Parametric Recoverability of Preferences

Yoram Halevy

University of Toronto

Dotan Persitz

Tel Aviv University

Lanny Zrill

Langara College

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.

I. Introduction

This paper studies the problem of recovering stable preferences from individual choices. The renewed interest in this problem emerges from the

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,

Electronically published July 3, 2018
[*Journal of Political Economy*, 2018, vol. 126, no. 4]

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 the 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 preference 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 nonsatiated 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 pref-

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 Oriel Nofekh provided excellent research assistance. Yoram Halevy gratefully acknowledges financial support from the Social Sciences and Humanities Research Council (SSHRC grants 435-2012-0577 and 410-2009-1062). Dotan Persitz gratefully acknowledges the support of the Israel Science Foundation (grant 1390/14) and the Henry Crown Institute for business research in Israel. Lanny Zrill gratefully acknowledges financial support from an SSHRC Doctoral Fellowship. Work on this project started when Zrill was a PhD student, Persitz was a postdoctoral fellow, and Halevy was a faculty member at the University of British Columbia. The experimental work was done under UBC's Behavioural Research Ethics Board approval H14-00510. Data are provided as supplementary material online.

erence relation).[1] 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 efficiency 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, acceptable, and locally nonsatiated utility functions.[2] 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 introduced 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 nonsatiated utility functions. If the data set does not satisfy GARP, each measure can be additively decomposed into

[1] If choices are inconsistent, the "revealed preference relation" refers to the ranking remaining after excluding cycles in some "minimal" way (see definition 1 below).

[2] A utility function is acceptable if the zero bundle is weakly worse than every other nonnegative bundle. See also definition 2.3 in Sec. 1 of the online appendix.

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 that 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" (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 measure, 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 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 constant relative risk aversion (CRRA) functional form, we recover parameters using nonlinear least squares (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) were 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 on the basis of their predictive success and apply it in an experimental setting similar to that of Choi et al. 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 these 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, subjects are 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 methods but also to observe subjects' choices in regions that may otherwise be unobservable. In particular, when subjects choose from linear budget sets, nonconvex 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 nonconvexity, 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 percent 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 percent of those subjects were more accurately predicted by the MMI recovery method. Moreover, when we focus our attention on only those subjects for whom the recovered parameters imply nonconvex preferences (i.e., local risk-seeking behavior), the MMI recovery method predicted more accurately in 62.5 percent of pairwise choices and for 75 percent of subjects for whom 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 in which 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 percent of the subjects are well approximated by expected utility compared to the general disappointment aversion functional form. In addition, we demonstrate nonnested 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 constant absolute risk aversion (CARA) utility index.

In the next section we generalize the standard definitions of revealed preference relations and extend Afriat's (1967) theorem to inconsistent data sets (theorem 1). In Section III, we introduce the main inconsistency indices discussed in the paper, and in Section IV, we introduce the money metric and the binary incompatibility measures and use them to characterize the inconsistency indices (theorem 2). In Section V, 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 VI, while the results are reported in Section VII. Section VIII demonstrates the use of our theoretical results for hypothesis testing and model selection. Section IX presents conclusions.

II. Preliminaries

Consider a decision maker (henceforth DM) who chooses bundles $x^i \in \mathfrak{R}^K_+$ $(i \in 1, \ldots, n)$ from budget menus $\{x : p^i x \leq p^i x^i, p^i \in \mathfrak{R}^K_{++}\}$. Let $D = \{(p^i, x^i)^n_{i=1}\}$ 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], and 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 $x^i R^0_{D,\mathbf{v}} x$, 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 $x^i P^0_{D,\mathbf{v}} x$, if $v^i p^i x^i > p^i x$;
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 $(x^j, x^k, \ldots, x^m)$ such that $x^i R^0_{D,\mathbf{v}} x^j$, $x^j R^0_{D,\mathbf{v}} x^k, \ldots, x^m R^0_{D,\mathbf{v}} x$.

[3] Throughout 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 for all i, $v_i = v'_i$, $\mathbf{v} \geqq \mathbf{v}'$ if for all i, $v_i \geq v'_i$, $\mathbf{v} \geq \mathbf{v}'$ if $\mathbf{v} \geqq \mathbf{v}'$, and $\mathbf{v} \neq \mathbf{v}'$ and $\mathbf{v} > \mathbf{v}'$ if for all i, $v_i > v'_i$.

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 Sec. 1.1 of the appendix).

FACT 1. Let $\mathbf{v}' \leq \mathbf{v}$. Then $R^0_{D,\mathbf{v}'} \subseteq R^0_{D,\mathbf{v}}$, $P^0_{D,\mathbf{v}'} \subseteq P^0_{D,\mathbf{v}}$, 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 $x^j P^0_{D,\mathbf{v}} x^i$.

When $\mathbf{v} = 1$, 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 $R^0_{D,\mathbf{v}}$ that contains no strict cycles while $R^0_{D,\mathbf{1}}$ 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 are in Secs. 1.2 and 1.3 of the appendix, 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 $x^i \in \mathfrak{R}^K_+$, $x^i R^0_{D,\mathbf{v}} x$ 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 that 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. Hence, $\mathbf{1}$-rationalizability reduces to the standard definition of rationalizability (Afriat 1967).[4]

Notice that $\mathbf{v}$-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 theorem was generalized in many directions. For example, Reny (2015) extended it to infinite data sets, Forges and Minelli (2009) to general budget sets, and Fujishige and Yang (2012) to indivisible goods. The following theorem generalizes Afriat's result to inconsistent data sets.

THEOREM 1. The following conditions are equivalent:

[4] Throughout the paper *rationalizability* means $\mathbf{1}$-*rationalizability*, D is *rationalizable* if it is $\mathbf{1}$-*rationalizable*, and D satisfies GARP if it satisfies GARP$_1$.

[5] For discussion and alternative proofs of the original theorem, see Diewert (1973), Varian (1982), Teo and Vohra (2003), Fostel, Scarf, and Todd (2004), and Geanakoplos (2013).

1. There exists a nonsatiated utility function that **v**-rationalizes the data.
2. The data satisfy $\text{GARP}_{\mathbf{v}}$.
3. There exists a continuous, monotone, and concave utility function that **v**-rationalizes the data.

Proof. See Section 1.4 of the appendix.[6]

III. Inconsistency Indices

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

DEFINITION 4. $f_n : [0, 1]^n \to [0, M]$, 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 $f : [0, 1]^n \to [0, M]$ be an aggregator function. *Varian's inconsistency index* is[10]

[6] Afriat (1973) uses the theorem of the alternative to provide a nonconstructive proof for the special case in which the coordinates of the adjustment vector are equal. Afriat (1987) states theorem 1 without a proof (theorem 6.3.I on p. 179). In his unpublished PhD dissertation, Houtman (1995, theorem 2.5) considers nonlinear 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.

[7] In 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.

[8] An aggregator function f_n is weakly decreasing if for every $\mathbf{v}, \mathbf{v}' \in [0, 1]^n$,

$$\mathbf{v} \geq \mathbf{v}' \Rightarrow f_n(\mathbf{v}) \leq f_n(\mathbf{v}'),$$
$$\mathbf{v} > \mathbf{v}' \Rightarrow f_n(\mathbf{v}) < f_n(\mathbf{v}').$$

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.

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

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

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

Varian (1990) suggested this index as a nonparametric 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 $\mathcal{I} = \{\mathbf{v} \in [0,1]^n : \mathbf{v} = v\mathbf{1}, \forall\ v \in [0,1]\}$ and a coordinate of a typical vector $\mathbf{v} \in \mathcal{I}$ by v.

DEFINITION 6. The *Afriat 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 $n - \Sigma_{i=1}^n v_i$. Again, we define this index with respect to an arbitrary aggregator function.

DEFINITION 7. Let $f : [0,1]^n \to [0,M]$ be an aggregator function. The *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 Section 1.5 of the appendix.

Afriat's and Houtman-Maks's inconsistency indices are considerably more prevalent in the empirical-experimental literature than Varian's inconsistency index, mainly because of computational considerations (discussed in Sec. 2.1 of the appendix).[11] However, definitions 5, 6, and 7 demonstrate that Afriat's and Houtman-Maks's 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 Section 2.1 of the appendix we 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

[11] The money pump inconsistency index proposed by Echenique, Lee, and Shum (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 Sec. 2.2 of the appendix. In Sec. 2.3 of the appendix we discuss an inconsistency index based on Euclidean distance rather than on revealed preference, related to an index mentioned in Beatty and Crawford (2011).

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], Masatlioğlu, Nakajima, and Ozbay [2012], and Cherepanov, Feddersen, and Sandroni [2013]) views the adjusted budget set as a consideration set that 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, and so forth. All the interpretations above take literally the existence of 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.

IV. 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 **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 Sec. 1.4 of the appendix).[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 toward risk, ambiguity, and time). Moreover, generically, a data set can be **v**-rationalized by more than

[12] Varian (1982) builds on the celebrated Afriat (1967) theorem to construct nonparametric bounds that partially identify the utility function, assuming that preferences are convex (see Halevy, Persitz, and Zrill 2017). His approach has been extended and developed in Blundell, Browning, and Crawford (2003, 2008) (see also sec. 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 nonconvex preferences.

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 trade-off 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 in which 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 IV.A we propose two loss functions that measure the incompatibility between the two rankings. Obviously, there are other loss functions that are not based on the incompatibility between the suggested utility function and the revealed preference relation. For example, NLLS 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 V and VII, we demonstrate empirically the difference between these two types of loss functions.

The main theoretical contribution of the paper is presented in Section IV.B. 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 VIII demonstrates the empirical implications of this decomposition to model selection.

A. *Incompatibility Indices*

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 $v^{*i}(D, u) = m(x^i, p^i, u)/p^i x^i$, where

$$m(x^i, p^i, u) = \min_{\{y \in \Re_+^K : u(y) \geq u(x^i)\}} p^i y.$$

Let $f: [0,1]^n \to [0, M]$ 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 nonsatiated, acceptable, and continuous utility functions on $\Re_+^K$.

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

Proof. See Section 1.6 of the appendix.

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. The function $\mathbf{v}^*(D, u)$ 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.[13]

If $\mathbf{v}^*(D, u) = 1$, 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 MMI measures the incompatibility between $\mathcal{U}$ and D.

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

[13] One 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 $R^0_{D,\mathbf{v}}$, 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. An 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 $\mathbf{v}^*(\cup_{i=1}^m D_i, u)$–rationalizes $\cup_{i=1}^m D_i$, where $\mathbf{v}^*(\cup_{i=1}^m D_i, u) = (\mathbf{v}^*(D_1, u)', \ldots, \mathbf{v}^*(D_m, u)')'$. Moreover, if $f_n(\cdot)$ is additive separable for every n, then $f_{\Sigma_{i=1}^m n_i}(\mathbf{v}^*(\cup_{i=1}^m D_i, u)) = \Sigma_{i=1}^m f_{n_i}(\mathbf{v}^*(D_i, u))$.

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

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 there does not exist 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 data set that includes only the ith observation from D. Then the ith element of the binary incompatibility vector tests whether the utility function rationalizes this data set. While the MMI is restricted to the classical environment of choice from linear budget sets, the BII may be easily applied to more general settings of choice from menus. The following proposition is the counterpart of proposition 1 for the BII.

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

Proof. See Section 1.7 of the appendix.

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

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

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 the proof in Sec. 1.8 of the appendix).

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 nonsatiated utility functions is a lower bound on the incompatibility indices for every subset of utility functions.

B. 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, we can recover preferences (within the parametric family $\mathcal{U}$) that approximate the consistent revealed preference information encoded in choices.[14] 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 nonsatiated, 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 the minimum aggregator, $I_M(D, f, \mathcal{U}^c)$ equals Afriat's 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 Section 1.9 of the appendix.

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 $f(\mathbf{v}) = 1 - \min_{i \in \{1, \ldots, n\}} v^i$, 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:

$$I_M(D, f, \mathcal{U}) = I_V(D, f) + [I_M(D, f, \mathcal{U}) - I_M(D, f, \mathcal{U}^c)],$$
$$I_B(D, f, \mathcal{U}) = I_{HM}(D, f) + [I_B(D, f, \mathcal{U}) - I_B(D, f, \mathcal{U}^c)].$$

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 Section 3 of the appendix.

Two reasons lead us to believe that such decomposition is essential for any method of recovering preferences of a DM who is inconsistent. First,

[14] Andreoni and Miller (2002) and Porter and Adams (2016) find that a great majority of the subjects satisfy GARP. However, other experimental studies (Choi et al. 2007, 2014; Fisman, Kariv, and Markovits 2007; Ahn et al. 2014) report that more than 75 percent of the subjects did not satisfy GARP.

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 can one 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.

V. Application to Choice under Risk

The goal of this section is to demonstrate the empirical applicability of the MMI as a criterion for recovering parametric preferences.[15] 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 (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 in which 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. estimate a disappointment aversion (DA) functional form introduced by Gul (1991) (for more details, see Sec. 4 of the appendix):

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

where

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

[15] In analyzing choices from budget menus, recovery based on the MMI retains more ranking information from the data than recovery based on the BII.

The parameter γ is the weight placed on the better outcome. For $\beta > 0$, the better outcome is underweighted relative to the objective probability (of .5) and the decision maker is *disappointment averse*. For $\beta < 0$, the better outcome is overweighted relative to the objective probability (of .5) and the decision maker is *elation seeking*. In the knife-edge case, where $\beta = 0$, the DA functional form 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 intertemporal 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 1 illustrates characteristic indifference curves for disappointment averse and elation seeking (locally nonconvex) subjects, respectively. Additionally, $w(x)$ is a standard utility function and is represented here by the CRRA functional form (we also report results in which the utility for wealth function is CARA, that is, $w(z) = -e^{-Az}$, 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,

$$\min_{\beta,\rho} \sum_{i=1}^{n} \left\| x^i - \arg \max_{x:\, p^i x \leq p^i x^i} (u(x; \beta, \rho)) \right\|, \tag{2}$$

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.[16]

[16] The 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 Sec. 2.1 of the appendix). The other three modules implement the NLLS, MMI (with various aggregators), and BII 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 constant elasticity of substitution 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 in the online data archive.

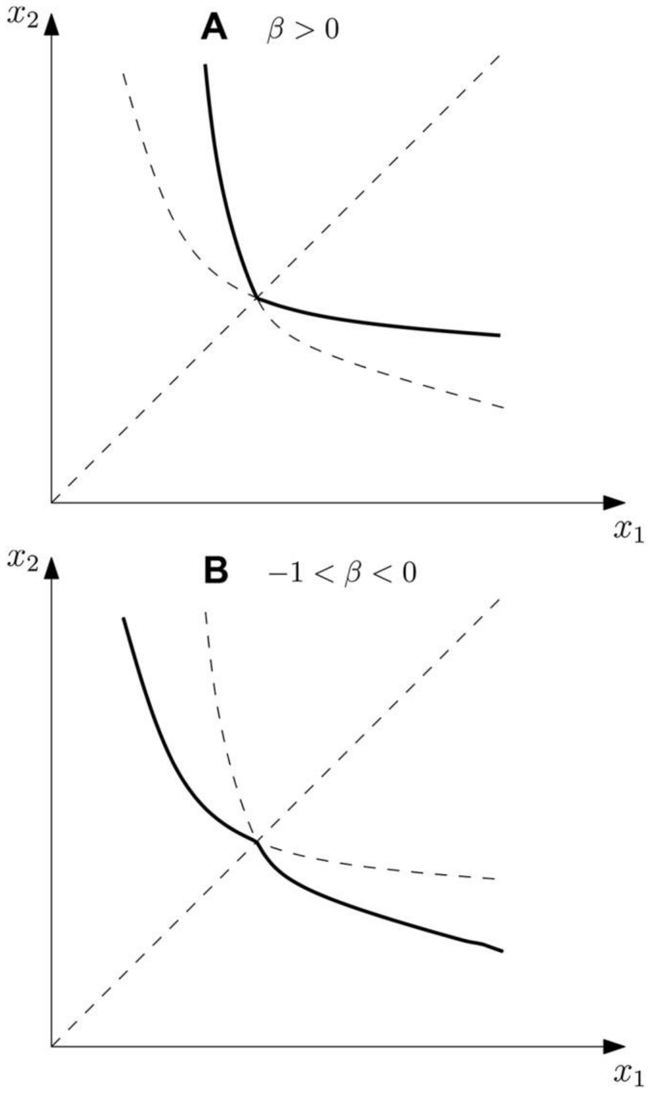

FIG. 1.—Typical indifference curves induced by Gul (1991) disappointment aversion function with $\beta \neq 0$. *A*, Disappointment aversion, $\beta > 0$. *B*, Elation seeking, $-1 < \beta < 0$.

A. Recovering Preferences for Inconsistent Subjects

In Section IV.B we prove the decomposition of the MMI 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 subjects who fail GARP.[17] Throughout the analysis, we exclude subjects with an unreliable Varian inconsistency index (nine out of 47 subjects).[18]

To illustrate, consider table 1, which compares the recovered parameters using the MMI 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 percent wasted income on average. On the other hand, the revealed preference information implied by subject 209's choices is nicely captured by the parametric family, since it implies incompatibility of only 5.63 percent, 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 percent $(I_M - I_V)$ 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 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 Sec. VIII).

B. Comparison of Recovered Parameters by Method

Figure 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 dif-

[17] Approximate 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.

[18] Computing the Varian inconsistency index is a hard computational problem (see the discussion in Sec. 2.1.2 of the appendix). The data of Choi et al. (2007) include 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 percent), and for four additional subjects we are able to provide a very good approximation. For the other nine subjects we report a weak approximation computed using an algorithm that overestimates 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.

TABLE 1
Comparing Consistent and Inconsistent Subjects

Subject	I_V	β	ρ	I_M
320	0	−.509	.968	.1322
209	.0288	.164	.352	.0563

ferences between the recovered parameters. However, when the preferences recovered by NLLS are nonconvex ($\beta < 0$), there seems to be no qualitative relation between the recovered parameters by the two methods.[19]

Moreover, the parameters recovered by NLLS in some of the nonconvex 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 Section 5 of the appendix.[20]

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.

VI. 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 rankings induced by each set of approximate preferences on these portfolios disagree. Therefore, each recovery method implied an 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.

[19] When β_{NLLS} is positive, then β_{NLLS} and β_{MMI} are significantly positively correlated ($p = .0283$), while when $\beta_{\text{NLLS}} < 0$, we cannot reject the null hypothesis of no linear correlation between them ($p = .1093$).

[20] Note 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.

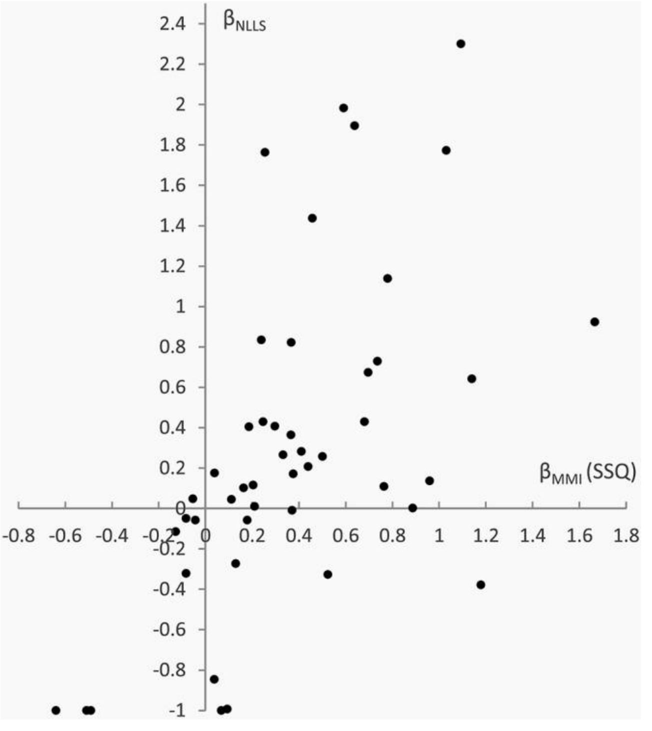

FIG. 2.—Disappointment aversion parameter: NLLS versus MMI (SSQ)

A. *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 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 Sec. 6.1 of the appendix 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. 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 to instantaneously 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 nine 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 : 1 exchange rate to Canadian dollars. Subjects were paid privately upon completion of the experiment, and their earnings averaged about C$19.53 in addition to a fixed fee of C$10.00 for showing up to the experiment on time.

B. *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 Section 6.1 of the appendix. 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 that lie on the boundary of the budget set to eliminate potential violations of monotonicity.[21]

The budget sets, and associated prices, were specifically chosen to address two issues. First, a sufficient overlap between budget sets is required

[21] Two 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 in which the values in both accounts are 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.

so that a GARP test will have sufficient power.[22] 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 Section 6.2 of the appendix.

C. 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 nine 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.[23]

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 a pairwise choice procedure to adjudicate between the two sets of recovered parameters, $\hat{\theta}_{\text{NLLS}} = \{\hat{\beta}_{\text{NLLS}}, \hat{\rho}_{\text{NLLS}}\}$ and $\hat{\theta}_{\text{MMI}} = \{\hat{\beta}_{\text{MMI}}, \hat{\rho}_{\text{MMI}}\}$. 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, $j \in \{\text{NLLS}, \text{MMI}\}$. In the case in which both $\hat{\beta}_{\text{NLLS}} > 0$ and $\hat{\beta}_{\text{MMI}} > 0$ (both recovered preferences are convex), we selected the safe portfolio to be the midpoint 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 in which at least one recovery method resulted in an elation seeking preference ($\hat{\beta}_{\text{NLLS}} < 0$ or $\hat{\beta}_{\text{MMI}} < 0$), part 2 of the experiment enabled us to identify the extent of nonconvexity of the underlying preferences, in addition to driving a wedge between the two sets of parameters. To achieve this additional goal we note that for locally nonconvex preferences the certainty equivalent may exceed the expected value for some risky portfolios. Therefore, the pairwise choice procedure searched for

[22] For a detailed analysis of a test that demonstrates that this set of budget sets is sufficiently powerful, see Sec. 6.2 of the appendix.

[23] A 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 part 1's interface was presented through a natural extension of a pairwise choice task. See the instructions in Sec. 6.1 of the appendix.

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 $x^S = E[x^R]$.[24] Similarly to the midpoint 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 direct evidence to the extent of nonconvexity of the underlying DM's preferences.[25]

To investigate the nature of local risk attitudes across subjects, the pairwise choice problems were constructed so that in six of them the risky portfolio was of *low variability* while in the other three problems, the risky portfolio was of *high variability*. For a detailed description of the algorithm that constructs the pairwise choices, see Section 6.3 of the appendix.

D. *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 on the basis of 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.

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 toward expected utility behavior ($\beta = 0$), a pattern observed for only about 40 percent of the subjects, as will be shown in Section VIII.B.

VII. Results: Pairwise Choice

The results of part 1 of the experiment exhibit patterns broadly similar to those reported in Section V for the data sets gathered by Choi et al. (2007)

[24] Since 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.

[25] The safe portfolio was the preferred alternative by the MMI recovery method in 927 of the 1,827 pairwise choices in our sample (50.7 percent).

(see Sec. 7 of the appendix).[26] We use these results extensively (together with the results of Choi et al.) in Section VIII 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 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), Section 8 of the appendix 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 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.

A. Results

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 1,827 observations (203 subjects times nine 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 percent) better predictor of subjects' choices both overall and for the two subclasses of portfolios separately (at an odds ratio of approximately 1.17).

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 nine 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 prob-

[26] The data gathered in the experiment are available in a separate Excel file named Halevy et al (2017)—Data. The disaggregated results of part 1 are available in a separate Excel file named Halevy et al (2017) Part 1—Results, both in the online data archive.

TABLE 2
Aggregate Results

	Observations	Correct Predictions by MMI	p-Value
Complete sample	1,827	986 (54.0%)	.0004
Low variability	1,218	652 (53.5%)	.0074
High variability	609	334 (54.8%)	.0093

ability of getting each one of these values at random is greater than 15 percent. 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 random prediction alone for a single subject is approximately 18 percent, 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 were 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.[27] As is evident from table 3, MMI is a significantly better predictor at the individual level as well (one-sided p-value .038), as it is a decisively better predictor for 45 percent more subjects than NLLS.[28]

B. Disappointment Aversion

1. Definite versus 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 whom both methods recover $\beta \geq 0$, whereas the indefinite disappointment averse (IDA) group is composed of those subjects for whom 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.

[27] The statistic for the multinomial likelihood ratio test is $-2\ln(L/R) = -2\Sigma_{i=1}^{k} x_i \ln(\pi_i/p_i)$, 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 nine degrees of freedom, has a p-value of approximately zero. Pearson's chi-squared test provides similar results.

[28] The 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 of being decisive.

TABLE 3
INDIVIDUAL-LEVEL RESULTS: 203 SUBJECTS

$X \geq 7$	$X \leq 2$	p-Value
61	42	.0378

In the aggregate analysis we treat the whole set of observations as a single data set with 1,350 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 the 5 percent level (but the advantage is not significant in the refined sample; see table 4 in Sec. 8 of the appendix). 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.

2. Definite Elation Seeking

Further, we focus on a subset of the 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 percent 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 < .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 percent of them (15 out of 20) the MMI produced the better prediction (p-value = .0207). These results suggest that the difference in predictive success between

TABLE 4
AGGREGATE RESULTS BY GROUP

	Observations	Correct Predictions by MMI	p-Value
DDA	1,350	706 (52.3%)	.0484
IDA	477	280 (58.7%)	.0001

TABLE 5
Individual-Level Results by Group

DDA (150)			IDA (53)		
$X \geq 7$	$X \leq 2$	p-Value	$X \geq 7$	$X \leq 2$	p-Value
38	30	.1981	23	12	.0448

the MMI and NLLS recovery methods can be attributed mostly (but not only) to subjects for whom the recovery methods resulted in apparent nonconvex preferences.

3. MMI versus NLLS When Preferences Are Nonconvex

The pairwise comparisons in part 2 of the experiment allow us to directly observe the subject's preferences in these nonconvex 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 nonconvex.

Specifically, for 21 of the 29 subjects in the DES group (72.4 percent) the disappointment aversion parameter recovered by the NLLS is more negative than the one recovered by the MMI.[29] 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 3 illustrates the choices from part 1 of the experiment for four 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.

C. *Illustrative Discussion*

To conclude this section we wish to suggest an informal explanation for our finding. Briefly, when choices are induced by nonconvex preferences for which the model is misspecified, the NLLS recovery method will most probably pick a set of parameters that implies greater nonconvexity 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 nonconvex region.

[29] For 19 of these 21 subjects the difference is more than 0.1. For six of the eight subjects for whom the parameter recovered by the NLLS is less negative than the one recovered by the MMI, the difference is less than 0.1.

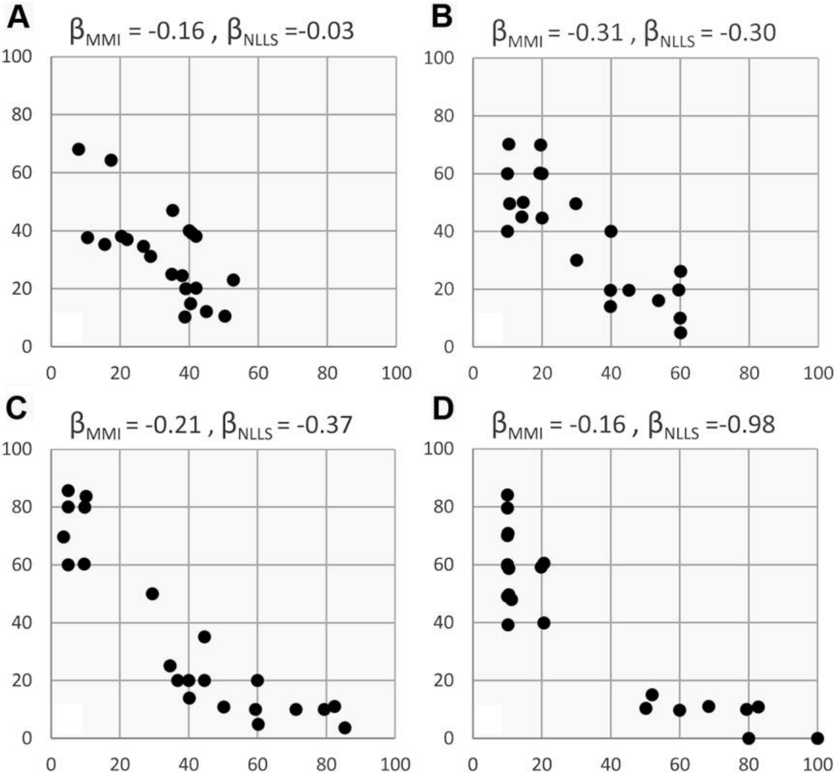

FIG. 3.—Patterns of choice: nonconvex preferences: *A*, subject 1203; *B*, subject 1512; *C*, subject 2203; *D*, subject 301.

To demonstrate the multiplicity of approximated preferences given the 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 4*A*. 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 4*B*.[30] 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 approximate this behavior within the DA model,

[30] Notice that the pattern of choice for these simulated subjects is very similar to that of subject 301 in fig. 3*D*. Not surprisingly, the recovered parameters for our simulated subject are also very similar to those of subject 301: $\beta_{\rm MMI} = -0.24$, $\rho_{\rm MMI} = 0.40$, $\beta_{\rm NLLS} = -0.91$, and $\rho_{\rm NLLS} = 1.55$.

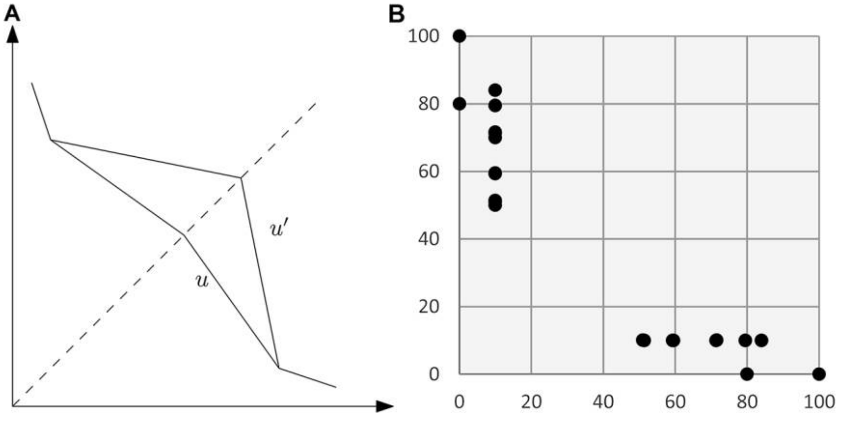

FIG. 4.—Two simulated subjects: *A*, typical indifference curves; *B*, choices given the linear choice problems presented in part 1 of the experiment.

which does not span this heuristic, NLLS resorts to substantial nonconvexity while the MMI can rationalize these choices without making strong claims on behavior that is unobservable using linear budget lines. For an informal demonstration, see Section 9 of the appendix.

VIII. 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 nonnested 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 nonnested model restrictions.[31]

A. *Evaluating Misspecification*

Using the decomposition of the MMI into the Varian inconsistency index (measure of consistency) and a residual that measures misspecification, 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 Section 2.1 of the appendix, 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 percent 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 inconsistency is larger than the portion of misspecification.[32]

B. *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.[33] Given the choice of functional form (disappointment

[31] For 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 (2017) Part 1—Results in the online data archive.

[32] Since those subjects for whom the Varian inconsistency index could not have been calculated properly were dropped, the sample slightly overrepresents the less inconsistent subjects.

[33] In the results files in the online data archive (Choi et al (2007)—Results and Halevy et al. (2017) Part 1—Results), we include descriptive statistics of the parameter frequencies in 1,000 resamplings 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 resamplings indeed recover confidence sets for the parameters, we merely interpret them as a measure for the sensitivity of the recovered parameters to extreme observations.

TABLE 6
Misspecification Using the Disappointment Aversion Functional Form

	Part 1 of the Experiment		Choi et al. (2007)	
	CRRA	CARA	CRRA	CARA
Number of subjects with at most 5% misspecification	136 (68%)	127 (63.5%)	26 (68.4%)	23 (60.5%)
Number 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%)
Original sample	203 subjects		47 subjects	
Consistent	92 (45%)		12 (26%)	
Dropped	3 (1.5%)		9 (195)	
Inconsistency level	At most 6%		At most 2.5%	

Note.—The sample includes all the subjects for whom the Varian inconsistency index was calculated exactly or with good approximation.

aversion with CRRA or CARA utility index), we use the ratio $[I_M(D, f, EU) - I_M(D, f, DA)]/[I_M(D, f, DA) - I_V(D, f)]$, 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 percent, 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 percent, implying that it is a reasonable model to capture their choices.

The results in table 7 demonstrate that choices of between one-third and one-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 a substantial increase in misspecification.

C. *Comparison of Nonnested Alternatives*

The MMI also allows the researcher to evaluate nonnested 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

TABLE 7
EVALUATING THE RESTRICTION TO EXPECTED UTILITY

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

NOTE.—The percentage of subjects for whom the additional misspecification implied by the expected utility restriction is less than 10 percent (the number of subjects who are well approximated by expected utility out of the number of subjects in the sample).

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 the 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 percent).

IX. 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 the data are from optimizing behavior (GARP), and a remainder that 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.

TABLE 8
CHOICE OF UTILITY INDEX

	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)

NOTE.—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). The full sample includes all subjects for whom the loss function was calculated. The restricted sample includes subjects whose Varian inefficiency index was calculated exactly or with good approximation and the difference in misspecification between the two indices is greater than 10 percent.

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 on the basis of an objective criterion, we design and execute an experiment that distinguishes between the methods on the basis of 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 nonconvex 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 a 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, choices made in pairwise comparisons reveal preferences in their purest form and permit their identification in scenarios in which 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 nonnested models. In the context of choice under risk, we demonstrate the relative importance of misspecification relative to inconsistency and that although a nonnegligible 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 nonconvexities are noncoincidental, as they result from a reasonable calculated procedure. We believe that an important next step in this research program is the integration of a stochastic component into the present deterministic model, while retaining the crucial distinction between inconsistency and misspecification.

References

Afriat, Sydney N. 1967. "The Construction of Utility Functions from Expenditure Data." *Internat. Econ. Rev.* 8 (1): 67–77.

———. 1972. "Efficiency Estimates of Production Functions." *Internat. Econ. Rev.* 13 (3): 568–98.

———. 1973. “On a System of Inequalities in Demand Analysis: An Extension of the Classical Method.” *Internat. Econ. Rev*. 14 (2): 460–72.

———. 1987. *Logic of Choice and Economic Theory*. Oxford: Clarendon.

Ahn, David, Syngjoo Choi, Douglas Gale, and Shachar Kariv. 2014. “Estimating Ambiguity Aversion in a Portfolio Choice Experiment.” *Quantitative Econ*. 5 (2): 195–223.

Alcantud, José Carlos R., Daniel L. Matos, and Carlos R. Palmero. 2010. “Goodness-of-Fit in Optimizing a Consumer Model.” *Math. and Computer Modelling* 52:1088–94.

Andreoni, James, and John Miller. 2002. “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism.” *Econometrica* 70 (2): 737–53.

Apesteguia, Jose, and Miguel A. Ballester. 2015. “A Measure of Rationality and Welfare.” *J.P.E*. 123 (6): 1278–1310.

Beatty, Timothy K. M., and Ian A. Crawford. 2011. “How Demanding Is the Revealed Preference Approach to Demand?” *A.E.R.* 101 (6): 2782–95.

Bernheim, B. Douglas, and Antonio Rangel. 2009. “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics.” *Q.J.E*. 124 (1): 51–104.

Blundell, Richard W., Martin Browning, and Ian A. Crawford. 2003. “Nonparametric Engel Curves and Revealed Preference.” *Econometrica* 71 (1): 205–40.

———. 2008. “Best Nonparametric Bounds on Demand Responses.” *Econometrica* 76 (6): 1227–62.

Cherchye, Laurens, Ian Crawford, Bram De Rock, and Frederic Vermeulen. 2009. “The Revealed Preference Approach to Demand.” In *Quantifying Consumer Preferences*, edited by Daniel J. Slottje, 247–79. Contributions to Economic Analysis, vol. 288. Bingley, UK: Emerald.

Cherepanov, Vadim, Timothy Feddersen, and Alvaro Sandroni. 2013. “Rationalization.” *Theoretical Econ*. 8 (3): 775–800.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv. 2007. “Consistency and Heterogeneity of Individual Behavior under Uncertainty.” *A.E.R.* 97 (5): 1921–38.

Choi, Syngjoo, Shachar Kariv, Wieland Muller, and Dan Silverman. 2014. “Who Is (More) Rational?” *A.E.R.* 104 (6): 1518–50.

Cox, James C. 1997. “On Testing the Utility Hypothesis.” *Econ. J*. 107 (443): 1054–78.

Dean, Mark, and Daniel Martin. 2015. “Measuring Rationality with the Minimum Cost of Revealed Preference Violations.” *Rev. Econ. and Statis*. 98 (3): 524–34.

Diewert, W. E. 1973. “Afriat and Revealed Preference Theory.” *Rev. Econ. Studies* 40 (3): 419–25.

Echenique, Federico, Sangmok Lee, and Matthew Shum. 2011. “The Money Pump as a Measure of Revealed Preference Violations.” *J.P.E*. 119 (6): 1201–23.

Epstein, Larry G. 1992. “Behaviour under Risk: Recent Developments in Theory and Application.” In *Advances in Economic Theory: Sixth World Congress*, vol. 2, edited by Jean-Jacques Laffont, 1–63. Cambridge: Cambridge Univ. Press.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits. 2007. “Individual Preferences for Giving.” *A.E.R.* 97 (5): 1858–76.

Forges, Francoise, and Enrico Minelli. 2009. “Afriat’s Theorem for General Budget Sets.” *J. Econ. Theory* 144 (1): 135–45.

Fostel, A., H. E. Scarf, and M. J. Todd. 2004. “Two New Proofs of Afriat’s Theorem.” *Econ. Theory* 24:211–19.

Fujishige, Satoru, and Zaifu Yang. 2012. “On Revealed Preference and Indivisibilities.” *Modern Economy* 3 (6): 752–58.

Geanakoplos, John. 2013. "Afriat from MinMax." *Econ. Theory* 54 (3): 443–48.

Greiner, Ben. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *J. Econ. Sci. Assoc.* 1 (1): 114–25.

Gross, John. 1995. "Testing Data for Consistency with Revealed Preference." *Rev. Econ. and Statis.* 77:701–10.

Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica* 59 (3): 667–86.

Halevy, Yoram, Dotan Persitz, and Lanny Zrill. 2017. "Non-parametric Bounds for Non-convex Preferences." *J. Econ. Behavior and Org.* 137:105–12.

Heufer, Jan. 2008. "A Geometric Measure for the Violation of Utility Maximization." Ruhr Economic Papers, no. 69, Technische Universität Dortmund.

———. 2009. "Essays on Revealed Preference: Contributions to the Theory of Consumer's Behavior." PhD thesis, Wirtschafts- und Sozialwissenschaftliche Fakultät, Lehrstuhl für Volkswirtschaftslehre (Mikroökonomie), Technische Universität Dortmund.

Heufer, Jan, and Per Hjertstrand. 2015. "Consistent Subsets: Computationally Feasible Methods to Compute the Houtman–Maks-Index." *Econ. Letters* 128: 87–89.

Houtman, Martijn. 1995. "Nonparametric Consumer and Producer Analysis." PhD thesis (no. 95-32), Faculty Econ. and Bus. Admin., Univ. Limburg, Maastricht.

Houtman, Martijn, and J. A. H. Maks. 1985. "Determining All Maximal Data Subsets Consistent with Revealed Preference." *Kwantitatieve Methoden* 19:89–104.

Manzini, Paola, and Marco Mariotti. 2007. "Sequentially Rationalizable Choice." *A.E.R.* 97 (5): 1824–39.

———. 2012. "Categorize Then Choose: Boundedly Rational Choice and Welfare." *J. European Econ. Assoc.* 10 (5): 1141–65.

Masatlioğlu, Yusufcan, Daisuke Nakajima, and Erkut Ozbay. 2012. "Revealed Attention." *A.E.R.* 102 (5): 2183–2205.

Porter, Maria, and Abi Adams. 2016. "For Love or Reward? Characterising Preferences for Giving to Parents in an Experimental Setting." *Econ. J.* 126 (598): 2424–45.

Reny, Philip J. 2015. "A Characterization of Rationalizable Consumer Behavior." *Econometrica* 83:175–92.

Rubinstein, Ariel, and Yuval Salant. 2012. "Eliciting Welfare Preferences from Behavioral Datasets." *Rev. Econ. Studies* 79 (1): 375–87.

Samuelson, Paul A. 1974. "Complementarity: An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory." *J. Econ. Literature* 12 (4): 1255–89.

Segal, Uzi, and Avia Spivak. 1990. "First Order versus Second Order Risk Aversion." *J. Econ. Theory* 51 (1): 111–25.

Simon, Herbert A. 1976. "From Substantive to Procedural Rationality." In *25 Years of Economic Theory: Retrospect and Prospect*, 65–86. Boston: Springer.

Smeulders, Bart, Frits C. R. Spieksma, Laurens Cherchye, and Bram De Rock. 2014. "Goodness-of-Fit Measures for Revealed Preference Tests: Complexity Results and Algorithms." *ACM Trans. Econ. and Computation* 2 (1): art. 3.

Teo, Chung Piaw, and Rakesh V. Vohra. 2003. "Afriat's Theorem and Negative Cycles." Discussion Paper no. 1377, Center Math. Studies Econ. and Management Sci., Northwestern Univ.

Tsur, Yacov. 1989. "On Testing for Revealed Preference Conditions." *Econ. Letters* 31:359–62.

Varian, Hal R. 1982. "The Nonparametric Approach to Demand Analysis." *Econometrica* 50 (4): 945–73.

———. 1985. "Non-parametric Analysis of Optimizing Behavior with Measurement Error." *J. Econometrics* 30:445–58.

———. 1990. "Goodness-of-Fit in Optimizing Models." *J. Econometrics* 46:125–40.

———. 1993. "Goodness-of-Fit for Revealed Preference Tests." Manuscript. http://econwpa.repec.org/eps/em/papers/9401/9401001.pdf.