



Revealed preference tests for linear probability-prize tradeoffs

Zachary Breig¹ · Paul Feldman²

Received: 13 October 2024 / Accepted: 1 September 2025
© The Author(s) 2025

Abstract

We provide necessary and sufficient conditions for expected utility and risk-averse expected utility to rationalize behavior when the decision maker faces linear probability-prize tradeoffs. The setting subsumes those proposed in Andreoni and Harbaugh (Unexpected utility: Five experimental tests of preferences for risk, 2009)—risk-reward budgets—and Crosetto and Filippin (J Risk Uncertain 47(1):31–65, 2013)—bomb risk elicitation tasks. The tests are intuitive and straightforward to carry out. They also make it possible to compute the Houtman–Maks Index to measure how close the models come to explaining the choice sets, although we emphasize the limitations of these measures in this setting. We implement these results empirically and show that only small portions of choice sets can be rationalized by risk-averse expected utility.

Keywords Risk preferences · Uncertainty · Revealed preference · Expected utility · Risk aversion

JEL Classification C91 · D81 · D91

Both authors contributed equally to this work. We are grateful to Mikhail Freer and seminar participants at the University of Queensland for helpful comments. We would like to give special thanks to Paolo Crosetto and Antonio Filippin for sharing their data. Part of the data used in this paper was gathered using funds from the National Science Foundation's dissertation grant #1824121.

✉ Zachary Breig
z.breig@uq.edu.au

Paul Feldman
paul.feldman@tamu.edu

¹ University of Queensland, Brisbane, Australia

² Texas A&M University, College Station, USA

1 Introduction

Evaluating and measuring risk preferences is fundamental to experimental economics and decision theory. Risk preferences are measured directly as an explanatory variable and indirectly as a potential confound. While researchers often attempt to measure risk aversion using experimental tasks, the revealed preference approach is to instead determine consistency of choices with existing models. In this paper, we link both approaches, showing how to test whether expected utility and risk-averse expected utility can explain behavior in a common experimental task used to measure risk aversion.

We study a commonly-used experimental risk elicitation: linear probability-prize tradeoffs. Decision makers must choose from amongst a set of lotteries where increasing the size of the prize decreases the likelihood of receiving that prize by a fixed amount. If the decision maker does not receive the prize, they receive nothing. This is a natural setting to elicit risk preferences because subjects are literally trading off “risks” against “rewards.” Our research questions are twofold. First, we ask what restrictions a choice set must satisfy in order to be consistent with standard economic models of choice under risk. Second, we empirically investigate how close choice sets from existing experiments are to satisfying these restrictions.

In our theoretical results, we characterize necessary and sufficient conditions for a choice set to be consistent with expected utility (in Theorem 1) and risk-averse expected utility (in Theorem 2). The conditions are straightforward and easy to check in a dataset. Our first theoretical result is that the only testable implication of expected utility is that chosen prizes are weakly increasing in the maximum prize available. That is, expected utility maximizers will choose (weakly) higher prizes when we increase the maximum prize. These theoretical results are closely related to Cherchye et al. (2021), which also studies probability-prize tradeoffs. It shows that in a first price auction where the distribution of bids is log-concave, the only implication of expected utility is that bids increase in the bidder’s value for the object. Because the linear budgets that we consider in this paper correspond to a uniform (and, thus, log-concave) distribution of bids, this result implies what we find in Theorem 1. Our second theoretical result, Theorem 2, is novel and provides a practical test for consistency with risk-averse expected utility, a model of interest in most applied research. We also show that, as a by-product, our theorems allow straightforward computation of Houtman-Maks Indices (which measure *how close* datasets are to being consistent with a model) for both expected utility and risk-averse expected utility in linear probability-prize tradeoffs. This is useful for comparing with prior revealed preference studies, although, as we later discuss, such indices can behave pathologically in our setting and should be interpreted cautiously.

The main contribution of our paper is linking these theoretical results to existing experimental procedures and datasets. We show how to carry out the tests and compute consistency measures using data from Breig and Feldman (2024) for the tasks proposed by Andreoni and Harbaugh (2009). We then complete the same exercise for the bomb risk elicitation task (BRET) using data from Crosetto and Filippin (2013). We find that while subjects have moderate levels of consistency with expected utility, only small portions of their choice sets can be rationalized by a risk-averse expected

utility function. The reason for this counter-intuitive result is that subjects should always select the same preferred prize from *all* budgets with the same maximal prize if they are risk-averse expected utility maximizers. We stress that the higher level of consistency with expected utility *does not* imply that choices were consistent with risk-seeking behavior; instead, the choices cannot jointly be explained by a single risk-averse utility function even though many choices can individually be explained by distinct risk-averse utility functions.

This paper contributes to the literature evaluating both revealed preferences and risk preferences. One strand of the literature has focused on decisions in which the subject trades off the size of rewards in two different states of the world, which can be thought of as the outcomes resulting from a coin flip (Choi et al., 2007, 2014; Halevy et al., 2018; Polisson et al., 2020; Dembo et al., 2021). In this setting, Echenique et al. (2023) propose an index that measures how far a dataset is from being rationalized by a risk-averse expected utility function. Their index captures the smallest perturbation required to render the data consistent with the model. In contrast, the index we develop measures the size of the largest subset of choices that is consistent with risk-averse expected utility.¹ These revealed preference approaches have also been used to study choices over three-outcome lotteries that can be represented in the normalized probability simplex (p_{high}, p_{low}) , as probabilities add up to one (Burghart, 2020; Feldman and Rehbeck, 2022). Linear budgets in the simplex trade off the chance of the middle outcome against the chance of both the low and high outcomes. Our approach focuses on the tradeoff between the size of a prize and the probability of receiving it, which has applications in the analysis of bargaining, criminal activity, and first-price auctions (Andreoni and Miller, 2002; Harbaugh et al., 2013; Cherchye et al., 2021).

More general revealed approaches have been discussed in the literature. Fishburn (1975) was the first to provide a revealed preference characterization for binary choices. Border (1992) and Kim (1996) extend the characterization to budgets. More recently, Payró (2019) generalized this line of work to the framework of preferences over menus with subjective state spaces, capturing behavioral attitudes like temptation (Gul and Pesendorfer, 2001; Dekel et al., 2009). In these frameworks, failure of consistency typically implies that one can construct a compound lottery (or menu) that weakly dominates the observed choices under some ex-ante distribution; the exact form of this dominance or equality relation depends on the structure of the choice objects (e.g., lotteries versus menus). These results are largely existential and do not provide straightforward computational tests. The one exception is Payró (2019), as it provides a way to turn this existential characterization into a set of non-linear inequalities which can be used to check for the existence of solutions numerically. Our contribution is a simple test that can be verified “manually” and that, in our restricted context, can be used to construct feasible choices and an ex-ante distribution that dominates the choices that violate expected utility.

¹ Our main theorems can also be used to compute one of the indices introduced by Echenique et al. (2023) in the context of linear probability-prize tradeoffs. Full details are available from the authors upon request.

Other elicitation methods, like the investment game from Gneezy and Potters (1997), certainty equivalents, and the gambles used in Eckel and Grossman (2002) are discrete/restricted versions of the state contingent payoffs. Most risky binary choices, as well as their corresponding choice lists, can be analyzed with some probability simplex.² For example, Holt and Laury (2002) choice lists can be summarized in the four-outcome simplex, and most probability equivalents can be represented in a normalized simplex. Hence, most known elicitation techniques have a budgetary analog, and revealed preferences may provide additional testable predictions for known models. Our results provide insights on how to leverage revealed preference results to derive specific tests for risk elicitation that are often used.

Our results can be readily applied to determine whether choice data from commonly used elicitation methods are consistent with standard models. The consistency measures that we propose are easy to compute and will allow researchers to evaluate how close subjects are to choosing according to expected utility.

2 Theory

We study choices from budgets of the form $x + \frac{M_t}{m_t}p = M_t$ for $t = 1, \dots, T$, where m_t and M_t are strictly positive and x and p are constrained to be weakly positive. With this formulation, m_t and M_t can be interpreted as the budget endpoints, where the former is the maximum available probability of receiving a prize and the latter is the maximum available prize. The agent selects the combination (x, p) that yields their preferred binary lottery $\{ \$x, p; \$0, 1 - p \}$. The choice variable x refers to the size of the positive prize that the decision maker will receive, while p refers to the likelihood that they will receive the prize. We define a dataset as $\mathcal{O} = \{(p_t, x_t, m_t, M_t)\}_{t=1}^T$, where p_t and x_t are the chosen probability and prize when the maximum probability is m_t and the maximum prize is M_t .

For reasons that will become apparent in our results below, it will be useful to “order” datasets in a particular way. We say that a dataset is “regular” if (1) $M_t \geq M_{t-1}$ for all $t \geq 2$ and (2) $x_{t'} \geq x_t$ for all $t' > t$ such that $M_t = M_{t'}$. These conditions ensure that the dataset is first sorted by M and then by x . Because this amounts to a simple relabeling of the dataset, and any dataset can be relabeled in this way, there is no loss of generality from restricting attention to regular datasets.

We say that a dataset \mathcal{O} is *rationalized by expected utility* if there exists a utility function $u : \mathbb{R} \rightarrow \mathbb{R}$ such that (1) u is strictly increasing, (2) $u(0) = 0$, and (3) for all t ,

$$p_t u(x_t) \geq p u(x)$$

²Note that Feldman and Ferraro (2023) shows that preferences over binary choices may not be *isolated* from the choice list in which they are embedded and thus choice-list dependent.

for all non-negative p and x such that $x + \frac{M_t}{m_t}p = M_t$. With this definition, we can state how an expected utility maximizer behaves when faced with linear probability-prize tradeoffs.

Theorem 1 *A regular dataset \mathcal{O} is rationalized by expected utility if and only if (1) $0 < x_t < M_t$ for all t and (2) $x_t \geq x_{t-1}$ for all $t \geq 2$.*

In essence, Theorem 1 states that in order for the decision maker's choices to be rationalized by expected utility, it must be the case that they do not make stochastically dominated choices (by choosing the endpoints of the budgets) and that their chosen prize is (weakly) increasing in the maximum prize. Furthermore, expected utility does not impose restrictions on how changes in the maximum available probability should change the decision maker's chosen prize.

Note that expected utility does not impose consistency in choice behavior; that is, an EU decision-maker could be indifferent among several options from the same budget and, therefore, make different choices across repetitions of that budget. However, the highest value of the prize chosen from one budget must be less than the lowest value of the prize chosen from any other budget with a higher maximum prize.

Another implication of Theorem 1 is that it is easy to find the size of the largest proportion of the decision maker's choices that are consistent with expected utility. This can be thought of as an Expected Utility "Houtman-Maks" index (EU HMI)—see Houtman and Maks (1985). When the dataset is "regular" as described above, finding (one of) the largest subset(s) of choices that are consistent with expected utility is equivalent to finding the longest increasing subsequence of chosen prizes, conditional on those prizes not being endpoints of the budget. The HMI is much simpler to compute like this than by finding the largest acyclic set.

While it is common to test for choices being consistent with expected utility, researchers may also be interested in determining whether choices are consistent with *risk-averse* expected utility, which would require a concave utility function.

We say that a dataset \mathcal{O} is *rationalized by risk-averse expected utility* if there exists a utility function $u : \mathbb{R} \rightarrow \mathbb{R}$ such that (1) u is strictly increasing and concave, (2) $u(0) = 0$, and (3) for all t ,

$$p_t u(x_t) \geq p_t u(x)$$

for all p and x such that $x + \frac{M_t}{m_t}p = M_t$. With this definition, we can state how a risk-averse expected utility maximizer behaves when faced with linear probability-prize tradeoffs.

Theorem 2 *A regular dataset \mathcal{O} is rationalized by a risk-averse expected utility function if and only if (1) $x_1 \leq \frac{1}{2}M_1$, (2) $0 < x_t$ for all t , and (3)*

$$0 \leq x_t - x_{t-1} \leq \frac{1}{2}(M_t - M_{t-1})$$

for all $t \geq 2$.

Clearly, in order to satisfy risk-averse expected utility, all of the requirements from Theorem 1 are required to hold (because risk-averse expected utility is a special case of expected utility). However, Theorem 2 has additional requirements. First, the choice from the budget with the lowest maximum prize must involve the subject allocating weakly less than half of their budget toward the prize. This is intuitive because allocating exactly half of the budget is the expected-value maximizing (i.e. risk neutral) choice. Second, any other choices must change the prize by less than half of the change in the maximum prize, with both changes in the same direction. A less obvious implication of this is that any choices from budgets that have the same maximum prize must have the same chosen prize (irrespective of the maximum probability). This is a very strict requirement that will be discussed further in Sect. 3. Two examples of the restrictions implied by risk-averse expected utility can be found in Fig. 1

Similar to Theorem 1, Theorem 2 has the benefit of making it easy to compute the size of the largest proportion of choices that are consistent with risk-averse expected utility (an index that we call Risk-Averse Expected Utility HMI (RAEU HMI)). With a slight restatement, the problem is equivalent to finding the longest path in a directed acyclic graph.³ Similar to the case of expected utility, this is easy to compute, although the following sections demonstrate that these scores can understate near-consistent behavior in this domain.

We emphasize that other papers (Green and Srivastava, 1986; Chambers et al., 2016; Polissou et al., 2020) have derived conditions—with or without direct tests—for

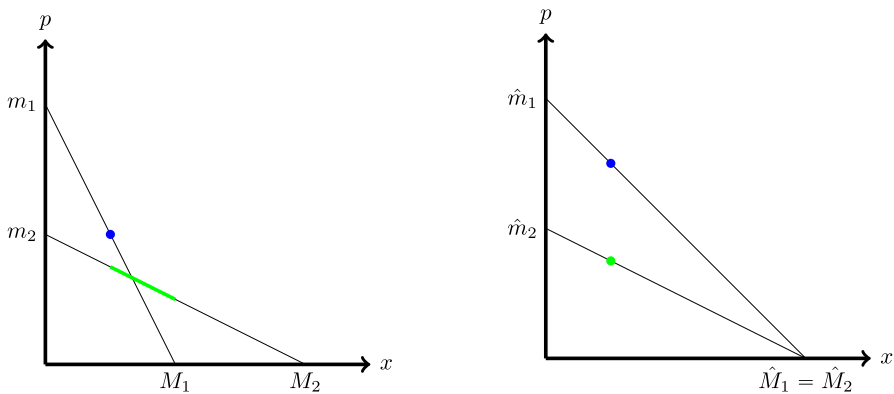


Fig. 1 Two pairs of budgets representing of the requirements of risk-averse expected utility. For each pair, we fix an arbitrary choice from the first budget (in blue). The green region in each second budget shows the choices from that budget that would be consistent with risk-averse expected utility. For the pair of budgets on the left, the difference in x from budget 1 to 2 must be between 0 and $\frac{1}{2}(M_2 - M_1)$. For the pair of budgets on the right, the chosen x must be the same from both budgets because $\frac{1}{2}(\hat{M}_2 - \hat{M}_1) = 0$

³To see this, define each choice that is not an endpoint and spends less than half of its budget on prize as a node in the graph. There is an edge from the node associated with choice i to the node associated with choice j if $j > i$ and $0 \leq x_j - x_i \leq \frac{1}{2}(M_j - M_i)$. The collection of nodes in any path of such a graph will represent a dataset that can be rationalized by risk-averse expected utility. Thus, finding the longest path is equivalent to finding the largest set of choices consistent with risk-averse expected utility.

datasets to be rationalizable by EU and RAEU in settings where contingent consumption demands are directly observable, e.g. Choi et al. (2007). Our theorems, unlike those, provide direct tests of EU and RAEU rationalizability that are directly applicable to the specific choice context that we examine. In the next section, we show explicitly how to apply our tests.

3 Empirical results

In this section, we apply our theoretical results to existing data from experiments. First, we show how our result can be applied using the Andreoni and Harbaugh (2009) task and data from Breig and Feldman (2024). We then apply the results to the bomb risk elicitation task, using data from the repeated treatment of Crosetto and Filippin (2013). To ensure our empirical results on HMI indices are comparable across experiments, we define HMI as the *proportion* of choices consistent with the model, rather than simply counting the number of choices, as was done in most prior studies.

3.1 Andreoni-Harbaugh task

We use data from Breig and Feldman (2024). In the initial part of the experiment, which is the source of the data we focus on, subjects made 50 choices from the class of budgets described in the previous section (we do *not* use any of the revised choices from the Breig and Feldman (2024) data). In each of the 50 choices, subjects used a slider to allocate their budget towards either increasing the size of a prize or increasing the likelihood of receiving said prize. Manipulating the slider modified an interactive graphical feature that showed the currently selected probability and prize using a pie chart and bar chart, respectively. Subjects confirmed their choice by typing it in before moving to the next task. The maximum prizes available across budgets varied from 30 Australian dollars to 200 Australian dollars, and the maximum probabilities ranged from 0.15 to 1. The full set of budgets can be seen in Table 1.

Breig and Feldman (2024) showed that 27/181 subjects choose in a way that is consistent with maximizing a strictly increasing utility function (i.e., they have an HMI of one). Of these 27 subjects, five were consistent with expected value maximization (allocating exactly 50% of each of their budgets towards the prize), and one allocated exactly 25% of each of their budgets towards the prize. It is easy to apply our theoretical results to find the corresponding numbers of decision-makers that are consistent with expected utility (7/181) and risk-averse expected utility (6/181). The

Table 1 Budget endpoints from the experiment described in Breig and Feldman (2024)

m	M	m	M	m	M	m	M	m	M
0.50	\$30	0.60	\$60	1.00	\$80	0.30	\$120	0.80	\$160
0.60	\$30	0.75	\$60	0.15	\$100	0.60	\$120	0.25	\$200
0.50	\$40	1.00	\$60	0.20	\$100	0.30	\$150	0.30	\$200
0.80	\$40	0.40	\$80	1.00	\$100	0.20	\$160	0.40	\$200
1.00	\$50	0.80	\$80	0.15	\$120	0.40	\$160	0.50	\$200

Subjects faced each budget twice, for a total of 50 choice problems

only subjects who were consistent with risk-averse expected utility were those who followed the simple decision rule of allocating a constant proportion of their budget to the prize in every choice problem.

Choices from 50 budgets prove to be a demanding test for expected utility and risk-averse expected utility. A natural question is how close subject's choices were to consistency with either model. To address this question, we compute the size of the largest subset of each subject's choices that is consistent with each model. We label these indices as EU HMI and RAEU HMI and show their distribution in Fig. 2.

There are a few things to note about the distributions found in Fig. 2. First, they both share the characteristic that there is a central distribution with a long tail of high scores. The scores that are equal to 1 are those subjects that are completely consistent with EU and RAEU, respectively. Since risk-averse expected utility is more restrictive than expected utility, the distribution of EU HMI (first-order) stochastically dominates RAEU HMI. However, it is stark how much lower RAEU scores are than EU. This is largely because of the restriction that for a dataset to be consistent with RAEU, the subject must choose the exact same positive prize for all budgets that have the same maximum prize. Because there are ten unique maximum prizes, this requires a large degree of consistency across the offered budgets. Given that chosen prizes should be identical in each set of budgets sharing a common maximum prize, we observe a low median RAEU HMI of 0.22.

In Fig. 3 we compare the distributions of HMI indices from Breig and Feldman (2024) to those that arise from random choice (Bronars, 1987). Using the experimenter's budgets, we create 1000 artificial datasets in which each artificial subject randomizes uniformly over the 101 available options in each budget. In general, the proportion of choices that are consistent in the dataset from real subjects is much higher than the proportion from randomly generated datasets. The exception to this

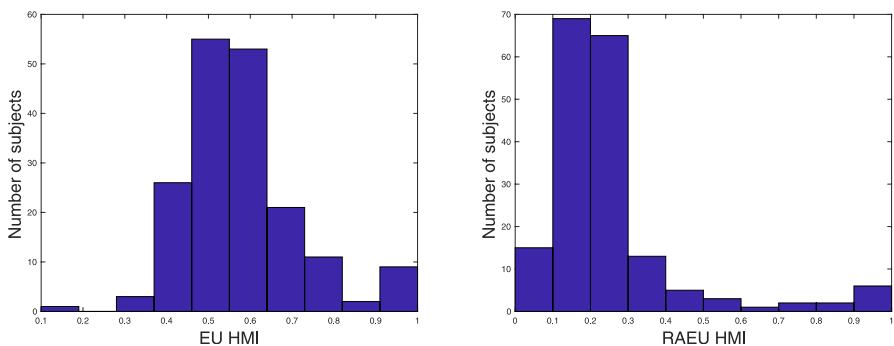


Fig. 2 Histogram of EU HMI and RAEU HMI in the data from Breig and Feldman (2024). Each subject's dataset consisted of 50 choices. We report the rationalizable proportion to allow for easier comparison between datasets of different sizes

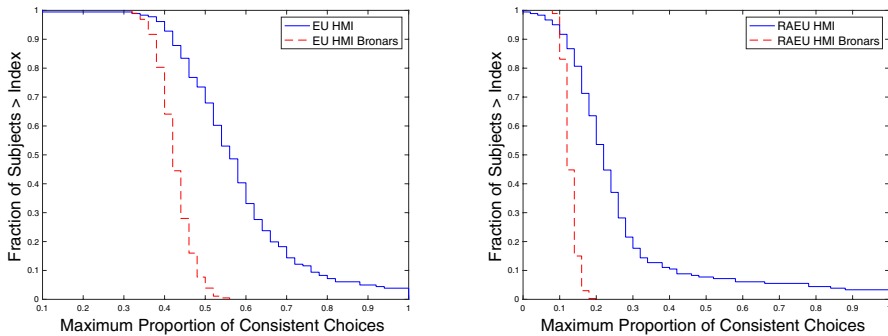


Fig. 3 EU HMI and RAEU HMI in the data from Breig and Feldman (2024)

observation is a small number of subjects that have RAEU HMI indices much lower than those that arise from random datasets.⁴

We show the relationship between various HMI scores and related consistency measures in Fig. 4.⁵ The Critical Cost Efficiency Index (CCEI) provides an alternative measure of distance from consistency—both under general utility maximization and, in the case of EU CCEI, under expected utility. Unlike HMI, which counts the proportion of consistent choices, CCEI captures the magnitude of the most severe violation.⁶ While these measures are correlated, their relationships are imperfect. Researchers should be mindful of the distinct assumptions and limitations of each index when choosing which to apply.

3.2 BRET task

Here, we discuss the bomb risk elicitation task that was first proposed in Crosetto and Filippin (2013). The bomb task provides a visual representation for lottery choices: the subject is offered a set of boxes and can choose how many to open. One box contains a “bomb”, and if opened leads to the subject earning nothing. All other boxes contain a fixed dollar amount that is added to the subject’s payment (conditional on the subject not opening the box with the bomb). Many researchers have adopted this as a way to elicit risk preferences due to its intuitiveness. It is worth restating this bomb elicitation in our notation. In particular, each budget that subjects face has a maximum probability of $m = 1$. The maximum prize M , on the other hand, is equal

⁴The 15 subjects with RAEU HMI scores less than or equal to 0.1 may be better described as being risk tolerant. Among these subjects, the median proportion of choices that were consistent with risk tolerance (i.e. more than 50% of the budget was allocated to the prize) was 38 out of 50.

⁵One may also wonder to what extent a utility function can rationalize behavior that is monotone with respect to first-order stochastic dominance (FOSD). See Breig and Feldman (2024) for a discussion of the distinction between utility rationalization and FOSD rationalization. In the setting of this paper, a dataset can be rationalized by a utility function that respects FOSD if it is consistent with any strictly increasing utility function and contains no choices at endpoints. This makes it straightforward to compute a FOSD HMI index. In the dataset from Breig and Feldman (2024), dominated choices are rare: for 160 out of 181 subjects, the FOSD HMI index equals the HMI index, with an average difference of less than 0.01.

⁶See Breig and Feldman (2024) for how to compute CCEI under linear probability-prize tradeoffs. To our knowledge, there is no known way to compute CCEI under risk-averse expected utility.

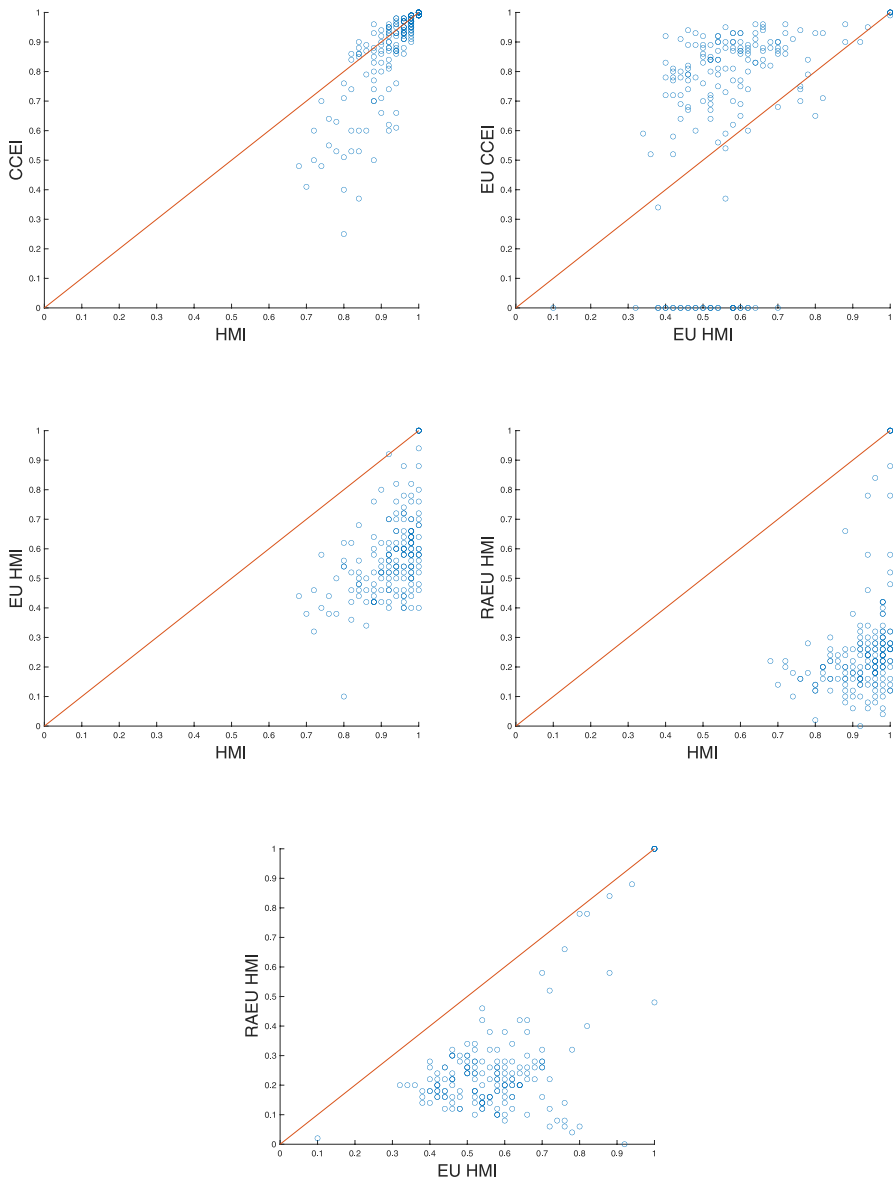


Fig. 4 Relationships between consistency measures. The 45° line is shown in all panels. In the bottom three panels, all observations lie (weakly) below this line by construction, since any set of observations that is consistent with a nested model must also be consistent with its nesting model

to the value of each opened box multiplied by the total number of boxes (usually 100).

While the BRET has become a popular device to measure preferences over risk, experiments rarely use within-repetitions of the BRET task. We use data from the “repeated” treatment used by Crosetto and Filippin (2013), in which 61 subjects

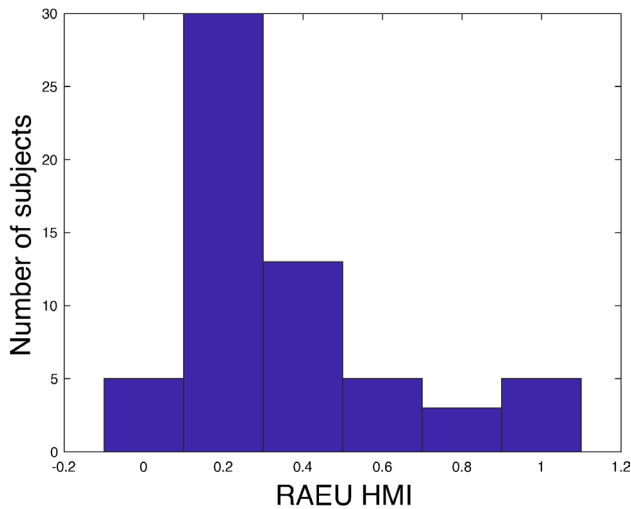


Fig. 5 RAEU HMI in the data from the repeated treatment of Crosetto and Filippin (2013). Each subject's dataset consisted of 5 choices. We report the rationalizable proportion to allow for easier comparison between datasets of different sizes

faced five repetitions with the same number of boxes (100) and value per box (10 Euro cents) across repetitions. As far as we know, no studies have been completed in which subjects face multiple repetitions of the BRET task with the box value differing across treatments.

We can now apply our revealed preference results to the BRET task. As we have already mentioned, the data we use includes five repetitions of the BRET task using the same box value for each repetition. Because the maximal prize does not vary across repetitions, the *only* implication of expected utility in this experiment is that subjects do not choose the “endpoints” of their budget, which in the case of BRET involves opening either no boxes or all of the boxes. Such behavior is extremely rare in the data from Crosetto and Filippin (2013): only one subject chooses an endpoint, and does so on two of their tasks. Thus, sixty subjects have an EU HMI of 1 and one subject has an EU HMI of 0.6.⁷

While expected utility has limited implications for choices without varying maximal budgets, risk-averse expected utility places much stronger restrictions on choice data. In particular, a subject's choices are only consistent with risk-averse expected utility if they involve opening less than half of the boxes and all choices are the same. This is empirically very restrictive for the same reason risk-averse expected utility is restrictive with the Andreoni-Harbaugh task: subjects do not often make the same choice across repetitions. This is shown in Fig. 5, which displays the RAEU HMI for

⁷ If we assume that “random choice” in the BRET tasks corresponds to opening a random number of boxes to open, it is straightforward to see that the HMI EU Bronars is equivalent to one-fifth of a random variable with binomial distribution and parameters $n = 5$ and $p = 99/101$.

the data from Crosetto and Filippin (2013). 35 of the 61 subjects have a RAEU HMI of either 0 or 0.2.⁸ The median RAEU HMI is 0.2.⁹

Comparing the disparate EU HMI results between the data from Crosetto and Filippin (2013) and Breig and Feldman (2024), one may be concerned that these results may be driven by the choice context. For comparison, we refer to the results from Polisson et al. (2020). They examine the rationalizability of choices in a context with two arrow securities and two states—subjects choose how much money to get in state one and how much money to get in state two, subject to a budget constraint that determines the total amount and the relative price between money in state one and two. Specifically, they examine data from Choi et al. (2007), Choi et al. (2014), and Halevy et al. (2018). The proportions of subjects whose choices could be rationalized by EU and RAEU, respectively, were 2/93 and 0/93 in Choi et al. (2007), 12/1,182 and 0/1,182 in Choi et al. (2014), and 59/207 and 18/207 in Halevy et al. (2018). Hence, overall low pass rates for both EU and the stricter RAEU are not uncommon and consistent with other choice contexts. We expect that the high pass rate for EU in the data from Crosetto and Filippin (2013) is purely an artifact of it using the same maximal prize in all budgets.

4 Conclusion

We conclude with a few observations. First, we encourage researchers that adopt the BRET to use multiple repetitions of the task with varied levels of box values. While almost all datasets are rationalizable by expected utility with a single observation, adding a single additional choice will allow for more information regarding the underlying preferences and stronger consistency tests.

Second, our results can be used to compare the BRET and the budget representation of linear probability-prize tradeoffs. Researchers using a single subject pool could fix the budget sets and then compare the degree of consistency with EU or RAEU across different framings. Higher consistency under one frame may suggest that it elicits preferences more effectively. However, we emphasize that this approach *presupposes* that subjects' true preferences are consistent with EU or RAEU, and researchers should carefully consider whether such an assumption is appropriate for their context.

Third, our theoretical and empirical results relate to a broader discussion in the revealed preference literature about the usefulness of various consistency indices

⁸Again under the assumption that random choice in the BRET tasks corresponds to opening a random number of boxes, we find that over 90% of random datasets give a RAEU HMI score of 0.2, with the remaining proportion nearly equally split between scores of 0 and 0.4.

⁹One might notice that the median RAEU HMI is roughly 20% of the total number of budgets in both Breig and Feldman (2024) and Crosetto and Filippin (2013). While this could be mistaken for a behavioral regularity, it more likely reflects a structural feature of the task design: in both datasets, the ratio of unique maximal prizes to total budgets is approximately 0.2. Since RAEU requires that prizes be exactly equal across all budgets included in the consistent set, it is natural that, for most subjects, only one budget from each group sharing the same maximal prize is included. This observation offers useful guidance for future experimental design: when too many linear probability-prize tradeoff budgets share the same maximal prize, it becomes mechanically unlikely that subjects will appear highly consistent with RAEU.

(Echenique et al., 2011; Apesteguia and Ballester, 2015; Allen and Rehbeck, 2021; Echenique, 2021). Our results demonstrate that Houtman-Maks Indices may have some undesirable characteristics when applied to RAEU in linear probability-prize tradeoffs. It is well known that the Houtman-Maks Index does not account for the degree of deviation from consistency—it treats all violations equally, regardless of how close the data is to being consistent (Varian, 1993). Our results show that this issue is amplified when the index is applied to more restrictive models than general utility maximization. In such cases, including risk-averse expected utility under linear probability-prize tradeoffs, a single choice may determine unique point predictions on other budgets. This implies that not only is inconsistency penalized indiscriminately, but consistency also becomes substantially more demanding under these models. For instance, a subject's choices under probability-prize tradeoffs can be arbitrarily close to maximizing expected value, but still have a very low RAEU HMI: a subject in the repeated treatment of Crosetto and Filippin (2013) who opened 46, 47, 48, 49, and 50 boxes would have a risk-averse expected utility HMI of only 0.2, even though the expected value maximizing choice is to open 50 boxes. Thus, some form of continuity in the rationality measure is one potential desideratum of consistency indices.

Proofs

Proof of Theorem 1

Proof of the “if” statement

The proof will be by construction. Suppose that for the regular dataset \mathcal{O} , $0 < x_t < M_t$ for all t and $x_t \geq x_{t-1}$ for all $t \geq 2$. We define

$$u^{\text{EU}}(x) = \begin{cases} \varepsilon x & \text{if } x < x_1 \\ 1 + \varepsilon(x - x_1) & \text{if } x_1 \leq x < x_2 \\ \frac{M_2 - x_1}{M_2 - x_2} + \varepsilon(x - x_2) & \text{if } x_2 \leq x < x_3 \\ \left(\frac{M_3 - x_2}{M_3 - x_3} \right) \left(\frac{M_2 - x_1}{M_2 - x_2} \right) + \varepsilon(x - x_3) & \text{if } x_3 \leq x < x_4 \\ \vdots \\ \left[\prod_{\tau=2}^{T-1} \frac{M_\tau - x_{\tau-1}}{M_\tau - x_\tau} \right] + \varepsilon(x - x_{T-1}) & \text{if } x_{T-1} \leq x < x_T \\ \left[\prod_{\tau=2}^T \frac{M_\tau - x_{\tau-1}}{M_\tau - x_\tau} \right] + \varepsilon(x - x_T) & \text{if } x_T \leq x \end{cases}$$

where the constant ε is taken to be an arbitrarily small but strictly positive number. Where the function u is continuous, its slope is ε .¹⁰ It is discontinuous for x_t such that $x_t > x_{t-1}$, with

$$\begin{aligned}\lim_{x \rightarrow x_t^-} u^{\text{EU}}(x) &= u^{\text{EU}}(x_{t-1}) + \varepsilon(x_t - x_{t-1}) \\ &\approx u^{\text{EU}}(x_{t-1}) \\ &< \frac{m_t - x_{t-1}}{m_t - x_t} u^{\text{EU}}(x_{t-1}) \\ &= u^{\text{EU}}(x_t).\end{aligned}$$

The second line follows from ε being close to 0 and the third line follows from $x_t > x_{t-1}$. Thus, u^{EU} is a strictly increasing piecewise linear function with $u(0) = 0$.

We will show that for all t ,

$$p_t u^{\text{EU}}(x_t) \geq p u^{\text{EU}}(x)$$

for all p and x such that $x + \frac{M_t}{m_t} p = M_t$. It will first be useful to prove a few lemmas that describe the characteristics of u^{EU} .

Lemma 1 *Suppose that conditions (1) and (2) of Theorem 1 hold and that u^{EU} is defined as above. Then for all $t < T$,*

$$(M_t - x_t) u^{\text{EU}}(x_t) \geq (M_t - x_{t+1}) u^{\text{EU}}(x_{t+1})$$

and for all $t > 1$.

$$(M_t - x_t) u^{\text{EU}}(x_t) \geq (M_t - x_{t-1}) u^{\text{EU}}(x_{t-1})$$

Proof Suppose that the first inequality stated in the lemma did not hold. Then

$$(M_t - x_t) u^{\text{EU}}(x_t) < (M_t - x_{t+1}) u^{\text{EU}}(x_{t+1}).$$

By the definition of u^{EU} , this implies that

$$(M_t - x_t) < (M_t - x_{t+1}) \left(\frac{M_{t+1} - x_t}{M_{t+1} - x_{t+1}} \right).$$

But this implies that $(M_t - x_t)(M_{t+1} - x_{t+1}) < (M_t - x_{t+1})(M_{t+1} - x_t)$, which simplifies to $(M_t - M_{t+1})(x_t - x_{t+1}) < 0$, a contradiction of the maintained assumptions that \mathcal{O} is regular and $x_t \geq x_{t-1}$ for all $t \geq 2$.

¹⁰The result can also be established under the assumption of a continuous utility function, although the resulting characterization is more complex. Full details are available from the authors upon request.

Similarly, suppose that

$$(M_t - x_t)u^{\text{EU}}(x_t) < (M_t - x_{t-1})u^{\text{EU}}(x_{t-1}).$$

Then using the definition of u^{EU} , we have that

$$(M_t - x_t) \left(\frac{M_t - x_{t-1}}{M_t - x_t} \right) < (M_t - x_{t-1}),$$

but this obviously holds with equality. \square

Lemma 2 *Suppose that conditions (1) and (2) of Theorem 1 hold and that u^{EU} is defined as above. Then for all t and t' ,*

$$(M_t - x_t)u^{\text{EU}}(x_t) \geq (M_t - x_{t'})u^{\text{EU}}(x_{t'}).$$

Proof By Lemma 1, we have that

$$(M_t - x_t)u^{\text{EU}}(x_t) \geq (M_t - x_{t+1})u^{\text{EU}}(x_{t+1})$$

for all $t < T$. Thus,

$$\begin{aligned} (M_t - x_t)u^{\text{EU}}(x_t) &\geq (M_t - x_{t+1})u^{\text{EU}}(x_{t+1}) \\ &= (M_{t+1} - x_{t+1})u^{\text{EU}}(x_{t+1}) + (M_t - M_{t+1})u^{\text{EU}}(x_{t+1}) \\ &\geq (M_{t+1} - x_{t+2})u^{\text{EU}}(x_{t+2}) + (M_t - M_{t+1})u^{\text{EU}}(x_{t+1}) \\ &= (M_t - x_{t+2})u^{\text{EU}}(x_{t+2}) + (M_t - M_{t+1})u^{\text{EU}}(x_{t+1}) + (M_{t+1} - M_t)u^{\text{EU}}(x_{t+2}) \\ &= (M_t - x_{t+2})u^{\text{EU}}(x_{t+2}) + (M_{t+1} - M_t)(u^{\text{EU}}(x_{t+2}) - u^{\text{EU}}(x_{t+1})) \\ &\geq (M_t - x_{t+2})u^{\text{EU}}(x_{t+2}). \end{aligned}$$

The first and second inequality follow from the fact that, as stated in Lemma 1,

$$(M_t - x_t)u^{\text{EU}}(x_t) \geq (M_t - x_{t+1})u^{\text{EU}}(x_{t+1})$$

for all $t < T$. The third inequality follows from the facts that $M_{t+1} \geq M_t$, $x_{t+2} \geq x_{t+1}$, and u^{EU} is increasing. This reasoning can be iterated to give the result for all $t' > t$.

Also by Lemma 1, we have that

$$(M_t - x_t)u^{\text{EU}}(x_t) \geq (M_t - x_{t-1})u^{\text{EU}}(x_{t-1}).$$

Similar reasoning as above gives us that

$$\begin{aligned}
(M_t - x_t)u^{\text{EU}}(x_t) &\geq (M_t - x_{t-1})u^{\text{EU}}(x_{t-1}) \\
&= (M_{t-1} - x_{t-1})u(x_{t-1}) + (M_t - M_{t-1})u(x_{t-1}) \\
&\geq (M_{t-1} - x_{t-2})u(x_{t-2}) + (M_t - M_{t-1})u(x_{t-1}) \\
&= (M_t - x_{t-2})u(x_{t-2}) + (M_t - M_{t-1})u(x_{t-1}) + (M_{t-1} - M_t)u(x_{t-2}) \\
&= (M_t - x_{t-2})u(x_{t-2}) + (M_t - M_{t-1})(u(x_{t-2}) - u(x_{t-1})) \\
&\geq (M_t - x_{t-2})u(x_{t-2}).
\end{aligned}$$

This again can be iterated, so the result holds for all $t' < t$. \square

Lemma 3 Suppose that conditions (1) and (2) of Theorem 1 hold and that u^{EU} is defined as above. Then for all t ,

$$(M_t - x_t)u^{\text{EU}}(x_t) \geq (M_t - x)u^{\text{EU}}(x).$$

Proof Notice that $(M_t - x)u^{\text{EU}}(x)$ is differentiable if $x \neq x_{\hat{t}}$ for any \hat{t} . If $x > x_1$ and we define t' as the largest value such that $x_{t'} \leq x$, we have that the derivative is

$$\varepsilon(M_t + x_{t'} - 2x) - u(x_{t'}),$$

which is always negative for ε small enough. Thus, in this case,

$$\begin{aligned}
(M_t - x)u^{\text{EU}}(x) &< (M_t - x_{t'})u^{\text{EU}}(x_{t'}) \\
&\leq (M_t - x_t)u^{\text{EU}}(x_t),
\end{aligned}$$

where the second inequality follows from Lemma 2. If instead $x < x_1$,

$$(M_t - x)u^{\text{EU}}(x) = \varepsilon(M_t x - x^2)$$

which has an upper bound that approaches zero as ε approaches 0. \square

We can now complete the proof of the “if” statement. Suppose that there were a t, x , and p such that

$$p_t u^{\text{EU}}(x_t) < p u^{\text{EU}}(x)$$

where $x + \frac{M_t}{m_t}p = M_t$. Then it must be the case that

$$\frac{m_t}{M_t}(M_t - x_t)u^{\text{EU}}(x_t) \geq \frac{m_t}{M_t}(M_t - x)u^{\text{EU}}(x)$$

and

$$(M_t - x_t)u^{\text{EU}}(x_t) \geq (M_t - x)u^{\text{EU}}(x).$$

However, this contradicts Lemma 3, so

$$p_t u(x_t) \geq p u(x)$$

for all p and x such that $x + \frac{M_t}{m_t} p = M_t$. Thus, \mathcal{O} is rationalized by expected utility.

Proof of the “only if” statement

First, suppose that $x_t = 0$ or $x_t = M_t$. Then \mathcal{O} cannot be rationalized by expected utility because if u is strictly increasing with $u(0) = 0$, then $p_t u(x_t) = 0$, which is strictly less than any other bundle on the interior of the budget line.

Next, suppose that \mathcal{O} was rationalized by expected utility. Then for all t and t' ,

$$(M_t - x_t)u(x_t) \geq (M_t - x_{t'})u(x_{t'})$$

and

$$(M_{t'} - x_t)u(x_t) \leq (M_{t'} - x_{t'})u(x_{t'}).$$

But together, these imply that

$$(M_t - M_{t'})u(x_t) \geq (M_t - M_{t'})u(x_{t'})$$

and $(M_t - M_{t'})(u(x_t) - u(x_{t'})) \geq 0$. Because u is required to be strictly increasing, this implies that $x_t \geq x_{t'}$ if $M_t > M_{t'}$. Thus, for a regular dataset, $x_t \geq x_{t-1}$ is a necessary condition for \mathcal{O} to be rationalized by expected utility.

Proof of Theorem 2

First, notice that if a concave utility function \hat{u} that rationalizes \mathcal{O} exists, then there also exists a concave utility function u that rationalizes \mathcal{O} where $u(x_t) = \hat{u}(x_t)$ for all t and $u(x_t)$ is piecewise linear. In this case, u gives the same utility for all of the chosen options but lower utility for any of the unchosen options. Thus, this proof will focus on whether there exists a piecewise linear utility function u that rationalizes \mathcal{O} and only has kinks at the values of x that are chosen.

Proof of the “if” statement

The proof will be by construction. We define

$$u^{\text{RAEU}}(x) = \begin{cases} \frac{x}{x_1} & \text{if } x \leq x_1 \\ \frac{M_2 - 2x_2 + x}{M_2 + x_1 - 2x_2} & \text{if } x_1 < x \leq x_2 \\ \left(\frac{M_3 - 2x_3 + x}{M_3 + x_2 - 2x_3} \right) \left(\frac{M_2 - x_2}{M_2 + x_1 - 2x_2} \right) & \text{if } x_2 < x \leq x_3 \\ \vdots & \\ \left(\frac{M_k - 2x_k + x}{M_k + x_{k-1} - 2x_k} \right) \left(\prod_{i=2}^{k-1} \frac{M_i - x_i}{M_i + x_{i-1} - 2x_i} \right) & \text{if } x_{k-1} < x \leq x_k \\ \vdots & \\ \left(\frac{M_T - 2x_T + x}{M_T + x_{T-1} - 2x_T} \right) \left(\prod_{i=2}^{T-1} \frac{M_i - x_i}{M_i + x_{i-1} - 2x_i} \right) & \text{if } x_{T-1} < x \leq x_T \\ \prod_{i=2}^{T-1} \frac{M_i - x_i}{M_i + x_{i-1} - 2x_i} + \varepsilon(x - x_T) & \text{if } x_T < x \end{cases}$$

where the constant ε is taken to be an arbitrarily small but strictly positive number. Notice that this is a continuous piecewise linear function with $u^{\text{RAEU}}(0) = 0$. u^{RAEU} only has kinks where $x = x_t$ for some t .

First, we show that this utility function is everywhere increasing. Because both x_t and ε are positive, u^{RAEU} is strictly increasing below x_1 and above x_T . Now consider the range (x_{t-1}, x_t) for some $t > 1$. On this range, the slope of u^{RAEU} is

$$\left(\frac{1}{M_t + x_{t-1} - 2x_t} \right) \left(\prod_{i=2}^{t-1} \frac{M_i - x_i}{M_i + x_{i-1} - 2x_i} \right).$$

Obviously, $M_i - x_i$ is strictly positive for all i . We can also note that if $M_t - M_{t-1} \geq 2(x_t - x_{t-1})$, then

$$\begin{aligned} M_t - 2x_t + x_{t-1} &\geq M_{t-1} - x_{t-1} \\ &> 0. \end{aligned}$$

Thus, $M_t + x_{t-1} - 2x_t$ is strictly positive for all t , and the slope must be positive.

Next we show that u^{RAEU} is concave. Because the function is piecewise linear and continuous, we only must show that the slope decreases at the kinks. Because $M_t - M_{t-1} \geq 2(x_t - x_{t-1})$ for all t and $x_1 \leq \frac{1}{2}M_1$, we have that $0 \leq M_1 - 2x_1 \leq M_2 - 2x_2$. But this implies that $M_2 + x_1 - 2x_2 \geq x_1$ and

$$\frac{1}{x_1} \geq \frac{1}{M_2 + x_1 - 2x_2}.$$

Thus, the slope decreases at x_1 . Next note that the slope on the interval (x_{t-1}, x_t) is

$$\left(\frac{1}{M_t + x_{t-1} - 2x_t} \right) \left(\prod_{i=2}^{t-1} \frac{M_i - x_i}{M_i + x_{i-1} - 2x_i} \right)$$

while the slope in the interval (x_t, x_{t+1}) is

$$\frac{M_t - x_t}{M_{t+1} + x_t - 2x_{t+1}} \left(\frac{1}{M_t + x_{t-1} - 2x_t} \right) \left(\prod_{i=2}^{t-1} \frac{M_i - x_i}{M_i + x_{i-1} - 2x_i} \right).$$

Because $M_{t+1} - M_t \geq 2(x_{t+1} - x_t)$, we have that $M_{t+1} + x_t - 2x_{t+1} \geq M_t - x_t$, so

$$\frac{M_t - x_t}{M_{t+1} + x_t - 2x_{t+1}} \leq 1.$$

This implies that the slope decreases at x_t for all $t < T$. Finally, because ε is taken to be arbitrarily small, the slope decreases at x_T .

Next, we show that for all t , $p_t u^{\text{RAEU}}(x_t) \geq p u^{\text{RAEU}}(x)$ for all p and x such that $x + \frac{M_t}{m_t} p = M_t$. Note that when we substitute in the budget constraint for p , we find that the concavity of u^{RAEU} implies that

$$\frac{m_t}{M_t} (M_t - x) u^{\text{RAEU}}(x)$$

is concave in x . Thus, to determine whether x_t maximizes this function, we only need to check whether it maximizes the function locally.

For $1 < t < T$, on the interval (x_{t-1}, x_t) , the derivative is

$$\frac{m_t}{M_t} \left(\prod_{i=2}^{t-1} \frac{M_i - x_i}{M_i + x_{i-1} - 2x_i} \right) \left(\frac{2(x_t - x)}{M_t + x_{t-1} - 2x_t} \right),$$

which is always positive. For $1 < t < T$, on the interval (x_t, x_{t+1}) , the derivative is

$$\frac{m_t}{M_t} \left(\prod_{i=2}^t \frac{M_i - x_i}{M_i + x_{i-1} - 2x_i} \right) \left(\frac{M_t - M_{t+1} + 2(x_{t+1} - x)}{M_{t+1} + x_t - 2x_{t+1}} \right).$$

This is always negative, because the maintained assumption that $M_{t+1} - M_t \geq 2(x_{t+1} - x_t)$ implies that $0 \geq M_t - M_{t+1} + 2(x_{t+1} - x_t)$. Thus, for $1 < t < T$, x_t maximizes

$$\frac{m_t}{M_t} (M_t - x) u^{\text{RAEU}}(x).$$

Finally, we consider the cases of $t = 1$ and $t = T$. Using the same computations as above, we can see that

$$\frac{m_T}{M_T}(M_T - x)u^{\text{RAEU}}(x)$$

is increasing on the interval (x_{T-1}, x_T) . For $x > x_T$, $u^{\text{RAEU}}(x) \approx 0$, so x_T must be the maximizer. Again using the computations from above, we can see that

$$\frac{m_1}{M_1}(M_1 - x)u^{\text{RAEU}}(x)$$

is decreasing on the interval (x_1, x_2) . For $x < x_1$, the derivative is

$$\frac{m_t}{M_t} \left(\frac{M_t - 2x}{x_1} \right),$$

so on that interval the derivative is always positive if $x_1 \leq \frac{1}{2}M_1$, which is the maintained assumption.

Proof of the “only if” statement

Suppose that some strictly increasing piecewise linear concave function u with kinks only at values of x_t rationalizes \mathcal{O} . We will show that if any of the conditions from the theorem are violated, then there is a contradiction.

First, suppose that $x_1 > \frac{1}{2}M_1$. But if u rationalizes \mathcal{O} , then it must be the case that

$$\frac{m_1}{M_1}(M_1 - x_1)u(x_1) \geq \frac{m_1}{M_1} \left(M_1 - \frac{1}{2}M_1 \right) u \left(\frac{1}{2}M_1 \right).$$

But because u is piecewise linear with no kinks between 0 and x_1 , this implies that

$$(M_1 - x_1)u(x_1) \geq \frac{1}{2}M_1 \left(\frac{M_1}{2x_1} \right) u(x_1)$$

and

$$(M_1 - x_1)x_1 \geq \frac{1}{4}M_1^2.$$

But this only holds for $x_1 = \frac{1}{2}M_1$, which is a contradiction.

Next, suppose that $x_t = 0$ for some t . If this were the case, then $p_t u(x_t) = 0$, which is strictly less than $pu(x)$ for any choice of p and x on the interior of the budget line. Thus, there is a contradiction.

The fact that $x_t - x_{t-1} \geq 0$ is a necessary condition for \mathcal{O} to be rationalized by risk-averse expected utility follows immediately from Theorem 1.

Finally, suppose that

$$M_t - M_{t-1} < 2(x_t - x_{t-1}).$$

Because u is linear on the interval from x_{t-1} to x_t , we know that on that interval,

$$u(x) = u(x_{t-1}) + \left(\frac{x - x_{t-1}}{x_t - x_{t-1}} \right) (u(x_t) - u(x_{t-1})).$$

Thus, u rationalizing \mathcal{O} implies that

$$\frac{M_{t-1}}{m_{t-1}}(M_{t-1} - x) \left[u(x_{t-1}) + \left(\frac{x - x_{t-1}}{x_t - x_{t-1}} \right) (u(x_t) - u(x_{t-1})) \right]$$

is weakly decreasing at $x = x_{t-1}$, which implies that

$$u(x_t) \leq \frac{M_{t-1} + x_t - 2x_{t-1}}{M_{t-1} - x_{t-1}} u(x_{t-1}).$$

Similarly, u rationalizing \mathcal{O} must also imply that

$$\frac{M_t}{m_t}(M_t - x) \left[u(x_{t-1}) + \left(\frac{x - x_{t-1}}{x_t - x_{t-1}} \right) (u(x_t) - u(x_{t-1})) \right]$$

must be weakly increasing at $x = x_t$, which implies that

$$u(x_t) \geq \frac{M_t - x_t}{M_t - 2x_t + x_{t-1}} u(x_{t-1}).$$

We can combine these inequalities and simplify to find that

$$(M_t - M_{t-1} - 2(x_t - x_{t-1}))(x_t - x_{t-1}) \geq 0.$$

Thus, if $2(x_t - x_{t-1}) > M_t - M_{t-1}$, there is a contradiction.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. Part of the data used in this paper was gathered using funds from the National Science Foundation's dissertation grant #1824121.

Declarations

Conflict of interest The authors have no other relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allen, R., & Rehbeck, J. (2021). Measuring rationality: Percentages vs expenditures. *Theory and Decision*, 91(2), 265–277.
- Andreoni, J., & Harbaugh, W. (2009). Unexpected utility: Five experimental tests of preferences for risk. *Unpublished Manuscript*.
- Andreoni, J., & Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2), 737–753.
- Apestequia, J., & Ballester, M. A. (2015). A measure of rationality and welfare. *Journal of Political Economy*, 123(6), 1278–1310.
- Border, K. C. (1992). Revealed preference, stochastic dominance, and the expected utility hypothesis. *Journal of Economic Theory*, 56(1), 20–42.
- Breig, Z., & Feldman, P. (2024). Revealing risky mistakes through revisions. *Journal of Risk and Uncertainty*, 68, 1–28.
- Bronars, S. G. (1987). The power of nonparametric tests of preference maximization. *Econometrica*, 693–698.
- Burghart, D. R. (2020). The two faces of independence: Betweenness and homotheticity. *Theory and Decision*, 88(4), 567–593.
- Chambers, C. P., Liu, C., & Martinez, S.-K. (2016). A test for risk-averse expected utility. *Journal of Economic Theory*, 163, 775–785.
- Cherchye, L., Demuynck, T., Rock, B., & Freer, M. (2021). Revealed preference analysis of expected utility maximization under prize-probability trade-offs. *Journal of Mathematical Economics*, 99, 102607.
- Choi, S., Fisman, R., Gale, D., & Kariv, S. (2007). Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*, 97(5), 1921–1938.
- Choi, S., Kariv, S., Müller, W., & Silverman, D. (2014). Who is (more) rational? *American Economic Review*, 104(6), 1518–1550.
- Crosetto, P., & Filippin, A. (2013). The “bomb” risk elicitation task. *Journal of Risk and Uncertainty*, 47(1), 31–65.
- Dekel, E., Lipman, B. L., & Rustichini, A. (2009). Temptation-driven preferences. *The Review of Economic Studies*, 76(3), 937–971.
- Dembo, A., Kariv, S., Polisson, M., & Quah, J. K.H. (2021). *Ever since allais*. Technical report, IFS Working Paper.
- Echenique, F. (2021). On the meaning of the critical cost efficiency index. [arXiv:2109.06354](https://arxiv.org/abs/2109.06354)
- Echenique, F., Imai, T., & Saito, K. (2023). Approximate expected utility rationalization. *Journal of the European Economic Association*, 21(5), 1821–1864.
- Echenique, F., Lee, S., & Shum, M. (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6), 1201–1223.
- Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4), 281–295.
- Feldman, P., & Rehbeck, J. (2022). Revealing a preference for mixtures: An experimental study of risk. *Quantitative Economics*, 13(2), 761–786.
- Feldman, P.J., & Ferraro, P.J. (2023). Certain Reversals. Working paper.
- Fishburn, P. C. (1975). Separation theorems and expected utilities. *Journal of Economic Theory*, 11(1), 16–34.
- Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2), 631–645.
- Green, R. C., & Srivastava, S. (1986). Expected utility maximization and demand behavior. *Journal of Economic Theory*, 38(2), 313–323.
- Gul, F., & Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69(6), 1403–1435.
- Halevy, Y., Persitz, D., & Zrill, L. (2018). Parametric recoverability of preferences. *Journal of Political Economy*, 126(4), 1558–1593.

- Harbaugh, W. T., Mocan, N., & Visser, M. S. (2013). Theft and deterrence. *Journal of Labor Research*, 34, 389–407.
- Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644–1655.
- Houtman, M., & Maks, J. (1985). Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve methoden*, 19(1), 89–104.
- Kim, T. (1996). Revealed preference theory on the choice of lotteries. *Journal of Mathematical Economics*, 26(4), 463–477.
- Payró, F. (2019). Revealed preference and the subjective state space hypothesis. *Journal of Mathematical Economics*, 82, 61–68.
- Polisson, M., Quah, J.K.-H., & Renou, L. (2020). Revealed preferences over risk and uncertainty. *American Economic Review*, 110(6), 1782–1820.
- Varian, H.R. (1993). Goodness-of-fit for revealed preference tests. Technical report.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.