or \mathcal{U}' depending on the magnitude of the loss index. Moreover, an important implication of Fact 5 is that if we impose an additional parametric restriction on preferences, the misspecification will necessarily (weakly) increase. If \mathcal{U}' is nested within \mathcal{U} , the difference between the value of the loss indices at \mathcal{U} and \mathcal{U}' is a measure of the marginal misspecification implied by the restriction of \mathcal{U} to \mathcal{U}' .

In this section we demonstrate the application of these insights for evaluating nested and non-nested model restrictions in the two experimental data sets. We perform a subject level analysis for the data collected in Part 1 of the experiment and the data collected by Choi et al. (2007). We begin by evaluating the misspecification implied by the Disappointment Aversion functional form (with *CRRA* and *CARA* utility functions). Then we demonstrate the evaluation of nested parametric restrictions by measuring the misspecification implied by restricting the functional form to expected utility. Finally, we compare the *CRRA* and *CARA* functional forms as an example for the evaluation of non-nested model restrictions.³²

8.1 Evaluating Misspecification

Using the decomposition of the Money Metric Index into the Varian Inconsistency Index (measure of consistency) and a residual which measures misspec-

³²For conciseness, throughout this Section we use the SSQ aggregator. Similar calculations are available using the MEAN aggregator in the results file “Choi et al (2007) - Results” and “Halevy et al (2016) Part 1 - Results”.

	Part 1 of the Experiment		Choi et al. (2007)	
Original Sample	203 subjects		47 subjects	
Consistent	92 (45%)		12 (26%)	
Dropped	3 (1.5%)		9 (19%)	
Inconsistency Level	at most 6%		at most 2.5%	
Utility index	CRRA	CARA	CRRA	CARA
# of Subjects with at most 5% misspecification	136 (68%)	127 (63.5%)	26 (68.4%)	23 (60.5%)
# of Subjects with at least 10% misspecification	4 (2%)	10 (5%)	3 (7.9%)	6 (15.8%)
Subjects for whom misspecification is more than 90% of the MMI	149 (74.5%)	153 (76.5%)	26 (68.4%)	27 (71.1%)
Subjects for whom misspecification is less than 50% of the MMI	0 (0 %)	0 (0 %)	1 (2.6%)	1 (2.6%)

The sample includes all the subjects for whom Varian Inconsistency Index was calculated exactly or with good approximation.

Table 6: Misspecification using the Disappointment Aversion functional form (with *CRRA* or *CARA*).

ification, we can calculate the misspecification for each subject.

One practical challenge is that the calculation of the Varian Inconsistency Index is computationally hard. However, as discussed in detail in Appendix B.1, we are able to calculate the exact values (or very good approximations) of this index for most of the subjects in the two samples.

Table 6 provides some descriptive statistics on the misspecification in the recovered preferences of subjects for whom the Varian Inconsistency Index was calculated exactly or with tight approximation. It demonstrates that for approximately two thirds of them, the Disappointment Aversion model entails less than 5% misspecification. In addition, Table 6 provides a preliminary evidence that, on an aggregate level, the Disappointment Aversion may be more misspecified with *CARA* than with *CRRA*.

The bottom two rows of Table 6 suggest that in both samples, the portion of misspecification in the loss index is considerably larger than the portion of inconsistency. In fact, there are almost no subjects for whom the portion of

	Part 1 of the Experiment	Choi et al. (2007)
<i>CRRA</i>	40.8% (80 of 196)	32.4% (11 of 34)
<i>CARA</i>	44.7% (85 of 190)	45.2% (14 of 31)

Percentage of subjects for which the additional misspecification implied by the expected utility restriction is less than 10% (number of subjects that are well approximated by expected utility out of the number of subjects in the sample).

Table 7: Evaluating the restriction to expected utility.

inconsistency is larger than the portion of misspecification.³³

8.2 Evaluating a Restriction to Expected Utility

Expected utility is nested within the disappointment aversion model, satisfying the restriction that $\beta = 0$. We evaluate whether or not this restriction is justified by examining the additional misspecification implied by this restriction.³⁴ Given the choice of functional form (Disappointment Aversion with *CRRA* or *CARA* utility index), we use the ratio $\frac{I_M(D,f,EU) - I_M(D,f,DA)}{I_M(D,f,DA) - I_V(D,f)}$ where *DA* stands for the Disappointment Aversion (unrestricted) model, *EU* stands for the expected utility model and *f* is the chosen aggregator.

If the restriction to expected utility implies a proportional increase in the misspecification of more than 10% then we tend to reject the expected utility specification. Included in the sample are subjects whose Varian Inconsistency Index was calculated exactly or with good approximation and whose measured misspecification of the disappointment aversion model was less than 10%, implying that it is a reasonable model to capture their choices.

The results in Table 7 demonstrate that choices of between one third and

³³Since those subjects for whom the Varian Inconsistency Index could not have been calculated properly were dropped, the sample slightly over-represents the less inconsistent subjects.

³⁴In the results files (“Choi et al (2007) - Results” and “Halevy et al (2016) Part 1 - Results”) we include descriptive statistics of the parameter frequencies in 1000 re-samplings of each individual data set in every reported recovery scheme. Potentially, we could have used these distributions to evaluate whether the restriction can be rejected. However, since we do not provide any proof that these re-samplings indeed recover confidence sets for the parameters, we merely interpret them as a measure for the sensitivity of the recovered parameters to extreme observations.

	Part 1 of the Experiment	Choi et al. (2007)
Full Sample*	71.4% (145 of 203)	80.9% (38 of 47)
Restricted Sample**	88% (103 of 117)	80% (24 of 30)

The percentage of subjects with lower misspecification using *CRRA* than *CARA* (number of subjects better approximated by *CRRA* than *CARA* out of the number of subjects in the sample).

*Includes all subjects for whom the loss function was calculated.

**Includes subjects whose Varian Inefficiency Index was calculated exactly or with good approximation and that the difference in misspecification between the two indices is greater than 10%.

Table 8: Choice of utility index.

half of the subjects are well approximated by the expected utility model, while for the others (more than half) the restriction to expected utility implies substantial increase in misspecification.

8.3 Comparison of Non-nested Alternatives

The Money Metric Index also allows the researcher to evaluate non-nested alternatives. Here, we compare two utility indices for the Disappointment Aversion functional form - *CRRA* and *CARA*. We can calculate the extent of misspecification implied by each functional form and select the functional form that represents a decision maker's preferences best on a subject by subject basis.

Table 8 reports that choices made by about three quarters of subjects are better approximated by the Disappointment Aversion model with *CRRA* than with *CARA* utility index.

This result strengthens if we restrict the samples to include only those subjects who are not too inconsistent (i.e. the Varian Inefficiency Index was calculated exactly or with good approximation) and the difference between the models is substantial (i.e. the difference in misspecification between the two models is greater than 10%).

9 Conclusions

This paper proposes a general methodology to structurally recover parameters (in the current study – preferences) based on minimizing the incompatibility between the ranking information encoded in choices and the ranking induced by a candidate structural model (here – utility function). We show that this incompatibility can be decomposed into an inconsistency index, which measures how far is the data from optimizing behavior (GARP), and a remainder which captures the model’s misspecification - which is in the researcher’s control. This approach is applicable to a variety of incompatibility indices and aggregator functions.

We demonstrate the proposed method in an environment of choice under risk and show that it may lead to different recovered parameters than standard NLLS, which represents recovery methods that minimize the distance between the observed data and the model’s prediction. In order to compare the two methods based on an objective criterion we design and execute an experiment that distinguishes between the methods based on their predictive success in out-of-sample pairwise comparisons. The results demonstrate that the proposed recovery method does a better job in predicting choices, especially when choices imply non-convex preferences – an environment in which minimizing the distance between observed and predicted choices is problematic. Although the goal of the experiment is to distinguish parametric recovery methods, it is fully based on subject’s choices: her choices in Part 1 (choice from linear budgets sets) determine the pairwise comparisons she will face in Part 2, and her choices in the latter part inform an outside observer which recovery method provides better predictions. Moreover, choice made in pairwise comparisons reveal preferences in their purest form, and permit their identification in scenarios where other elicitation methods can only provide bounds.

The empirical analysis followed the theoretical decomposition result, which allows a researcher to evaluate the change in misspecification implied by nested and non-nested models. In the context of choice under risk, we demonstrate

the relative importance of misspecification relative to inconsistency, and that although a non-negligible minority of the subjects are well approximated by the expected utility model, the choices of the majority of subjects are better approximated by a more general model of non-expected utility.

The current investigation includes theoretical foundations, empirical implications and experimental evaluation, but we view it only as a necessary first-step in integrating insights from revealed preference theory into otherwise standard structural recovery problems in Economics. The model selected here is simple (utility maximization), yet central in Economics and Finance. The implied non-convexities are non-coincidental, as they result from a reasonable calculated procedure. We believe that an important next step in this research program is the integration of stochastic component into the present deterministic model, while retaining the crucial distinction between inconsistency and misspecification.

A Proofs

A.1 Fact 1.

Let $\mathbf{v}' \leq \mathbf{v}$. Then: $R_{D,\mathbf{v}'}^0 \subseteq R_{D,\mathbf{v}}^0$, $P_{D,\mathbf{v}'}^0 \subseteq P_{D,\mathbf{v}}^0$ and $R_{D,\mathbf{v}'} \subseteq R_{D,\mathbf{v}}$.

Proof. For example, the proof of $R_{D,\mathbf{v}'}^0 \subseteq R_{D,\mathbf{v}}^0$ is as follows: If $x = x^i$ the statement holds by Definition 1.1. Otherwise, if $x^i R_{D,\mathbf{v}'}^0 x$ then $v'^i p^i x^i \geq p^i x$. $\mathbf{v}' \leq \mathbf{v}$ implies that for every observation i , $v'^i \leq v^i$. Therefore, $v^i p^i x^i \geq p^i x$, meaning $x^i R_{D,\mathbf{v}}^0 x$. \square

A.2 Fact 2.

Every D satisfies $GARP_0$.

Proof. For every pair of observed bundles x^i and x^j , $x^j P_{D,0}^0 x^i$ is false since for every bundle x , $p^j x \geq 0 = 0 \times p^j x^j$ ($P_{D,0}^0$ is the empty relation). \square

A.3 Fact 3.

Let $\mathbf{v}, \mathbf{v}' \in [0, 1]^n$ and $\mathbf{v} \geq \mathbf{v}'$. If D satisfies $GARP_{\mathbf{v}}$ then D satisfies $GARP_{\mathbf{v}'}$.

Proof. By Fact 1, for every pair of observed bundles x^i and x^j , $x^i R_{D,\mathbf{v}'} x^j$ implies $x^i R_{D,\mathbf{v}} x^j$. By Definition 2, since D satisfies $GARP_{\mathbf{v}}$ for every pair of observed bundles x^i and x^j , $x^i R_{D,\mathbf{v}'} x^j$ implies not $x^j P_{D,\mathbf{v}}^0 x^i$. By Fact 1, for every pair of observed bundles x^i and x^j , $x^i R_{D,\mathbf{v}'} x^j$ implies not $x^j P_{D,\mathbf{v}'}^0 x^i$. Therefore, D satisfies $GARP_{\mathbf{v}'}$. \square

A.4 Theorem 1.

Notation. Let \succeq be a binary relation. Then, \succ is defined as $x \succ y$ if and only if $x \succeq y$ and not $y \succeq x$, while \sim is defined as $x \sim y$ if and only if $x \succeq y$ and $y \succeq x$. Denote by X/\sim the set of all equivalence classes on X induced by \sim . Also, denote by \succeq^* the transitive closure of \succeq and by \succeq^c the relation where $x \succeq^c y$ if and only if $y \succeq x$.

Definition 12. Let $\mathbf{v} \in [0, 1]^n$. A transitive and reflexive binary relation \succeq *v-rationalizes-by-relation* D , if $R_{D, \mathbf{v}}^0 \subseteq \succeq$ and $P_{D, \mathbf{v}}^0 \subseteq \succ$.

Notation. Let $x \in \mathfrak{R}_+^K$ and $\delta > 0$. $B_\delta(x) = \{y \in \mathfrak{R}_+^K : \|y - x\| < \delta\}$.

Definition 13. A utility function $u : \mathfrak{R}_+^K \rightarrow \mathfrak{R}$ is

1. *Locally non-satiated* if $\forall x \in \mathfrak{R}_+^K$ and $\forall \epsilon > 0$, $\exists y \in B_\epsilon(x)$ such that $u(x) < u(y)$.
2. *Continuous* if $\forall x \in \mathfrak{R}_+^K$ and $\forall \epsilon > 0$ there exists $\delta > 0$ such that $y \in B_\delta(x)$ implies $u(y) \in B_\epsilon(u(x))$.
3. *Acceptable* if $\forall x \in \mathfrak{R}_+^K$, $u(0) \leq u(x)$.³⁵
4. *Monotone* if $\forall x, y \in \mathfrak{R}_+^K$, $x \leq y$ implies $u(x) \leq u(y)$.
5. *Concave* if $\forall x, y \in \mathfrak{R}_+^K$ and $0 \leq \alpha \leq 1$: $\alpha u(x) + (1 - \alpha)u(y) \leq u(\alpha x + (1 - \alpha)y)$.

Lemma 1. Let \succeq be transitive and reflexive binary relation on a set X . Then, there exists a complete, transitive and reflexive binary relation \succeq' on X such that $\succeq \subseteq \succeq'$ and $\succ \subseteq \succ'$.

Proof. Construct the mapping $\Pi : X \rightarrow X/\sim$ where each element of X is mapped into its equivalence class (the Canonical Projection Map). Consider the relation $\bar{\succeq}$ on X/\sim where $x \succ y$ implies $\Pi(x) \bar{\succeq} \Pi(y)$. $\bar{\succeq}$ is reflexive and transitive since \succeq is reflexive and transitive. Also, $\bar{\succeq}$ is antisymmetric since if $x \sim y$ then $\Pi(x) = \Pi(y)$. By Szpilrajn (1930)'s Extension Theorem, there is a complete, transitive, reflexive and antisymmetric binary relation, $\bar{\succeq}'$, such that $\bar{\succeq} \subseteq \bar{\succeq}'$. Consider now the relation \succ' on X where $\Pi(x) \bar{\succeq}' \Pi(y)$ implies $x \succ' y$. \succ' is complete, reflexive and transitive since $\bar{\succeq}'$ is complete, reflexive and transitive. Also, suppose $x \succ y$, then, by the first construction, $\Pi(x) \bar{\succeq} \Pi(y)$,

³⁵For every $D = \{(p^i, x^i)_{i=1}^n\}$ and for every $\mathbf{v} \in [0, 1]^n$, $\forall i \in 1, \dots, n : x^i R_{D, \mathbf{v}}^0 0$ (where 0 is the zero bundle). Therefore, a necessary condition for a binary relation \succeq to *v-rationalize-by-relation* D is that for every observed bundle $x \in \mathfrak{R}_+^K$, $x \succeq 0$. Similarly, for a utility function $u(x)$ to *v-rationalize* D it must be that for every observed bundle $x \in \mathfrak{R}_+^K$, $u(x) \geq u(0)$.

by the Extension Theorem $\Pi(x) \succeq' \Pi(y)$ and by the second construction $x \succeq' y$. Therefore $\succeq \subseteq \succeq'$. Similarly, $\succ \subseteq \succ'$. \square

Lemma 2. *Let R and P be two arbitrary binary relations on X . The following statements are equivalent:*

1. *There exists a transitive and reflexive binary relation \succeq on X such that $R \subseteq \succeq$ and $P \subseteq \succ$.*
2. *There exists a complete, transitive and reflexive binary relation \succeq' on X such that $R \subseteq \succeq'$ and $P \subseteq \succ'$.*
3. *$(R \cup P)^* \cap P^c = \emptyset$.*

Proof. By Lemma 1, the first two statements are equivalent. Next, suppose (1) holds. Then, $(R \cup P) \subseteq (\succeq \cup \succ)$ and therefore $(R \cup P)^* \subseteq (\succeq \cup \succ)^*$. Also, $P^c \subseteq \succ^c$. Therefore, $(R \cup P)^* \cap P^c \subseteq (\succeq \cup \succ)^* \cap \succ^c$. Since $\succ \subseteq \succeq$ and since \succeq is transitive we get $(R \cup P)^* \cap P^c \subseteq \succeq \cap \succ^c$. But, $\succeq \cap \succ^c = \emptyset$ and hence $(R \cup P)^* \cap P^c = \emptyset$.

Last, suppose (3) holds. We construct a transitive and reflexive binary relation \succeq on X such that $R \subseteq \succeq$ and $P \subseteq \succ$. Let \succeq be such that $x \succeq y$ if and only if $x(R \cup P)^*y$ or $x = y$. \succeq is reflexive by definition and transitive since $(R \cup P)^*$ is transitive. Moreover, since $R \subseteq (R \cup P)^*$ and $P \subseteq (R \cup P)^*$ then $R \subseteq \succeq$ and $P \subseteq \succeq$. It is left to show that $P \subseteq \succ$. Suppose xPy . Since $P \subseteq \succeq$ then $x \succeq y$. Moreover, since xPy then yP^cx and since $(R \cup P)^* \cap P^c = \emptyset$ we get that it cannot be that $y(R \cup P)^*x$. In particular, it cannot be that yPx and therefore $x \neq y$. Thus, by the definition of \succeq , it cannot be that $y \succeq x$. Therefore, $x \succ y$ and we conclude that $P \subseteq \succ$. \square

Lemma 3. *Let $\mathbf{v} \in [0, 1]^n$ and let $D = \{(p^i, x^i)_{i=1}^n\}$ be a finite data set of choices from budget sets. The following statements are equivalent:*

1. *There exists a transitive and reflexive binary relation \succeq on \mathfrak{R}_+^K such that \succeq \mathbf{v} -relation-rationalizes D .*
2. *There exists a complete, transitive and reflexive binary relation \succeq' on \mathfrak{R}_+^K such that \succeq' \mathbf{v} -relation-rationalizes D .*
3. *D satisfies $GARP_{\mathbf{v}}$.*

Proof. By Lemma 2 and Definition 12 if $X = \mathfrak{R}_+^K$, $R = R_{D,\mathbf{v}}^0$ and $P = P_{D,\mathbf{v}}^0$, then the first two statements are equivalent and both are also equivalent to $(R_{D,\mathbf{v}}^0 \cup P_{D,\mathbf{v}}^0)^* \cap P_{D,\mathbf{v}}^{0,c} = \emptyset$. But, $(R_{D,\mathbf{v}}^0 \cup P_{D,\mathbf{v}}^0)^* \cap P_{D,\mathbf{v}}^{0,c} = \emptyset$ holds if and only if for every pair of bundles x and y , $xR_{D,\mathbf{v}}y$ implies not $yP_{D,\mathbf{v}}^0x$. If x is an unobserved bundle or y is an unobserved bundle then by Definition 1, $xR_{D,\mathbf{v}}y$ implies not $yP_{D,\mathbf{v}}^0x$. Therefore, $(R_{D,\mathbf{v}}^0 \cup P_{D,\mathbf{v}}^0)^* \cap P_{D,\mathbf{v}}^{0,c} = \emptyset$ holds if and only if for every pair of observed bundles x and y , $xR_{D,\mathbf{v}}y$ implies not $yP_{D,\mathbf{v}}^0x$. Hence, by Definition 2, $(R_{D,\mathbf{v}}^0 \cup P_{D,\mathbf{v}}^0)^* \cap P_{D,\mathbf{v}}^{0,c} = \emptyset$ holds if and only if D satisfies $GARP_{\mathbf{v}}$. \square

Lemma 4. Let $D = \{(p^i, x^i)_{i=1}^n\}$ be a finite data set and let $\{(z_i : \mathfrak{R}_+^K \rightarrow \mathfrak{R})_{i=1}^n\}$ be a family of real functions. Define the following two binary relations on $\{(x^i)_{i=1}^n\}$: $x^i R x^j \Leftrightarrow z_i(x^j) \leq 0$ and $x^i P x^j \Leftrightarrow z_i(x^j) < 0$. If there exists a transitive and reflexive binary relation $\bar{\succeq}$ on $\{(x^i)_{i=1}^n\}$ such that $R \subseteq \bar{\succeq}$ and $P \subset \bar{\succ}$ then there exists a function $f(x) = \min_{i \in \{1, \dots, n\}} f_i + \lambda_i z_i(x)$ such that $\lambda_i > 0$ and $f(x^i) \geq f_i$.

Proof. By Lemma 2, there exists a complete, transitive and reflexive binary relation \succeq on $\{(x^i)_{i=1}^n\}$ such that $R \subseteq \succeq$ and $P \subset \succ$. Since \succeq is complete and transitive and $\{(x^i)_{i=1}^n\}$ is finite we can partition the observed bundles and rank them according to \succeq . Let $I = \{1, \dots, n\}$. Then $E_1 = \{i \in I \mid \nexists y \in \{(x^i)_{i \in I}\}, y \succ x^i\}$ is the set of indices of those observed bundles that are not dominated by any other observed bundle according to \succeq . Similarly, from the remaining observed bundles, $E_2 = \{i \in I/E_1 \mid \nexists y \in \{(x^i)_{i \in I/E_1}\}, y \succ x^i\}$, is the set of indices of those observed bundles that are not dominated according to \succeq by any other observed bundle, and so $E_3 = \{i \in I/(E_1 \cup E_2) \mid \nexists y \in \{(x^i)_{i \in I/(E_1 \cup E_2)}\}, y \succ x^i\}$, etc. Denote the number of classes by l . Transitivity guarantees that there are no empty classes while completeness assures that for every $k \in 1, \dots, l$ and for every pair of observed bundles $x, y \in E_k$ it must be that $x \sim y$.

The following procedure uses this partition and the functions $\{(z_i : \mathfrak{R}_+^K \rightarrow \mathfrak{R})_{i=1}^n\}$ to construct a mapping $(f_i, \lambda_i) : \{(x^i)_{i=1}^n\} \rightarrow \mathfrak{R}^2$ such that $\lambda_i > 0$ and $f(x^i) \geq f_i$ where $f(x) = \min_{i \in I} \{f_i + \lambda_i z_i(x)\}$:

1. For every $i \in E_1$, set $f_i = 1$ and $\lambda_i = 1$. Also, set $k = 1$. If $l = 1$ the procedure terminates, otherwise continue.

2. Set $k := k + 1$.

3. Denote $B_k = \cup_{m=1}^{k-1} E_m$.

4. Calculate $\alpha_k = \min_{i \in B_k} \min_{j \in E_k} \min \left\{ f_i + \lambda_i z_i(x^j), f_i \right\}$.

5. Choose some $f < \alpha_k$ and set $f_j = f$ for every $j \in E_k$.

6. Calculate $\beta_k = \max_{i \in B_k} \max_{j \in E_k} \frac{f_i - f_j}{z_j(x^i)}$.

7. Choose some $\lambda > \beta_k$ and set $\lambda_j = \lambda$ for every $j \in E_k$.

8. If $k < l$ return to step 2, otherwise the procedure terminates.

Stage 1 guarantees that for every $i \in E_1$, $\lambda_i = 1$ and $f_i = 1$. Suppose $i \in E_1$, $l \geq 2$ and $k \in \{2, \dots, l\}$. Then $i \in B_k$ and for every $j \in E_k$, $x^i \succ x^j$ (since \succeq is complete). Steps 4 and 5 guarantee that $x^i \succ x^j$ implies that $f_i > f_j$ or $f_i - f_j > 0$. In addition, $x^i \succ x^j$ implies that $z_j(x^i) > 0$ (otherwise $x^j R x^i$ and therefore $x^j \succeq x^i$). Therefore, steps 6 and 7 guarantee that for every observation $i \in I$, $\lambda_i > 0$. It is left to show that for every observation $i \in I$, $f(x^i) \geq f_i$. That is, $\min_{j \in I} [f_j + \lambda_j z_j(x^i)] \geq f_i$ or, equivalently, for every pair of observations $i, j \in I$, $f_j + \lambda_j z_j(x^i) \geq f_i$. First, if $x^j \succ x^i$ steps 4 and 5 guarantee that $f_j + \lambda_j z_j(x^i) > f_i$. If $x^j \sim x^i$ then $z_j(x^i) \geq 0$ (otherwise $x^j P x^i$ and therefore $x^j \succ x^i$) and in addition by step 5, $f_j = f_i$. Since for every $j \in I$, $\lambda_j > 0$ we get that $x^j \sim x^i$ implies $f_j + \lambda_j z_j(x^i) \geq f_i$. Last, if $x^i \succ x^j$ then $z_j(x^i) > 0$ and $f_i - f_j > 0$ and steps 6 and 7 guarantee that $\lambda_j > \frac{f_i - f_j}{z_j(x^i)}$. Therefore, $x^i \succ x^j$ implies $f_j + \lambda_j z_j(x^i) > f_i$. Thus, for every observation $i \in I$, $f(x^i) \geq f_i$. If $l = 1$ then for every pair of observations $i, j \in I$, we have $x^i \sim x^j$. Therefore, for every pair of observations $i, j \in I$, $z_j(x^i) \geq 0$. In addition, for every $i \in I$, $\lambda_i = 1$ and $f_i = 1$. Hence, for every $i \in I$, $f(x^i) \geq f_i$. \square

Lemma 5. *If u is a locally non satiated utility function that \mathbf{v} -rationalizes $D = \{(p^i, x^i)_{i=1}^n\}$, then $x^i P_{D, \mathbf{v}}^0 x$ implies $u(x^i) > u(x)$.*

Proof. If $x^i P_{D, \mathbf{v}}^0 x$ then $x^i R_{D, \mathbf{v}}^0 x$. Since $u(\cdot)$ \mathbf{v} -rationalizes D , $x^i R_{D, \mathbf{v}}^0 x$ implies $u(x^i) \geq u(x)$. Suppose that $u(x^i) = u(x)$. Since $v^i p^i x^i > p^i x$, $\exists \epsilon > 0$ such that $\forall y \in B_\epsilon(x) : v^i p^i x^i > p^i y$. By local non-satiation $\exists y' \in B_\epsilon(x)$ such

that $u(y') > u(x) = u(x^i)$. Thus, y' is a bundle such that $v^i p^i x^i > p^i y'$ and $u(y') > u(x^i)$, in contradiction to $u(\cdot)$ \mathbf{v} -rationalizing D . Therefore, $u(x^i) > u(x)$. \square

We proceed to the proof of Theorem 1,

Proof. First, suppose there exists a locally non-satiated utility function $u(\cdot)$ that \mathbf{v} -rationalizes D . If, in negation, D does not satisfy $GARP_{\mathbf{v}}$ then, by Definition 2, there are two observed bundles x^i, x^j such that $x^i R_{D,\mathbf{v}} x^j$ and $x^j P_{D,\mathbf{v}}^0 x^i$. By Definition 1.3, $x^i R_{D,\mathbf{v}} x^j$ implies that there exists a sequence of observed bundles (x^k, \dots, x^m) such that $x^i R_{D,\mathbf{v}}^0 x^k, \dots, x^m R_{D,\mathbf{v}}^0 x^j$. Therefore, by Definition 3, $x^i R_{D,\mathbf{v}} x^j$ implies $u(x^i) \geq u(x^k) \geq \dots \geq u(x^m) \geq u(x^j)$, meaning $x^i R_{D,\mathbf{v}} x^j$ implies $u(x^i) \geq u(x^j)$. However, by Lemma 5, since $u(\cdot)$ is a locally non-satiated utility function that \mathbf{v} -rationalizes D , $x^j P_{D,\mathbf{v}}^0 x^i$ implies $u(x^j) > u(x^i)$. Contradiction. Therefore, D satisfies $GARP_{\mathbf{v}}$.

Since the third statement implies the first statement, it is left to be shown that if D satisfies $GARP_{\mathbf{v}}$ then there exists a continuous, concave, acceptable and monotone utility function that \mathbf{v} -rationalizes D .

By Lemma 3 and by Definition 12, we have to show that for every data set D and adjustments vector \mathbf{v} , if \succeq is a transitive and reflexive binary relation on \mathfrak{R}_+^K such that $R_{D,\mathbf{v}}^0 \subseteq \succeq$ and $P_{D,\mathbf{v}}^0 \subseteq \succ$ then there exists a continuous, concave, acceptable and monotone utility function that \mathbf{v} -rationalizes D .

Define $z_i(x) = \frac{1}{v_i} p^i x - p^i x^i$ if $x \neq x^i$ and zero otherwise. Then, $x^i R_{D,\mathbf{v}}^0 x \Leftrightarrow z_i(x) \leq 0$ and $x^i P_{D,\mathbf{v}}^0 x \Leftrightarrow z_i(x) < 0$. Thus, by Lemma 4, there exists a function $f(x) = \min_{i \in \{1, \dots, n\}} f_i + \lambda_i z_i(x)$ such that $\lambda_i > 0$ and $f(x^i) \geq f_i$.

Next we show that $f(\cdot)$ \mathbf{v} -rationalizes D . Suppose $x^i R_{D,\mathbf{v}}^0 x$. By the definition of f we get $f(x) \leq f_i + \lambda_i z_i(x)$. Since, $\lambda_i > 0$ and since $x^i R_{D,\mathbf{v}}^0 x$ we get $\lambda_i z_i(x) \leq 0$ and therefore $f(x) \leq f_i$. However, $f(x^i) \geq f_i$. Therefore, $x^i R_{D,\mathbf{v}}^0 x$ implies $f(x^i) \geq f(x)$, that is $f(\cdot)$ \mathbf{v} -rationalizes D .

The functions z_i are discontinuous at x^i when $v_i < 1$. Therefore, f is continuous everywhere except maybe at the observed bundles. We use f to construct a continuous utility function \hat{f} that \mathbf{v} -rationalizes D . Let $\hat{z}_i(x) = \lim_{y \rightarrow x} z_i(y)$ then $\hat{z}_i(x) \geq z_i(x)$ for $x = x^i$ and $\hat{z}_i(x) = z_i(x)$ otherwise.

Construct $\hat{f}(x) = \min_{i \in \{1, \dots, n\}} f_i + \lambda_i \hat{z}_i(x)$ where f_i and λ_i are the same as in f and therefore $\lambda_i > 0$ and $f(x^i) \geq f_i$. Note that $\hat{z}_j(x^i) \geq z_j(x^i) = 0$ for all $j \in \{1, \dots, n\}$ implies $\hat{f}(x^i) \geq f(x^i) \geq f_i$. If $x \neq x^i$ then $z_i(x) \leq 0$ implies $\hat{z}_i(x) \leq 0$ and therefore $\hat{f}(x) \leq f_i$. Hence, for every bundle $x \neq x^i$ such that $z_i(x) \leq 0$ we get $\hat{f}(x) \leq f_i$. Thus, for every bundle x such that $x^i R_{D, \mathbf{v}}^0 x$ we get $\hat{f}(x) \leq f_i$, that is \hat{f} \mathbf{v} -rationalizes D . Obviously, $\hat{z}_i(x)$ is continuous and therefore for every observation $i \in I$, $f_i + \lambda_i \hat{z}_i(x)$ is continuous. Since the minimum of any finite number of continuous functions is continuous we get that $\hat{f}(x) = \min_{i \in \{1, \dots, n\}} f_i + \lambda_i \hat{z}_i(x)$ is continuous.

For every $i \in I$, since $\hat{z}_i(x)$ is linear with positive slope, the zero bundle, $x = 0$, minimizes $f_i + \lambda_i \hat{z}_i(x)$. Therefore, $\hat{f}(0) = \min_{x \in \mathbb{R}_+^K} \hat{f}(x)$. Hence, \hat{f} satisfies acceptability. Also, since $\hat{z}_i(x)$ is increasing monotonically, for every observation $i \in I$, $f_i + \lambda_i \hat{z}_i(x)$ is increasing monotonically and therefore \hat{f} is monotonic. $\hat{z}_i(x)$ is linear and therefore for every observation $i \in I$, $f_i + \lambda_i \hat{z}_i(x)$ is linear. Since the minimum of a set of linear functions is concave, \hat{f} is concave. \square

A.5 Fact 4.

$I_V(D, f)$, $I_A(D)$ and $I_{HM}(D, f)$ always exist.

Proof. The aggregator function $f(\cdot)$ is bounded. In addition, by Fact 2, the sets $\{\mathbf{v} \in [0, 1]^n : D \text{ satisfies } GARP_{\mathbf{v}}\}$, $\{\mathbf{v} \in \mathcal{I} : D \text{ satisfies } GARP_{\mathbf{v}}\}$ and $\{\mathbf{v} \in \{0, 1\}^n : D \text{ satisfies } GARP_{\mathbf{v}}\}$ are non-empty. Hence, $I_V(D, f)$, $I_A(D)$ and $I_{HM}(D, f)$ always exist. \square

A.6 Proposition 1.

Let $D = \{(p^i, x^i)_{i=1}^n\}$, $u \in \mathcal{U}^c$ and $\mathbf{v} \in [0, 1]^n$. $u(\cdot)$ \mathbf{v} -rationalizes D if and only if $\mathbf{v} \leq \mathbf{v}^*(D, u)$.

Proof. First, let us show that if $u(\cdot)$ \mathbf{v} -rationalizes D then $\mathbf{v} \leq \mathbf{v}^*(D, u)$. Suppose that \mathbf{v} is such that $u(\cdot)$ \mathbf{v} -rationalizes D and for observation i , $v^i > v^{*i}(D, u)$. By Definition 3, $u(x^i) \geq u(x)$ for all x such that $v^i p^i x^i \geq p^i x$.

By Definition 8 and since $v^i > v^{*i}(D, u)$ we get that $v^i p^i x^i > m(x^i, p^i, u) = p^i x^*$ where $x^* \in \operatorname{argmin}_{\{y \in \mathfrak{R}_+^K : u(y) \geq u(x^i)\}} p^i y$. It follows that $\exists \epsilon > 0$ such that $\forall y \in B_\epsilon(x^*) : v^i p^i x^i > p^i y$. By local non-satiation $\exists y' \in B_\epsilon(x^*)$ such that $u(y') > u(x^*) \geq u(x^i)$. Thus, y' is a bundle such that $v^i p^i x^i > p^i y'$ and $u(y') > u(x^i)$ contradicting that $u(\cdot)$ \mathbf{v} -rationalizes D .

Next, let us show that if $\mathbf{v} \leq \mathbf{v}^*(D, u)$ then $u(\cdot)$ \mathbf{v} -rationalizes D . We begin by establishing that $u(\cdot)$ $\mathbf{v}^*(D, u)$ -rationalizes D . Suppose, in negation, that for some observation $(p^i, x^i) \in D$ there exists a bundle x such that $x^i R_{D, \mathbf{v}^*(D, u)}^0 x$ and $u(x^i) < u(x)$. If $x = 0$ then we get a contradiction by acceptability. If $x \neq 0$ then by Definition 1.1, $v^{*i}(D, u) p^i x^i \geq p^i x$. By Definition 8, $m(x^i, p^i, u) \geq p^i x$. By continuity of $u(\cdot)$ there exists $\gamma > 0$ such that $u(x^i) < u((1 - \gamma)x)$. However, since $p^i(1 - \gamma)x < m(x^i, p^i, u)$, we reach a contradiction to Definition 8.

Finally, since $u(\cdot)$ $\mathbf{v}^*(D, u)$ -rationalizes D , for every observation $(p^i, x^i) \in D$, $v^{*i}(D, u) p^i x^i \geq p^i x$ implies $u(x^i) \geq u(x)$. Since $\mathbf{v} \leq \mathbf{v}^*(D, u)$, for every observation $(p^i, x^i) \in D$, $v^{*i}(D, u) p^i x^i \geq v^i p^i x^i$. Therefore, for every observation $(p^i, x^i) \in D$, $v^i p^i x^i \geq p^i x$ implies $u(x^i) \geq u(x)$. Hence, $u(\cdot)$ \mathbf{v} -rationalizes D . \square

A.7 Proposition 2

Let $D = \{(p^i, x^i)_{i=1}^n\}$, $u \in \mathcal{U}^c$ and $\mathbf{b} \in \{0, 1\}^n$. $u(\cdot)$ \mathbf{b} -rationalizes D if and only if $\mathbf{b} \leq \mathbf{b}^*(D, u)$.

Proof. First, let us show that if $u(\cdot)$ \mathbf{b} -rationalizes D then $\mathbf{b} \leq \mathbf{b}^*(D, u)$. Suppose, in negation, that \mathbf{b} is such that $u(\cdot)$ \mathbf{b} -rationalizes D and for observation i , $b^i = 1$ while $b^{*i}(D, u) = 0$. By Definition 10, $b^{*i}(D, u) = 0$ implies that there exists $y \in \mathfrak{R}_+^K$ such that $p^i x^i \geq p^i y$ and $u(y) > u(x^i)$. Thus, $x^i R_{D, \mathbf{b}}^0 y$ does not imply $u(x^i) \geq u(y)$, contradicting that $u(\cdot)$ \mathbf{b} -rationalizes D .

Next, let us show that if $\mathbf{b} \leq \mathbf{b}^*(D, u)$ then $u(\cdot)$ \mathbf{b} -rationalizes D . Since, $\mathbf{b} \leq \mathbf{b}^*(D, u)$, for every observation $(p^i, x^i) \in D$, $b^i = 1$ implies $b^{*i}(D, u) = 1$. By Definition 10, this means that $b^i p^i x^i \geq p^i x$ implies $u(x^i) \geq u(x)$. Otherwise, if $b^i = 0$ by the acceptability of $u(\cdot)$, $b^i p^i x^i \geq p^i x$ implies $u(x^i) \geq$

$u(x)$. Therefore, $b^i p^i x^i \geq p^i x$ implies $u(x^i) \geq u(x)$ and by Definition 1.1 $x^i R_{D, \mathbf{b}}^0 x$ implies $u(x^i) \geq u(x)$. Hence, by Definition 3 $u(\cdot)$ \mathbf{b} -rationalizes D . \square

A.8 Fact 5.

For every $\mathcal{U}' \subseteq \mathcal{U} : I_M(D, f, \mathcal{U}) \leq I_M(D, f, \mathcal{U}')$ and $I_B(D, f, \mathcal{U}) \leq I_B(D, f, \mathcal{U}')$.

Proof. $\mathcal{U}' \subseteq \mathcal{U}$ implies $\inf_{u \in \mathcal{U}'} f(\mathbf{v}^*(D, u)) \geq \inf_{u \in \mathcal{U}} f(\mathbf{v}^*(D, u))$ and therefore $I_M(D, f, \mathcal{U}) \leq I_M(D, f, \mathcal{U}')$ and similarly for the Binary Incompatibility Index. \square

A.9 Theorem 2.

Proof. We begin with the proof of part (1). First, we show that $I_V(D, f) \leq I_M(D, f, \mathcal{U}^c)$. If $I_V(D, f) = 0$ then by definitions 4 and 9 we get $I_V(D, f) \leq I_M(D, f, \mathcal{U}^c)$. Otherwise, if $I_V(D, f) > 0$, suppose that $I_V(D, f) > I_M(D, f, \mathcal{U}^c)$. Then, there exists $u \in \mathcal{U}^c$ such that $f(\mathbf{v}^*(D, u)) < I_V(D, f)$. By Proposition 1, $u(\cdot)$ $\mathbf{v}^*(D, u)$ -rationalizes D . By Theorem 1 D satisfies $GARP_{\mathbf{v}^*(D, u)}$. However, since D satisfies $GARP_{\mathbf{v}^*(D, u)}$ and $f(\mathbf{v}^*(D, u)) < I_V(D, f)$, $I_V(D, f)$ cannot be the infimum of $f(\cdot)$ on the set of all $\mathbf{v} \in [0, 1]^n$ such that D satisfies $GARP_{\mathbf{v}}$. Contradiction.

For the converse direction note that by Theorem 1, D satisfies $GARP_{\mathbf{v}}$ if and only if there exists $u \in \mathcal{U}^c$ that \mathbf{v} -rationalizes D . By Proposition 1, $\mathbf{v} \preceq \mathbf{v}^*(D, u)$. Since $f(\cdot)$ is weakly decreasing $f(\mathbf{v}^*(D, u)) \leq f(\mathbf{v})$. Therefore, by Definition 9, D satisfies $GARP_{\mathbf{v}}$ implies that $I_M(D, f, \mathcal{U}^c) \leq f(\mathbf{v})$. Since $I_V(D, f) = \inf_{\mathbf{v} \in [0, 1]^n : D \text{ satisfies } GARP_{\mathbf{v}}} f(\mathbf{v})$ we have $I_V(D, f) \geq I_M(D, f, \mathcal{U}^c)$. Hence, $I_V(D, f) = I_M(D, f, \mathcal{U}^c)$.

To prove part (2) we first show that $I_{HM}(D, f) \leq I_B(D, f, \mathcal{U}^c)$. If $I_{HM}(D, f) = 0$ by definitions 4 and 11 we get $I_{HM}(D, f) \leq I_B(D, f, \mathcal{U}^c)$. Otherwise, if $I_{HM}(D, f) > 0$ suppose that $I_{HM}(D, f) > I_B(D, f, \mathcal{U}^c)$. Then, there exists $u \in \mathcal{U}^c$ such that $f(\mathbf{b}^*(D, u)) < I_{HM}(D, f)$. By Proposition 2 $u(\cdot)$ $\mathbf{b}^*(D, u)$ -rationalizes D . By Theorem 1, D satisfies $GARP_{\mathbf{b}^*(D, u)}$. How-

ever, since D satisfies $GARP_{\mathbf{b}^*(D,u)}$ and $f(\mathbf{b}^*(D,u)) < I_{HM}(D,f)$, $I_{HM}(D,f)$ cannot be the infimum of $f(\cdot)$ on the set of all $\mathbf{v} \in \{0,1\}^n$ such that D satisfies $GARP_{\mathbf{v}}$. Contradiction.

Second, by Theorem 1, D satisfies $GARP_{\mathbf{b}}$ if and only if there exists $u \in \mathcal{U}^c$ that \mathbf{b} -rationalizes D . By Proposition 2, $\mathbf{b} \leq \mathbf{b}^*(D,u)$. Since $f(\cdot)$ is weakly decreasing $f(\mathbf{b}^*(D,u)) \leq f(\mathbf{b})$. Therefore, by Definition 11, D satisfies $GARP_{\mathbf{b}}$ implies that $I_B(D,f,\mathcal{U}^c) \leq f(\mathbf{b})$. Since $I_{HM}(D,f) = \inf_{\mathbf{v} \in \{0,1\}^n: D \text{ satisfies } GARP_{\mathbf{v}}} f(\mathbf{v})$ we have $I_{HM}(D,f) \geq I_B(D,f,\mathcal{U}^c)$. Hence, $I_{HM}(D,f) = I_B(D,f,\mathcal{U}^c)$.

We conclude with the proof of part (3). By part (1), since $f(\mathbf{v}) = 1 - \min_{i \in \{1, \dots, n\}} v_i$ is continuous and weakly decreasing then for every finite data set D , $I_V(D,f) = I_M(D,f,\mathcal{U}^c)$. By Definition 6, since $\mathcal{I} \subset [0,1]^n$ then if $f(\mathbf{v}) = 1 - \min_{i \in \{1, \dots, n\}} v_i$ we get $I_V(D,f) \leq I_A(D)$. Suppose that $I_V(D,f) < I_A(D)$, then there exists $\hat{\mathbf{v}} \in [0,1]^n$ such that D satisfies $GARP_{\hat{\mathbf{v}}}$ and $f(\hat{\mathbf{v}}) < I_A(D)$. By Fact 3, for every $\mathbf{v} \in [0,1]^n$ such that D satisfies $GARP_{\mathbf{v}}$ there exists $\mathbf{v}' \in \mathcal{I}$ such that D satisfies $GARP_{\mathbf{v}'}$ where $v'_i = \min_{i \in \{1, \dots, n\}} v_i$. Hence, there exists $\hat{\mathbf{v}}' \in \mathcal{I}$ such that D satisfies $GARP_{\hat{\mathbf{v}}'}$ and $f(\hat{\mathbf{v}}') = f(\hat{\mathbf{v}}) < I_A(D)$. Contradiction. \square

B Inconsistency Indices

This appendix provides detailed information regarding inconsistency indices mentioned or related to this work. Section B.1 describes the theoretical and practical computational issues concerning the indices analyzed in Theorem 2. Three important alternative inconsistency indices based on revealed preferences are discussed in Section B.2. A fourth alternative, which is not based on revealed preferences, is discussed in Section B.3.³⁶

³⁶We do not discuss indices based on the number of violations of the revealed preference axioms (see Swofford and Whitney (1987); Famulari (1995) and Harbaugh et al. (2001)) or indices based on the distance of the observed Slutsky matrix from the set of rational Slutsky matrices (see Jerison and Jerison, 1993; Aguiar and Serrano, 2015a,b).

B.1 Computation

Theorem 2 relates three inconsistency indices to loss functions used in the recovery of parametric preferences. Since an inconsistency index is constant (given a data set), its value is inconsequential to the selection of the best approximating function within a parametric family. However, the value of the index is necessary in order to determine the decomposition of the loss between the subject’s inconsistency and the researcher’s inaccuracy in her choice of functional form. Therefore, a practical consideration in the choice of a loss function is the computability of the corresponding inconsistency index.

B.1.1 Afriat’s Inconsistency Index

Theorem 3 in Afriat (1973) suggests an NP-Hard algorithm to calculate Afriat’s inconsistency index. Based on a similar idea, Smeulders et al. (2014) provide a polynomial time algorithm to calculate this index. Houtman and Maks (1987) describe an efficient binary search routine that approximates Afriat’s inconsistency index with an arbitrary accuracy in polynomial time.

In the supplemented code package we follow Houtman and Maks (1987). Let GL denote a lower bound on the index (initialized to zero) and let GU denote an upper bound on the index (initialized to one). At each iteration we cut the difference between the bounds by half, by testing the data for $GARP_{\frac{GU+GL}{2}}$ and updating the upper bound in case of a failure and the lower bound otherwise. l iterations guarantee an accuracy of approximately $\log_{10} 2^l \approx 0.3l$ significant decimal digits (we implement $l = 30$). Finally, we report GL .

B.1.2 Varian’s Inconsistency Index

The problem of finding the exact value of Varian’s Inconsistency Index is equivalent to solving the *minimum cost feedback arc set problem*.³⁷ Karp (1972) shows that this problem is NP-Hard and therefore finding the exact value of Varian’s Inconsistency Index is also NP-Hard (as mentioned in Varian

³⁷Given a directed and weighted graph, find the “cheapest” subset of arcs such that its removal turns the graph into an acyclic graph.

(1990)).³⁸ Moreover, [Smeulders et al. \(2014\)](#) show that no polynomial time algorithm can achieve a constant factor approximation (a ratio of $o(n^{1-\delta})$). [Tsur \(1989\)](#), [Varian \(1993\)](#) and [Alcantud et al. \(2010\)](#) suggest approximation algorithms that overestimate the actual Varian’s Inconsistency Index.

Our calculation of Varian’s Inconsistency Index in the supplemented code package attempts to take advantage of the moderate size of the analyzed datasets (at most 50 observations per subject). Denote the number of GARP violations by m and the set of all GARP violations by $M = \{h_1, \dots, h_m\}$ (each element is an ordered sequence of observations). For every violation h_i , denote the set of budget line adjustments that can potentially prevent it by H_i (each element is an ordered pair of an observation and an adjustment percentage).

If $\sum_{i=1}^m |H_i| < K_1$ then we take a “brute force” approach (we implement $K_1 = 26$). For each subset of $\cup_{i=1}^m H_i$, we construct the corresponding adjustment vector \mathbf{v} and check whether $GARP_{\mathbf{v}}$ is satisfied. We report three versions of Varian’s Inconsistency Index, each minimizing a different aggregator function - the Minimum aggregator ($1 - \min_{i \in \{1, \dots, n\}} v_i$), the MEAN aggregator ($\frac{1}{n} \sum_{i=1}^n (1 - v^i)$) and the SSQ aggregator ($\sqrt{\frac{1}{n} \sum_{i=1}^n (1 - v^i)^2}$).

Otherwise, we take advantage of the small commodity space ($K = 2$). [Rose \(1958\)](#) shows that in this case WARP is satisfied if and only if SARP is satisfied. Denote the set of WARP violations by W (each element, w_i , is an unordered pair of observations). If $|W| \leq K_2$ we take a similar approach, on budget adjustments that can prevent the WARP violations (we implement $K_2 = 12$). For each of $\cup_{i=1}^{|W|} w_i$, we construct the corresponding adjustment vector \mathbf{v} and check whether $GARP_{\mathbf{v}}$ is satisfied. We report the minimum of the three aggregators mentioned above. We observe that resolving WARP violations provides a very good approximation to the actual Varian’s Inconsistency Index.

Finally, if $\sum_{i=1}^m |H_i| \geq K_1$ and $|W| > K_2$ we implement Algorithm 3 of [Alcantud et al. \(2010\)](#). This algorithm initializes the vector of adjustments, \mathbf{v} , to $\mathbf{1}$. Then, a loop is implemented that ends only when the data satis-

³⁸[Smeulders et al. \(2014\)](#) show a similar result for the generalized mean aggregator function.

fies $GARP_{\mathbf{v}}$. Inside the loop, the matrix A is maintained where the cell in the i^{th} row and the j^{th} column contains $\frac{p_j x_i}{v_j p_j x_j}$ if $x_i R_{\mathbf{v}, D} x_j$ and $x_j P_{\mathbf{v}, D}^0 x_i$ and zero otherwise. In each iteration, the maximal element of A is picked and substituted into the corresponding element in the vector of adjustments. We report the three aggregators mentioned above operated on the resulted vector of adjustments \mathbf{v} .

For the data collected in the first part of our experiment, where each subject made 22 choices from linear budget lines, we are able to calculate the Varian Inconsistency Index exactly for 91.6% (186 out of 203) of the subjects. We fail to calculate a reliable index for only 3 subjects (we provide good approximation for 14 subjects). Since [Choi et al. \(2007\)](#) collected 50 observations per subject, the success rate of our algorithm is somewhat lower. We are able to calculate the index exactly for 72.3% (34 out of 47) of the subjects and to provide good approximation for 4 other subjects. We fail to calculate a reliable result for 9 subjects.

B.1.3 Houtman-Maks Inconsistency Index

[Boodaghians and Vetta \(2015\)](#) show that there exists a polynomial time algorithm to calculate the Houtman-Maks Inconsistency Index for the two commodities case ($K = 2$).³⁹ In addition, they follow [Houtman and Maks \(1985\)](#) and [Smeulders et al. \(2014\)](#) to show that for three commodities or more, calculating the Houtman-Maks Inconsistency Index is NP-Hard. [Smeulders et al.](#)

³⁹[Rose \(1958\)](#) shows that in the two commodity case ($K = 2$) WARP is satisfied if and only if SARP is satisfied. Let G be an undirected graph where each node is a chosen bundle and two nodes are linked if they constitute a pair that violates WARP. [Boodaghians and Vetta \(2015\)](#) use [Rose \(1958\)](#) to prove that in the two commodity case calculating the Houtman-Maks Inconsistency Index is equivalent to finding the minimal vertex cover of G (the smallest set of nodes S such that every edge in G has an endpoint in S). Next, a graph is perfect if the chromatic number (the smallest number of colors needed to color all nodes where no two adjacent vertices share the same color) of every induced subgraph equals the size of the largest clique (a set of fully connected nodes) of that subgraph. [Boodaghians and Vetta \(2015\)](#) show that G is perfect and recall that finding the minimal vertex cover of a perfect graph is solvable in polynomial time. Hence, they conclude that the calculation of the Houtman-Maks Inconsistency Index in two commodities case is also solvable in polynomial time.

(2014) show that no polynomial time algorithm can achieve a constant factor approximation (a ratio of $o(n^{1-\delta})$) for this Index (see also the discussion in [Boodaghians and Vetta \(2015\)](#) following Lemma 2.1).

Our calculation of the Houtman-Maks Inconsistency Index in the supplemented code package begins with an exhaustive search approach.⁴⁰ Given a dataset D of size n , denote by D_m the set of all subsets of D of size $m < n$. Also, denote $M = \min_{m \in \{1, \dots, n-1\}} m \text{ s.t. } |\cup_{l=m}^{n-1} D_l| < K_3$. The algorithm first goes over every element in D_{n-1} , then over every element in D_{n-2} , etc. The algorithm terminates either after an adjusted dataset that satisfies *GARP* is found, or after every element in D_M was checked (we implement $K_3 = 10^8$).⁴¹ If the algorithm terminated without finding a subset that satisfies *GARP*, we use a modified complementary package⁴² where the Houtman-Maks Inconsistency Index problem for the case of two goods is represented as an integer linear program which is solved by an approximation algorithm provided by Matlab. This solution is an upper bound since the removals suggested by the linear program might not be minimal.⁴³

For the data collected in the first part of our experiment, where each subject made 22 choices from a linear budget line, we are able to calculate the Houtman-Maks Inconsistency Index for all subjects. For the data collected by [Choi et al. \(2007\)](#) (50 observations per subject) we failed to calculate the exact index for 7 subjects (14.9%).

⁴⁰Algorithm 1 in [Gross and Kaiser \(1996\)](#) is a different, more efficient algorithm, for an exact calculation of the Houtman-Maks Inconsistency Index.

⁴¹For example, if the data set includes 50 observations then all subsets of size 46 or more are tested while if the data set is of size 22 then all subsets are checked (in fact for every dataset of size 23 or less, all subsets will be examined).

⁴²Downloaded from Daniel Martin’s personal website on November 5th 2011. The modifications are mainly due to the simplifications enabled by the result of [Rose \(1958\)](#) for the case of two commodities. The second algorithm in [Heufer and Hjertstrand \(2015\)](#) is closely related to Martin’s implementation.

⁴³Another, more efficient approximation is implemented by the algorithm suggested by Algorithm 2 in [Gross and Kaiser \(1996\)](#) and Algorithm 1 in [Heufer and Hjertstrand \(2015\)](#).

B.2 Alternative Indices Based on Revealed Preferences

The Money Pump Index (MPI, [Echenique et al., 2011](#)) and the Minimum Cost Index (MCI, [Dean and Martin, 2015](#)) are recently proposed alternatives to the Varian, Afriat and Houtman-Maks Inconsistency Indices. In this section we describe and discuss these indices and their relation to those characterized by [Theorem 2](#). In addition, we discuss an additional possible inconsistency index, and highlight the challenges in its application.

B.2.1 Money Pump Index

The premise of the MPI is that every violation of GARP corresponds to a cycle of observed bundles.⁴⁴ Each cycle can be interpreted as a sequence of trades, resulting in a sure loss of money, that the DM will accept. The MPI of a cycle is the monetary loss, relative to the total income in the cycle, incurred by one sequence of these trades. The MPI of a data set is an aggregation of these losses.⁴⁵ The MPI is the only inconsistency index mentioned in the current study that does not minimize any loss function, but rather calculates some measure of severity for each GARP violation.⁴⁶ In addition, the MPI takes into account every link in a cycle, rather than focusing only on the weakest link as the other indices analyzed here.

B.2.2 Minimum Cost Index

The MCI is based on the fact that SARP is satisfied if and only if the direct revealed preference relation is acyclic. [Dean and Martin \(2015\)](#) suggest to remove direct revealed preference relations between observed bundles until $R_{D,1}^0$ becomes acyclic. They calculate the cost of removing the ordered pair $(x^{k_i}, x^{k_{i+1}})$ from $R_{D,1}^0$ by $\frac{p^{k_i} x^{k_i} - p^{k_{i+1}} x^{k_{i+1}}}{\sum_{k \in \{1, \dots, n\}} p^k x^k}$, and propose the MCI as the minimal

⁴⁴A sequence of observed bundles $x^{k_1}, x^{k_2}, \dots, x^{k_l}$ in dataset D is a cycle of length l if $x^{k_1} R_{D,1}^0 x^{k_2}, \dots, x^{k_{l-1}} R_{D,1}^0 x^{k_l}$ and $x^{k_l} P_{D,1}^0 x^{k_1}$.

⁴⁵[Echenique et al. \(2011\)](#) suggest the mean and the median aggregators, while [Smeulders et al. \(2013\)](#) recommend, due to computational complexity concerns, the minimum or the maximum aggregators.

⁴⁶Counting the violations of the revealed preference axioms is a similar approach in this respect, see [Swofford and Whitney \(1987\)](#); [Famulari \(1995\)](#); [Harbaugh et al. \(2001\)](#).

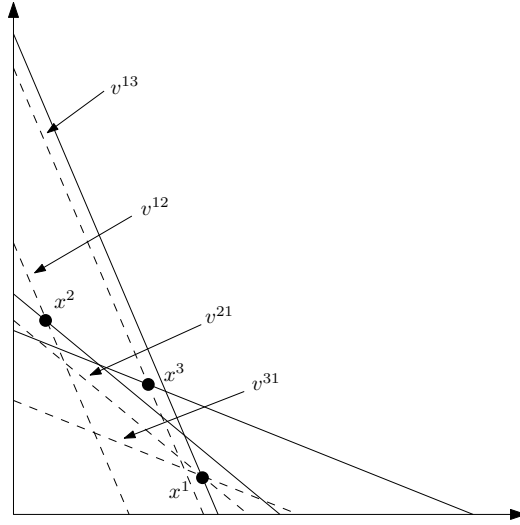


Figure B.1: Comparing Inconsistency Indices.

cost of removals that make $R_{D,1}^0$ acyclic. The MCI does not take into account that a budget line adjustment required to remove one relation may also remove additional relations. In comparison, such inter-dependencies between cycles are accounted for by the Varian Inconsistency Index.

B.2.3 MCI and MPI vs. other Indices

[Echenique et al. \(2011, Section III.B\)](#) and [Dean and Martin \(2015, Section 2.1\)](#) provide thorough discussions on the relative merits of the MPI and the MCI, respectively. Here, we provide an example that highlights a property common to both indices. Note that the MPI is defined over cycles of observations and the MCI over pairs of observations, while the Varian Inconsistency Index is defined observation-by-observation. As a consequence, the latter internalizes the effect of a single adjustment on all cycles or pairs (in which this observation is involved), while the former two do not. The most important implication of this property, in the context of parametric recovery of preferences, is that it is not clear that there exist corresponding measures of incompatibility that can be decomposed into these inconsistency indices (MPI or MCI) and misspecification measures, in the spirit of [Theorem 2](#).

Consider the data set demonstrated in Figure B.1. This data set is of size 3, $D = \{(p^1, x^1), (p^2, x^2), (p^3, x^3)\}$ where $p^i x^i = 1$. The strict direct revealed preference relation $P_{D,1}^0$ (and hence also $R_{D,1}^0$) includes the ordered pairs (x^1, x^2) , (x^2, x^1) , (x^1, x^3) and (x^3, x^1) and therefore the data set is inconsistent with GARP. A budget set adjustment $v^{ij} p^i x^i$, where v^{ij} is such that $v^{ij} p^i x^i = p^j x^j$, is the dashed line denoted by v^{ij} .

We first attend to the Varian Inconsistency Index. There are three possible minimal adjustment vectors \mathbf{v} such that $GARP_{\mathbf{v}}$ is satisfied: $\mathbf{v}_A = (v^{12}, 1, 1)$, $\mathbf{v}_B = (v^{13}, v^{21}, 1)$ and $\mathbf{v}_C = (1, v^{21}, v^{31})$. Note that in \mathbf{v}_A , where the budget line of Observation 1 is adjusted to x^2 , both cycles $((x^1, x^2, x^1)$ and $(x^1, x^3, x^1))$, are broken at once. Therefore $I_V(D, f) = \min\{f(\mathbf{v}_A), f(\mathbf{v}_B), f(\mathbf{v}_C)\}$ and if f is the MEAN aggregator of $\mathbf{1} - \mathbf{v}$, then $I_V(D, f) = \min\{\frac{1-v^{12}}{3}, \frac{2-v^{21}-\max\{v^{13}, v^{31}\}}{3}\}$. Alternatively, if we use the minimum aggregator ($f(\mathbf{v}) = 1 - \min_{i \in \{1, \dots, n\}} v_i$) we get that $I_V(D, f) = 1 - \max\{v^{12}, \min\{v^{13}, v^{21}\}, \min\{v^{21}, v^{31}\}\}$. By Theorem 2.3, $I_A(D) = 1 - \max\{v^{12}, \min\{v^{13}, v^{21}\}, \min\{v^{21}, v^{31}\}\}$, as well. There are two minimal adjustment vectors for the Houtman-Maks Inconsistency Index: $\mathbf{v}_{A'} = (0, 1, 1)$ and $\mathbf{v}_{C'} = (1, 0, 0)$. Therefore, $I_{HM}(D, f) = \min\{f(\mathbf{v}_{A'}), f(\mathbf{v}_{C'})\}$. If f is anonymous then $I_{HM}(D, f) = f(\mathbf{v}_{A'})$.

The MPI takes into account three cycles - (x^1, x^2, x^1) , (x^1, x^3, x^1) and $(x^2, x^1, x^3, x^1, x^2)$. For each cycle it accounts for all the links. Therefore, the measure for (x^1, x^2, x^1) is $\frac{2-v^{12}-v^{21}}{2}$, the measure for (x^1, x^3, x^1) is $\frac{2-v^{13}-v^{31}}{2}$ and the measure for $(x^2, x^1, x^3, x^1, x^2)$ is $\frac{4-v^{12}-v^{21}-v^{13}-v^{31}}{4}$ and using the MEAN aggregator we get $MPI = \frac{4-v^{12}-v^{21}-v^{13}-v^{31}}{4} \geq I_V(D, f)$.

The MCI ignores the fact that adjusting the budget line of Observation 1 to x^2 resolves also the cycle that includes x^1 and x^3 . Therefore, $MCI = \frac{2-\max\{v^{12}, v^{21}\}-\max\{v^{13}, v^{31}\}}{3} \geq I_V(D, f)$.

B.2.4 Area-based Measures

A natural alternative to the incompatibility indices discussed in the current study is an Intersection Incompatibility Index, which is based on the area bounded between the upper contour set of the indifference curve passing

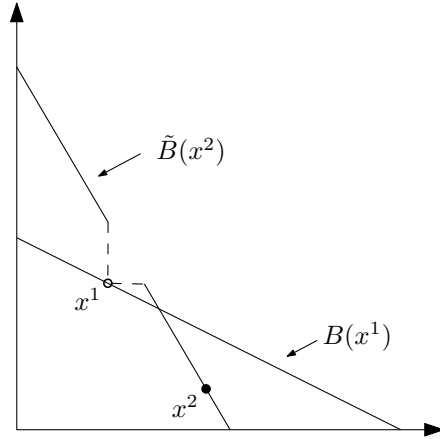


Figure B.2: Modified budget sets

through the chosen bundle and the set of feasible alternatives.

A related measure is introduced in the online appendix (Part D.3) of [Apestegua and Ballester \(2015\)](#) in which they extend their Minimal Swaps Index to the case of infinite number of alternatives.⁴⁷ Their proposal is based on the Lebesgue measure of the bounded area and the sum aggregator over observations. They define the Consumer Setting Swaps Index as the infimum of this sum over the set of all continuous, strictly monotone and quasi-concave utility functions.

In light of Theorem 2, one needs to have, in addition, a corresponding measure of inconsistency, so that when the set of utility functions is restricted, this index measures the inconsistency embedded in choices, while the remainder of the Intersection Incompatibility Index represents the misspecification implied by the chosen parametric family.

One option is to define an index of inconsistency based on the area of intersection between the revealed preferred set and the budget set corresponding to

⁴⁷For the case of finite number of alternatives, [Apestegua and Ballester \(2015\)](#) define the Swaps Index of a given preference relation to be the minimal number of swaps required to reconcile the observations with the ranking induced by the given preference. The Minimal Swaps Index minimizes the Swaps Index over all possible rankings. Applying the current paper's terminology, the Swaps Index is an incompatibility measure. However, since [Apestegua and Ballester \(2015\)](#) domain includes a finite number of alternatives and therefore a finite number of rankings, the Minimal Swaps Index becomes an inconsistency measure, in the spirit of Theorem 2.

an observed choice. Define the *revealed preferred* set of a given bundle as *only* those bundles that are either revealed preferred or those that monotonically dominate a bundle that is revealed preferred to the given bundle. Hence, as illustrated in Figure B.2, violations of consistency are removed by modifying budget sets so as to eliminate the area of overlap between the budget set and those bundles which are revealed preferred. These violations can be measured and aggregated to construct the Area Inconsistency Index.⁴⁸

Nevertheless, the Area Inconsistency Index is not ideal. First, currently, there does not exist an elegant theoretical analog to Theorem 1 with respect to the modified budget sets in Figure B.2 as there does for the specific type of adjustments utilized in calculating the Varian and the Houtman-Maks Inconsistency Indices. Therefore a decomposition result may be difficult to achieve. Second, computing the inconsistency index suggested above would not be any easier than computing the Varian or Houtman-Maks Inconsistency Indices, problems which are NP-hard (see Appendix B.1 above). Third, we conjecture that any recovery procedure related to the Area Inconsistency Index would be biased towards non-convex preferences due to the geometric characteristics of the suggested budget line adjustments. Finally, the Area Inconsistency Index lacks intuitive interpretation that the considered indices enjoy. All these are surmountable difficulties, that we think are worthwhile pursuing in future work.

B.3 Distance-based Indices

The common method for parametric recovery of individual preferences minimizes some loss function of the distance between observed and predicted bundles. Similar to the money metric and binary incompatibility measures, the result of this method can also be decomposed into an inconsistency and misspecification measures.

⁴⁸Heufer (2008, 2009, Section 9.2.3) suggests, in the spirit of of Varian’s (1982) non-parametric bounds, a similar inconsistency index with the additional external assumption of convexity of preferences. Apesteguia and Ballester (2015) provide a simple example in their online appendix in which they implement a measure that corresponds to Heufer’s index, assuming its equivalence to their Consumer Setting Swaps Index.

One example of such decomposition can be based on an inconsistency index suggested by [Beatty and Crawford \(2011\)](#). This index measures the Euclidean distance between the observed data set and the set of potential data sets that satisfy GARP.⁴⁹ It can be shown that a generalization of the proposed index equals the infimum of the appropriate loss function calculated over all continuous and locally non-satiated utility functions. Therefore, the difference between the minimal loss calculated over a subset of utility functions and the proposed inconsistency index results in a natural measure of misspecification.

However, this method ignores the fact that making a choice from a menu reveals that the chosen alternative is preferred to *every* other feasible alternative, not only to the predicted one. In addition, this measure entails an additional assumption on the ranking of unchosen alternatives. It requires that the closer is a bundle to the choice, the higher it is ranked. Such ranking can be justified only by the auxiliary assumption that the choices were generated through a maximization of convex preferences, which is not part of revealed preference theory. Therefore, if choices were generated by a maximization of non-convex preferences then this additional assumption will lead to an erroneous ranking of unchosen alternatives, as demonstrated by the results of the experiment reported in sections [6](#) and [7](#).

C Decomposition: Graphical Example

Figure [C.1](#) demonstrates the decomposition graphically. Consider the data set: $D = \{(p^1, x^1), (p^2, x^2)\}$. The data set is inconsistent with GARP since $x^1 R_{D,1} x^2$ and $x^2 P_{D,1}^0 x^1$. Note that the dashed line $v^2 p^2 x^2$, together with the original budget line from which x^1 was chosen, represent graphically the adjustments that lead D to satisfy $GARP_{(1,v^2)}$. If $v^2 \geq v^1$, for any anonymous aggregator, the Varian Inconsistency Index is $I_V(D, f) = f((1, v^2))$ and the

⁴⁹Let $R_i = \{x \in \mathfrak{R}_+^K : p^i x = p^i x^i\}$ be the set of bundles that cost $p^i x^i$ at prices p^i . Then, the set of all potential data sets given data set D is $\{\{(p^i, x)_{i=1}^n\} : x \in R_i\}$. [Beatty and Crawford \(2011\)](#) propose $1 - \frac{d}{d^{max}}$ as an inconsistency index where d is the Euclidean distance between the data set and the closest element in the set of potential data sets that satisfy GARP and it is normalized by d^{max} to restrict the index to $[0, 1]$.

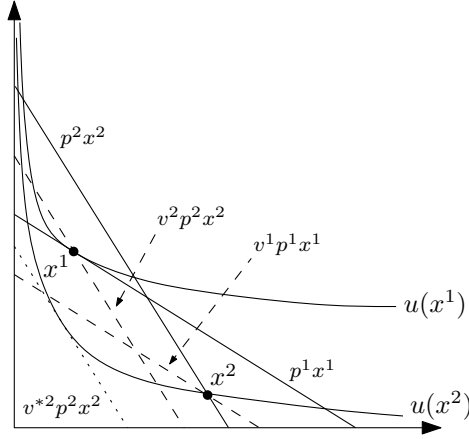


Figure C.1: Decomposition

Houtman-Maks Inconsistency Index is $I_{HM}(D, f) = f((1, 0))$.

Now consider the monotonic and continuous function u . Since $\{(p^1, x^1)\}$ is rationalizable by this utility function, then $v^{*1}(D, u) = b^{*1}(D, u) = 1$. In addition, $v^{*2}(D, u)$ is the minimal expenditure required to achieve utility level of $u(x^2)$ under prices p^2 , which is represented graphically by the dotted line $v^{*2}p^2x^2$ while $b^{*2}(D, u) = 0$ since u does not rationalize $\{(p^2, x^2)\}$.

Thus, $I_M(D, f, \{u\}) = f((1, v^{*2}(D, u)))$ and since $v^{*2}(D, u)$ is smaller than v^2 , it implies that $I_M(D, f, \{u\})$ is weakly greater than $I_V(D, f)$. Since in this specific example, no other adjustments are required, the difference between the original budget line from which x^2 was chosen and the dashed line - $v^2p^2x^2$, represents graphically the inconsistency implied by D , while the difference between the dashed line and the dotted line - $v^{*2}p^2x^2$, represents the misspecification implied by u . Their sum is the goodness of fit measured by the money metric index. However, $I_B(D, f, \{u\}) = I_{HM}(D, f)$, meaning that no misspecification is implied by the binary incompatibility index since u rationalizes $\{(p^1, x^1)\}$ which is the largest subset of D that can be rationalized by any utility function as suggested by the Houtman-Maks Inconsistency Index.⁵⁰

⁵⁰If one considers an alternative utility function u' such that $\{(p^1, x^1)\}$ is not rationalizable by u' (but suppose $v^{*2}(D, u') = v^{*2}(D, u)$), this would not affect the inconsistency indices but would imply weakly higher loss indices than those measured for u (e.g. $I_B(D, f, \{u'\}) = f(\mathbf{0})$).

D Disappointment Aversion Preferences

Let $p = (p_1, x_1; \dots; p_n, x_n)$ be a lottery such that $x_1 \leq \dots \leq x_n$. Assuming (for simplicity) that $ce(p) \notin \text{supp}(p)$, the support of p can be partitioned into elation and disappointment sets: there exists a unique j such that for all $i < j : (x_i, 1) \prec p$ and for all $i \geq j : (x_i, 1) \succ p$. Let $\alpha = \sum_{i=j}^n p_i$. Gul's elation/disappointment decomposition is then given by $r = (x_1, r_1; \dots; x_{j-1}, r_{j-1})$, $q = (x_j, q_j; \dots; x_n, q_n)$ such that $r_i = \frac{p_i}{1-\alpha}$ and $q_i = \frac{p_i}{\alpha}$. Note that $p = \alpha q + (1 - \alpha) r$. Then:

$$u_{DA}(p) = \gamma(\alpha) E(v, q) + (1 - \gamma(\alpha)) E(v, r)$$

and $\exists -1 < \beta < \infty$ such that

$$\gamma(\alpha) = \frac{\alpha}{1 + (1 - \alpha)\beta}$$

where $v(\cdot)$ is a utility index and $E(v, \mu)$ is the expectation of the functional v with respect to measure μ . If $\beta = 0$ disappointment aversion reduces to expected utility, if $\beta > 0$ the DM is disappointment averse ($\gamma(\alpha) < \alpha$ for all $0 < \alpha < 1$), and if $\beta < 0$ the DM is elation seeking ($\gamma(\alpha) > \alpha$ for all $0 < \alpha < 1$). Gul (1991) shows that the DM is averse to mean preserving spreads if and only if $\beta \geq 0$ and v is concave. That is, if v is concave then, by Yaari (1969), preferences are convex if and only if the DM is weakly disappointment averse.

For binary lotteries: Let $(x_1, p; x_2, 1 - p)$ be a lottery. The elation component is x_2 and the disappointment component is x_1 and $\alpha = 1 - p$ (in our case $\alpha = 0.5$). Therefore:

$$u_{DA}(x_1, p; x_2, 1 - p) = \gamma(1 - p) v(x_2) + (1 - \gamma(1 - p)) v(x_1)$$

and since $\gamma(0) = 0, \gamma(1) = 1$ and $\gamma(\cdot)$ is increasing, $\gamma(\cdot)$ can be viewed as a weighting function, and DA is a special case of Rank Dependent Utility (Quiggin, 1982).

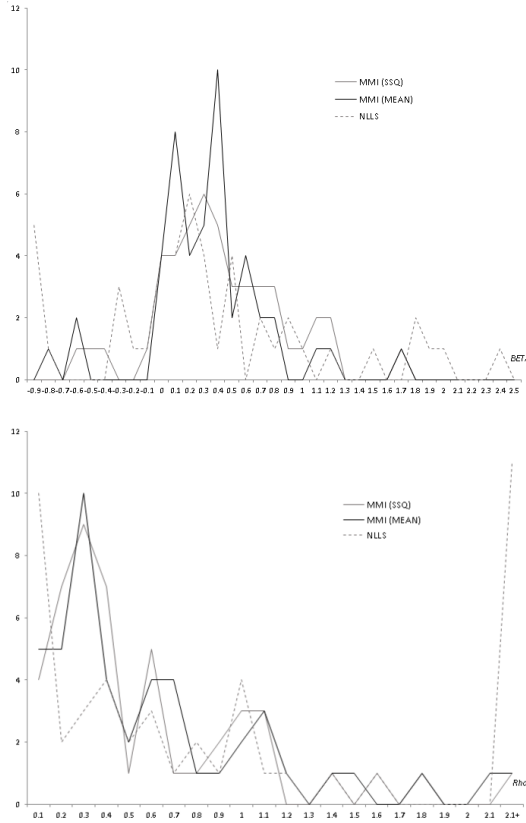


Figure E.1: The distribution of the recovered β (upper) and ρ (lower) by MMI (SSQ), MMI (MEAN) and NLLS in [Choi et al. \(2007\)](#).

E CRRA Parameters: Distributions

Figure [E.1](#) provides the distribution of the recovered parameters for the Disappointment Aversion functional form with the CRRA utility index by three recovery methods - NLLS, MMI (SSQ) and MMI (MEAN). Both distributions provide some evidence as to the extreme values recovered by NLLS.

Consider for example, the distribution of the disappointment aversion parameter (upper panel of [Figure E.1](#)). The NLLS recovers $\beta < -0.9$ or $\beta > 1.3$ for 11 subjects, while the MMI methods recover such extreme values only once. Similar pattern can be easily observed in the lower panel for the CRRA parameter.

F The Experiment

F.1 Instructions

Welcome

Welcome to the experiment. Please silence your cell phone and put it away for the duration of the experiment. Additionally, please avoid any discussions with other participants. At any time, if you have any questions please raise your hand and an experiment coordinator will approach you.

Please note: If you want to review the instructions at any point during the experiment, simply click on this window (the instructions window). To return to the experiment, please click on the experiment icon on the task bar.

Study Procedures

This is an experiment in individual decision making. The study has two parts and the second part will begin immediately following completion of the first part. Before Part 1, the instructions will be read aloud by the experiment coordinator and you will be given an opportunity to practice. The practice time will allow you to familiarize yourself with the experimental interface and ask any questions you may have. We describe the parts of the experiment in reverse order, beginning with Part 2 now.

Part 2

You will be presented with 9 independent decision problems that share a common form. In each round you will be given a choice between a pair of allocations of tokens between two accounts, labeled x and y. Each choice will involve choosing a point on a two-dimensional graph that represents the values in the two accounts. The x-account is represented by the x-axis and the y-account is represented by the y-axis.

For all rounds, in Option 1 the amount allocated to the x-account and y-account will differ, and in Option 2 the amount allocated to each account will be the same. For both options, the values allocated to each account will be displayed beside the point corresponding to each option on the graph, as well as, in the dialog box labeled “Options” on the right-hand side of the screen. Figure F.1 illustrates some examples of types of choices you may face.

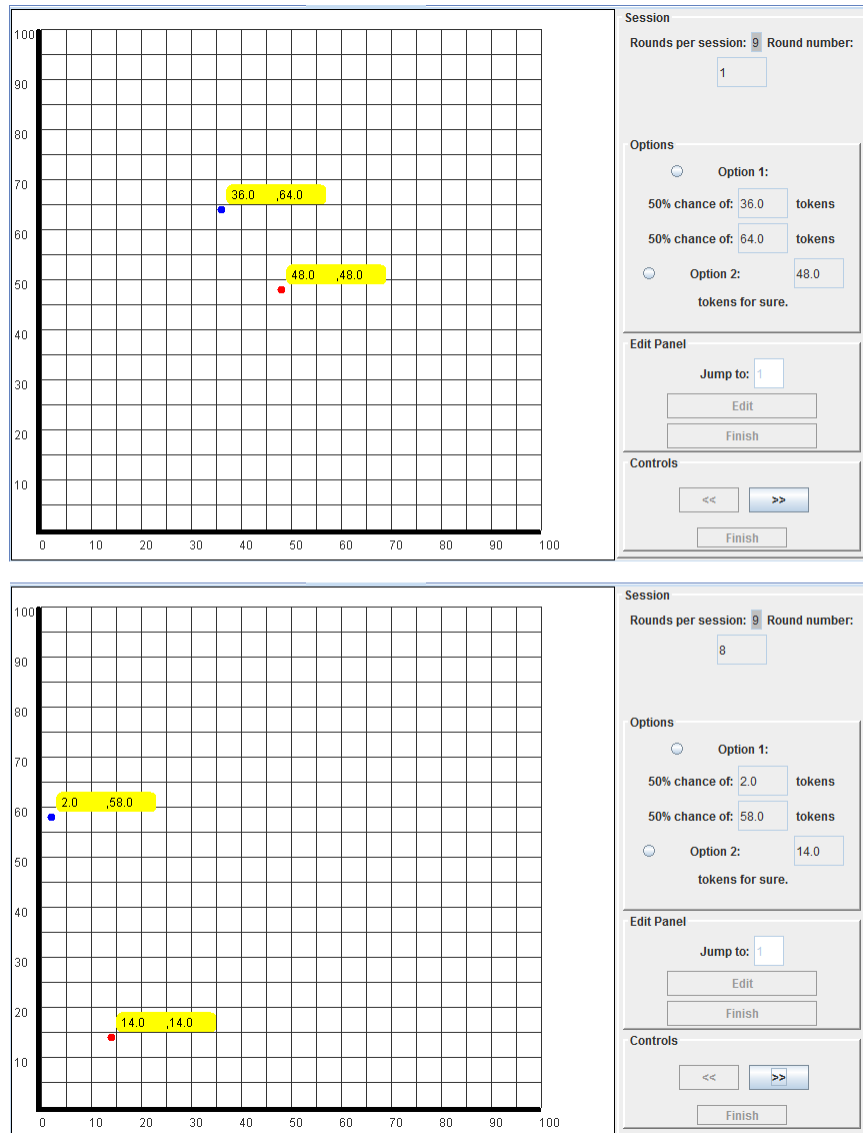


Figure F.1: Pairwise Choices

For the round that is selected for payment, your payment is determined by the number of tokens allocated to each account. At the end of the experiment, you will toss a fair coin to randomly select one of the two accounts, x or y. For each participant, each account is equally likely to be chosen. That is, there is a 50% chance account x will be selected and a 50% chance account y will be chosen. You will only receive the amount of tokens you allocated to the account that was chosen. The round for which you will be paid will be selected randomly at the conclusion of the experiment and each round is equally likely to be chosen. Remember that tokens are valued at the following conversion rate: 2 tokens = \$1.

Please Note: Only one round (from both parts combined) will be selected for payment and your payment will be determined only after completion of both parts.

Each round begins with the computer selecting a pair of allocations. For example, as illustrated in Figure F.2, Option 1, if selected, implies a 50% chance of winning 32.0 tokens and a 50% chance of winning 58.0 tokens, where as Option 2, if selected, implies winning 43.0 tokens for sure.

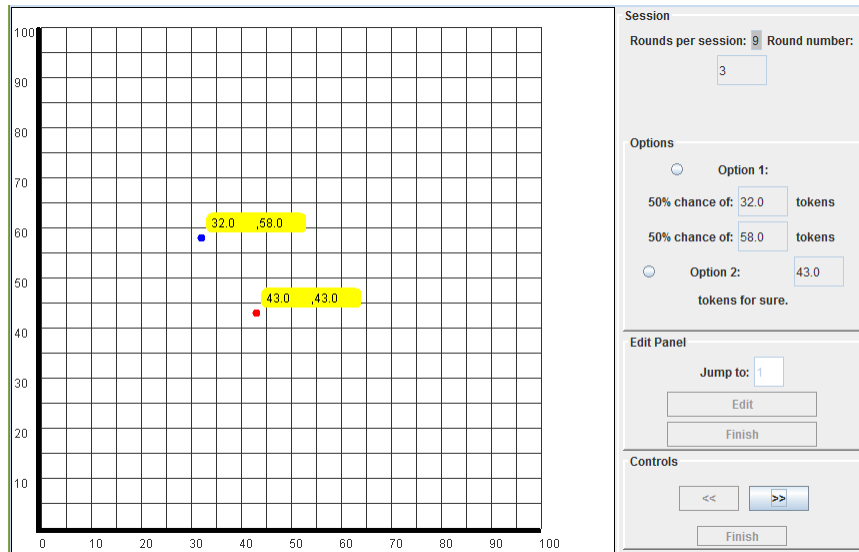


Figure F.2: Pairwise Choices - Example

In some cases, the two options will be so close to each other that it will be difficult to distinguish between them graphically. In this case, you may refer to the “Options” box on the right-hand side of the screen where the values associated with each option are listed. Additionally, it may be difficult to select your preferred option by clicking on the graph itself, so instead you may use the radio buttons in the “Options” box to make your selection. Figure F.3 provides an example of this situation.

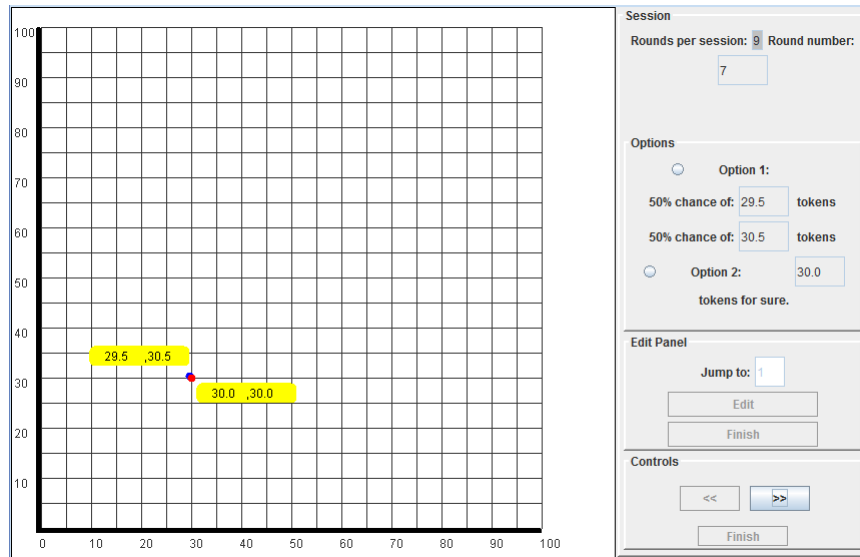


Figure F.3: Pairwise Choices - Overlapping Points

In all rounds, you may select a particular allocation in either of two ways: 1) You may use the mouse to move the pointer on the computer screen to the option that you desire, and when you are ready to make your decision, simply left-click near that option, or 2) You may select your preferred option using the radio buttons on the right-hand side of the screen, and when you are ready to make your decision, simply left-click on the radio button that corresponds to your choice. In either case, a dialog box, illustrated in Figure F.4, will ask you to confirm your decision by clicking “OK”.

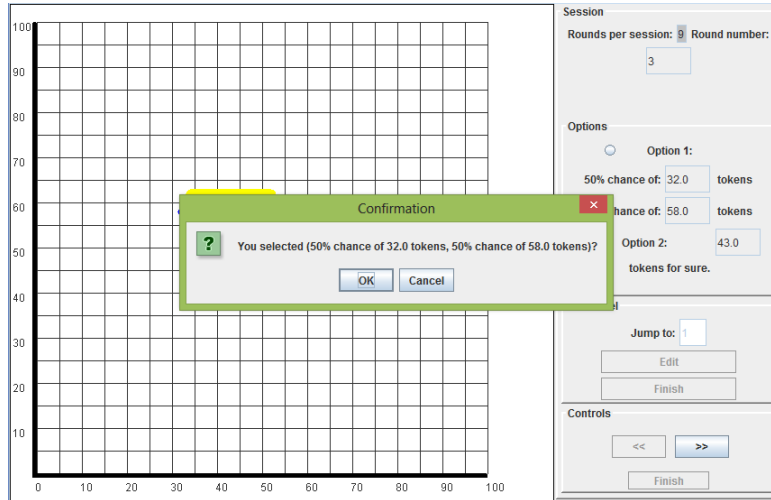


Figure F.4: Pairwise Choices - Confirmation Screen

If you wish to revise your choice simply click “Cancel” instead. After you click “OK”, your choice will be highlighted in green and the screen will darken, as illustrated in Figure F.5, indicating that your choice is confirmed. You may proceed to the next round by clicking on the “>>” button located in the lower right-hand corner of the screen in the box labeled “Controls”. Please note that you will be given an opportunity to review and edit your choices upon completion of Part 2 of the experiment.

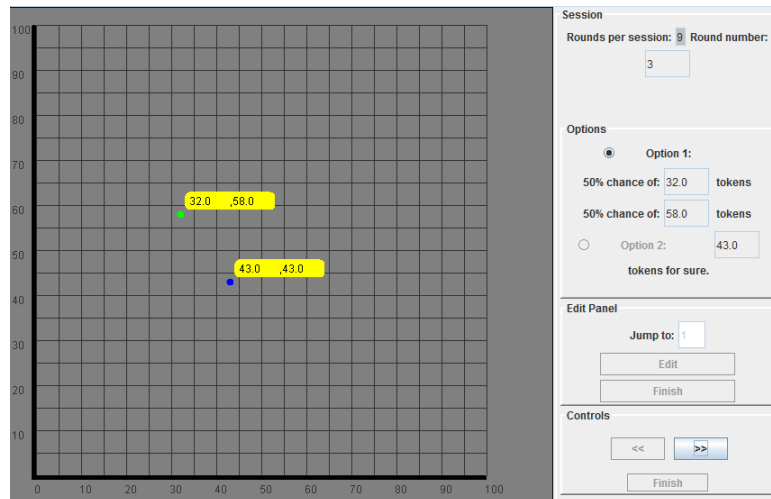


Figure F.5: Pairwise Choices - Confirmed Choice

Next you will be asked to make an allocation in another independent decision problem. This process will be repeated until all 9 rounds are completed. At the end of the last round, you must click the “Finish” button, located in the lower right-hand corner of the screen in the box labeled “Controls”, and you will be given an opportunity to review your choices. You may use the navigation buttons to move between choices or the “Jump to” feature in the “Edit Panel” to navigate to a specific round. If you are content with your choices, you may exit the review by clicking on the “Finish” button. At this stage you may no longer go back to review and/or edit your choices. Instead, click “OK” to complete the experiment.

Part 1

In Part 1, you will be presented with 22 independent decision problems that are very similar to those in Part 2. However, rather than selecting an allocation from among only two options, now you will have many options to choose from. In each round your available options will be illustrated by a straight line on the graph and you will make your choice by selecting a point on this line. As in Part 2, your payoff in the round that is selected for payment is determined by the number of tokens allocated to each account. Examples of different lines you may face are illustrated in Figure F.6.

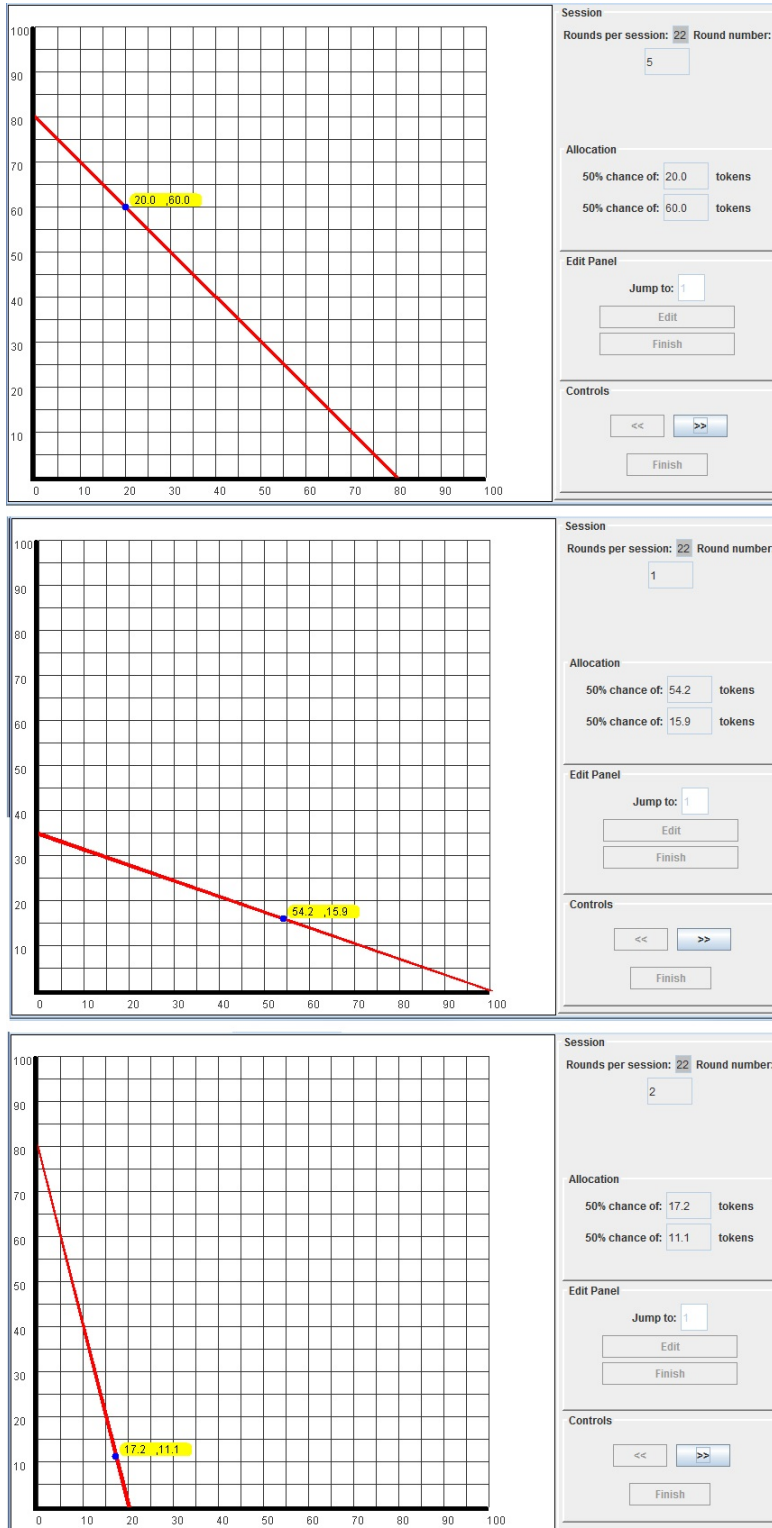


Figure F.6: Budget Lines
71

Figure F.7 illustrates the differences and similarities between the problems in Part 1 and Part 2. In Part 2, you are offered the choice between only two options, A and B. On the other hand, if we were to draw a straight line between these options and allow one to choose any point on this line, then this would increase the number of available choices. Notice, however, that the two original options are still available as well as many more. Hence, the problems in Part 1 are conceptually the same as in Part 2, but with many more possible allocations.

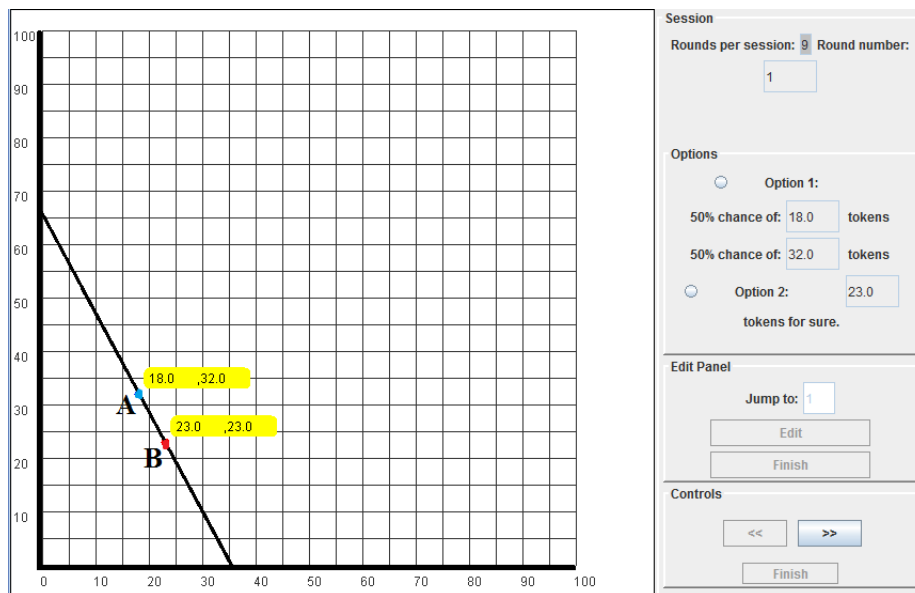


Figure F.7: Budget Lines - Relationship to Pairwise Choice

The following two examples further illustrate the nature of the problem. If, in a particular round, you were to select an allocation where the amount in one of the accounts is zero, for example if you allocate all tokens to account x and \$0 to account y (or vice versa), then in the event that this round is chosen for payment there is a 50% chance you will receive nothing at all, and a 50% chance you will receive the highest possible payment available in that round. In contrast, if you were to select an allocation where the amount in accounts x and y are equal, then in the event that this round is chosen for payment you will receive this amount regardless of which account is chosen

by the coin toss.

Each round begins with the computer selecting a line. As in Part 2, the lines selected for you in different rounds are independent of each other. For example, as illustrated in Figure F.8, choice A represents an allocation in which you allocate approximately 9.4 tokens in the x-account and 60.7 tokens in the y-account. Another possible allocation is choice B, in which you allocate 22.6 tokens in the x-account and 33.6 tokens in the y-account.

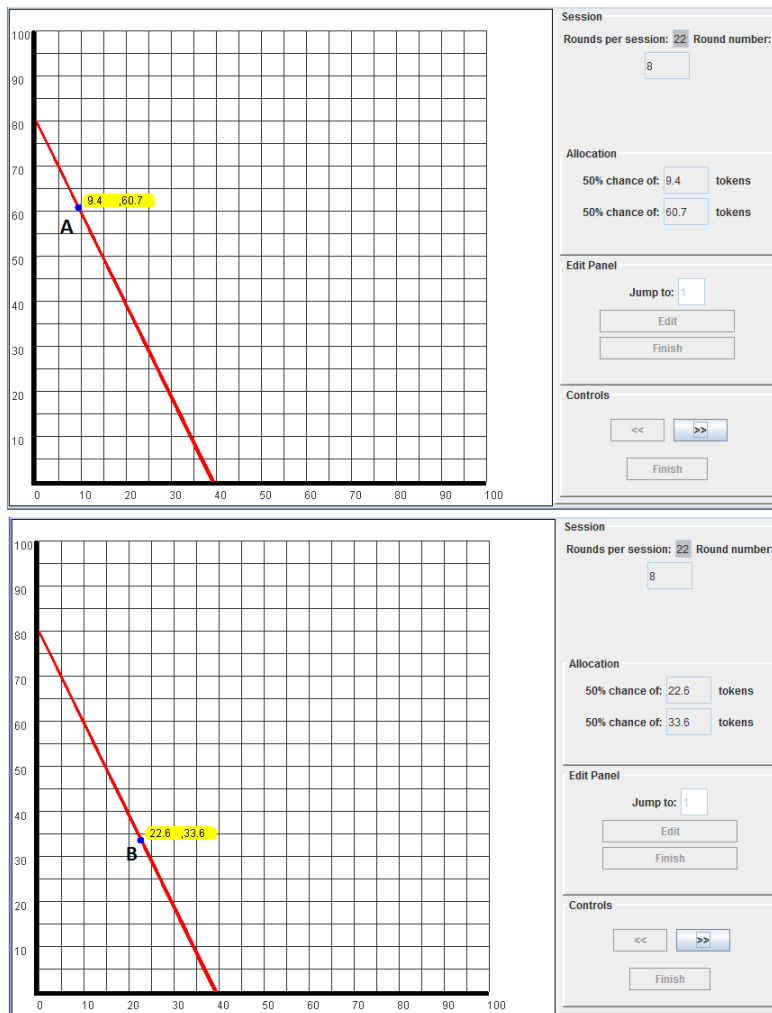


Figure F.8: Budget Lines - Examples

To choose an allocation, use the mouse to move the pointer on the computer screen to the allocation that you desire. On the right hand side of the program dialog window you will be able to see the exact allocation where the pointer is located. Please note that, in each choice, you may only choose an allocation which lies on the line provided. Additionally, if you select an allocation that is close to the x-axis or the y-axis, you will be asked if you would like to select an allocation on the boundary or if you intended for your choice to be as originally selected. Similarly, if you select an allocation that is close to the middle, (roughly the same amounts in each account), you will be asked if you would like to select an allocation where the amounts in both accounts are exactly equal or if you intended for your choice to be as originally selected. The dialog boxes associated with these scenarios are illustrated in Figure F.9.

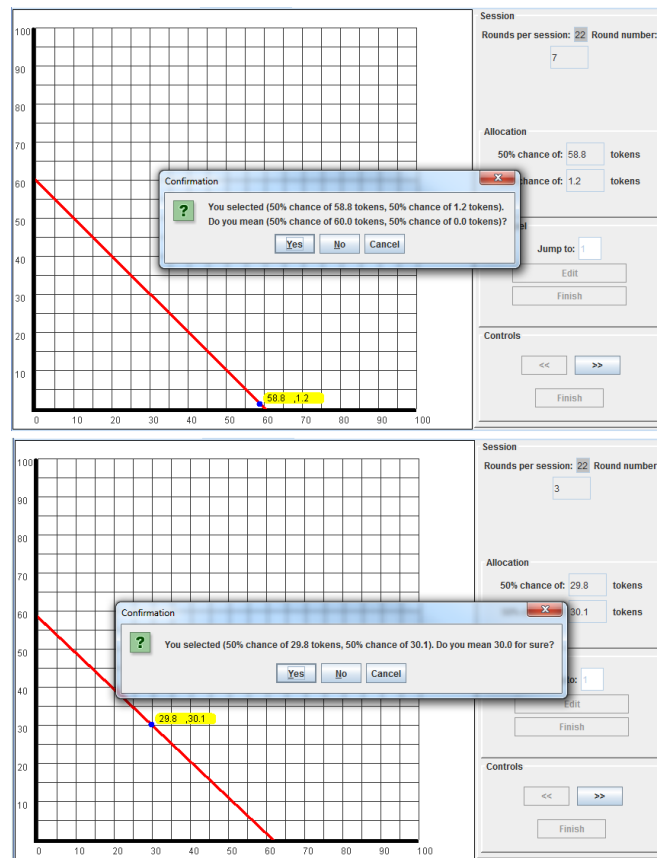


Figure F.9: Budget Lines - Special Cases

The controls to confirm your choices and navigate between rounds are identical to those described above for Part 2. Once you have finished with all 22 rounds, you will be given an opportunity to review your choices. You may conclude your review by clicking on the finish button in the “Edit Panel” at any time. Once complete, please click on the instructions window in order to move on to Part 2.

Please remember that there are no “right” or “wrong” choices. Your preferences may be different from other participants, and as a result your choices can be different. Please note that as in all experiments in Economics, the procedures are described fully and all payments are real.

Compensation

After completing both parts of the experiment you will be informed of your payment via an on-screen dialog box. Payments are determined as follows:

The computer will randomly select one decision round from both parts (combined) to carry out. The round selected depends solely on chance and it is equally likely that any particular round will be chosen. The payment dialog box will inform you of which round was randomly chosen as well as your choice in that round. At this point please raise your hand and an experiment coordinator will provide you with a fair coin, e.g. a quarter. To determine your final payoff, please flip the coin. If it lands heads, you will be paid according to the amount of tokens in the x-account and if it lands tails, you will be paid according to the amount of tokens in the y-account. For both parts of the experiment, tokens are valued at the following conversion rate:

$$\mathbf{2 \text{ tokens} = \$1}$$

You will receive your payment, along with the \$10 show-up bonus, privately

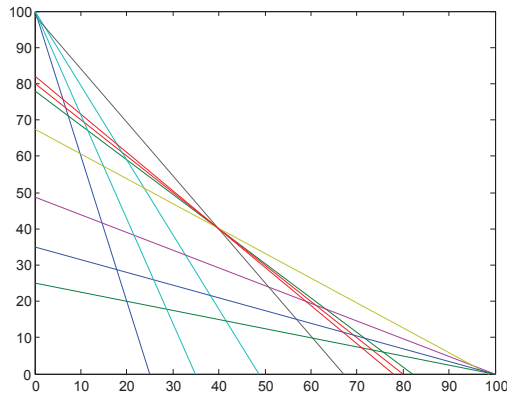


Figure F.10: Subset of the budget lines shown in Part 1

before you leave the lab. You will be asked to sign a receipt acknowledging receipt of your payment, after which time you may leave.

F.2 Choice of Budget Lines

Section 6.2 describes the set of budget sets chosen for the first part of the experiment as a result of two considerations: sufficient power and first-order risk aversion/seeking identification. The 22 budget lines were divided into two subsets of 11 budget lines such that each subset was composed of the same price ratios, where the only difference was the wealth level. For each of the two subsets, 5 of the 11 price ratios had relatively moderate slopes, where as the other 6 were much steeper. Figure F.10 shows the set of 11 budget lines for the higher wealth level.

To corroborate that this set of budget sets submits the subjects to a sufficiently powerful test of consistency, we conducted a power test (following Bronars (1987)) by constructing 1000 simulated data sets.⁵¹ First, not a single simulated data set passed GARP while in the experimental data 44.4% were found to be consistent. Second, in the simulation, 1.3% (4.5%) of the data sets had Afriat Inconsistency Index below 0.05 (0.1) while in the experimental data 86% (93.7%) of the subjects exhibited this level of inconsistency.

⁵¹The results of the power test are available in a separate Excel file named “Halevy et al (2016) Part 1 - Power Test”.

Third, the Houtman-Maks Inconsistency Index (calculated exactly) suggests that 91.9% (57.9%) of the simulated data sets require at least 4 (6) observations to be discarded to satisfy GARP. However, in the experimental data, only 8.7% (0.05%) of the data sets require as many observations to be dropped to achieve consistency. Finally, while we were able to calculate the Varian Inconsistency Index exactly (or with very good approximation) for 98.6% of the experimental data sets, this was feasible for only 25.9% of the simulated data sets. In fact, even within this set, using the MEAN aggregator, while 57.1% of the simulated data sets exhibited Varian Inconsistency Index greater than 0.05, only 2.9% of the experimental data sets showed similar levels of inconsistency.

F.3 The Construction of the Pairwise Choices

In Section 6.3 we describe the basic logic behind the algorithm used to construct the pairwise choices for Part 2 of the experiment. Here, we provide a more detailed description of this algorithm.

Each pairwise choice is constructed using the following search algorithm. First, we fix an expected value for the risky portfolio. Then, we search over the line that connects all the portfolios with the same expected value until a risky portfolio, x^R , is found that satisfies certain stopping conditions. The starting point for the search as well as the stopping conditions are chosen to construct a sufficiently rich set of choices that are appropriate for addressing the research questions.

To investigate the nature of *local* risk attitudes across subjects we designated 6 out of the 9 pairwise choices to this task by beginning our search for x^R at certainty and progressing along the equal expected value line in the direction of increasingly variable portfolios until the stopping rule is satisfied. In the case where both methods recover $\beta \geq 0$, the stopping rule requires that the difference in certainty equivalents exceeds one token. Hence, to construct these *low-variability portfolios* we search for the lowest variance portfolio among all those with the same expected value such that there is sufficient difference in certainty equivalents between recovered parameters. For sets of parameters

where the difference between certainty equivalents does not exceed one token for *all* low-variability portfolios we reduce this threshold incrementally until a valid pairwise comparison is chosen. In all cases the safe portfolio, x^S , is chosen as the mid-point between the certainty equivalents of the risky portfolio, x^R (see Section 6.3).

For subjects where either or both methods recover $\beta < 0$, we use a different stopping rule. In these cases the search terminates as soon as a risky portfolio is found such that the certainty equivalent corresponding to one method exceeds the expected value of the portfolio and the certainty equivalent corresponding to the other method is less. Here we choose the safe portfolio as the expected value of the risky portfolio, i.e. $x^S = E[x^R]$.

The remaining 3 out of 9 pairwise choices are constructed such that the risky portfolio is close to, but not literally on, the axis. We refer to these pairwise choices as *high-variability portfolios*. We avoid offering corner choices as they can be difficult to rationalize with the CRRA functional form. We choose risky portfolios as close to the axis as possible by starting with a portfolio that includes a minimum payoff of two tokens and searching towards the certainty line. The stopping condition is that the difference in certainty equivalents is at least one token. High-variability portfolios are chosen in the same manner regardless of the recovered values for β .⁵²

G Part 1: Comparison to Choi et al. (2007)

This Appendix compares the results of Part 1 of the experiment and the data collected by Choi et al. (2007). Table 9 summarizes the inconsistency indices and the parameters recovered for the Disappointment Aversion with CRRA utility index. We attribute the slight differences to the difference in instructions, interface, the number of rounds and to the variability and range of the price ratios.

⁵²The six low-variability portfolios have expected values of 50, 45, 40, 35, 30, and 25 tokens, where as the three high-variability portfolios have expected values of 50, 40, and 30 tokens.

	Choi et al. (2007)				Part 1 of the Experiment			
Consistent Subjects	12 (25.5%)				91 (44.8%)			
Median (mean) Afriat Inconsistency Index*	0.045 (0.0881)				0.0126 (0.0374)			
Median (mean) Houtman-Maks Inconsistency Index*	0.06 (0.079)				0.0909 (0.097)			
Median (mean) Varian Inconsistency Index (SSQ)**	0.006 (0.007)				0.0027 (0.0084)			
	MMI (SSQ)		NLLS		MMI (SSQ)		NLLS	
	β	ρ	β	ρ	β	ρ	β	ρ
Complete Sample (Median)	0.3326	0.3559	0.171	0.5799	0.39	0.3764	0.3343	0.3674
Consistent Subjects Only (Median)	0.4121	0.7319	0.0058	1.277	0.4065	0.4137	0.3443	0.5597
# Subjects with Non-Convex Preferences	8 (17%)		15 (31.9%)		37 (18.2%)		45 (22.2%)	
Subjects with $\beta \geq 0$ (Median)	0.3759	0.295	0.4058	0.3404	0.4668	0.3022	0.4654	0.1964
Subjects with $\beta < 0$ (Median)	-0.1047	0.8691	-0.3275	3.8642	-0.1575	0.8008	-0.8941	4.0782

* Computed on inconsistent subjects.

** Computed on inconsistent subjects with reliable index.

Table 9: Choi et al. (2007) vs. Part 1 of the Experiment.

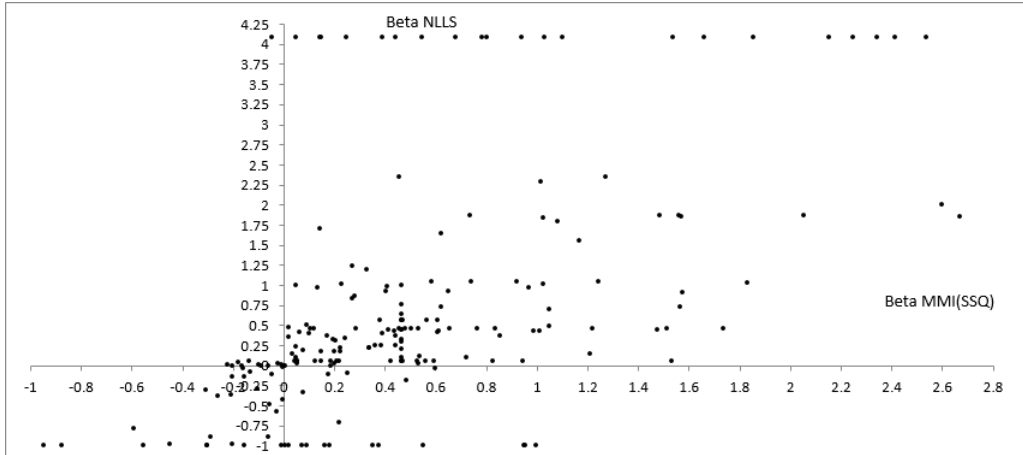


Figure G.1: Disappointment Aversion Parameter: NLLS vs. MMI (SSQ).

Figure G.1 replicates Figure 5.2 for the data collected in Part 1 of the experiment. Also here, when the NLLS recovery method recovers convex preferences then in most cases the MMI method recovers convex preferences as well, while when the preferences recovered by the NLLS are non-convex, there seem to be no qualitative relation between the recovered parameters by the two methods.

H Pairwise Choice: Refined Results

The complete sample includes subjects and choices that arguably should not be included in a comparison between the MMI and the NLLS recovery methods.

In Section 7 we report the results using the full sample while in this Appendix we refine the sample and recalculate the results reported in the main text using the refined sample.

H.1 The Refinement

We find two reasons to consider dropping an observation from the sample. First, the subject's choices may be too inconsistent to believe that there exists some underlying stable preference that guides her choices. Second, since the pairwise choices the subject encountered in Part 2 of the experiment were generated automatically, in some cases the two proposed portfolios were too similar for the subject to be able to thoughtfully distinguish between them.⁵³ Hence, our refinement scheme apply two criteria - inconsistency and similarity.

The *inconsistency refinement* removes two subjects whose Afriat Inconsistency Index greater than 0.2.⁵⁴

The *similarity refinement* removes observations for which there is little difference between the portfolios constructed in Part 2 of the experiment. We consider a pairwise choice to be *indefinitive* if the two sets of parameters imply similar local risk attitude (either $\min \{CE_{MMI}(x^R), CE_{NLLS}(x^R)\} > E[x^R]$ or $\max \{CE_{MMI}(x^R), CE_{NLLS}(x^R)\} < E[x^R]$) and the difference in implied certainty equivalents is very small ($|CE_{MMI}(x^R) - CE_{NLLS}(x^R)| < 0.5$).

	# of Observations	Correct Predictions by MMI (%)	p-value
Refinement	1489	804 (54.0%)	0.0011

Table 10: Preliminary Results - Aggregate Level Analysis (refined sample)

Refined Sample		
$X \geq 7$	$X \leq 2$	p-value
41	25	0.032

Table 11: Preliminary Results - Individual Level Analysis (refined sample)

H.2 Results: Refined Sample

Table 10 recalculates the aggregate level analysis reported in Table 2 in Section 7 for the refined sample. These results are almost identical to the results reported for the complete sample.

In the individual level analysis, for the similarity refinement we remove all subjects who confronted one or more indefinite pairwise comparison in Part 2. Thus, the remaining 131 subjects are deemed sufficiently rational and exhibit a sufficient difference in predictions between recovery methods to admit a reasonable comparison.

Table 11 recalculates the individual level analysis reported in Table 3 in Section 7 for the refined sample. As the results reported for the complete sample, Table 11 also provide statistically significant evidence for the predictive superiority of the MMI recovery method over the NLLS recovery method.

⁵³While in some of these cases, the similarity can be traced back to the NLLS and the MMI recovering very similar parameters, in other cases it may be a consequence of the substitutability between the two parameters, β and ρ , with respect to the subject's local risk attitude.

⁵⁴In fact, these two subjects also have the highest number of GARP violations. Moreover, we provide an exact calculation of the Varian Inconsistency Index for all but three subjects (for whom we report overestimates, see Appendix B.1.2). These three subjects include the pair with the extreme Afriat Inconsistency Index values. The approximated Varian Inconsistency Index values for these two subjects are substantially greater than 0.1 for the minimum, MEAN and the SSQ aggregators. No other subjects have Varian Inconsistency Index greater than 0.1.

	# of Observations	# Correct Predictions by MMI	% Correct Predictions by MMI	p-value
DDA	1025	528	51.5%	0.1744
IDA	464	276	59.5%	<0.0001

Table 12: Refined Sample Results by Group - Aggregate Level Analysis.

DDA			IDA		
$X \geq 7$	$X \leq 2$	p-value	$X \geq 7$	$X \leq 2$	p-value
19	16	0.3679	22	9	0.0147

Table 13: Refined Sample Results by Group - Individual Level Analysis.

H.3 Disappointment Aversion: Refined Sample

The Definite Disappointment Averse (DDA) group is composed of those subjects for which both methods recover $\beta \geq 0$, whereas Indefinite Disappointment Averse (IDA) group is composed of those subjects for which β is negative for one or both recovery methods. After the inconsistency refinement we are left with 148 subjects in the DDA group and 53 subjects in the IDA group.

In the aggregate analysis we treat the whole set of observations as a single data set with 1332 observations for the DDA group and 477 for the IDA group. Then we remove all the indefinite pairwise comparisons. Table 12 demonstrates that, also when using the refined sample, the MMI recovery method remains a better predictor in both cases, but while its advantage is insignificant in the DDA group, it is highly significant in IDA group.

In the individual level analysis, using the refinement we are left with 84 subjects in the DDA group and 47 subjects in the IDA group. Table 13 demonstrates that also here, although the MMI recovery method predicts better than the NLLS recovery method in both DDA and IDA, the difference in predictive accuracy within the DDA group is insignificant. However, this difference within the IDA group is substantial and statistically significant.

Next, consider the Definite Elation Seeking (DES) group that includes those subjects for whom both recovery methods recover $\beta < 0$. After the refinement is applied, for the aggregate analysis the DES group includes 248 observations. The MMI recovery method predicted correctly 156 of the choice

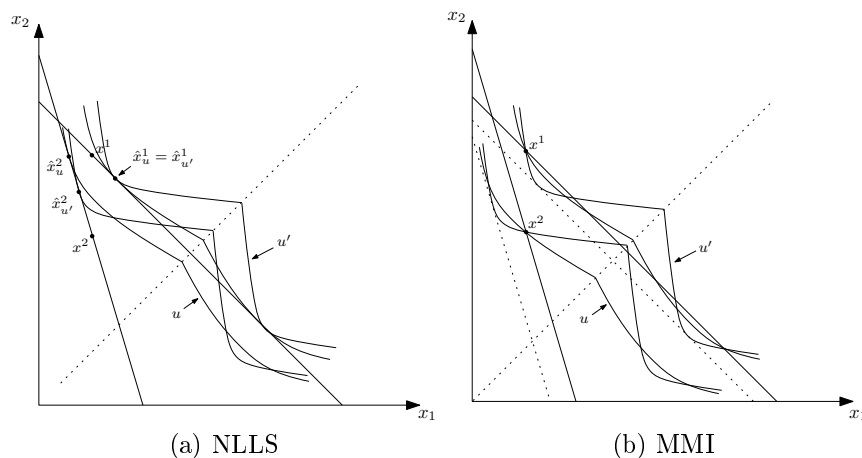


Figure I.1: MMI vs NLLS - Non-convex Preferences Recovery

problems, which amount to 62.9% of the observations. Hence, the difference between the recovery methods within the DES group is even more substantial than in the whole IDA group and it is highly statistically significant (p -value smaller than 0.0001).

The individual results for the DES group are similar - for 16 out of the 25 subjects that survive the refinement one method predicted correctly more than two thirds of the pairwise choices. It turns out that in 13 of the 16 cases, it was the MMI (81.3%, p -value 0.0106).⁵⁵

I Recovery of Non-Convex Preferences

Figure I.1 demonstrates how the MMI and NLLS may recover different sets of parameters for the same data set. Suppose we take two observations, x^1 and x^2 , and try to determine which of two utility functions – u and u' , is a better fit for the data. Define \hat{x}_v^i as the utility maximizing choice from budget line i given utility function v .

The left panel shows that the NLLS recovery method selects u' over u , as

⁵⁵We exclude Subject 1702 from the DES group since $\beta_{NLLS} \approx 0$. For similar reason we excluded also the definitive observations of Subject 604 from the previously mentioned aggregate analysis.

the distance between the utility maximizing bundle and the observed choice is identical at x^1 , and smaller for u' at x^2 . This arises from the lower price elasticity (higher non-convexity) implied by u' . The right panel demonstrates that the MMI selects u over u' using minimal budget set adjustment. The farther the observed portfolio is from the certainty line, the smaller is the adjustment required for the “flatter” (less non-convex) u compared to u' .

References

- Afriat, Sydney N.**, “The Construction of Utility Functions from Expenditure Data,” *International Economic Review*, February 1967, 8 (1), 67–77.
- , “Efficiency Estimates of Production Functions,” *International Economic Review*, October 1972, 13 (3), 568–598.
- , “On a System of Inequalities in Demand Analysis: An Extension of the Classical Method,” *International Economic Review*, June 1973, 14 (2), 460–472.
- , *Logic of Choice and Economic Theory*, Oxford: Clarendon Press, 1987.
- Aguiar, Victor and Roberto Serrano**, “Slutsky Matrix Norms and Revealed Preference Tests of Consumer Behaviour,” *Available at SSRN 2553995*, 2015b.
- Aguiar, Victor H and Roberto Serrano**, “Slutsky Matrix Norms: The Size, Classification, and Comparative Statics of Bounded Rationality,” Technical Report, Mimeo 2015a.
- Ahn, David, Syngjoo Choi, Shachar Kariv, Douglas Gale, David Ahn, Syngjoo Choi, Douglas Gale, and Shachar Kariv**, “Estimating Ambiguity Aversion in a Portfolio Choice Experiment,” *Quantitative Economics*, 2014, 5 (2), 195–223.

- Alcantud, José Carlos R., Daniel L. Matos, and Carlos R. Palmero**, “Goodness-of-fit in Optimizing a Consumer Model,” *Mathematical and Computer Modelling*, 2010, *52*, 1088–1094.
- Andreoni, James and John Miller**, “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism,” *Econometrica*, March 2002, *70* (2), 737–753.
- Apestequia, Jose and Miguel A. Ballester**, “A Measure of Rationality and Welfare,” *Journal of Political Economy*, December 2015, *123* (6), 1278–1310.
- Beatty, Timothy K. M. and Ian A Crawford**, “How Demanding is the Revealed Preference Approach to Demand?,” *The American Economic Review*, 2011, *101* (6), 2782–2795.
- Bernheim, B Douglas and Antonio Rangel**, “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics,” *The Quarterly Journal of Economics*, 2009, *124* (1), 51–104.
- Blundell, Richard W., Martin Browning, and Ian A. Crawford**, “Non-parametric Engel Curves and Revealed Preference,” *Econometrica*, 2003, *71* (1), 205–240.
- , – , **and** – , “Best Nonparametric Bounds on Demand Responses,” *Econometrica*, 2008, *76* (6), 1227–1262.
- Boodaghians, Shant and Adrian Vetta**, “Testing Consumer Rationality using Perfect Graphs and Oriented Discs,” in “International Conference on Web and Internet Economics” Springer 2015, pp. 187–200.
- Bronars, Stephen G**, “The power of nonparametric tests of preference maximization,” *Econometrica*, May 1987, *55* (3), 693–698.
- Cherchye, Laurens, Ian Crawford, Bram De Rock, and Frederic Vermeulen**, “The Revealed Preference Approach to Demand,” in Daniel J.

- Slottje, ed., *Quantifying Consumer Preferences (Contributions to Economic Analysis, Volume 288)*, Emerald Group Publishing Limited, 2009, pp. 247–279.
- Cherepanov, Vadim, Timothy Feddersen, and Alvaro Sandroni**, “Rationalization,” *Theoretical Economics*, September 2013, *8* (3), 775–800.
- Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv**, “Consistency and Heterogeneity of Individual Behavior under Uncertainty,” *American Economic Review*, December 2007, *97* (5), 1921–1938.
- , **Shachar Kariv, Wieland Muller, and Dan Silverman**, “Who is (More) Rational?,” *American Economic Review*, June 2014, *104* (6), 1518–1550.
- Cox, James C.**, “On Testing the Utility Hypothesis,” *The Economic Journal*, 1997, *107* (443), 1054–1078.
- Dean, Mark and Daniel Martin**, “Measuring Rationality with the Minimum Cost of Revealed Preference Violations,” *Review of Economics and Statistics*, July 2015, *98* (3), 524–534.
- Diewert, W. E.**, “Afriat and Revealed Preference Theory,” *Review of Economic Studies*, July 1973, *40* (3), 419–425.
- Echenique, Federico, Sangmok Lee, and Matthew Shum**, “The Money Pump as a Measure of Revealed Preference Violations,” *Journal of Political Economy*, 2011, *119* (6), 1201–1223.
- Epstein, Larry G.**, “Behavior Under Risk: Recent Developments in Theory and Applications,” in “Advances in Economic Theory,” Vol. 2 1992, pp. 1–63.
- Famulari, Melissa**, “Nonparametric Test of Demand Theory,” *Review of Economics and Statistics*, May 1995, *77* (2), 372–382.
- Fisman, Raymond, Shachar Kariv, and Daniel Markovits**, “Individual Preferences for Giving,” *American Economic Review*, May 2007, *97* (5), 1858–1876.

- Forges, Françoise and Enrico Minelli**, “Afriat’s Theorem for General Budget Sets,” *Journal of Economic Theory*, 2009, *144* (1), 135–145.
- Fostel, A., H. E. Scarf, and M. J. Todd**, “Two New Proofs of Afriat’s Theorem,” *Economic Theory*, 2004, *24*, 211–219.
- Fujishige, Satoru, Zaifu Yang et al.**, “On revealed preference and indivisibilities,” *Modern Economy*, 2012, *3* (06), 752.
- Geanakoplos, John**, “Afriat from MinMax,” *Economic Theory*, April 2013, *54* (3), 443–448.
- Greiner, Ben**, “Subject pool recruitment procedures: organizing experiments with ORSEE,” *Journal of the Economic Science Association*, 2015, *1* (1), 114–125.
- Gross, John**, “Testing data for consistency with revealed preference,” *The Review of Economics and Statistics*, 1995, pp. 701–710.
- **and Dan Kaiser**, “Two simple algorithms for generating a subset of data consistent with WARP and other binary relations,” *Journal of Business & Economic Statistics*, 1996, *14* (2), 251–255.
- Gul, Faruk**, “A Theory of Disappointment Aversion,” *Econometrica*, May 1991, *59* (3), 667–686.
- Halevy, Yoram, Dotan Persitz, and Lanny Zrill**, “On Non-parametric Bounds for Non-Convex Preferences,” *Working Paper*, 2016.
- Harbaugh, William T, Kate Krause, and Timothy R Berry**, “GARP for kids: On the development of rational choice behavior,” *The American economic review*, 2001, *91* (5), 1539–1545.
- Heufer, Jan**, “A Geometric Measure for the Violation of Utility Maximization,” *Working Paper*, 2008.

- , “Essays on Revealed Preference: Contributions to the Theory of Consumer’s Behavior.” PhD dissertation, Technische Universität Dortmund, Wirtschafts- und Sozialwissenschaftliche Fakultät, Lehrstuhl für Volkswirtschaftslehre (Mikroökonomie) 2009.
- **and Per Hjertstrand**, “Consistent subsets: Computationally feasible methods to compute the Houtman–Maks-index,” *Economics Letters*, 2015, *128*, 87–89.
- Houtman, Martijn**, “Nonparametric Consumer and Producer Analysis.” PhD dissertation, Faculty of Economics and Business Administration, University of Limburg, Maastricht, the Netherlands 1995. Dissertation no. 95-32.
- **and J.A.H. Maks**, “Determining all Maximal Data Subsets Consistent with Revealed Preference,” *Kwantitatieve Methoden*, 1985, *19*, 89–104.
- **and** – , “The Existence of Homothetic Utility Functions Generating Dutch Consumer Data,” Mimeo 1987.
- Jerison, David and Michael Jerison**, “Approximately Rational Consumer Demand,” *Economic Theory*, April 1993, *3* (2), 217–241.
- Karp, Richard M.**, “Reducibility Among Combinatorial Problems,” in R. E. Miller and J. W. Thatcher, eds., *Complexity of Computer Computations*, New York: Plenum, 1972, pp. 167–174.
- Manzini, Paola and Marco Mariotti**, “Sequentially Rationalizable Choice,” *American Economic Review*, 2007, *97* (5), 1824–1839.
- **and** – , “Categorize then Choose: Boundedly Rational Choice and Welfare,” *Journal of the European Economic Association*, 2012, *10* (5), 1141–1165.
- Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Ozbay**, “Revealed Attention,” *American Economic Review*, 2012, *102* (5), 2183–2205.

- Porter, Maria and Abi Adams**, “For love or reward? Characterising preferences for giving to parents in an experimental setting,” *The Economic Journal*, 2015. Forthcoming.
- Quiggin, John**, “A Theory of Anticipated Utility,” *Journal of Economic Behavior and Organization*, 1982, 3, 323–343.
- Reny, Philip J.**, “A Characterization of Rationalizable Consumer Behavior,” *Econometrica*, 2015, 83, 175–192.
- Rose, Hugh**, “Consistency of preference: the two-commodity case,” *The Review of Economic Studies*, 1958, 25 (2), 124–125.
- Rubinstein, Ariel and Yuval Salant**, “Eliciting Welfare Preferences from Behavioral Datasets,” *Review of Economic Studies*, 2012, 79 (1), 375–387.
- Samuelson, Paul A.**, “Complementarity: An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory,” *Journal of Economic Literature*, 1974, 12 (4), 1255–1289.
- Segal, Uzi and Avia Spivak**, “First Order versus Second Order Risk Aversion,” *Journal of Economic Theory*, June 1990, 51 (1), 111–125.
- Simon, Herbert A.**, “From Substantive to Procedural Rationality,” in “25 Years of Economic Theory,” Springer US, 1976, pp. 65–86.
- Smeulders, Bart, Frits CR Spieksma, Laurens Cherchye, and Bram De Rock**, “Goodness-of-fit Measures for Revealed Preference Tests: Complexity Results and Algorithms,” *ACM Transactions on Economics and Computation*, 2014, 2 (1), 3.
- , **Laurens Cherchye, Frits CR Spieksma, and Bram De Rock**, “The money pump as a measure of revealed preference violations: A comment,” *Journal of Political Economy*, 2013, 121 (6), 1248–1258.
- Swofford, James L and Gerald A Whitney**, “Nonparametric Tests of Utility Maximization and Weak Separability for Consumption, Leisure and Money,” *The Review of Economics and Statistics*, 1987, 69 (3), 458–464.

- Szpilrajn, Edward**, “Sur l’extension de l’ordre partiel,” *Fundamenta mathematicae*, 1930, *16* (1), 386–389.
- Teo, Chung Piaw and Rakesh V. Vohra**, “Afriat’s Theorem and Negative Cycles,” 2003.
- Tsur, Yacov**, “On Testing For Revealed Preference Conditions,” *Economics Letters*, 1989, *31*, 359–362.
- Varian, Hal R.**, “The Nonparametric Approach to Demand Analysis,” *Econometrica*, July 1982, *50* (4), 945–973.
- , “Non-Parametric Analysis of Optimizing Behavior with Measurement Error,” *Journal of Econometrics*, 1985, *30*, 445–458.
- , “Goodness-of-fit in Optimizing Models,” *Journal of Econometrics*, 1990, *46*, 125–140.
- , “Goodness-of-fit for Revealed Preference Tests,” *Working Paper*, 1993.
- Yaari, Menahem E.**, “Some Remarks on Measures of Risk Aversion and on Their Uses,” *Journal of Economic Theory*, 1969, *1*, 315–329.