School of Economics and Finance Discussion Papers

Revealed preferences over risk and uncertainty

Matthew Polisson, John K.-H. Quah and Ludovic Renou

School of Economics and Finance Discussion Paper No. 1706
29 Apr 2017 (revised 16 Apr 2019)
JEL Classification: C14, C60, D11, D12, D81
Keywords: expected utility, rank dependent utility, disappointment aversion, generalized axiom of revealed preference, first order stochastic dominance, risk aversion, Afriat efficiency, intertemporal choice

University of St Andrews

School of Economics and Finance
Online Discussion Paper Series
issn 2055-303X
http://ideas.repec.org/s/san/wpecon.html
info: econ@st-andrews.ac.uk

Revealed preferences over risk and uncertainty

Matthew Polisson, John K.-H. Quah, and Ludovic Renou*

April 16, 2019

Abstract: We develop a nonparametric procedure, called the *lattice method*, for testing the consistency of budgetary choice data with a broad class of models of choice under risk and under uncertainty. Our method can allow for risk loving and elation seeking attitudes, or it can be adapted to require risk aversion. It can also be used to calculate, via Afriat's efficiency index, the magnitude of violations from a particular model. We evaluate the performance of different models under risk (including expected utility, disappointment aversion, rank dependent utility, and stochastically monotone utility) in the data collected from several recent portfolio choice experiments.

Keywords: expected utility, rank dependent utility, disappointment aversion, generalized axiom of revealed preference, first order stochastic dominance, risk aversion, Afriat efficiency, intertemporal consumption

JEL classification numbers: C14, C60, D11, D12, D81

* Email addresses: matthew.polisson@st-andrews.ac.uk; john.quah@jhu.edu; l.renou@qmul.ac.uk.

This research has made use of the ALICE High Performance Computing Facility at the University of Leicester. Part of this research was carried out while Matthew Polisson was visiting Caltech and UC Berkeley, and while Ludovic Renou was visiting the University of Queensland, and they would like to thank these institutions for their hospitality and support. Ludovic Renou would like to acknowledge financial support from the French National Research Agency (ANR), under the grant CIGNE (ANR-15-CE38-0007-01). We are grateful to Ray Fisman, Yoram Halevy, Shachar Kariv, Felix Kubler, and Collin Raymond for helpful discussions and comments, as well as to seminar audiences at Adelaide, Bonn, Bristol, Brown, Caltech, Chapman (ESI), EUI, George Mason (ICES), Glasgow, IFN (Stockholm), Johns Hopkins, Lancaster, Leicester, NUS, NYU Stern, Oregon, Oxford, PSE, QUT, Queen Mary, Queensland, St Andrews, Toulouse, UBC, UC Berkeley, UC Irvine, and UC San Diego, and to participants at many workshops and conferences.

1. Introduction

This paper is a methodological contribution to the empirical investigation of decision making under risk and under uncertainty. While the expected utility (EU) model is the most widely used model for decision making in these contexts, there is a large literature developing models that aim to give a better account of observed choice behavior. An empirical literature that tests the EU and other models on experimental data has also emerged alongside these theoretical developments. These experiments often employ elicitation procedures in which subjects are in effect making repeated choices between two risky or uncertain outcomes; the data obtained in this way consist of a finite number of *binary choices*, which can then be used to partially recover a subject's preference. A more recent strand of experiments employs a different elicitation procedure, which we shall call the *budgetary choice* procedure. In these experiments, subjects are asked to choose a preferred option from a potentially infinite set of alternatives. For example, a subject could be presented with a portfolio problem where she has to allocate her budget between two assets with state-contingent payoffs. An early experiment of this kind, the data from which we analyze in this paper, is found in Choi *et al.* (2007);[1] other examples include Loomes (1991), Gneezy and Potters (1997), Bayer *et al.* (2013), Ahn *et al.* (2014), Choi *et al.* (2014), Hey and Pace (2014), Cappelen *et al.* (2015), and Halevy, Persitz, and Zrill (2018).

For reasons which we explain in Section 1.1, the nonparametric evaluation of data collected through a budgetary choice procedure requires a new methodological approach. The contribution of this paper is twofold: (1) we develop a new empirical method that could be used to analyze data (be it experimental or field data) collected from portfolio decisions, and (2) we apply this new method to evaluate the performance of different models of choice under risk using data from a number of recent portfolio choice experiments. Our method allows us to determine whether a data set is consistent with the EU model or some of its generalizations, *without* making parametric assumptions on the Bernoulli function (such as constant relative risk aversion) or on other features of the model. This is empirically important because if we happen to find that a data set is incompatible with a given model, then we can safely conclude that this incompatibility is attributable to the model itself rather than

[1] See this paper also for an account of the advantages of a budgetary choice approach.

a poorly selected parametric form. Since the test also yields a utility function (or functions) that best fits the data, these can be used to make out-of-sample predictions.

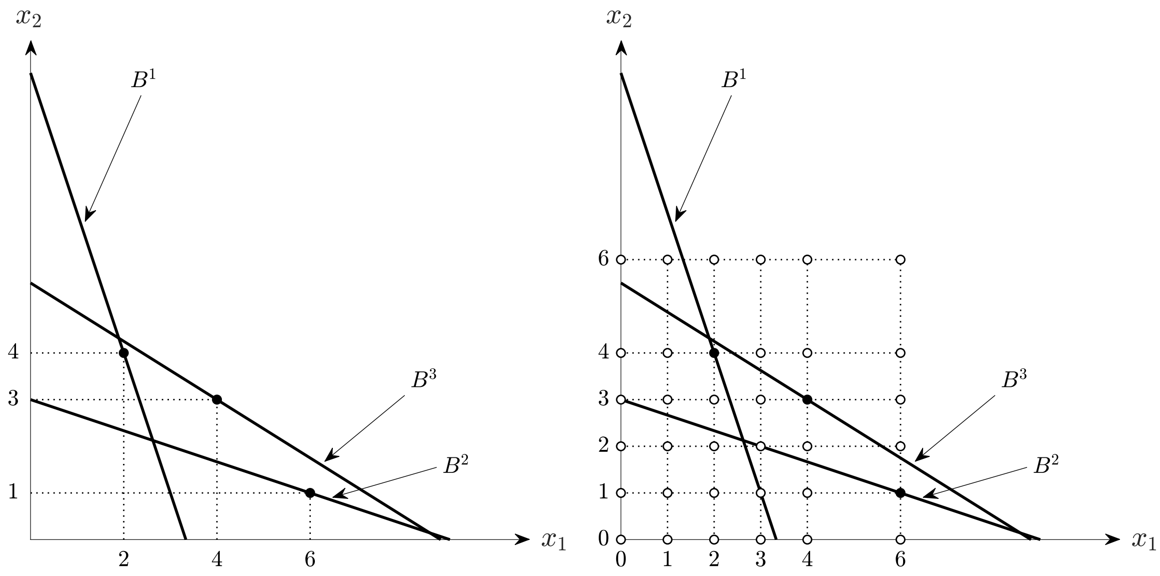
Our method can also be applied to test models of intertemporal choice (such as discounted utility) and other models which are formally similar to the EU model and its generalizations. Budgetary choice procedures are increasingly used in experiments to study intertemporal consumption (see, for example, Andreoni and Sprenger (2012) and Imai and Camerer (2016)).

1.1 Testing EU and other models on a finite lattice

A feature of the budgetary choice procedure is that instead of requiring a subject to choose one alternative or another, it allows her to calibrate a response and to choose something 'in between'. But this feature is also the crucial reason why the nonparametric analysis of data collected from this procedure requires a new empirical method, whereas no such method is necessary for binary choices. Indeed, suppose that we make a finite number of observations, where at observation t a subject chooses a lottery that gives a monetary payoff x_s^t in state s over one that gives y_s^t in state s (for $s = 1, 2, \ldots, \bar{s}$), and where the probability of state s is known to be $\pi_s > 0$. Imagine that we would like to test if this data set is consistent with the EU model. Checking for exact consistency with the EU model simply involves finding a strictly increasing Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\sum_{s=1}^{\bar{s}} \pi_s u(x_s^t) \geqslant \sum_{s=1}^{\bar{s}} \pi_s u(y_s^t)$ holds at every observation t. This amounts to solving a finite set of linear inequalities,[2] and it is computationally straightforward to ascertain if a solution exists. However, it is clear that this method will no longer work when the subject is choosing from classical budget sets, since even a single observed choice from a budget set reveals an infinite set of binary preferences between the chosen bundle and alternatives in the budget.

We now give a short and intuitive explanation of how our new method works. Consider a data set with three observations and two states, as depicted in Figure 1a; the horizontal axis corresponds to consumption in state 1, and the vertical axis to consumption in state 2. The subject chooses the contingent consumption bundle $\mathbf{x}^1 = (2, 4)$ from budget set B^1, $\mathbf{x}^2 = (6, 1)$ from B^2, and $\mathbf{x}^3 = (4, 3)$ from B^3, where B^1, B^2, and B^3 are classical linear budget sets.[3] Assume that the probability of state s is commonly known to be π_s.

[2] The unknowns to be solved are $\{u(r) : r = x_s^t \text{ or } y_s^t \text{ for some } t \text{ and } s\}$.
[3] Note that B^1 consists of the budget line *and* all the bundles below the line.



(a) Data set with three observations

(b) Data set with lattice inserted

Figure 1: The lattice method

This data set is said to be consistent with (or rationalizable by) the EU model if there is a strictly increasing Bernoulli function u such that $\pi_1 u(2) + \pi_2 u(4) \geqslant \pi_1 u(x_1) + \pi_2 u(x_2)$ for all $\mathbf{x} = (x_1, x_2)$ in B^1, and similarly at the other two observations.

We show (in Theorem 1, Section 2) that this data set can be rationalized by the EU model if it can be rationalized on an appropriately modified consumption set. Specifically, let $\mathcal{X}$ be the set of consumption levels that are observed to have been chosen at some observation and in some state, plus zero; in this example $\mathcal{X} = \{0, 1, 2, 3, 4, 6\}$. Then for the data set to be EU-rationalizable, it is *sufficient* (and obviously necessary) for it to be EU-rationalizable on the finite set $\mathcal{X}^2$, i.e., there is a strictly increasing function $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ such that the expected utility of $\mathbf{x}^1 = (2, 4)$ is greater than any other bundle in $B^1 \cap \mathcal{X}^2$, and so forth.[4] We refer to $\mathcal{X}^2$ as the *lattice generated by* $\mathcal{O}$; it is depicted by the open circles in Figure 1b.[5] Therefore, checking if a data set is EU-rationalizable involves checking if there is a solution to a *finite* set of linear inequalities, a problem which is computationally feasible.

The lattice method can also be used to check for consistency with other models of choice under risk (such as the rank dependent utility (RDU) model (Quiggin, 1982) and the dis-

[4] For example, since $(1, 6) \in B^1 \cap \mathcal{X}^2$, $\bar{u}$ must satisfy $\pi_1 \bar{u}(2) + \pi_2 \bar{u}(4) > \pi_1 \bar{u}(1) + \pi_2 \bar{u}(6)$. The full set of inequalities $\bar{u}$ must satisfy is displayed in Table 1 in Section 2.2.

[5] We call $\mathcal{X}^2$ a lattice because that is what it looks like in Figure 1b. It is also a lattice in the formal mathematical sense, but our results do not use lattice theory in any way.

appointment aversion (DA) model (Gul, 1991)) and under uncertainty (such as the maxmin expected utility model (Gilboa and Schmeidler, 1989)). The basic test we just described requires the Bernoulli function to be strictly increasing and continuous, but not necessarily concave, so a risk loving EU-maximizer would pass the test, which is as it should be if model consistency is the principal concern. But perhaps a researcher is (also) interested in whether the subject behaves as a risk averse EU-maximizer, or more generally whether the subject is consistent with the EU, DA, or RDU models with a concave Bernoulli function; these hypotheses could be tested as well. In all of these tests, the basic idea is to identify a finite number of consumption bundles within each constraint set such that the superiority (according to some utility function drawn from the model under consideration) of the observed choice over this finite set is sufficient to guarantee its superiority over *all* alternatives in the constraint set. Lastly, note that while the constraint sets in Figure 1 are depicted as linear budget sets, this restriction is not crucial to our method.

1.2 Empirical implementation and findings

We implement our empirical method on three data sets obtained from the well known portfolio choice experiments in Choi *et al.* (2007), Choi *et al.* (2014), and Halevy, Persitz, and Zrill (2018). In doing so, we are able to demonstrate the versatility and practicality of the lattice method, and also to reveal certain empirical features common to all three data sets. In the Choi *et al.* (2007) experiment, each subject was asked to purchase Arrow-Debreu securities under different budget constraints. There were two states of the world, and it was commonly known that states occurred either symmetrically (each with probability 1/2) or asymmetrically (one with probability 1/3 and the other with probability 2/3); the experimental designs in Choi *et al.* (2014) and Halevy, Persitz, and Zrill (2018) closely resemble the symmetric design in Choi *et al.* (2007).

We use the lattice method developed in this paper to test the model performance of the EU, DA, and RDU models. We also check whether a subject's observations are consistent with the maximization of some locally nonsatiated utility function on the contingent consumption space. This is the most permissive utility model possible and forms the backdrop to our empirical analysis; Afriat's (1967) Theorem tells us that compatibility with utility maximization can be assessed by testing the generalized axiom of revealed preference (GARP).[6] The GARP test could be strengthened to test for consistency with the maximization of a utility function that is *stochastically monotone*, in the sense that if a bundle dominates another with respect to first order stochastic dominance, then it must have higher utility; a test for stochastically monotone utility maximization has recently been developed by Nishimura, Ok, and Quah (2017), and we implement it in this paper for the first time. The EU, DA, and RDU models are all special cases of stochastically monotone utility maximization, which is in turn more stringent than locally nonsatiated utility maximization.

In a rich budgetary choice environment with many observations on behavior, a data set would typically not pass GARP (let alone more stringent requirements) exactly. It is, however, possible to quantify a data set's departure from rationalizability by a given model using the *critical cost efficiency index* (Afriat, 1973); this index is widely used in the empirical revealed preference literature, including in Choi *et al.* (2007) and Choi *et al.* (2014), while Halevy, Persitz, and Zrill (2018) implements a variant of this index first proposed by Varian (1990). The efficiency index runs from 1 to 0, with the index equal to 1 if a data set passes the test exactly. We adopt this measure of rationality throughout our empirical implementation.[7]

For each subject in each of the experiments, we are able to calculate the subject's efficiency indices for the different models under consideration. A negligible number of subjects pass GARP exactly. In the case of the primarily undergraduate subjects in Choi *et al.* (2007) and Halevy, Persitz, and Zrill (2018), more than 80% would pass GARP if we set a threshold of 0.9 for the efficiency index. In the case of the large scale (representative sample) experiment in Choi *et al.* (2014), the efficiency indices for GARP are distinctly lower, with almost 60% passing GARP at the 0.9 efficiency threshold. The following highlights some salient features of the data collected from the three experiments.

- A significant minority of subjects either violate GARP and/or stochastic monotonicity; the decisions of these subjects cannot be explained by the EU, DA, or RDU models since all of them respect first order stochastic dominance.

[6] This term and its acronym were coined by Varian (1982), who also provided a proof of Afriat's Theorem.

[7] We also carry out some analysis with Varian's version of the index which is reported in the Online Appendix. The calculation of Varian's index is more computationally demanding than calculating Afriat's, so our analysis with that index does not cover all of the models under consideration.

- Around half of the subjects who pass GARP (at some reasonable efficiency threshold) would also be compatible with the EU model; for these subjects, the EU model seems a good model of behavior, provided that some allowance is made for optimization errors.
- We find no evidence that the DA model accounts for the behavior of a significant proportion of subjects not accounted for by the EU model.
- On the other hand, there is some evidence that the RDU model could explain a significant part of the population not behaving as EU-maximizers.

Since our testing procedure also produces, for each subject, a rationalizing utility function belonging to a given model, that recovered utility function could then be used to make out-of-equilibrium predictions. We carry out a simple exercise of this type, using a rank dependent utility function estimated from a subject's portfolio decisions, to make predictions on the subject's choice when she is independently presented with a choice between two lotteries. Our objective here is to not check how often the procedure makes correct predictions, since the data we have access to do not allow us to explore that question in a meaningful way, but simply to illustrate the potential usefulness of our nonparametric methods for this purpose.

1.3 Relationship with the revealed preference literature

Our paper is related to the revealed preference literature originating from Afriat's (1967) Theorem, which characterizes price and demand observations that are consistent with the maximization of a locally nonsatiated utility function (see also Diewert (1973) and Varian (1982)). A natural follow up to Afriat's contribution is to characterize those data sets which are rationalizable by more specialized utility functions. Among these papers are those which characterize state price and contingent consumption demand observations that are consistent with the EU model[8] and (in more recent papers) some of its generalizations; these include Varian (1983a, 1983b, 1988), Green and Srivastava (1986), Diewert (2012), Bayer *et al.* (2013), Echenique and Saito (2015), Chambers, Liu, and Martinez (2016), and

[8] The EU model requires the utility function on the contingent consumption space to be additively separable. There are results which characterize data sets that are rationalizable by a *weakly* separable utility function (see Varian (1983a) and Quah (2014)).

Chambers, Echenique, and Saito (2016).[9] The principal difference between our results and this literature is that we do not rely on the sufficiency of first order conditions. This has two important implications: (i) the models we consider need not induce a convex preference over the contingent consumption space (e.g., we allow for risk loving behavior under EU or elation seeking behavior under DA/RDU), and (ii) we can weaken the requirement that the constraint set is a linear budget set. For reasons which we give in Section 3, allowing for nonlinear constraint sets enables our method to be used to calculate Afriat's efficiency index.

1.4 Organization of the paper

Section 2 describes how the lattice method can be used to test the EU, DA, and RDU models. Further applications of the lattice method to test models of decision making under uncertainty or over time, can be found in the Online Appendix. In Section 3 we explain the Afriat and Varian efficiency indices and how the lattice method is useful in their computation. Section 4 extends the lattice method to require the concavity of the Bernoulli function. The empirical implementation is found in Section 5. The Online Appendix contains further discussion of theoretical results and empirical findings.

2. The lattice method

We assume that there is a finite set of states, denoted by $S = \{1, 2, \ldots, \bar{s}\}$. The contingent consumption space is $\mathbb{R}^{\bar{s}}_+$; for a typical consumption bundle $\mathbf{x} \in \mathbb{R}^{\bar{s}}_+$, the sth entry, x_s, specifies the consumption level in state s.[10] There is a finite data set

$$\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$$

consisting of T observations, where $\mathbf{x}^t \in B^t$ and $B^t \subset \mathbb{R}^{\bar{s}}_+$. We could interpret this as data collected from an experiment where the subject chooses the bundle $\mathbf{x}^t$ from the constraint

[9] There is also a closely related literature on recovering expected utility from asset or contingent consumption demand *functions*, where, in effect the data set is assumed to be infinite (see, for example, Dybvig and Polemarchakis (1981) and Kubler, Selden, and Wei (2014)).

[10] Our results do depend on the realization in each state being one-dimensional (which can be interpreted as a monetary payoff, but not a bundle of goods). This case is the one most often considered in applications and experiments and is also the assumption in a number of recent papers, including Kubler, Selden, and Wei (2014), Echenique and Saito (2015), and Chambers, Echenique, and Saito (2016). The papers by Varian (1983a, 1983b), Green and Srivastava (1986), Bayer *et al.* (2013), and Chambers, Liu, and Martinez (2016) allow for multi-dimensional realizations. However, and crucially, in all of these papers, convexity of the agent's preference over contingent consumption and linear budget sets are also required.

set B^t at observation t and this will indeed be our interpretation throughout the paper.[11] We assume that B^t is a compact set.

We denote the *upper boundary* of B^t by ∂B^t; an element $\mathbf{x} \in B^t$ is said to be in ∂B^t if there is no $\mathbf{x}' \in B^t$ such that $\mathbf{x}' > \mathbf{x}$.[12] The *downward extension of* B^t is the set

$$\underline{B}^t = \{\mathbf{y} \in \mathbb{R}^{\bar{s}}_+ : \mathbf{y} \leqslant \mathbf{x} \text{ for some } \mathbf{x} \in B^t\}.$$

Obviously, $\underline{B}^t$ contains B^t. The most important example of a constraint set is the *classical* or *linear budget set*. At price vector $\mathbf{p} \in \mathbb{R}^{\bar{s}}_{++}$ and wealth $w > 0$, the classical budget set is $\mathscr{B}(\mathbf{p}, w) = \{\mathbf{x} \in \mathbb{R}^{\bar{s}}_+ : \mathbf{p} \cdot \mathbf{x} \leqslant w\}$. By a *classical data set*, which we denote by $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$, we mean a data set where, at observation t, the subject chooses $\mathbf{x}^t$ from

$$B^t = \mathscr{B}(\mathbf{p}, \mathbf{p}^t \cdot \mathbf{x}^t) = \{\mathbf{x} \in \mathbb{R}^{\bar{s}}_+ : \mathbf{p} \cdot \mathbf{x} \leqslant \mathbf{p}^t \cdot \mathbf{x}^t\}. \tag{1}$$

Thus $\mathcal{O}$ can also be written as $\{(\mathbf{x}^t, \mathscr{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t))\}_{t=1}^T$. Note that the upper boundary of $\mathscr{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t)$ is simply the *budget plane*, i.e., $\partial\mathscr{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t) = \{\mathbf{x} \in \mathbb{R}^{\bar{s}}_+ : \mathbf{p}^t \cdot \mathbf{x} = \mathbf{p}^t \cdot \mathbf{x}^t\}$, while the downward extension of $\mathscr{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t)$ is itself. The experiments conducted by Choi *et al.* (2007), Choi *et al.* (2014), and Halevy, Persitz, and Zrill (2018), the data from which we analyze in Section 5, involve subjects choosing from classical budget sets with two states.

Bear in mind, however, that our formulation only requires B^t to be compact; in particular, it does not have to be a linear budget set. For example, a nonlinear budget set occurs when a subject chooses contingent consumption through a portfolio of securities in an incomplete market (i.e., loosely speaking, when the number of securities is fewer than the number of states); in this case, the budget set will not be linear, but it will be compact so long as the security prices do not admit arbitrage.[13] In this paper, the crucial application requiring B^t to be nonlinear is in *quantifying approximate rationalizability* when a data set $\mathcal{O}$ cannot be exactly rationalized; as we explain in Section 3, this requires testing the rationalizability of

[11] Obviously nothing in principle forecloses the possibility of applying our method to observational budgetary choice data of the type found in insurance or financial decision problems.

[12] For the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\bar{s}}$, we write $\mathbf{x} \geqslant \mathbf{y}$ if $x_s \geqslant y_s$ for all s, and $\mathbf{x} > \mathbf{y}$ if $\mathbf{x} \geqslant \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$; if $x_s > y_s$ for all s, we write $\mathbf{x} \gg \mathbf{y}$.

[13] Indeed, there is $\mathbf{p}^t \gg 0$ such that $B^t = \{\mathbf{x} \in \mathbb{R}^{\bar{s}}_+ : \mathbf{p}^t \cdot \mathbf{x} \leqslant \mathbf{p}^t \cdot \mathbf{x}^t\} \cap \{Z + \boldsymbol{\omega}\}$, where Z is the span of assets available to the agent and $\boldsymbol{\omega}$ is the agent's endowment of contingent consumption. Both B^t and $\mathbf{x}^t$ will be known to the observer, if he knows the asset prices, the agent's holding of securities, the asset payoffs in every state, and the agent's endowment of contingent consumption $\boldsymbol{\omega}$.

a modified data set that has nonlinear constraint sets, even if the true constraint sets are linear budget sets.

Before presenting the formal results, we provide an informal explanation of the general approach we adopt in ascertaining whether a data set $\mathcal{O}$ is compatible with a given model of decision making under risk or uncertainty. We first notice that most models of decision making under risk and under uncertainty have two essential components: a *Bernoulli function* $u : \mathbb{R}_+ \to \mathbb{R}_+$ and an *aggregator function* $\phi : \mathbb{R}^{\bar{s}}_+ \to \mathbb{R}$, so that the utility of a bundle $\mathbf{x}$ is $\phi(u(x_1), \ldots, u(x_{\bar{s}}))$. For a *given* aggregator ϕ, the first step is to test whether there exists a Bernoulli function u, which rationalizes the data, i.e., which guarantees that, at each observation t, the utility of the chosen bundle $\mathbf{x}^t$ is weakly greater than the utility of any other bundle $\mathbf{x} \in B^t$. Theorem 1 (in Section 2.1) provides that test. However, a model of decision making under risk or uncertainty may correspond to a *family* of aggregators ϕ. If so, there is a second step that involves testing whether there is an aggregator in the family of aggregators under consideration *and* a Bernoulli function u that together rationalize the data. We now turn to the formal exposition.

Let $\{\phi(\cdot, t)\}_{t=1}^T$ be a collection of functions, where $\phi(\cdot, t) : \mathbb{R}^{\bar{s}}_+ \to \mathbb{R}$ is continuous and strictly increasing.[14] The data set $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is said to be *rationalizable by* $\{\phi(\cdot, t)\}_{t=1}^T$ if there exists a continuous and strictly increasing function $u : \mathbb{R}_+ \to \mathbb{R}_+$, which we shall refer to as the *Bernoulli function*, such that

$$\phi(\mathbf{u}(\mathbf{x}^t), t) \geqslant \phi(\mathbf{u}(\mathbf{x}), t) \text{ for all } \mathbf{x} \in B^t, \tag{2}$$

where $\mathbf{u}(\mathbf{x}) = (u(x_1), u(x_2), \ldots, u(x_{\bar{s}}))$. In other words, the observed choice behavior is consistent with the hypothesis that, at observation t, the subject has chosen a bundle from B^t that maximizes the utility function $\phi(\mathbf{u}(\cdot), t) : \mathbb{R}^{\bar{s}}_+ \to \mathbb{R}$.[15]

The function $\phi(\cdot, t)$ aggregates the vector of 'utils' $\mathbf{u}(\mathbf{x})$ into a single number. Of course, the most familiar formula for $\phi(\cdot, t)$ arises in the *expected utility* (EU) model; in this case, if the probability of state s at observation t is objectively known to be $\pi^t_s > 0$,

$$\phi(u_1, u_2, \ldots, u_{\bar{s}}, t) = \sum_{s=1}^{\bar{s}} \pi^t_s u_s. \tag{3}$$

[14] By strictly increasing, we mean that $\phi(\mathbf{z}, t) > \phi(\mathbf{z}', t)$ if $\mathbf{z} > \mathbf{z}'$.

[15] In keeping with the more empirically oriented parts of the revealed preference literature, this definition allows for the possibility that there are other bundles $\mathbf{x}$ in B^t that maximize $\phi(\mathbf{u}(\cdot), t)$.

Other models will lead to different formulations of ϕ (as we explain in greater detail in Section 2.4). Note that the objective probabilities need not vary across observations; in that case, ϕ would be independent of t.

Two requirements are imposed on the Bernoulli function u. Continuity is an important technical condition because it guarantees that $\phi(\mathbf{u}(\cdot), t)$ is continuous, which in turn guarantees that the agent's utility maximization problem always has a well behaved solution on compact constraint sets.[16] The other requirement on u is that it is strictly increasing. Notice that *some* assumption of this type is necessary: in particular, if we allow u to be a constant function then *every* data set $\mathcal{O}$ is rationalizable because the subject would be indifferent across all bundles in $\mathbb{R}^{\bar{s}}_{+}$. Requiring u to be strictly increasing is reasonable since its argument is typically interpreted as money. This assumption, together with the assumption that $\phi(\cdot, t)$ is a strictly increasing function guarantees that

$$\phi(\mathbf{u}(\mathbf{x}^t), t) > \phi(\mathbf{u}(\mathbf{x}), t) \text{ for all } \mathbf{x} \in \underline{B}^t \backslash \partial B^t.^{17} \tag{4}$$

In other words, a bundle that is not on the boundary of the constraint set has strictly lower utility than the chosen bundle $\mathbf{x}^t$.

Note that we do not require u to be concave. There are two reasons for this. First, it is desirable to have a test with a minimum of auxiliary assumptions, so that any rejection of the model would be decisive and could not attributed to the effect of the ancillary assumption. This is relevant because we know that concavity of u in the EU model imposes observable restrictions on portfolio choices over and above those implied by the EU model (see Section A4 of the Online Appendix).[18] Second, while the concavity of u is an oft-imposed assumption because it facilitates theoretical and empirical analysis, it is *not* a fundamental part of EU theory or many of its later generalizations. Indeed, departures from concavity have even been exploited to explain certain empirical phenomena; an early paper of that type is Friedman and Savage (1948). In prospect theory, the nonconcavity of the Bernoulli function around

[16] To be precise it guarantees that the optimal solutions form a nonempty compact set and is (in the case of demand) an upper hemicontinuous correspondence of prices.

[17] Indeed, if $\mathbf{x} \in \underline{B}^t \backslash \partial B^t$, then there is $\mathbf{y} \in B^t$ such that $\mathbf{y} \geqslant \mathbf{x}$ and by the optimality of $\mathbf{x}^t$, $\phi(\mathbf{u}(\mathbf{x}^t), t) \geqslant \phi(\mathbf{u}(\mathbf{y}), t)$. If $\mathbf{y} = \mathbf{x}$ then $\mathbf{y} \in B^t \backslash \partial B^t$, so $\phi(\mathbf{u}(\mathbf{x}^t), t) > \phi(\mathbf{u}(\mathbf{y}), t) = \phi(\mathbf{u}(\mathbf{x}), t)$. Otherwise, $\mathbf{y} > \mathbf{x}$ and by the strict increasing property, $\phi(\mathbf{u}(\mathbf{x}^t), t) \geqslant \phi(\mathbf{u}(\mathbf{y}), t) > \phi(\mathbf{u}(\mathbf{x}), t)$.

[18] This situation is unlike that in Afriat's Theorem, where it is known that concavity has no additional observable restrictions.

a reference point also plays a crucial role.[19] So there is advantage in having a test that is agnostic about the curvature of u.

2.1 The Main Result

Let Y be any subset of $\mathbb{R}_+$. Given a Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$, the function $\bar{u} : Y \to \mathbb{R}_+$ is the *restriction* of u to Y, if the functions agree on Y, i.e., $\bar{u}(r) = u(r)$ for all $r \in Y$. In the other direction, a Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$ is said to *extend* a function $\bar{u} : Y \to \mathbb{R}_+$ if the two functions agree on Y.

Given a data set $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$, we define

$$\mathcal{X}^* = \{0\} \cup \{x' \in \mathbb{R}_+ : x' = x_s^t \text{ for some } t,\ s\}; \tag{5}$$

besides zero, $\mathcal{X}^*$ contains those levels of consumption that are chosen at some observation and in some state. Since the data set is finite, so is $\mathcal{X}^*$. Let $\mathcal{X}$ be a finite subset of $\mathbb{R}_+$ containing $\mathcal{X}^*$. We define $\mathcal{L} = \mathcal{X}^{\bar{s}}$ and shall refer to $\mathcal{L}$ as the *lattice* associated with $\mathcal{O}$, because $\mathcal{L}$ has this appearance when $\bar{s} = 2$ (see Figure 1b).[20] Suppose that $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is rationalizable by $\{\phi(\cdot, t)\}_{t=1}^T$ with the Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$. If $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ is the restriction of u to $\mathcal{X}$, then

$$\phi(\bar{\mathbf{u}}(\mathbf{x}^t), t) \geqslant \phi(\bar{\mathbf{u}}(\mathbf{x}), t) \ \text{ for all } \mathbf{x} \in \underline{B}^t \cap \mathcal{L} \tag{6}$$

(where $\bar{\mathbf{u}}(\mathbf{x}) = (\bar{u}(x_1), \bar{u}(x_2), \ldots, \bar{u}(x_{\bar{s}}))$) and

$$\phi(\bar{\mathbf{u}}(\mathbf{x}^t), t) > \phi(\bar{\mathbf{u}}(\mathbf{x}), t) \ \text{ for all } \mathbf{x} \in \left(\underline{B}^t \backslash \partial B^t\right) \cap \mathcal{L}. \tag{7}$$

This follows immediately from (2) and (4) since $u(r) = \bar{u}(r)$ for all $r \in \mathcal{X}$ and $B^t \cap \mathcal{L} \subset B^t$ and $\left(\underline{B}^t \backslash \partial B^t\right) \cap \mathcal{L} \subset \underline{B}^t \backslash \partial B^t$. In other words, if u rationalizes the data set $\mathcal{O}$ then it will continue to rationalize the data set if the consumption space is restricted to the lattice $\mathcal{L}$. Our main theorem says that the converse of this statement is also true.[21]

[19] An extension to gains/losses around a reference point is presented in Section A3 of the Online Appendix.

[20] $\mathcal{L}$ is also a lattice in the mathematical sense, but our results do not rely on lattice theory.

[21] We cannot replace $\underline{B}^t$ with B^t in (6) and (7). For example, suppose $\mathbf{x}^1 = (1, 0)$ is chosen from $B^1 = \{(x_1, x_2) \in \mathbb{R}_+^2 : 2x_1 + x_2 = 2\}$ and $\mathbf{x}^2 = (0, 1)$ is chosen from $B^2 = \{(x_1, x_2) \in \mathbb{R}_+^2 : x_1 + 2x_2 = 2\}$ (so the constraint sets are straight lines). These observations cannot be rationalized by any increasing utility function and, in particular, cannot be rationalized in the sense of Theorem 1 (with ϕ constant across t). However, since $\mathcal{L} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, $B^1 \cap \mathcal{L} = \{(1, 0)\}$ and $B^2 \cap \mathcal{L} = \{(0, 1)\}$, conditions (6) and (7) hold if $\underline{B}^t$ is replaced with B^t. On the other hand $\left(\underline{B}^1 \backslash \partial B^1\right) \cap \mathcal{L}$ contains $(0, 1)$ and $\left(\underline{B}^2 \backslash \partial B^2\right) \cap \mathcal{L}$ contains $(1, 0)$, so (7) requires $\phi(\bar{\mathbf{u}}(\mathbf{x}^1)) > \phi(\bar{\mathbf{u}}(\mathbf{x}^2))$ and $\phi(\bar{\mathbf{u}}(\mathbf{x}^1)) < \phi(\bar{\mathbf{u}}(\mathbf{x}^2))$, which plainly cannot happen. This allows us to conclude, correctly, that this data set is not rationalizable.

Theorem 1. *Suppose that* $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ *is rationalizable by the collection of continuous and strictly increasing functions* $\{\phi(\cdot, t)\}_{t=1}^T$ *with the Bernoulli function* $u : \mathbb{R}_+ \to \mathbb{R}_+$. *Let* $\mathcal{X}$ *be a finite set in* $\mathbb{R}_+$ *that contains* $\mathcal{X}^*$ *(as defined by (5)) and let* $\mathcal{L} = \mathcal{X}^{\bar{s}}$. *Then the restriction of* u *to* $\mathcal{X}$, $\bar{u} : \mathcal{X} \to \mathbb{R}_+$, *satisfies conditions (6) and (7).*

Conversely, suppose that given $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ *and a collection of continuous and strictly increasing functions* $\{\phi(\cdot, t)\}_{t=1}^T$, *there is a strictly increasing function* $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ *that satisfies conditions (6) and (7). Then there is a Bernoulli function* $u : \mathbb{R}_+ \to \mathbb{R}_+$ *that extends* $\bar{u}$ *and with which* $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ *is rationalizable by* $\{\phi(\cdot, t)\}_{t=1}^T$.

This theorem tells us that testing for the rationalizability of $\mathcal{O}$ is equivalent to testing for rationalizability in the case where the agent's consumption space is considered to be $\mathcal{L}$ rather than $\mathbb{R}_+^{\bar{s}}$, which (crucially) reduces the rationality requirements to a *finite* number of optimality conditions involving the observed choices and alternatives (see (6) and (7)), and with the Bernoulli function defined on $\mathcal{X}$ rather than $\mathbb{R}_+$.

The intuition for Theorem 1 ought to be strong. Given $\bar{u}$ satisfying (6) and (7), we can define the step function $\hat{u} : \mathbb{R}_+ \to \mathbb{R}_+$ where $\hat{u}(r) = \bar{u}([r])$, with $[r]$ being the largest element of $\mathcal{X}$ weakly lower than r, i.e., $[r] = \max\{r' \in \mathcal{X} : r' \leqslant r\}$. Notice that $\phi(\hat{\mathbf{u}}(\mathbf{x}^t), t) = \phi(\bar{\mathbf{u}}(\mathbf{x}^t), t)$ and, for any $\mathbf{x} \in \underline{B}^t$, $\phi(\hat{\mathbf{u}}(\mathbf{x}), t) = \phi(\bar{\mathbf{u}}([\mathbf{x}]), t)$, where $[\mathbf{x}] = ([x_1], [x_2], \ldots, [x_{\bar{s}}])$ in $\underline{B}^t \cap \mathcal{L}$. Clearly, if $\bar{u}$ obeys (6) and (7) then $\mathcal{O}$ is rationalized by $\{\phi(\cdot, t)\}_{t=1}^T$ and $\hat{u}$ (in the sense that (2) holds). This falls short of the claim in the theorem only because $\hat{u}$ is neither continuous nor strictly increasing;[22] the proof in the Appendix shows how one could in fact construct a function with these additional properties.

Note that Theorem 1 gives some leeway on how $\mathcal{X}$ is chosen. If we are simply interested in testing for the rationalizability of $\mathcal{O}$ by a given model, then we could pick $\mathcal{X} = \mathcal{X}^*$, but sometimes it is advantageous to let $\mathcal{X}$ be a strictly larger set (see Section 2.3 on making out-of-sample predictions). Note also that in checking the conditions (6) and (7) we can confine ourselves to checking those bundles $\mathbf{x}'$ in $\underline{B}^t \cap \mathcal{L}$ which are not dominated by some other bundle in $\underline{B}^t \cap \mathcal{L}$. This is because if $\mathbf{x}' > \mathbf{x}''$ and property (6) or (7) holds for $\mathbf{x} = \mathbf{x}'$, it will also hold for $\mathbf{x} = \mathbf{x}''$ since both ϕ and $\bar{u}$ are strictly increasing.

[22] Recall that the Bernoulli function u is continuous and strictly increasing by definition.

Theorem 1 provides us with a convenient way of testing for rationalizability by the EU model. Consider an experiment where the probability of any state can be (possibly) varied across observations and where these probabilities are announced, so that both the observer and the subject know that the probability of state s at observation t is $\pi_s^t > 0$. The data set $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is *EU-rationalizable* if there is a Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\sum_{s=1}^{\bar{s}} \pi_s^t u(x_s^t) \geqslant \sum_{s=1}^{\bar{s}} \pi_s^t u(x_s)$ for all $\mathbf{x} \in B^t$. Theorem 1 tell us that $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is EU-rationalizable if and only if there is a strictly increasing function $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ for which (6) and (7) hold, with ϕ given by (3), i.e.,

$$\sum_{s=1}^{\bar{s}} \pi_s^t \bar{u}(x_s^t) \geqslant \sum_{s=1}^{\bar{s}} \pi_s^t \bar{u}(x_s) \text{ for all } \mathbf{x} \in \underline{B}^t \cap \mathcal{L} \text{ and} \tag{8}$$

$$\sum_{s=1}^{\bar{s}} \pi_s^t \bar{u}(x_s^t) > \sum_{s=1}^{\bar{s}} \pi_s^t \bar{u}(x_s) \text{ for all } \mathbf{x} \in \left(\underline{B}^t \backslash \partial B^t\right) \cap \mathcal{L}. \tag{9}$$

This is a system of linear inequalities, and solving it is both formally possible (in the sense that there is an algorithm that can decide within a known number of steps whether or not it has a solution) and computationally feasible.

As an example of how this works in practice, consider again the data set depicted in Figure 1 in Section 1. Suppose that it is commonly known that the probability of state s ($s = 1, 2$) at observation t is π_s^t. Since the three observed choices are $\mathbf{x}^1 = (2, 4)$, $\mathbf{x}^2 = (6, 1)$ and $\mathbf{x}^3 = (4, 3)$, $\mathcal{X}^* = \{0, 1, 2, 3, 4, 6\}$. Choosing $\mathcal{X} = \mathcal{X}^*$, EU-rationalizability can be tested by checking for a solution to the conditions listed in Table 1. In the top left panel are the strict inequalities guaranteeing that $\bar{u}$ is strictly increasing. The other panels list the conditions for the optimality of $\mathbf{x}^1$ in B^1, $\mathbf{x}^2$ in B^2, and $\mathbf{x}^3$ in B^3.[23] For example, at observation 1, the observed choice is $\mathbf{x}^1 = (2, 4)$ and there are 18 bundles in $B^1 \cap \mathcal{L}$ (besides $\mathbf{x}^1$), of which only the two bundles $(1, 6)$ and $(3, 1)$ are undominated,[24] with the former in the interior of the budget set and the latter on the upper boundary. The first inequality in the top right panel is imposed by (9) and the second inequality by (8). Similarly, the reader can check that there are two undominated bundles in $B^2 \cap \mathcal{L}$ and two in $B^3 \cap \mathcal{L}$, leading to

[23] Recall that since each B^t is a classical linear budget set, we have $\underline{B}^t = B^t$ for all t.

[24] See the remarks at the end of Section 2.1.

Monotonicity of $\bar{u}$	Optimality of $\mathbf{x}^1 = (2,4)$
$\bar{u}(6) > \bar{u}(4) > \bar{u}(3) > \bar{u}(2)$	$\pi_1^1\bar{u}(2) + \pi_2^1\bar{u}(4) > \pi_1^1\bar{u}(1) + \pi_2^1\bar{u}(6)$
$\bar{u}(2) > \bar{u}(1) > \bar{u}(0)$	$\pi_1^1\bar{u}(2) + \pi_2^1\bar{u}(4) \geqslant \pi_1^1\bar{u}(3) + \pi_2^1\bar{u}(1)$
Optimality of $\mathbf{x}^2 = (6,1)$	**Optimality of $\mathbf{x}^3 = (4,3)$**
$\pi_1^2\bar{u}(6) + \pi_2^2\bar{u}(1) \geqslant \pi_1^2\bar{u}(0) + \pi_2^2\bar{u}(3)$	$\pi_1^3\bar{u}(4) + \pi_2^3\bar{u}(3) > \pi_1^3\bar{u}(2) + \pi_2^3\bar{u}(4)$
$\pi_1^2\bar{u}(6) + \pi_2^2\bar{u}(1) \geqslant \pi_1^2\bar{u}(3) + \pi_2^2\bar{u}(2)$	$\pi_1^3\bar{u}(4) + \pi_2^3\bar{u}(3) > \pi_1^3\bar{u}(6) + \pi_2^3\bar{u}(1)$

Table 1: Conditions on $\bar{u}$ for EU-rationalizability

the inequality conditions displayed in the bottom panels. EU-rationalizability holds if there is a $\bar{u}$ that solves the linear inequalities displayed in Table 1.[25]

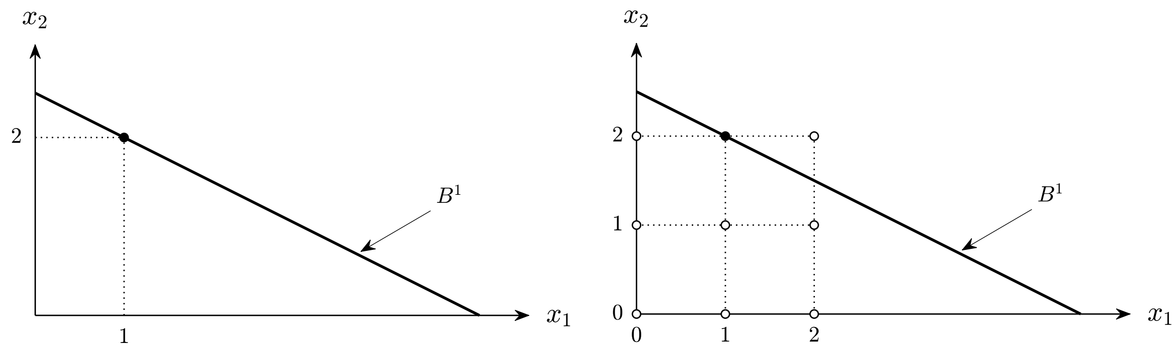
At this point it is worth emphasizing that requiring a data set to be EU-rationalizable is certainly more stringent than simply requiring it to be rationalizable by a locally nonsatiated utility function (on the contingent consumption space $\mathbb{R}_+^{\bar{s}}$). Indeed, while a data set with a single observation $(\mathbf{x}^1, \mathbf{p}^1)$ must necessarily be rationalizable in that sense, even a single observation can be incompatible with the EU model.

Example 1. Suppose that at the price vector $\mathbf{p}^1 = (1,2)$, the subject chooses the bundle $\mathbf{x}^1 = (1,2)$. This subject is buying more of the more expensive good, which is incompatible with the maximization of expected utility when the two states are equiprobable. It would, of course, fail the lattice test. Indeed, let $\mathcal{X} = \mathcal{X}^* = \{0,1,2\}$. In Figure 2a, we depict $\mathbf{x}^1$ chosen from $B^1 = \{\mathbf{x} \in \mathbb{R}_+^2 : x_1 + 2x_2 \leqslant 5\}$, and with the lattice $\mathcal{L} = \mathcal{X}^2$ inserted in Figure 2b. Clearly, $(2,1) \in \mathcal{L} \cap (B^1 \backslash \partial B^1)$; comparing $\mathbf{x}^1 = (1,2)$ with $(2,1)$, condition (9) requires $0.5\bar{u}(1) + 0.5\bar{u}(2)$ to be strictly greater than itself, which is impossible.[26] □ ,

Afriat's Theorem characterizes classical data sets that are rationalizable by locally nonsatiated utility functions. Readers who are familiar with Afriat's Theorem will notice some similarity between it and Theorem 1, in the sense that both results involve revealed prefer-

[25] If $\pi_s^t = 1/2$ for all t, s, the reader can verify that one solution to this problem is $\bar{u}(0) = 0$, $\bar{u}(1) = 1$, $\bar{u}(2) = 4$, $\bar{u}(3) = 6$, $\bar{u}(4) = 8$, and $\bar{u}(6) = 9$. So the data set depicted by Figure 1 is EU-rationalizable.

[26] If the state probabilities are not known to the observer then it is impossible to disprove expected utility with one observation. Instead the observation in Example 1 would tell us that state 2 is more probable than state 1. This means that if there is another observation where the subject buys more of state 1 consumption even if it is more expensive, an observer could conclude that the agent is *not* maximizing expected utility. This is the essential idea in Epstein (2000).



(a) Observation 1 (b) Observation 1 with lattice

Figure 2: The lattice method applied to Example 1

ence relationships (such as (6) and (7), or the EU versions (8) and (9)), between the chosen bundle $\mathbf{x}^t$ and a finite subset of the budget set B^t. In the case of Theorem 1 this subset is $\underline{B}^t \cap \mathcal{L}$ whereas in the case of Afriat's Theorem, the comparison is with $\underline{B}^t \cap \mathcal{D}$, where $\mathcal{D} = \{\mathbf{x}^t\}_{t=1}^T$. Enlarging the set of comparisons from $\mathcal{D}$ to $\mathcal{L}$ is necessitated by the special structure imposed on the utility function by Theorem 1. This is clear from Example 1, where the observation is *not* EU-rationalizable, even though the inequalities (8) and (9) are trivially satisfied in $\underline{B}^1 \cap \mathcal{D}$, since $\mathcal{D} = \{\mathbf{x}^1\}$.[27]

There is a further connection between the two results. In the case of Afriat's Theorem, the revealed preference relations can be formulated as a no-cycling condition among the elements of $\mathcal{D}$ called the generalized axiom of revealed preference (GARP). By appealing to a result of Fishburn (1975), it is possible to characterize EU-rationalizabilty in terms of a condition that is stronger than GARP but similar to it in the sense that it forbids a generalized notion of a revealed preference cycle on the set $\mathcal{L}$. These observations are discussed in greater detail in Section A1 of the Online Appendix.

2.3 Making out-of-sample predictions

Once it has been ascertained that a subject's behavior is consistent with a given model, it would be natural to exploit this compatibility by using the same model to make predictions of that subject's out-of-sample behavior. We now explain how this can also be done using Theorem 1. To simplify our discussion, we only explain this in the context of the EU model;

[27] For another result which involves comparing $\mathbf{x}^t$ with a subset of B^t that is larger than $B^t \cap \mathcal{D}$, see Quah's (2014) characterization of data sets which are rationalizable by weakly separable utility functions.

making predictions when some other model is assumed can be carried out in a similar fashion. The procedure outlined here is implemented in Section 5.2.

Suppose $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is collected from a subject who is EU-rationalizable (with objective probabilities $\pi_s^t > 0$ for all t and s). Using the information from $\mathcal{O}$ and assuming that the subject is behaving as an EU-maximizer, how can we predict the subject's preference between two lotteries: lottery A, which pays out a_i with probability $\alpha_i > 0$ (for $i = 1, 2$, with $\alpha_1 + \alpha_2 = 1$), and lottery B which pays out b_i with probability $\beta_i > 0$ (for $i = 1, 2$ with $\beta_1 + \beta_2 = 1$)? In formal terms, a strict preference for A over B is consistent with the EU model if there is a Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$ that EU-rationalizes $\mathcal{O}$ and satisfies $\alpha_1 u(a_1) + \alpha_2 u(a_2) > \beta_1 u(b_1) + \beta_2 u(b_2)$. Whether or not u exists can be easily answered using Theorem 1. In this case, it is convenient to choose $\mathcal{X}$ to be strictly larger than $\mathcal{X}^*$. Specifically, let $\mathcal{X} = \mathcal{X}^* \cup \{a_1, a_2, b_1, b_2\}$. Since $\mathcal{O}$ is EU-rationalizable, there must be a strictly increasing function $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ that solves the inequalities (8) and (9). Furthermore, Theorem 1 tells us that $\bar{u}$ has an extension u, with domain $\mathbb{R}_+$, that rationalizes $\mathcal{O}$. Therefore, to ascertain whether a strict preference for A over B is consistent with the EU model, a necessary and sufficient test is whether there is a strictly increasing function $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ that, in addition to (8) and (9), obeys

$$\alpha_1 \bar{u}(a_1) + \alpha_2 \bar{u}(a_2) > \beta_1 \bar{u}(b_1) + \beta_2 \bar{u}(b_2). \tag{10}$$

This test is easy to implement since (10) is a linear inequality. Note that because there are potentially multiple Bernoulli functions that EU-rationalize the data, it is entirely possible that *both* a preference for A over B and a preference for B over A is consistent with the EU model: in this case, there will be an increasing function $\bar{u}$ that solves (8), (9) and (10), and another one that solves (8), (9) and (10), the last with the inequality reversed.

2.4 Testing other models using the lattice method

So far, we have considered tests of EU-rationalizability in the case where the probability of each state is known to both the agent and the observer. Our test could be extended to the case where no objective probabilities can be attached to each state. A data set $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is *rationalizable by subjective expected utility* (SEU) if there exists probability distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_{\bar{s}}) \gg 0$ and a Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$ such that, at every observation t, we have $\sum_{s=1}^{\bar{s}} \pi_s u(x_s^t) \geqslant \sum_{s=1}^{\bar{s}} \pi_s u(x_s)$ for all $\mathbf{x} \in B^t$. In this case, ϕ is independent of t and is required to belong to the family Φ_{SEU} such that $\phi \in \Phi_{SEU}$ if $\phi(\mathbf{u}) = \sum_{s=1}^{\bar{s}} \pi_s u_s$ for some $\boldsymbol{\pi} \gg 0$. By Theorem 1, $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ can be rationalized by some $\phi \in \Phi_{SEU}$ if and only if *there is a strictly increasing* $\bar{u}$ *such that (8) and (9) holds* for some $\boldsymbol{\pi} \gg \mathbf{0}$. These conditions form a system of inequalities bilinear in the unknowns $\{\pi_s\}_{s=1}^{\bar{s}}$ and $\{\bar{u}(r)\}_{r \in \mathcal{X}}$.

For many of the standard models of decision making under risk, under uncertainty, or over time, the rationalizability problem has a structure similar to that of SEU in the sense that it involves finding a Bernoulli function u and a function ϕ belonging to some family Φ that together rationalize the data, and this problem can in turn be transformed via Theorem 1 into a problem of solving a system of bilinear inequalities. In Section A3 of the Online Appendix, we use Theorem 1 to devise such tests for various models of contingent choice, including choice acclimating personal equilibrium (Kőszegi and Rabin, 2007), maxmin expected utility (Gilboa and Schmeidler, 1989), variational preferences (Maccheroni, Marinacci, and Rustichini, 2006), and a model with budget dependent reference points. We also explain how we could test models of choice over time on data from budgetary allocations, such as those collected by Andreoni and Sprenger (2012). A model of discounted utility (with or without present bias) is formally very similar to the subjective expected utility model.

Even though solving a bilinear problem may be computationally intensive, the Tarski-Seidenberg Theorem tells us that this problem is decidable, in the sense that there is a known algorithm that can determine in a finite number of steps whether or not a solution exists. Nonlinear tests are not new to the revealed preference literature; for example, they appear in tests of weak separability (Varian, 1983a), in tests of maxmin expected utility and other models of ambiguity (Bayer *et al.*, 2013), and in tests of Walrasian general equilibrium (Brown and Matzkin, 1996). Solving these problems can be computationally straightforward in some cases because of certain special features of the model/environment or when the number of observations is small. The tests that we develop simplify dramatically and are easily implementable in practice when there are only two states (though they remain nonlinear).

The two-state case, while special, is very common in applied theoretical settings and laboratory experiments. For example, to implement the SEU test, we simply condition on the probability of state 1 (and hence on the probability of state 2), and then perform a linear test to check whether there is a strictly increasing function $\bar{u}$ solving (8) and (9). If not, we choose another probability for state 1, implement, and repeat (if necessary). Even a uniform grid search of up to two decimal places on the probability of state 1 will lead to no more than 99 linear tests, which can be implemented with very little difficulty.[28]

Rank dependent utility (RDU). In Section 5, we report the findings of an empirical test of the RDU model (Quiggin, 1982) when there are two states, so we explain this case of the model, and its corresponding test, in detail here. The Online Appendix contains a treatment of the multi-state case (see Sections A2 and A4).

Consider an experiment where the probability of states 1 and 2 are objectively known and given by $\pi_1 > 0$ and $\pi_2 > 0$. With no loss of generality, assume that $\pi_1 \geqslant \pi_2$. In the RDU model, the subject behaves as though these probabilities are distorted: if state s is the less favorable state, i.e., the state where the payout is smaller, then the probability given to state s is ρ_s, with this distortion respecting the rank of the objective probabilities, i.e.,

$$1 > \rho_1 > \rho_2 > 0 \text{ if } \pi_1 > \pi_2 \text{ and } 1 > \rho_1 = \rho_2 > 0 \text{ if } \pi_1 = \pi_2. \tag{11}$$

The utility of (x_1, x_2) when $x_1 \leqslant x_2$ is $V(x_1, x_2) = \rho_1 u(x_1) + (1 - \rho_1)u(x_2)$ and the utility of (x_1, x_2) when $x_1 > x_2$ is $V(x_1, x_2) = (1 - \rho_2)u(x_1) + \rho_2 u(x_2)$. Converting this into the framework of Theorem 1, we are testing rationalizability in the case where ϕ has the form:

$$\phi(u_1, u_2) = \begin{cases} \rho_1 u_1 + (1 - \rho_1)u_2 & \text{if } u_1 \leqslant u_2 \\ (1 - \rho_2)u_1 + \rho_2 u_2 & \text{if } u_1 > u_2 \end{cases} \tag{12}$$

By Theorem 1, a necessary and sufficient condition for $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ to be RDU-rationalizable is for there to be (ρ_1, ρ_2) obeying (11) and a strictly increasing $\bar{u}$ such that with ϕ defined by (12), the conditions (6) and (7) admits a solution. Given the formula for ϕ, this test involves solving a set of inequalities that are bilinear in the unknowns $\{\bar{u}(r)\}_{r \in \mathcal{X}}$ and $\{\rho_1, \rho_2\}$. When implementing this test, we let ρ_1 and ρ_2 take different values on a very fine grid in $[0, 1]^2$, subject to (11), and (for each case) perform the corresponding linear test

[28] While we have not found it necessary to use them in our implementation in this paper, there are solvers available for mixed integer nonlinear programs (for example, as surveyed in Bussieck and Vigerske (2010)) that are potentially useful for implementing bilinear tests more generally.

Monotonicity of $\bar{u}$	Optimality of $\mathbf{x}^1 = (2, 4)$
$\bar{u}(6) > \bar{u}(4) > \bar{u}(3) > \bar{u}(2)$ $\bar{u}(2) > \bar{u}(1) > \bar{u}(0)$	$\rho_1\bar{u}(2) + (1-\rho_1)\bar{u}(4) > \rho_1\bar{u}(1) + (1-\rho_1)\bar{u}(6)$ $\rho_1\bar{u}(2) + (1-\rho_1)\bar{u}(4) \geqslant (1-\rho_2)\bar{u}(3) + \rho_2\bar{u}(1)$
Optimality of $\mathbf{x}^2 = (6, 1)$	**Optimality of $\mathbf{x}^3 = (4, 3)$**
$(1-\rho_2)\bar{u}(6) + \rho_2\bar{u}(1) \geqslant \rho_1\bar{u}(0) + (1-\rho_1)\bar{u}(3)$ $(1-\rho_2)\bar{u}(6) + \rho_2\bar{u}(1) \geqslant (1-\rho_2)\bar{u}(3) + \rho_2\bar{u}(2)$	$(1-\rho_2)\bar{u}(4) + \rho_2\bar{u}(3) > \rho_1\bar{u}(2) + (1-\rho_1)\bar{u}(4)$ $(1-\rho_2)\bar{u}(4) + \rho_2\bar{u}(3) > (1-\rho_2)\bar{u}(6) + \rho_2\bar{u}(1)$

Table 2: Conditions on $\bar{u}$ for RDU-rationalizability given (ρ_1, ρ_2)

to search for a solution in $\{\bar{u}(r)\}_{r\in\mathcal{X}}$; $\mathcal{O}$ is RDU-rationalizable if such a solution exists for some value of (ρ_1, ρ_2).

As an illustration, consider again the data set displayed in Figure 1. In Table 2 we collect the relevant inequalities for rationalizability by ϕ as defined by (12); the data set can be rationalized by ϕ (for specific values of ρ_1 and ρ_2) if and only if there is $\bar{u}$ that satisfies the inequalities displayed in that table. Comparing this test with the test for EU-rationalizability (displayed in Table 1 in Section 2.2), notice that there is no change to $\mathcal{X}$ or to $\mathcal{L}$, nor is there a change to the relevant comparisons at each observation (for example at observation 1, $(2, 4)$ is compared against $(1, 6)$ and $(3, 1)$ in both tables). The only difference between them is in the functional form, with the EU-form in Table 1 and the RDU-form in Table 2.

Disappointment Aversion (DA). We also implement a lattice test of the DA model (Gul,1991). When there are two states, the DA model is a special case of RDU, with a further restriction on ρ_1 and ρ_2. Specifically, there is $\beta \in (-1, \infty)$ such that, for $s = 1, 2$,

$$\rho_s = \frac{(1+\beta)\pi_s}{1+\pi_s\beta}. \tag{13}$$

Note that this restriction has bite only if $\pi_1 \neq \pi_2$, so the RDU and DA models coincide when $\pi_1 = \pi_2$. If $\beta = 0$, the agent simply maximizes expected utility. If $\beta > 0$, we have $\rho_s > \pi_s$; the agent attaches a probability to state s that is higher than the objective probability when state s is the less favorable state and the agent is said to be *disappointment averse*. If $\beta < 0$, then $\rho_s < \pi_s$, and the agent is *elation seeking*. Similar to RDU, we test the DA model by letting β take on different values and performing the associated linear test.[29]

[29] In practice, we let ρ_1 take on different values on $(0, 1)$, which maps to different values of β and then to ρ_2.

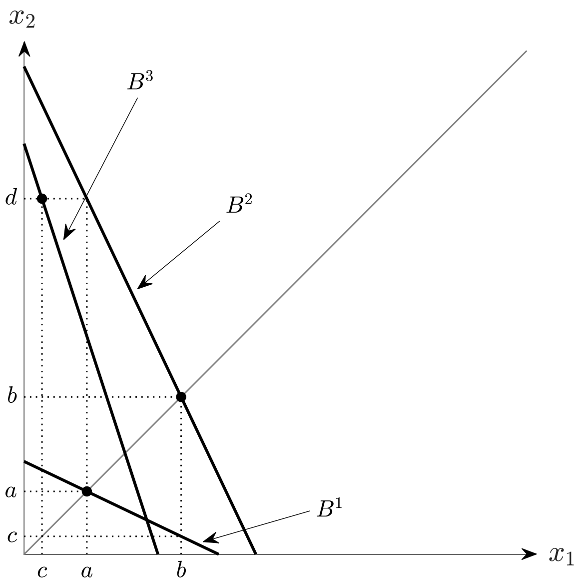


Figure 3: RDU-rationalizable but not EU-rationalizable data set

While it is well known that the RDU and EU models lead to different predictions, it not immediately clear that they are observationally distinct in the context of observations drawn from linear budgets. We end this section with an example of a data set that is RDU-rationalizable but not EU-rationalizable.

Example 2. Suppose the data set consists of three observations as depicted in Figure 3 where $\mathbf{x}^1 = (a, a)$, $\mathbf{x}^2 = (b, b)$, and $\mathbf{x}^3 = (c, d)$. Note that (b, c) is on the first observation's budget line and (a, d) is on the second observation's budget line. The price at observation t is $\mathbf{p}^t = (1, q^t)$, where $q^1 > 1 > 1/q^1 \geqslant q^2 > q^3$. Consequently the first budget line is the flattest and the third budget line is the steepest.

We claim that these observations are not EU-rationalizable if the two states are equiprobable. Suppose that they are, for some Bernoulli function u. Then the first observation tells us that $2u(a) \geqslant u(b) + u(c)$, since (b, c) is available when (a, a) is chosen. Similarly, from the second observation, we know that $2u(b) \geqslant u(a) + u(d)$. Together this gives

$$u(b) - u(d) \geqslant u(a) - u(b) \geqslant u(c) - u(a),$$

from which we obtain $u(a) + u(b) \geqslant u(c) + u(d)$. But this is contradicted by observation 3 where (c, d) is chosen even though (a, b) is in the interior of the budget set.

However, these observations are RDU-rationalizable. This should be quite intuitive because the demand pattern involves stickiness on the 45 degree line over a range of prices, with the demand deviating away from $x_1 = x_2$ only when (at the third observation) state 2 consumption is sufficiently cheap. Indeed, suppose $V(x_1, x_2) = \rho u(x_1) + (1-\rho)u(x_2)$ when $x_1 \leqslant x_2$ and $V(x_1, x_2) = (1-\rho)u(x_1) + \rho u(x_2)$ when $x_1 > x_2$, with $\rho \in (1/2, 1)$, so the agent displays disappointment aversion. It is straightforward to check that if u is strictly concave, then the agent's utility is maximized at $x_1 = x_2$ at observations 1 and 2 so long as $\rho \geqslant 1/(1+q^2)$ and $\rho \geqslant q^1/(1+q^1)$. Since we assume that $1/q^1 \geqslant q^2$, the first inequality is tighter than the second. Let us set $\rho = 1/(1+q^2)$. It remains for us to find a Bernoulli function that rationalizes the third observation. It suffices to find $u : \mathbb{R}_+ \to \mathbb{R}_+$ such that $u' > 0$ and $u'' < 0$ and $u'(c)$ and $u'(d)$ satisfy the first order condition

$$\frac{\rho u'(c)}{(1-\rho)u'(d)} = \frac{1}{q^2}\frac{u'(c)}{u'(d)} = \frac{1}{q^3}.$$

Since $d > c$ and $q^2 > q^3$, such a u must exist.

3. Measuring departures from rationalizability

The revealed preference tests presented in the previous section are 'sharp', in the sense that a data set either passes the test for a given model or it fails. This either/or feature of the tests is not particular to our results but is true of all classical revealed preference tests, including Afriat's. It would, of course, be desirable to develop a way of measuring *the extent* to which a given class of utility functions succeeds or fails in rationalizing a data set, and the most common approach adopted in the revealed preference literature to address this issue was developed by Afriat (1972, 1973) and Varian (1990) in the context of classical data sets, i.e., data sets with classical budget sets (see (1)).[30,31] The basic idea is that if a consumer's choice fails to maximize utility, then it is natural to compare what he spent with what he *could* have spent in order to achieve the same utility level. This gives us a metric to quantify the utility loss in expenditure terms. We now give an account of this approach and explain how Theorem 1 is useful for calculating this metric.

[30] For examples where Afriat-Varian type indices are used to measure a model's fit, see Mattei (2000), Harbaugh, Krause, and Berry (2001), Andreoni and Miller (2002), Choi *et al.* (2007), Beatty and Crawford (2011), Choi *et al.* (2014), and Halevy, Persitz, and Zrill (2018). See also Echenique, Lee, and Shum (2011), which develops and applies a related index called the money pump index.

[31] Varian (1990) and Halevy, Persitz, and Zrill (2018) discuss why such measures may be more suitable than other measures such as the sum of squared errors between observed and predicted demands.

Let $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$ be a classical data set. For any number $e^t \in [0,1]$, we define

$$\begin{aligned} B^t(e^t) &= \mathscr{B}(\mathbf{p}^t, e^t\mathbf{p}^t \cdot \mathbf{x}^t) \cup \{\mathbf{x}^t\} \\ &= \{\mathbf{x} \in \mathbb{R}_+^{\bar{s}} : \mathbf{p}^t \cdot \mathbf{x} \leqslant e^t\mathbf{p}^t \cdot \mathbf{x}^t\} \cup \{\mathbf{x}^t\}. \end{aligned} \tag{14}$$

Notice that when $e^t = 1$, this set coincides with the true budget set $\mathscr{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t)$ (see (1)). If $e^t < 1$, then $B^t(e^t)$ is a shrunken version of this set that retains the observed choice $\mathbf{x}^t$, but removes all bundles for which total expenditure at $\mathbf{p}^t$ is strictly higher than $e^t\mathbf{p}^t \cdot \mathbf{x}^t$, i.e., those bundles $\mathbf{x}$ where $\mathbf{p}^t \cdot \mathbf{x} > e^t\, \mathbf{p}^t \cdot \mathbf{x}^t$. Clearly $B^t(e^t)$ shrinks with the value of e^t.

Let $\mathcal{U}$ be a collection of utility functions defined on $\mathbb{R}_+^{\bar{s}}$ belonging to a given family; for example, $\mathcal{U}$ could be the family of locally nonsatiated utility functions (which was the family considered by Afriat (1972, 1973) and Varian (1990)). We say that the modified data set $\mathcal{O}(\mathbf{e}) = \{(\mathbf{x}^t, B^t(e^t))\}_{t=1}^T$ is *rationalizable by* $\mathcal{U}$ if there is $U \in \mathcal{U}$ such that $U(\mathbf{x}^t) \geqslant U(\mathbf{x})$ for all $\mathbf{x} \in B^t(e^t)$. Notice that if $\mathcal{O}(\mathbf{e})$ is rationalizable by $\mathcal{U}$, then so is $\mathcal{O}(\mathbf{e}')$ for any $\mathbf{e}' < \mathbf{e}$; in other words, shrinking budget sets will make it easier for rationalizability to hold. Notice also that so long as $U(\mathbf{x}) \geqslant U(\mathbf{0})$ for all $\mathbf{x} \geqslant \mathbf{0}$ for some $U \in \mathcal{U}$, then $\mathcal{O}(\mathbf{e})$ is rationalizable at $\mathbf{e} = \mathbf{0}$ (since $B^t(0) = \{\mathbf{0}, \mathbf{x}^t\}$ for all t). Of course, in general $\mathcal{O}(\mathbf{e})$ will be rationalizable by $\mathcal{U}$ without shrinking budget sets so drastically. This suggests that if $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$ is not itself rationalizable by $\mathcal{U}$, then one way of measuring the severity of this failure is to measure the extent to which budget sets need to shrink to obtain rationalizability. This is the key idea behind the indices put forward by Afriat and Varian.

Afriat's proposal is to shrink all budget sets by the *same* factor e (so $\mathbf{e} = (e, e, \ldots, e)$) and to find the largest number e at which $\mathcal{O}(\mathbf{e})$ is rationalizable by $\mathcal{U}$. Afriat refers to

$$\sup\{e : \mathcal{O}(e, e, \ldots, e) \text{ is rationalizable by } \mathcal{U}\}$$

as the *critical cost efficiency index*. Of course if $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$ is itself rationalizable by $\mathcal{U}$, then this index equals 1. If this index equals $e^* < 1$, then it means that there is some utility function in $\mathcal{U}$ for which the observed choice $\mathbf{x}^t$ is superior to every bundle that costs $e^*\mathbf{p}^t \cdot \mathbf{x}^t$ or less, but rationality is limited because there is some observation t' and a bundle $\mathbf{y}$ costing more than $e^*\mathbf{p}^{t'} \cdot \mathbf{x}^{t'}$ but strictly less than $\mathbf{p}^{t'} \cdot \mathbf{x}^{t'}$ that gives higher utility than $\mathbf{x}^{t'}$.

The alternative measure proposed by Varian (1990) allows different budget sets to shrink by different factors; Varian's *inconsistency index* is defined as the smallest sum of square differences between the *efficiency vector* **e** and the vector $(1,1,\ldots,1)$; formally, it is

$$\inf \sum_{t=1}^{T}(1-e^t)^2 \text{ subject to } \mathcal{O}(\mathbf{e}) \text{ being rationalizable by } \mathcal{U}.$$

Of course, if $\mathcal{O}$ is rationalizable by $\mathcal{U}$, then there is no inconsistency and Varian's index equals 0. There is a sense in which the Varian efficiency index is more discriminating than the one proposed by Afriat, but Afriat's measure is the one more commonly used because it is much easier to compute: while Afriat's measure simply requires searching for a threshold e at which the modified data set is just rationalizable by $\mathcal{U}$,[32] calculating Varian's index requires searching through all efficiency vectors **e**.[33]

In our empirical analysis in Section 5, we use Afriat's efficiency index, because it is easy to compute and because it facilitates comparison with other papers, which mostly use the same index. But we also carry calculate Varian's index for some (though not all) of the models we consider (see Seciton A8 in the Online Appendix).

3.1 Testing the EU-rationalizability of $\mathcal{O}(\mathbf{e})$

Whether one is calculating Afriat's index or Varian's, it will require checking if the modified data set $\mathcal{O}(\mathbf{e}) = \{(\mathbf{x}^t, B^t(e^t))\}_{t=1}^T$ is rationalizable by $\mathcal{U}$, at a given efficiency vector $\mathbf{e} = (e^1, e^2, \ldots, e^T)$. When $\mathcal{U}$ is the family of all locally nonsatiated utility functions, Afriat (1972, 1973) provides a necessary and sufficient condition for the rationalizability of $\mathcal{O}(\mathbf{e}) = \{(\mathbf{x}^t, B^t(e^t))\}_{t=1}^T$ (which we describe in greater detail in Section A5 of the Online Appendix). In the case where $\mathcal{U}$ is the family of expected utility or rank dependent utility functions, Theorem 1 provides us with a way of testing rationalizability within these families.

To be specific, consider a data set $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$ collected from an experiment in which state s occurs with probability $\pi_s > 0$ at every observation. Suppose that for a given $\mathbf{e} = (e^t)_{t=1}^T$, we would like to check whether $\mathcal{O}(\mathbf{e})$ is EU-rationalizable or, in the language of this section, whether it is rationalizable by $\mathcal{U}$, where $U : \mathbb{R}_+^{\bar{s}} \to \mathbb{R}_+$ is in $\mathcal{U}$ if

[32] The binary search algorithm works as follows. We first set the lower and upper bounds on e^* to $e_L = 0$ and $e_H = 1$, respectively. We then check whether the data set passes or fails the test at $e = (e_L + e_H)/2$; if it passes the test, then we update both e^* and its lower bound to $(e_L + e_H)/2$; if it fails the test, then we update e^* to e_L and the upper bound on e^* to $(e_L + e_H)/2$. We then repeat the procedure, selecting and testing a new midpoint. The algorithm terminates when the lower and upper bounds on e are sufficiently close, in our case within 10^{-6} of one another.

[33] In fact, it is known that calculating Varian's index is an NP hard problem (Smeulders *et al.*, 2014).

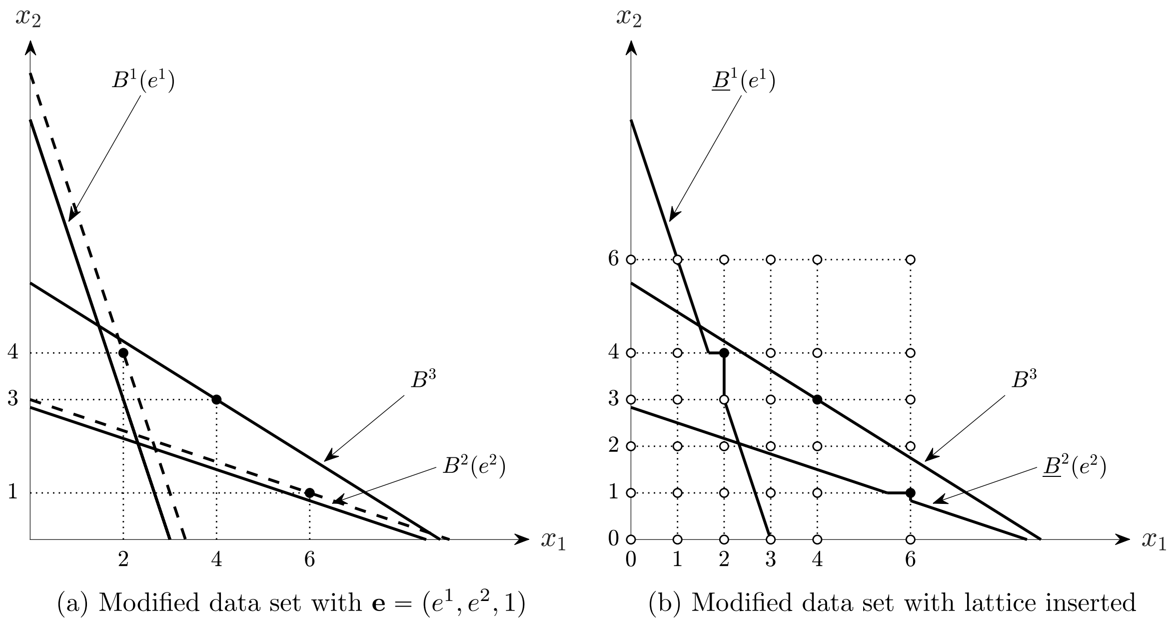


Figure 4: The lattice method applied to a modified data set

$U(\mathbf{x}) = \sum_{s=1}^{\bar{s}} \pi_s u(x_s)$ for some Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$. By Theorem 1, a necessary and sufficient condition for rationalizability is that there is a strictly increasing function $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ that satisfies (8) and (9), which form a set of linear conditions. Note that the constraint set at observation t is $B^t(e^t)$ (which is *not* a linear or even convex budget set) and its downward extension is $\{\mathbf{x} \in \mathbb{R}_+^{\bar{s}} : \mathbf{p}^t \cdot \mathbf{x} \leqslant e^t \mathbf{p}^t \cdot \mathbf{x}^t\} \cup \{\mathbf{x} \in \mathbb{R}_+^{\bar{s}} : \mathbf{x} \leqslant \mathbf{x}^t\}$.

As an illustration, we return to the example first depicted in Figure 1a and suppose we shrink B^1 and B^2 by e^1, $e^2 < 1$ respectively but leave B^3 as it is (so that $\mathbf{e} = (e^1, e^2, 1)$), as shown in Figure 4a. The downward extensions of $B^1(e^1)$, $B^2(e^2)$, are depicted in Figure 4b, along with B^3 (which is unchanged and coincides with its downward extension). In this case, $\mathcal{X}^* = \{0, 1, 2, 3, 4, 6\}$, we can choose $\mathcal{X} = \mathcal{X}^*$, and (by Theorem 1) $\mathcal{O}(\mathbf{e})$ is EU-rationalizable if and only if there is $\bar{u}$ that solves the inequalities in Table 3. We again have the strict monotonicity conditions in the top left panel, with the other panels listing the optimality conditions applicable to $\mathbf{x}^1$, $\mathbf{x}^2$, and $\mathbf{x}^3$. There are 17 bundles in $\underline{B}^1(e^1) \cap \mathcal{L}$ (besides the observed choice $\mathbf{x}^1 = (2, 4)$); two of them, $(1, 6)$ and $(3, 0)$, are undominated and both lie on the upper boundary, which leads (by condition (8)) to the weak inequalities displayed in the top right panel. There is just one undominated bundle, $(2, 2)$, in $\underline{B}^2(e^2) \cap \mathcal{L}$ (besides $\mathbf{x}^2$); this leads to the strict inequality displayed in the bottom left panel (by condition (9)). The optimality conditions on $\mathbf{x}^3$ are unchanged from those displayed in Table 1.

Monotonicity of $\bar{u}$	Optimality of $\mathbf{x}^1 = (2,4)$
$\bar{u}(6) > \bar{u}(4) > \bar{u}(3) > \bar{u}(2)$	$\pi_1^1\bar{u}(2) + \pi_2^1\bar{u}(4) \geqslant \pi_1^1\bar{u}(1) + \pi_2^1\bar{u}(6)$
$\bar{u}(2) > \bar{u}(1) > \bar{u}(0)$	$\pi_1^1\bar{u}(2) + \pi_2^1\bar{u}(4) \geqslant \pi_1^1\bar{u}(3) + \pi_2^1\bar{u}(0)$
Optimality of $\mathbf{x}^2 = (6,1)$	**Optimality of $\mathbf{x}^3 = (4,3)$**
$\pi_1\bar{u}(6) + \pi_2\bar{u}(1) > \pi_1\bar{u}(2) + \pi_2\bar{u}(2)$	$\pi_1\bar{u}(4) + \pi_2\bar{u}(3) > \pi_1\bar{u}(2) + \pi_2\bar{u}(4)$
	$\pi_1\bar{u}(4) + \pi_2\bar{u}(3) > \pi_1\bar{u}(6) + \pi_2\bar{u}(1)$

Table 3: Conditions on $\bar{u}$ for the EU-rationalizability of $\mathcal{O}(\mathbf{e})$

Checking for the RDU-rationalizability of $\mathcal{O}(\mathbf{e})$ involves a similar procedure, with the functional form modified in the way explained in Section 2.4.

3.2 Approximate smooth rationalizability

While Theorem 1 guarantees that there is a Bernoulli function u that extends $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ and rationalizes the data when the required conditions are satisfied, the Bernoulli function is not necessarily smooth. The smoothness of u is commonly assumed in applications of expected utility and related models and its implications can appear to be stark. For example, suppose that it is commonly known that states 1 and 2 occur with equal probability and that we observe the agent choosing $(1,1)$ at a price vector (p_1, p_2), with $p_1 \neq p_2$. This observation is incompatible with a smooth EU model; indeed, given that the two states are equiprobable, the slope of the indifference curve at $(1,1)$ must equal -1 and thus it will not be tangential to the budget line and will not be a local optimum. On the other hand, it is trivial to check that this observation is EU-rationalizable in our sense. In fact, one could even find a *concave* Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$ for which $(1,1)$ maximizes expected utility. (Such a u will be continuous and strictly increasing, but have a kink at 1.)

These two facts are reconcilable. Given any strictly increasing and continuous function u defined on a compact interval of $\mathbb{R}_+$, there is a strictly increasing and smooth function $\tilde{u}$ that is uniformly and arbitrarily close to u on that interval. Thus if a Bernoulli function u rationalizes $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ by $\{\phi(\cdot,t)\}_{t=1}^T$, then for any efficiency threshold $e \in (0,1)$, then there is a *smooth* Bernoulli function $\tilde{u}$ that rationalizes $\mathcal{O}' = \{(\mathbf{x}^t, B^t(e))\}_{t=1}^T$ by $\{\phi(\cdot,t)\}_{t=1}^T$. In other words, if a data set is rationalizable by a Bernoulli function, then it can also be rationalized by a smooth Bernoulli function, for any efficiency threshold arbitrarily close to 1. In this sense, imposing a smoothness requirement on the Bernoulli function does not radically alter a model's ability to explain a data set.

4. Concave Bernoulli functions

A common assumption in applications of expected utility (EU) theory is that agents are risk averse, which is equivalent to the concavity of the Bernoulli function. The necessary and sufficient conditions that we have developed for EU-rationalizability (with Theorem 1) neither require nor guarantee that the Bernoulli function is concave. This distinction is significant because there are data sets which can be rationalized by the EU model, but only with nonconcave Bernoulli functions. This will be made readily apparent in the empirical implementation in Section 5, but we also provide an intuitive example of such a phenomenon, in a classical data set with two observations, in Section A4 of the Online Appendix.

In this section we provide a test for *concave EU-rationalizability*, i.e., EU-rationalizability with a concave Bernoulli function. (Recall that, by definition, Bernoulli functions are continuous and strictly increasing.) Unfortunately, we do not, in this case, have a result like Theorem 1 which is applicable to observations drawn from general compact constraint sets. Our procedure works in a narrower set of environments: it allows us to test for the concave EU-rationalizability of a classical data set $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$, and also when it is modified by $\mathbf{e} = (e^t)_{t=1}^T$, i.e., the data set $\mathcal{O}(\mathbf{e}) = \{(\mathbf{x}^t, B^t(e^t)\}_{t=1}^T$ (with $B^t(e^t)$ defined by (14)). Note that there is already a test of concave EU-rationalizability for classical data sets (see Varian (1983a) and Green and Srivastava (1986)), but that test makes use of the *sufficiency* of the first order conditions, which in turn relies crucially on the linearity of the classical budget sets; since $B^t(e^t)$ is *not* a convex set, that method does not obviously extend to testing for the concave EU-rationalizability of $\mathcal{O}(\mathbf{e})$. The added value of our approach lies in its applicability to modified data sets, and this extension is important because (as we explain in Section 3) it enables us to calculate the critical cost efficiency index in the event that $\mathcal{O}$ itself is not concave EU-rationalizable.

Throughout this section we confine our discussion to the case where there are just two states of the world with known probabilities π_1 and π_2. The results carry over to the multi-state case, which is covered in Section A4 of the Online Appendix. That section also contains a test for the rank dependent utility model with a concave Bernoulli function.

Let $\bar{r} > 0$ be a number such that $\bar{r} p_s^t \geqslant p^t \cdot x^t$ for all t and s. In other words, $\bar{r}$ is a consumption level that beyond which nothing is affordable to the subject in any state and at any observation. Clearly, for the purposes of rationalizing a data set, the behavior of the Bernoulli function beyond $\bar{r}$ is of no relevance. We define

$$\mathcal{X}^{**} = \{0\} \cup \{r \in \mathbb{R}_+ : \quad r = x_s^t \text{ for some } t \text{ and } s\} \cup \{\bar{r}\} \tag{15}$$

and let $\mathcal{X} \subset \mathbb{R}_+$ be a *finite* set containing $\mathcal{X}^{**}$.[34]

For any function $\bar{h} : \mathcal{X} \to \mathbb{R}_+$, we define the *piecewise linear extension* of $\bar{h}$ (or simply *linear extension* of $\bar{h}$ for short) as the function $\bar{h}_\ell : \mathbb{R}_+ \to \mathbb{R}_+$ that is linear between adjacent points in $\mathcal{X}$, with $\bar{h}_\ell(r) = \bar{h}(r)$ for all $r \in \mathcal{X}$.[35,36] Thus if $r < r'$ are two adjacent points in $\mathcal{X}$ and a lies between them with $a = \lambda r + (1-\lambda) r'$ for $\lambda \in (0,1)$, then $\bar{h}_\ell(a) = \lambda \bar{h}(r) + (1-\lambda)\bar{h}(r')$.

The essence of our test is as follows. Suppose $\mathcal{O}(\mathbf{e})$ is EU-rationalizable by the concave Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$. Let $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ be the restriction of u to $\mathcal{X}$ and let $\bar{u}_\ell : \mathbb{R}_+ \to \mathbb{R}_+$ be the linear extension of $\bar{u}$. We claim that $\bar{u}_\ell$ also rationalizes $\mathcal{O}(\mathbf{e})$, which implies that in searching for concave Bernoulli functions that EU-rationalize $\mathcal{O}(\mathbf{e})$, we can confine our search to linear extensions of functions defined on $\mathcal{X}$. We then show that with Bernoulli functions of this type, checking that the expected utility of $\mathbf{x}^t$ is greater than a certain finite set in $B^t(e^t)$ is sufficient to guarantee that the same holds for *all* bundles in $B^t(e^t)$.

We first show that if u rationalizes $\mathcal{O}(\mathbf{e})$, then so does $\bar{u}_\ell$. Indeed, by definition, $\bar{u}_\ell(r) = u(r)$ for all $r \in \mathcal{X}$ and the concavity of u guarantees that $u(r) \geqslant \bar{u}_\ell(r)$ for all $r \in [0, \bar{r}]$. Thus $\pi_1 u(x_1^t) + \pi_2 u(x_2^t) = \pi_1 \bar{u}_\ell(x_1^t) + \pi_2 \bar{u}_\ell(x_2^t)$ for all $\mathbf{x}^t = (x_1^t, x_2^t)$ and, for any other bundle $\mathbf{x} = (x_1, x_2)$, we have $\pi_1 u(x_1) + \pi_2 u(x_2) \geqslant \pi_1 \bar{u}_\ell(x_1) + \pi_2 \bar{u}_\ell(x_2)$. Since

$$\pi_1 u(x_1^t) + \pi_2 u(x_2^t) \geqslant \pi_1 u(x_1) + \pi_2 u(x_2) \text{ for all } \mathbf{x} = (x_1, x_2) \in B^t(e^t),$$

we also have

$$\pi_1 \bar{u}_\ell(x_1^t) + \pi_2 \bar{u}_\ell(x_2^t) \geqslant \pi_1 \bar{u}_\ell(x_1) + \pi_2 \bar{u}_\ell(x_2) \text{ for all } \mathbf{x} = (x_1, x_2) \in B^t(e^t).$$

[34] The definition of $\mathcal{X}^{**}$ is similar to $\mathcal{X}^*$ (see (5)), but the latter does not include $\bar{r}$.
[35] Two points r and r' are adjacent in $\mathcal{X}$ if there is no point in $\mathcal{X}$ between r and r'.
[36] Strictly speaking, $\bar{h}_\ell$ is not uniquely defined for $r > \bar{r}$, but the value of $\bar{h}_\ell$ beyond $\bar{r}$ is irrelevant.

We define $\mathcal{N} = \{(a,b) \in \mathbb{R}^2_+ : \ a,\, b \in [0, \bar{r}]$ and either a or b is in $\mathcal{X}\}$. The set $\mathcal{N}$ looks a like a net containing $\mathcal{L} = \mathcal{X}^2$. If the concave Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$ rationalizes $\mathcal{O}(\mathbf{e})$, then, for all t,

$$\pi_1 \bar{u}_\ell(x_1^t) + \pi_2 \bar{u}_\ell(x_2^t) \geqslant \pi_1 \bar{u}_\ell(x_1) + \pi_2 \bar{u}_\ell(x_2) \text{ for all } \mathbf{x} = (x_1, x_2) \in \mathcal{N} \cap \partial\mathscr{B}(\mathbf{p}^t, e^t\mathbf{p}^t \cdot \mathbf{x}^t). \quad (16)$$

(Recall that $\partial\mathscr{B}(\mathbf{p}^t, e^t\mathbf{p}^t \cdot \mathbf{x}^t)$ is the upper boundary of the budget set $\mathscr{B}(\mathbf{p}^t, e^t\mathbf{p}^t \cdot \mathbf{x}^t)$ and is equal to the budget line, i.e., $\{\mathbf{x} \in \mathbb{R}^2_+ : \mathbf{p}^t \cdot \mathbf{x} = e^t\mathbf{p}^t \cdot \mathbf{x}^t\}$.) Inequality (16) is obviously true since $\bar{u}_\ell$ also rationalizes $\mathcal{O}(\mathbf{e})$, $\mathcal{N} \cap \partial\mathscr{B}(\mathbf{p}^t, e^t\mathbf{p}^t \cdot \mathbf{x}^t)$ is a subset of $B^t(e^t)$ (see (14)) and $\mathbf{x}^t$ is optimal in $B^t(e^t)$.

Proposition 1. *Suppose the data set $\mathcal{O}(\mathbf{e})$ is EU-rationalizable with probability $(\pi_1, \pi_2) \gg (0,0)$ by a concave Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$. Let $\mathcal{X}$ be a finite set in $\mathbb{R}_+$ containing $\mathcal{X}^{**}$ (as defined by (15)). Then the restriction of u to $\mathcal{X}$, $\bar{u} : \mathcal{X} \to \mathbb{R}_+$, has the following properties:* (i) $\bar{u}(r) < \bar{u}(r')$ *if $r < r'$;* (ii) *for any three adjacent points $r < r' < r''$ in $\mathcal{X}$,*

$$\frac{\bar{u}(r') - \bar{u}(r)}{r' - r} \geqslant \frac{\bar{u}(r'') - \bar{u}(r')}{r'' - r'};$$

and (iii) *$\bar{u}_\ell : \mathbb{R}_+ \to \mathbb{R}_+$, the linear extension of $\bar{u}$, satisfies (16) at all t.*

Conversely, if $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ satisfies (i), (ii), *and* (iii), *then its linear extension $\bar{u}_\ell$ is a strictly increasing and concave Bernoulli function that EU-rationalizes $\mathcal{O}(\mathbf{e})$.*

If $\mathcal{O}(\mathbf{e})$ is EU-rationalizable by a concave Bernoulli function u, then (i) holds because u is increasing, (ii) holds because u is concave, and we have already shown that (iii) is necessary. The proof of the converse is in the Appendix. The import of this proposition is that it provides us with an easy-to-implement test, since the conditions (i) to (iii) translate into a finite set of linear inequalities on a finite set of unknowns $\{\bar{u}(r)\}_{r\in\mathcal{X}}$, and checking whether or not a solution exists is a straightforward matter.

As an illustration of how this test works, we consider the data set $\mathcal{O}(\mathbf{e})$ previously depicted in Figure 4a. Given that the three observed choices are $(2,4)$, $(6,1)$, and $(4,3)$, and choosing $\bar{r} = 10$, we obtain $\mathcal{X}^{**} = \{0,1,2,3,4,6,10\}$. Letting $\mathcal{X} = \mathcal{X}^{**}$, the test involves setting up a collection of linear inequalities in the unknowns $\{\bar{u}(r)\}_{r\in\mathcal{X}}$ (corresponding to conditions (i) to (iii)) and checking if it has a solution. Conditions (i) and (ii) are clear enough, so let us explain condition (iii), which guarantees the optimality of the observed choice $\mathbf{x}^t$ over

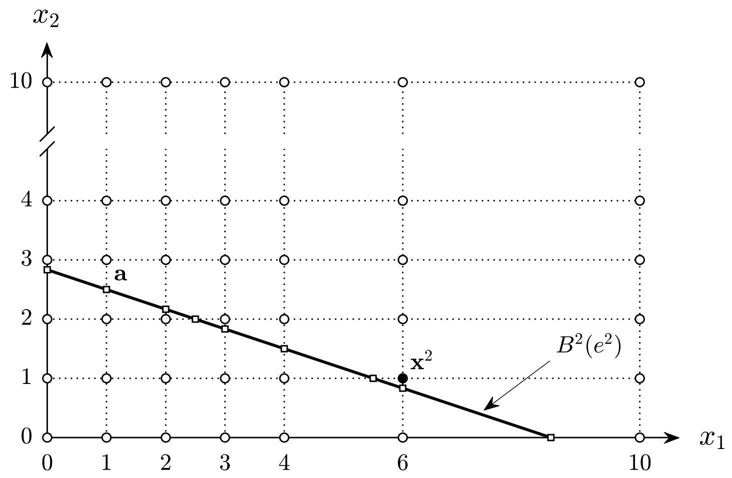


Figure 5: Test for concave EU-rationalizability

a finite set of alternatives in $B^t(e^t)$. To be specific, consider its restrictions on the second observation. In Figure 5 we zoom in on $B^2(e^2)$, where $\mathcal{N}$ is indicated by the dashed lines dividing $\mathbb{R}^2_+$. There are nine bundles in $\mathcal{N} \cap \partial\mathscr{B}^2(\mathbf{p}^2, e^2\mathbf{p}^2 \cdot \mathbf{x}^2)$, indicated by small squares on the budget line.[37] Condition (iii) requires that the expected utility of $\mathbf{x}^2$, computed with $\bar{u}_\ell$, be higher than the expected utility of those nine bundles. This translates into nine linear inequalities in the unknowns $\{\bar{u}(r)\}_{r\in\mathcal{X}}$. For example, the bundle $\mathbf{a} = (1, 2\lambda + 3(1-\lambda))$ for some $\lambda \in (0,1)$ (which can easily be computed). The expected utility of $\mathbf{a}$ is

$$\pi_1\bar{u}_\ell(1) + \pi_2\bar{u}_\ell(2\lambda + (1-\lambda)3) = \pi_1\bar{u}(1) + \pi_2[\lambda\bar{u}(2) + (1-\lambda)\bar{u}(3)],$$

since $\bar{u}_\ell$ is piecewise linear. Condition (iii) requires

$$\pi_1\bar{u}(6) + \pi_2\bar{u}(1) \geqslant \pi_1\bar{u}(1) + \pi_2\lambda\bar{u}(2) + \pi_2(1-\lambda)\bar{u}(3),$$

One could work out the other eight inequalities in a similar fashion.

[37] Their coordinates can be easily computed from $\mathcal{N}$ and $\partial\mathscr{B}^2(\mathbf{p}^2, e^2\mathbf{p}^2 \cdot \mathbf{x}^2)$.

5. Implementation

We study the data collected from the well known portfolio choice experiment in Choi *et al.* (2007), and from two other similar, more recent, experiments in Choi *et al.* (2014) and Halevy, Persitz, and Zrill (2018). The new tests developed in this paper allow us to evaluate, with a completely nonparametric approach, the empirical performance of different models of decision making under risk.

5.1 Model performance

The experiment in Choi *et al.* (2007) was performed on 93 undergraduate subjects at the University of California, Berkeley. Every subject was asked to make consumption choices on 50 decision problems under risk. Each subject divided her budget between two Arrow-Debreu securities, with each security paying one token if the corresponding state was realized, and zero otherwise. In a **symmetric treatment** applied to 47 subjects, each state of the world occurred with probability 1/2, and in a (balanced) **asymmetric treatment** applied to 46 subjects, the probabilities of the states were 1/3 and 2/3. These probabilities were objectively known. Lastly, income was normalized to one, and the state prices were chosen at random and varied across rounds and subjects. In their analysis, Choi *et al.* (2007) first tested whether each subject's behavior is consistent with maximizing a locally nonsatiated utility function by performing a GARP test (or, strictly speaking, a modified version of the GARP test which characterizes rationalizability at a given (Afriat) cost efficiency threshold). Those subjects who passed GARP at a sufficiently high efficiency threshold were then fitted individually to a two-parameter version of the disappointment aversion model of Gul (1991).[38]

The lattice method developed in this paper makes it possible to analyze the same data using purely revealed preference techniques. By applying the tests developed in Sections 2 and 3, we can calculate the efficiency index at which a subject's choice behavior is rationalizable by the expected utility (EU), disappointment aversion (DA), and rank dependent utility (RDU) models. We can also do the same with the additional requirement that the Bernoulli function is concave, using the results in Section 4 and in Section A4 of the Online Appendix; we shall refer to these models as CEU, CDA, and CRDU.

[38] One parameter governed the distortion of state probabilities, and the other the degree of absolute/relative risk aversion.

It is well known that all of these models are contained within the larger class of *stochastically monotone* utility functions; these utility functions give strictly higher utility to the bundle $\mathbf{x}$ compared to $\mathbf{y}$ whenever $\mathbf{x}$ first order stochastically dominates $\mathbf{y}$ (with respect to the objective state probabilities) and gives them the same utility whenever they are stochastically equivalent. In the Choi *et al.* (2007) experiment, there are just two states. In this case it is straightforward to check that when $\pi_1 = \pi_2 = 1/2$, a utility function is stochastically monotone if and only if it is strictly increasing and symmetric, and when $\pi_2 > \pi_1$, a utility function U is stochastically monotone if and only if it is strictly increasing and $U(a, b) > U(b, a)$ whenever $b > a$. Lastly, stochastically monotone utility functions are contained within the still larger class of locally nonsatiated utility functions.

Afriat's Theorem tells us that the rationalizability by locally nonsatiated utility maximization is observationally characterized by GARP. A test of stochastically monotone utility maximization was recently developed by Nishimura, Ok, and Quah (2017); this test has features similar to GARP and we shall refer to it as F-GARP (where 'F' stands for first order stochastic dominance). In both cases, it is also known that the axioms can be extended to test for rationalizability on modified data sets and can therefore be used to calculate the critical cost efficiency index (see Section 3) at which the data set is rationalizable. An explanation of these axioms can be found in Section A5 of the Online Appendix.

To recap, for each subject in Choi *et al.* (2007), we calculate the critical cost efficiency index at which that subject is consistent with a given model. There are a total of 8 models under consideration (locally nonsatiated utility maximization (GARP), stochastically monotone utility maximization (F-GARP), and RDU, DA, EU, CRDU, CDA, and CEU maximization). Therefore, to each subject under the asymmetric treatment, we assign 8 efficiency indices (one for each model), while to each subject under the symmetric treatment, we assign 6 indices (since in the symmetric case, the RDU and DA models are identical, and the CRDU and CDA models are identical). When one model is, by definition, more stringent than another, its efficiency index must be weakly lower. So for a given subject, the efficiency index corresponding to GARP will be the highest, and the index corresponding to CEU will be the lowest. More generally, for each subject, the efficiency indices must be ordered the

<table>
<tr><th colspan="4">$\pi_1 = 1/2$</th><th colspan="4">$\pi_1 \neq 1/2$</th></tr>
<tr><td></td><td>$e = 0.90$</td><td>$e = 0.95$</td><td>$e = 1.00$</td><td></td><td>$e = 0.90$</td><td>$e = 0.95$</td><td>$e = 1.00$</td></tr>
<tr><td>GARP</td><td>38/47 (81%)</td><td>32/47 (68%)</td><td>12/47 (26%)</td><td>GARP</td><td>37/46 (80%)</td><td>29/46 (63%)</td><td>4/46 (9%)</td></tr>
<tr><td>F-GARP</td><td>30/47 (64%)</td><td>23/47 (49%)</td><td>1/47 (2%)</td><td>F-GARP</td><td>33/46 (72%)</td><td>26/46 (57%)</td><td>3/46 (7%)</td></tr>
<tr><td rowspan="2">RDU</td><td rowspan="2">30/47 (64%)</td><td rowspan="2">23/47 (49%)</td><td rowspan="2">1/47 (2%)</td><td>RDU</td><td>33/46 (72%)</td><td>24/46 (52%)</td><td>2/46 (4%)</td></tr>
<tr><td>DA</td><td>20/46 (43%)</td><td>12/46 (26%)</td><td>1/46 (2%)</td></tr>
<tr><td>EU</td><td>30/47 (64%)</td><td>18/47 (38%)</td><td>1/47 (2%)</td><td>EU</td><td>18/46 (39%)</td><td>12/46 (26%)</td><td>1/46 (2%)</td></tr>
<tr><td rowspan="2">CRDU</td><td rowspan="2">24/47 (51%)</td><td rowspan="2">12/47 (26%)</td><td rowspan="2">0/47 (0%)</td><td>CRDU</td><td>24/46 (52%)</td><td>14/46 (30%)</td><td>1/46 (2%)</td></tr>
<tr><td>CDA</td><td>13/46 (28%)</td><td>6/46 (13%)</td><td>1/46 (2%)</td></tr>
<tr><td>CEU</td><td>23/47 (49%)</td><td>10/47 (21%)</td><td>0/47 (0%)</td><td>CEU</td><td>11/46 (24%)</td><td>5/46 (11%)</td><td>0/46 (0%)</td></tr>
</table>

Table 4: Pass rates by efficiency threshold (Choi *et al.*, 2007)

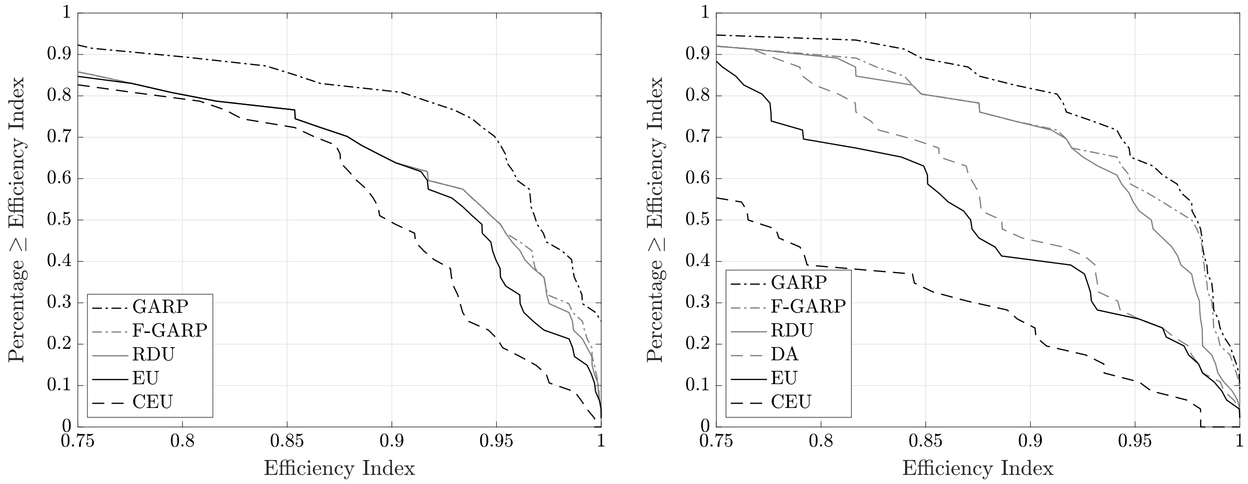
same way the models are nested, i.e.,

$$e_{CEU} \leqslant e_{EU}; \quad e_{CDA} \leqslant e_{DA}; \quad e_{CEU} \leqslant e_{CDA} \leqslant e_{CRDU} \leqslant e_{RDU};$$
$$e_{EU} \leqslant e_{DA} \leqslant e_{RDU} \leqslant e_{F-GARP} \leqslant e_{GARP}$$

Basic rationalizability. Table 4 gives pass rates for the different models at three different thresholds of the critical cost efficiency index: 0.9, 0.95, and 1, with the last corresponding to exact rationalizability.[39] Across both treatments, 16 out of 93 subjects obey GARP exactly and are therefore consistent with locally nonsatiated utility maximization, with subjects in the symmetric treatment performing distinctly better than those in the asymmetric treatment. Of the 16 subjects who pass GARP, only 4 pass F-GARP, and still fewer subjects are rationalizable by the more stringent models. Given that we observe 50 decisions for every subject, it is not altogether surprising that so many subjects should have violated GARP (let alone the more stringent conditions). The picture changes significantly once we allow for some error in the form of cost inefficiencies: about 81% of the subjects pass GARP at efficiency thresholds exceeding 0.9, and 66% at thresholds exceeding 0.95, suggesting that a large fraction of the sample *does* indeed behave in a way that is broadly compatible with utility maximization.[40]

[39] The efficiency indices corresponding to GARP were also calculated by Choi *et al.* (2007). The indices corresponding to all other models are new.

[40] Furthermore, we know that the experiment provides a high-powered test of utility maximization, in the sense that we can safely dismiss the possibility that this outcome would have occurred randomly. Indeed, as Choi *et al.* (2007) have already pointed out, these pass rates are very different from what arises if one instead calculates efficiency indices for (uniformly) randomly generated budgetary data, following Bronars



(a) $\pi_1 = 1/2$ (b) $\pi_1 \neq 1/2$

Figure 6: Distributions of efficiency indices (Choi *et al.*, 2007)

In Figure 6, we depict (separately) the distributions of efficiency indices across subjects for 5 models under the symmetric treatment, and 6 models under the asymmetric treatment.[41] The models are nested by definition, so one would expect the efficiency distributions to be stacked, as indeed they are. In both panels, the topmost curve represents the distribution of efficiency indices corresponding to GARP (in other words, rationalizability by locally nonsatiated utility maximization), and the bottommost curve represents the distribution of indices corresponding to CEU maximization, which is the most stringent model.

The EU model. We can see from Table 4 that around half of all subjects passing GARP are then consistent with the EU model (at the 0.9 or 0.95 efficiency thresholds).[42] One might worry that the experimental design is insufficiently discriminating or powerful, so that, at a given efficiency threshold, random GARP-consistent data sets would have passed the EU test at the same rate, but this is far from the case. We can confirm that the pass rate for the EU model on a large collection of randomly generated GARP-consistent data sets is effectively zero. In fact, we can say even more. In Table 5, we report the results from a large collection of randomly generated data sets, all of which pass F-GARP at the given efficency

(1987); in that case, the proportion of synthetic (random) subjects passing GARP at efficiency thresholds exceeding 0.9 is very close to zero. (See Figure 4 in Choi *et al.* (2007).)

[41] We exclude the CRDU and CDA distributions in order to avoid congestion; the interested reader can find those distributions in Section A6.3 of the Online Appendix.

[42] The precise ratios are $(30 + 18)/(38 + 37) = 64\%$ at the 0.9 threshold, and $(18 + 12)/(32 + 29) = 49\%$ at the 0.95 threshold.

$\pi_1 = 1/2$			$\pi_1 \neq 1/2$		
	$e = 0.90$	$e = 0.95$		$e = 0.90$	$e = 0.95$
RDU	25%	1%	RDU	1%	0%
			DA	0%	0%
EU	13%	1%	EU	0%	0%

Table 5: Pass rates for random F-GARP-consistent data (Choi *et al.*, 2007)

threshold (either 0.9 or 0.95).[43] At the 0.9 threshold, the pass rate for the EU model in these randomly generated data is 13% under the symmetric treatment and 0% under the asymmetric treatment; at the 0.95 threshold, the pass rates are effectively zero under both treatments. In other words, the observed EU pass rates are substantially higher than what would have arisen had the subjects been merely maximizing some stochastically monotone utility function. The CEU pass rates (where concavity is imposed on the Bernoulli function) are lower than the EU pass rates, but these too are substantially higher than the EU pass rates on randomly generated F-GARP-consistent data.[44]

That said, it is worth emphasizing that there *is* a significant difference in the performance of the GARP and EU tests. Had the distributions for GARP and EU in Figure 6 been very close, we could have concluded that while subjects make mistakes when choosing from budget sets (since they fall short of complete consistency with basic rationality), they are nonetheless consistent with the EU model once that has been accounted for. However, since the distributions are distinct, that is not the case for a significant number of subjects.

We can have a sense of the *preference misspecification*, i.e., the extent to which the EU model misspecifies each subject's preference, by looking at the difference in the efficiency index between GARP and the EU model.[45] The median difference is 0.027 for subjects in the symmetric treatment and 0.075 for those in the asymmetric treatment. It exceeds 0.05 for 17 out of 47 subjects in the symmetric treatment and 30 out of 46 subjects in the

[43] Section A6.1 of the Online Appendix describes the procedure we use to randomly generate GARP (or F-GARP) consistent data at a given efficiency threshold.

[44] Naturally, the CEU pass rates on random F-GARP-consistent data sets must be even lower than the EU pass rates reported in Table 5, since CEU is a more stringent model.

[45] We are broadly following Halevy, Persitz, and Zrill (2018) in using this measure of preference misspecification. Halevy, Persitz, and Zrill (2018) compare the change in the money metric index (which is essentially Varian's inconsistency index) between GARP and a parametric version of the RDU/DA model and interpret that difference as a measure of the misspecification.

asymmetric treatment.

The DA and RDU models. Can these models play a useful role in explaining behavior which is not captured by the EU model? Under the symmetric treatment, there is little scope for these models to capture subjects not already compatible with the EU model, since the latter is already accommodating most of the subjects who pass F-GARP. (Note the closeness of the F-GARP and EU distributions in Figure 6a.)

However, the RDU model appears to capture subject behavior more successfully than the EU model under the asymmetric treatment, covering almost all F-GARP-consistent behavior (see Table 4 or Figure 6b).[46] When we require the Bernoulli function to be concave, the rank dependent utility model can no longer account for nearly all F-GARP-consistent behavior, but even then it captures many more subjects than expected utility (compare CRDU with CEU). On the other hand, that is not true of the DA model. Under both the symmetric and asymmetric treatments, the DA pass rates are only slightly higher than the EU pass rates, and the same is true when comparing CDA with CEU.

Table 6 summarizes these observations. We record (as a fraction of all subjects within each treatment) the pass rates for the CEU and EU models. We also report the marginal contributions of the RDU and DA models (relative to EU or CEU) in explaining the data. For example, under the asymmetric treatment, at the 0.9 threshold, 15 subjects out of 46 pass RDU but fail EU (see the row beginning RDU $\setminus$ EU); using this information, we can form a 95% binomial proportion confidence interval on the probability that a subject is rationalizable by RDU but not by EU, which turns out to be [0.195, 0.480].[47]

We conduct similar tests on the data collected by Choi *et al.* (2014) and Halevy, Persitz, and Zrill (2018). In both experiments, subjects allocated investment between two Arrow-Debreu securities, with commonly known equiprobable states; thus the designs closely resembled the symmetric treatment in Choi *et al.* (2007).[48]

[46] We discuss the probability distortions needed to rationalize the data in Section A6.2 of the Online Appendix.

[47] All confidence intervals in this table are exact, and calculated using the Clopper-Pearson method.

[48] Since the DA and RDU models coincide when states are equiprobable, the interesting distinction we find between them in the asymmetric treatment in Choi *et al.* (2007) could not be further investigated. We think there is a case for including the asymmetric treatment in future experiments, or even to have the same subject choosing under different state probabilities.

$\pi_1 = 1/2$				
	$e = 0.9$		$e = 0.95$	
	Sample Prop.	Conf. Interval	Sample Prop.	Conf. Interval
CEU	23/47 (0.489)	[0.341, 0.639]	10/47 (0.213)	[0.107, 0.357]
EU	30/47 (0.638)	[0.485, 0.773]	18/47 (0.383)	[0.245, 0.536]
EU \ CEU	7/47 (0.149)	[0.062, 0.283]	8/47 (0.170)	[0.076, 0.308]
RDU \ EU	0/47 (0.000)	[0.000, 0.075]	5/47 (0.106)	[0.035, 0.231]
CRDU \ CEU	1/47 (0.021)	[0.001, 0.113]	2/47 (0.043)	[0.005, 0.145]
$\pi_1 \neq 1/2$				
	$e = 0.9$		$e = 0.95$	
	Sample Prop.	Conf. Interval	Sample Prop.	Conf. Interval
CEU	11/46 (0.239)	[0.126, 0.388]	5/46 (0.109)	[0.036, 0.236]
EU	18/46 (0.391)	[0.251, 0.546]	12/46 (0.261)	[0.143, 0.411]
EU \ CEU	7/46 (0.152)	[0.063, 0.289]	7/46 (0.152)	[0.063, 0.289]
DA \ EU	2/46 (0.043)	[0.005, 0.148]	0/46 (0.000)	[0.000, 0.077]
RDU \ EU	15/46 (0.326)	[0.195, 0.480]	12/46 (0.261)	[0.143, 0.411]
CDA \ CEU	2/46 (0.043)	[0.005, 0.148]	1/46 (0.022)	[0.001, 0.115]
CRDU \ CEU	13/46 (0.283)	[0.160, 0.435]	9/46 (0.196)	[0.094, 0.339]

Table 6: Confidence intervals on preference types (Choi *et al.*, 2007)

Analysis of the Choi *et al.* (2014) data set This experiment was conducted on 1,182 CentERpanel adult members, where the latter is meant to be representative of the Dutch-speaking population of the Netherlands. Each subject made allocation decisions on 25 linear budget sets; since this is just half the number of decisions in Choi *et al.* (2007), the pass rates should be higher if the subject population is the same, but it is not. As has already been noted by Choi *et al.* (2014), the pass rates for GARP at any efficiency threshold are instead *lower* those in Choi *et al.* (2007). This observation can now be extended further: the pass rates are across the board lower for all tests, and not just GARP. This is clear if we compare Figure 7 with Figure 6a. That said, it is also clear from these two figures that certain qualitative features of the data are the same in both experiments. In particular, around half the subjects who pass GARP at a given efficiency threshold are also consistent with the EU model. There is a significant difference in the pass rates between GARP and F-GARP, and the EU model manages to explain a very large share of subjects who pass F-GARP. In this experiment (but unlike in Choi *et al.* (2007)), the CEU model also manages to account for many subjects who pass F-GARP. Since the rank dependent utility (equivalently, disappointment aversion) model is more stringent than F-GARP, the model's contribution,

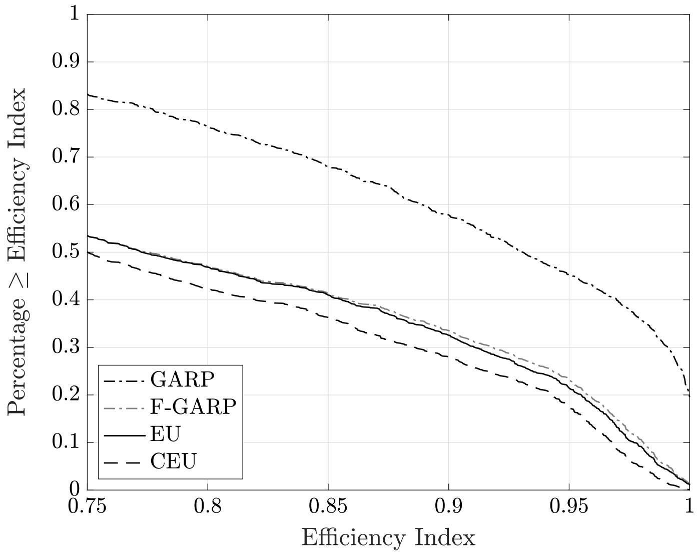


Figure 7: Distributions of efficiency indices (Choi *et al.*, 2014)

as measured by the proportion of subjects who obey RDU but not EU, or CRDU but not CEU (at some reasonable threshold), is modest. This echoes our finding for the symmetric treatment in Choi *et al.* (2007).

More details of our analysis of the Choi *et al.* (2014) data can be found in the Section A7 of the Online Appendix. We also explore in that section the relationship between a subject's efficiency indices (for different models) and various socio-economic variables and outcomes (such as age, education, and wealth), extending the analysis in Choi *et al.* (2014).

Analysis of the Halevy, Persitz, and Zrill (2018) data set. This experiment was conducted on 207 primarily undergraduate subjects at the University of British Columbia, with a set of portfolio choice problems forming the first part of a two-part experiment. (We discuss the second part of the experiment in Section 5.2.) Each subject made allocation decisions on 22 linear budget sets; since this number is lower than in Choi *et al.* (2007) (where each subject made decisions on 50 budget sets) and the sample population is similar, one would expect the pass rates in Halevy, Persitz, and Zrill (2018) to be generally higher than those in Choi *et al.* (2007), and this is what we find. However, the relative performance of the different models (relative to one another) is broadly similar across the two experiments. The EU model performs well: significantly more than half the subjects who pass GARP (at some reasonable efficiency threshold) also pass the test for the EU model. That said, there

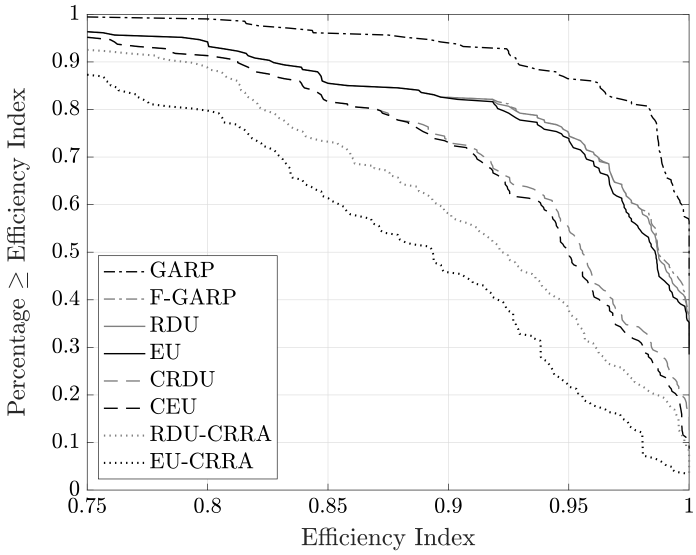


Figure 8: Distributions of efficiency indices (Halevy, Persitz, and Zrill, 2018)

is still a distinct difference in performance between the models; indeed a significant number of subjects who pass GARP fail F-GARP (and thus EU). The RDU model makes only a modest contribution relative to EU, and similarly, consistency with the CRDU model is only slightly higher than with the CEU model. These observations are clear in Figure 8, which depicts the distributions of efficiency indices for different models in this data set.

While parametric models are easy to use and have other advantages, they will by definition, fit a data set less well than their nonparametric counterparts, so there is some advantage in developing ways to assess the size of that loss of fit. Halevy, Persitz, and Zrill (2018) evaluated the basic rationality of their subjects (through GARP) and also the goodness-of-fit of parametric versions of the rank dependent and expected utility models. Following their example, we calculate the efficiency indices for the rank-dependent and expected utility models, with the Bernoulli functions confined to the CRRA class.[49] Their distributions are depicted in Figure 8 (see the RDU-CRRA and EU-CRRA curves).

[49] The algorithm for calculating the critical cost efficiency index for the EU-CRRA and RDU-CRRA models is straightforward and does not involve the lattice method. Consider, for example, the EU-CRRA case. At the efficiency vector $\mathbf{e} = (e, e, \ldots, e)$ for $e \in (0, 1]$, we can determine if the modified data set $\mathcal{O}(\mathbf{e})$ is consistent with EU-CRRA for a *given* coefficient of relative risk aversion η. We denote the EU-CRRA utility function of the bundle $\mathbf{x} \in \mathbb{R}^2_+$ by $U(\mathbf{x}; \eta)$ and the EU-CRRA indirect utility at price $\mathbf{p}$ and income m by $V((\mathbf{p}, m); \eta)$ (the formula for which can be easily calculated). $\mathcal{O}(\mathbf{e})$ is rationalized by $U(\cdot; \eta)$ if and only if $U(\mathbf{x}^t; \eta) \geqslant V((\mathbf{p}^t, e\mathbf{p}^t \cdot \mathbf{x}^t); \eta)$ for $t = 1, 2, \ldots, T$. This can be checked for a given η, and by letting η take different values we can establish if $\mathcal{O}(\mathbf{e})$ is EU-CRRA-rationalizable for a given $\mathbf{e}$. Lastly, we perform a binary search over $(0, 1]$ in order to determine the critical value of e, as described in Footnote 32.

Two things are clear from this exercise. First, the misspecification involved in using a parametric model appears to be high. This is suggested by Figure 8 where the pass rate for GARP is much higher than that for RDU-CRRA or EU-CRRA.[50] We can also measure the size of the preference misspecification by the difference in the efficiency index between GARP and RDU-CRRA for each subject. This difference exceeds 0.05 for 118 (out of 207) subjects and exceeds 0.10 for 76 subjects. These observations are broadly in line with those made by Halevy, Persitz, and Zrill (2018).[51] Second, this misspecification is considerably worse than that for the corresponding *non*parametric model; indeed, the difference in the efficiency index between GARP and CRDU exceeds 0.05 for 82 (out of 207) subjects and exceeds 0.10 for 38 subjects.

Section A8 of the Online Appendix provides more analysis of the data collected by Halevy, Persitz, and Zrill (2018). Note that Halevy, Persitz, and Zrill (2018) make use of Varian's inconsistency index, the calculation of which is feasible in the case of GARP and the parametric models they consider. We have not used Varian's index because its calculation for *all* of the nonparametric models that we consider is simply too computationally demanding. We did, however, calculate Varian's index (exactly) for GARP and F-GARP and also have good approximations for the EU model. In all three cases, the indices of Afriat and Varian are highly correlated among subjects.

Comparison with empirical findings in other papers. There is a large empirical literature that evaluates the performance of different models of choice under risk using experimental or field data, and our results appear to be broadly in line with the findings obtained in earlier studies, even though the very different empirical methods employed make formal comparisons difficult. In particular, other papers have concluded that the rank dependent utility model performs well (see, for example, Bruhin, Fehr-Duda, and Epper (2010) and Barseghyan *et al.* (2013) and their references), which is something we also notice, at least in the asymmetric treatment in Choi *et al.* (2007). We find that the expected utility model captures a significant portion of subjects, though by no means everyone, which is broadly

[50] For example, at any efficiency threshold of 0.9, more than 90% of subjects pass GARP but fewer than 60% are consistent with RDU-CRRA.

[51] Using Varian's index rather than Afriat's, Halevy, Persitz, and Zrill (2018) reach a similar conclusion that the contribution to inconsistency from parametric misspecification is large relative to that from the failure of basic rationality.

consistent with the not altogether uncommon finding that this canonical model puts in a respectable performance (see, for example, Hey and Orme (1994)). Lastly, the relatively poor performance of the disappointment aversion model has also been noted in some other studies such as Hey and Orme (1994) and Barseghyan *et al.* (2013).

5.2 Out-of-sample predictions

We use the data from Choi *et al.* (2007) and Halevy, Persitz, and Zrill (2018) to make out-of-sample predictions. For different reasons, neither experiment is ideal for our purpose, but these applications are still indicative of the potential usefulness of our procedure.

The Halevy, Persitz, and Zrill (2018) data. In the second part of the experiment in Halevy, Persitz, and Zrill (2018), each subjects was asked to make a number of choices between two lotteries: lottery A, which pays off some amount for sure, and lottery B, which has two unequal but equiprobable payoffs.[52] The objective was to adjudicate between two competing parametric recovery approaches, one using nonlinear least squares (NLLS) and another using a money metric index (MMI) (which is essentially Varian's inconsistency index). The recovery exercise first involved fitting (using either procedure) the RDU-CRRA model to each subject's portfolio choice data (from the first part of the experiment). This yielded two parameters for each subject: β, which governs the probability distortion (see Section 2.4), and η, the coefficient of relative risk aversion. With these parameters, one could then predict the subject's choice between any two lotteries A and B. The lotteries A and B were not randomly chosen, but instead tailor-made to each subject so that the two recovery methods would lead to *different predictions* for each pair.

Our nonparametric approach could also be used to predict lottery choices. First, for a given model, say RDU, we calculate a subject's efficiency index. Suppose the index is e^*; then we test whether a preference for A over B is consistent with the RDU model given the data set $\{(\mathbf{x}^t, B^t(e^*))\}_{t=1}^T$ (in the sense explained in Section 2.3). Since $\{(\mathbf{x}^t, B^t(e^*))\}_{t=1}^T$ is RDU-rationalizable by definition, either a preference for A over B or its reverse will be consistent. It is also possible that both are consistent, in which case the model cannot discriminate between these lotteries; note that this is where a nonparametric model differs

[52] For example, lottery A might pay 50 for sure and lottery B might pay 60 and 40 with equal probability.

from a parametric model, since the recovered parameters are typically unique in the latter (which lead to unique predictions).

In the experiment, each subject (of 207) was asked to make 9 pairwise choices, giving 1,863 binary comparisons in total. Using the RDU model, we find that only 78 of these cases (around 4%) are discriminating; among these 78 cases, 48 are correctly predicted, about 62%. Under CRDU, discrimination is sharpened, with 949 (51%) being discriminating, and among these, 507 (53%) are correctly predicted. Lastly, under RDU-CRRA, which is of course a parametric model, all 1,863 binary comparisons are discriminating and 1,051 of these (56%) are correctly predicted.[53] The final result is in line with Halevy, Persitz, and Zrill (2018), which correctly predicts 54% of 1,827 binary choices when the RDU-CRRA parameters are recovered using the MMI method (which means, given the special way lotteries were chosen in that experiment, that the NLLS method is correct for 46% of the choices).[54,55]

The Choi *et al.* (2007) data. The special way in which lotteries are chosen in the Halevy, Persitz, and Zrill (2018) experiment gives us only limited information on whether our nonparametric procedure is discriminating enough to be useful for making out-of-sample predictions. To explore this issue further, we conduct a simple but instructive exercise using the data collected for the symmetric treatment in Choi *et al.* (2007). We first identify those subjects with RDU efficiency indices exceeding 0.9. For each subject, we choose an observation at random; call it t'. We then randomly choose a bundle $\mathbf{y}$ that is undominated[56] by $\mathbf{x}^{t'}$ and satisfies $\mathbf{p}^t \cdot \mathbf{y} = 0.9\, \mathbf{p}^{t'} \cdot \mathbf{x}^{t'}$. We then ask whether we could 'predict' the choice between $\mathbf{x}^{t'}$ and $\mathbf{y}$ from the remaining 49 portfolio choice observations (using the procedure set out in Section 2.3). Since we have chosen subjects with efficiency indices above 0.9, a preference for $\mathbf{x}^{t'}$ over $\mathbf{y}$ *must* be consistent with the 49 portfolio choice decisions. The issue is whether a preference for $\mathbf{y}$ over $\mathbf{x}^{t'}$ is *also* consistent with the 49 observations; if so, it means that the nonparametric procedure has failed to be discriminating 'out-of-sample.'

Under RDU, 30 of 47 subjects have efficiency indices exceeding 0.9. For each of these

[53] In a sense, RDU-CRRA is both correct more often and wrong more often than CRDU.

[54] Notice that 4 subjects were removed in Halevy, Persitz, and Zrill (2018), giving a sample of 203 subjects.

[55] Since the lottery pairs in this experiment were chosen in a very special way, the reader should take care not to conclude that these methods, whether parametric or otherwise, make correct predictions only half the time when presented with a 'random' pair of lotteries.

[56] By undominated, we mean that if $\mathbf{y} = (a, b)$ then $\mathbf{x}^{t'} \ngeq (a, b)$ and $\mathbf{x}^{t'} \ngeq (b, a)$.

subjects, we perform two independent (i.e., drawn from different budgets) predictive exercises and find that 50 of 60 (83%) of these are predictively discriminating. In the case of CRDU, 24 of 47 subjects have efficiency indices exceeding 0.9 and we find that 46 of 48 (96%) of the predictive exercises are discriminating. Obviously this simple exercise is no more than indicative, but it does suggest that our nonparametric procedure is capable of making sharp predictions out-of-sample.

Appendix

Proof of Theorem 1. We require the following lemma.

Lemma 1. *Let $\{C^t\}_{t=1}^T$ be a finite collection of constraint sets in $\mathbb{R}_+^{\bar{s}}$ that are compact and downward closed (i.e., if $\mathbf{x} \in C^t$ then so is $\mathbf{y} \in \mathbb{R}_+^{\bar{s}}$ such that $\mathbf{y} < \mathbf{x}$) and let the functions $\{\phi(\cdot, t)\}_{t=1}^T$ be continuous and increasing in all dimensions. Suppose that there is a finite set $\mathcal{X}$ of $\mathbb{R}_+$, a strictly increasing function $\bar{u} : \mathcal{X} \to \mathbb{R}_+$, and $\{M^t\}_{t=1}^T$ such that the following holds:*

$$M^t \geqslant \phi(\bar{\mathbf{u}}(\mathbf{x}), t) \text{ for all } \mathbf{x} \in C^t \cap \mathcal{L} \text{ and} \tag{17}$$

$$M^t > \phi(\bar{\mathbf{u}}(\mathbf{x}), t) \text{ for all } \mathbf{x} \in (C^t \backslash \partial C^t) \cap \mathcal{L}, \tag{18}$$

where $\mathcal{L} = \mathcal{X}^{\bar{s}}$ and $\bar{\mathbf{u}}(\mathbf{x}) = (\bar{u}(x_1), \bar{u}(x_2), \ldots, \bar{u}(x_{\bar{s}}))$. Then there is a Bernoulli function $u : \mathbb{R}_+ \to \mathbb{R}_+$ that extends $\bar{u}$ such that

$$M^t \geqslant \phi(\mathbf{u}(\mathbf{x}), t) \text{ for all } \mathbf{x} \in C^t \text{ and} \tag{19}$$

$$\text{if } \mathbf{x} \in C^t \text{ and } M^t = \phi(\mathbf{u}(\mathbf{x}), t), \text{ then } \mathbf{x} \in \partial C^t \cap \mathcal{L} \text{ and } M^t = \phi(\bar{\mathbf{u}}(\mathbf{x}), t). \tag{20}$$

Remark: The property (20) needs some explanation. Conditions (17) and (18) allow for the possibility that $M^t = \phi(\bar{\mathbf{u}}(\mathbf{x}'), t)$ for some $\mathbf{x}' \in \partial C^t \cap \mathcal{L}$; we denote the set of points in $\partial C^t \cap \mathcal{L}$ with this property by X'. Clearly any extension u will preserve this property, i.e., $M^t = \phi(\mathbf{u}(\mathbf{x}'), t)$ for all $\mathbf{x}' \in X'$. Property (20) says that we can choose u such that for all $\mathbf{x} \in C^t \backslash X'$, we have $M^t > \phi(\mathbf{u}(\mathbf{x}), t)$.

Proof: We shall prove this result by induction on the dimension of the space containing the constraint sets. It is trivial to check that the claim is true if $\bar{s} = 1$. In this case, $\mathcal{L}$ consists of a finite set of points on $\mathbb{R}_+$ and each C^t is a closed interval with 0 as its minimum. Now let us suppose that the claim holds for $\bar{s} = m$ and we shall prove it for $\bar{s} = m + 1$. If, for each t, there is a strictly increasing and continuous utility function $u^t : \mathbb{R}_+ \to \mathbb{R}_+$ extending $\bar{u}$ such that (19) and (20) hold, then the the same conditions will hold for the increasing and continuous function $u = \min_t u^t$. So we can focus our attention on constructing u^t for a single constraint set C^t.

Suppose $\mathcal{X} = \{0, r^1, r^2, r^3, \ldots, r^I\}$, with $r^0 = 0 < r^i < r^{i+1}$, for $i = 1, 2, \ldots, I-1$. Let $\bar{r} = \max\{r \in \mathbb{R}_+ : (r, 0, 0, \ldots, 0) \in C^t\}$ and suppose that $(r^i, 0, 0, \ldots, 0) \in C^t$ if and only if $i \leqslant N$ (for some $N \leqslant I$). Consider the collection of sets of the form $D^i = \{\mathbf{y} \in \mathbb{R}^m_+ : (r^i, \mathbf{y}) \in C^t\}$ (for $i = 1, 2, \ldots, N$); this is a finite collection of compact and downward closed sets in $\mathbb{R}^m_+$. By the induction hypothesis applied to $\{D^i\}_{i=1}^N$, with $\{\phi(\bar{u}(r^i), \cdot, t)\}_{i=1}^N$ as the collection of functions, there is a strictly increasing function $u^* : \mathbb{R}_+ \to \mathbb{R}_+$ extending $\bar{u}$ such that

$$M^t \geqslant \phi(\bar{u}(r^i), \mathbf{u}^*(\mathbf{y}), t) \text{ for all } (r^i, \mathbf{y}) \in C^t \text{ and} \tag{21}$$

$$\text{if } (r^i, \mathbf{y}) \in C^t \text{ and } M^t = \phi(\bar{u}(r^i), \mathbf{u}^*(\mathbf{y}), t), \text{ then } (r^i, \mathbf{y}) \in \partial C^t \cap \mathcal{L} \text{ and } M^t = \phi(\bar{\mathbf{u}}(r^i, \mathbf{y}), t). \tag{22}$$

For each $r \in [0, \bar{r}]$, define

$$U(r) = \{u \leqslant u^*(r) : \max\{\phi(u, \mathbf{u}^*(\mathbf{y}), t) : (r, \mathbf{y}) \in C^t\} \leqslant M^t\}.$$

This set is nonempty; indeed $\bar{u}(r^k) = u^*(r^k) \in U(r)$, where r^k is the largest element in $\mathcal{X}$ that is weakly smaller than r. This is because, if $(r, \mathbf{y}) \in C^t$ then so is $(r^k, \mathbf{y})$, and (21) guarantees that $\phi(\bar{u}(r^k), \mathbf{u}^*(\mathbf{y}), t) \leqslant M^t$. The downward closedness of C^t and the fact that u^* is increasing also guarantees that $U(r) \subseteq U(r')$ whenever $r < r'$. Now define $\tilde{u}(r) = \sup U(r)$; the function $\tilde{u}$ has a number of significant properties. (i) *For* $r \in \mathcal{X}$, $\tilde{u}(r) = u^*(r) = \bar{u}(r)$ (by the induction hypothesis). (ii) $\tilde{u}$ *is a nondecreasing function* since U is nondecreasing. (iii) $\tilde{u}(r) > \bar{u}(r^k)$ *if* $r > r^k$*, where* r^k *is largest element in* $\mathcal{X}$ *smaller than* r. Indeed, because C^t is compact and ϕ continuous, $\phi(\tilde{u}(r), \mathbf{u}^*(\mathbf{y}), t) \leqslant M^t$ for all $(r, \mathbf{y}) \in C^t$. By way of contradiction, suppose $\tilde{u}(r) = \bar{u}(r^k)$ and hence $\tilde{u}(r) < u^*(r)$. It follows from the definition of $\tilde{u}(r)$ that, for any sequence u_n, with $\tilde{u}(r) < u_n < u^*(r)$ and $\lim_{n\to\infty} u_n = \tilde{u}(r)$, there is $(r, \mathbf{y}_n) \in C^t$ such that $\phi(u_n, \mathbf{u}^*(\mathbf{y}_n), t) > M^t$. Since C^t is compact, we may assume with no loss of generality that $\mathbf{y}_n \to \hat{\mathbf{y}}$ and $(r, \hat{\mathbf{y}}) \in C^t$, from which we obtain $\phi(\tilde{u}(r), \mathbf{u}^*(\hat{\mathbf{y}}), t) = M^t$. Since C^t is downward closed, $(r^k, \hat{\mathbf{y}}) \in C^t$ and, since $\bar{u}(r^k) = u^*(r^k)$, we have $\phi(\mathbf{u}^*(r^k, \hat{\mathbf{y}}), t) = M^t$. This can only occur if $(r^k, \hat{\mathbf{y}}) \in \partial C^t \cap \mathcal{L}$ (because of (22)), but it is clear that $(r^k, \hat{\mathbf{y}}) \notin \partial C^t$ since $(r^k, \hat{\mathbf{y}}) < (r, \hat{\mathbf{y}})$. (iv) *If* $r_n < r^i$ *for all* n *and* $r_n \to r^i \in \mathcal{X}$, *then* $\tilde{u}(r_n) \to u^*(r^i)$. Suppose to the contrary, that the limit is $\hat{u} < u^*(r^i) = \bar{u}(r^i)$. Since u^* is continuous, we can assume, without loss of generality, that $\tilde{u}(r_n) < u^*(r_n)$. By the compactness of C^t, the continuity of ϕ, and the definition of $\tilde{u}$, there is $(r_n, \mathbf{y}_n) \in C^t$ such that $\phi(\tilde{u}(r_n), \mathbf{u}^*(\mathbf{y}_n), t) = M^t$. This leads to $\phi(\hat{u}, \mathbf{u}^*(\mathbf{y}'), t) = M^t$, where $\mathbf{y}'$ is an accumulation point of $\mathbf{y}_n$ and $(r^i, \mathbf{y}') \in C^t$. But since ϕ is strictly increasing, we obtain $\phi(u^*(r^i), \mathbf{u}^*(\mathbf{y}'), t) > M^t$, which contradicts (21).

Given the properties of $\tilde{u}$, we can find a continuous and strictly increasing function u^t such that u^t extends $\bar{u}$, i.e., $u^t(r) = \bar{u}(r)$ for $r \in \mathcal{X}$, $u^t(r) < u^*(r)$ for all $r \in \mathbb{R}_+ \backslash \mathcal{X}$ and $u^t(r) < \tilde{u}(r) \leqslant u^*(r)$ for all $r \in [0, \bar{r}] \backslash \mathcal{X}$. (In fact we can choose u^t to be smooth everywhere except possibly on $\mathcal{X}$.) We claim that (19) and (20) are satisfied for C^t. To see this, note that for $r \in \mathcal{X}$ and $(r, \mathbf{y}) \in C^t$, the induction hypothesis guarantees that (21) and (22) hold and they will continue to hold if u^* is replaced by u^t. In the case where $r \notin \mathcal{X}$ and $(r, \mathbf{y}) \in C^t$, since $u^t(r) < \tilde{u}(r)$ and ϕ is increasing, we obtain $M^t > \phi(\mathbf{u}^t(r, \mathbf{y}), t)$. **QED**

Proof of Theorem 1: This follows immediately from Lemma 1 if we set $C^t = \underline{B}^t$, and $M^t = \phi(\bar{\mathbf{u}}(\mathbf{x}^t), t)$. If $\bar{u}$ obeys conditions (6) and (7) then it obeys conditions (17) and (18). The rationalizability of $\mathcal{O}$ by $\{\phi(\cdot, t)\}_{t \in T}$ then follows from (19). **QED**

Proof of Proposition 1 (sufficiency). Suppose there is $\bar{u} : \mathcal{X} \to \mathbb{R}_+$ satisfying (i) to (iii) and let $\bar{u}_\ell$ be its linear extension. Then (i) guarantees that $\bar{u}_\ell$ is strictly increasing on $\mathbb{R}_+$ and (ii) guarantees that $\bar{u}_\ell$ is concave. We claim that with $\bar{u}_\ell$ as the Bernoulli function, $\mathbf{x}^t$ has higher expected utility than any bundle in $B^t(e^t)$. By definition, $\bar{u}_\ell$ is linear between adjacent values of $\mathcal{X}$; it follows that the map from (a, b) to its expected utility $\pi_1 \bar{u}_\ell(a) + \pi_2 \bar{u}_\ell(b)$ is also linear for all $(a, b) \in [r, r'] \times [m, m']$, where r and r' are adjacent points in $\mathcal{X}$ (and similarly m and m'). A linear map is maximized at an extreme point; thus if $([r, r'] \times [m, m']) \cap \partial \mathscr{B}(\mathbf{p}^t, e^t \mathbf{p}^t \cdot \mathbf{x}^t)$ is nonempty then there is a bundle (a^*, b^*) maximizing expected utility in this set with either $a^* \in \{r, r'\}$ or $b^* \in \{m, m'\}$. More generally, there must be a bundle (a^{**}, b^{**}) that maximizes expected utility in $\mathscr{B}(\mathbf{p}^t, e^t \mathbf{p}^t \cdot \mathbf{x}^t)$ and is contained in $\partial \mathscr{B}(\mathbf{p}^t, e^t \mathbf{p}^t \cdot \mathbf{x}^t) \cap \mathcal{N}$. It follows that (16) is sufficient to guarantee the optimality of $\mathbf{x}^t$ in

$B^t(e^t)$. **QED**

References

AFRIAT, S. N. (1967): "The Construction of Utility Functions from Expenditure Data," *International Economic Review*, 8(1), 67–77.

AFRIAT, S. N. (1972): "Efficiency Estimation of Production Functions," *International Economic Review*, 13(3), 568–598.

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

AHN, D., S. CHOI, D. GALE, S. KARIV (2014): "Estimating Ambiguity Aversion in a Portfolio Choice Experiment," *Quantitative Economics*, 5(2), 195–223.

ANDREONI, J., and C. SPRENGER (2012): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102(7), 3333–3356.

ANDREONI, J., and J. MILLER (2002): "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 70(2), 737–753.

BARSEGHYAN, L., F. MOLINARI, T. O'DONOGHUE, and J. C. TEITELBAUM (2013): "The Nature of Risk Preferences: Evidence from Insurance Choices," *American Economic Review*, 103(6), 2499–2529.

BAYER, R.-C., S. BOSE, M. POLISSON, and L. RENOU (2013): "Ambiguity Revealed," *IFS Working Papers*, W13/05, Institute for Fiscal Studies.

BEATTY, T. K. M., and I. A. CRAWFORD (2011): "How Demanding Is the Revealed Preference Approach to Demand?," *American Economic Review*, 101(6), 2782–2795.

BRONARS, S. G. (1987): "The Power of Nonparametric Tests of Preference Maximization," *Econometrica*, 55(3), 693–698.

BROWN, D. J., and R. L. MATZKIN (1996): "Testable Restrictions on the Equilibrium Manifold," *Econometrica*, 64(6), 1249–1262.

BRUHIN, A., H. FEHR-DUDA, and T. EPPER (2010): "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," *Econometrica*, 78(4), 1375–1412.

BUSSIECK, M. R., and S. VIGERSKE (2010): "MINLP solver software," *Wiley Encyclopedia of Operations Research and Management Science*.

CAPPELEN, A. W., S. KARIV, E. Ø. SØRENSEN, and B. TUNGODDEN (2015): "Is There a Development Gap in Rationality?," Unpublished manuscript, UC Berkeley.

CHAMBERS, C. P., C. LIU, and S.-K. MARTINEZ (2016): "A Test for Risk-Averse Expected Utility," *Journal of Economic Theory*, 163, 775–785.

CHAMBERS, C. P., F. ECHENIQUE, and K. SAITO (2016): "Testing Theories of Financial Decision Making," *Proceedings of the National Academy of Sciences*, 113(15), 4003–4008.

CHOI, S., R. FISMAN, D. GALE, and S. KARIV (2007): "Consistency and Heterogeneity of Individual Behavior under Uncertainty," *American Economic Review*, 97(5), 1921–1938.

CHOI, S., S. KARIV, W. MÜLLER, and D. SILVERMAN (2014): "Who Is (More) Rational?," *American Economic Review*, 104(6), 1518–1550.

DIEWERT, W. E. (1973): "Afriat and Revealed Preference Theory," *Review of Economic Studies*, 40(3), 419–425.

DIEWERT, W. E. (2012): "Afriat's Theorem and Some Extensions to Choice Under Uncertainty," *Economic Journal*, 122(560), 305–331.

DYBVIG, P., and H. POLEMARCHAKIS (1981): "Recovering Cardinal Utility," *Review of Economic Studies*, 48(1), 159–166.

ECHENIQUE, F., and K. SAITO (2015): "Savage in the Market," *Econometrica*, 83(4), 1467–1495.

ECHENIQUE, F., S. LEE, and M. SHUM (2011): "The Money Pump as a Measure of Revealed Preference Violations," *Journal of Political Economy*, 119(6), 1201–1223.

EPSTEIN, L. G. (2000): "Are Probabilities Used in Markets?," *Journal of Economic Theory*, 91(1), 86–90.

FISHBURN, P. C. (1975): "Separation Theorems and Expected Utilities," *Journal of Economic Theory*, 11(1), 16–34.

FRIEDMAN, M., and L. J. SAVAGE (1948): "Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, 56(4), 279–304.

GILBOA, I., and D. SCHMEIDLER (1989): "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, 18(2), 141–153.

GNEEZY, U., and J. POTTERS (1997): "An Experiment on Risk Taking and Evaluation Periods," *Quarterly Journal of Economics*, 112(2), 631–645.

GREEN, R. C., and S. SRIVASTAVA (1986): "Expected Utility Maximization and Demand Behavior," *Journal of Economic Theory*, 38(2), 313–323.

GUL, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59(3), 667–686.

HALEVY, Y., D. PERSITZ, and L. ZRILL (2018): "Parametric Recoverability of Preferences," *Journal of Political Economy*, 126(4), 1558–1593.

HARBAUGH, W. T., K. KRAUSE, and T. R. BERRY (2001): "GARP for Kids: On the Development of Rational Choice Behavior," *American Economic Review*, 91(5), 1539–1545.

HEY, J. D., and C. ORME (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), 1291–1326.

HEY, J. D., and N. PACE (2014): "The Explanatory and Predictive Power of Non Two-Stage-Probability Theories of Decision Making under Ambiguity," *Journal of Risk and Uncertainty*, 49(1), 1–29.

IMAI, T., and C. F. CAMERER (2018): "Estimating Time Preferences from Budget Set Choices Using Optimal Adaptive Design," Unpublished manuscript, LMU Munich.

KŐSZEGI, B., and M. RABIN (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97(4), 1047–1073.

KUBLER, F., L. SELDEN, and X. WEI (2014): "Asset Demand Based Tests of Expected Utility Maximization," *American Economic Review*, 104(11), 3459–3480.

LOOMES, G. (1991): "Evidence of a New Violation of the Independence Axiom," *Journal of Risk and Uncertainty*, 4(1), 91–108.

MACCHERONI, F., M. MARINACCI, and A. RUSTICHINI (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74(6), 1447–1498.

MATTEI, A. (2000): "Full-Scale Real Tests of Consumer Behavior Using Experimental Data," *Journal of Economic Behavior and Organization*, 43(4), 487–497.

NISHIMURA, H., E. A. OK, and J. K.-H. QUAH (2017): "A Comprehensive Approach to Revealed Preference Theory," *American Economic Review*, 107(4), 1239–1263.

QUAH, J. K.-H. (2014): "A Test for Weakly Separable Preferences," *Department of Economics Discussion Paper Series*, 708, University of Oxford.

QUIGGIN, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3(4), 323–343.

SMEULDERS, B., F. C. R. SPIEKSMA, L. CHERCHYE, and B. DE ROCK (2014): "Goodness-of-Fit Measures for Revealed Preference Tests: Complexity Results and Algorithms," *ACM Transactions on Economics and Computation*, 2(1), 3:1–3:16.

VARIAN, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50(4), 945–973.

VARIAN, H. R. (1983a): "Non-Parametric Tests of Consumer Behaviour," *Review of Economic Studies*, 50(1), 99–110.

VARIAN, H. R. (1983b): "Nonparametric Tests of Models of Investor Behavior," *Journal of Financial and Quantitative Analysis*, 18(3), 269–278.

VARIAN, H. R. (1988): "Estimating Risk Aversion from Arrow-Debreu Portfolio Choice," *Econometrica*, 56(4): 973–979.

VARIAN, H. R. (1990): "Goodness-of-Fit in Optimizing Models," *Journal of Econometrics*, 46(1–2), 125–140.