Forthcoming in the JEEA.

Approximate Expected Utility Rationalization

Federico Echenique Taisuke Imai Kota Saito *

November 16, 2022

Abstract

We propose a new measure of deviations from expected utility theory. For any positive number e, we give a characterization of the datasets with a rationalization that is within e (in beliefs, utility, or perceived prices) of expected utility theory, under the assumption of risk aversion. The number e can then be used as a measure of how far the data is to expected utility theory. We apply our methodology to data from three large-scale experiments. Many subjects in these experiments are consistent with utility maximization, but not with expected utility maximization. Our measure of distance to expected utility is correlated with the subjects' demographic characteristics.

*Echenique: Department of Economics, University of California, Berkeley, fede@econ.berkeley.edu. Imai: Department of Economics, LMU Munich, taisuke.imai@econ.lmu.de. Saito: Division of the Humanities and Social Sciences, California Institute of Technology, saito@caltech.edu. We are very grateful to Nicola Persico, who posed questions to us that led to some of the results in this paper, and to Jose Apesteguía, Miguel Ballester, Geoffroy de Clippel, Dan Friedman, Yves Le Yaouanq, Pietro Ortoleva, Matthew Polisson, John Quah, and Kareen Rozen for helpful comments. We are also grateful for the feedback provided by numerous audiences at FUR 2018, ESA World Meetings 2018, 2018 European Summer Meeting of the Econometric Society, CESifo Area Conference on Behavioural Economics 2018, Measuring Individual Well-Being Workshop, and 2019 European Summer Symposium in Economic Theory. This research is supported by Grant SES-1558757 from the National Science Foundation. The authors also acknowledge financial support by the NSF through the grants CNS-1518941 (Echenique) and SES-1919263 (Saito), and the Deutsche Forschungsgemeinschaft through CRC TRR 190 (Imai).

1 Introduction

Revealed preference theory has traditionally, through its 80-year history, dealt with the empirical content of *general* utility maximization. The idea is to understand which observed choice behaviors are consistent with the hypothesis of utility maximization. Recent research has, in contrast, turned to the empirical content of *specific* utility theories. Mostly the focus has been on expected utility (EU): EU is the workhorse model of economics under risk and uncertainty, and recent theoretical work seeks to characterize the observable choice behaviors that are consistent with expected utility maximization. The EU hypothesis has been subject to substantial empirical scrutiny, but mainly in experiments targeted to test its axiomatic foundations. Empirical work using market data to test EU is much less common. At the same time, a number of recent empirical revealed preference studies use data on choices under risk and uncertainty, in which subjects choose state-contingent consumption from budget sets, and the state is unknown. These studies have focused on testing for "rationality," or consistency with general utility maximization, but we shall argue that the same data can be used to test for the more narrow theory of EU maximization. In our paper, we seek to bridge the gap between the theoretical understanding of expected utility theory, and the machinery needed to analyze experimental data on choices under risk and uncertainty, so as to use the data on choices in market settings to test for consistency with EU maximization.

Imagine an agent making economic decisions, choosing contingent consumption given market prices and income. Revealed preference theory studies the consistency of such choices with utility maximization. The best-known result in the literature says that an agent's behavior is consistent if and only if it satisfies the Generalized Axiom of Revealed Preference, GARP. Consistency, however, is a black-or-white question. The agent's choices either satisfy GARP or do not. Our paper is about EU and about shades of grey. Our contribution is to describe the degree to which choices are consistent with EU. We propose a *measure* of the degree of a choice dataset's consistency with EU, and we use our measure to analyze several large datasets on choices under risk.

Consistency with GARP is a black-or-white question, but revealed preference theory has developed measures of the degree of consistency with general utility maximization. The most widely used measure is the Critical Cost Efficiency Index (CCEI) proposed by Afriat (1972). The basic idea in the CCEI is to fictitiously decrease an agent's budget so that fewer options are revealed preferred to a given choice. The CCEI has been widely used to analyze experimental data on choices from budget sets. See, for example, Choi et al. (2007), Ahn et al. (2014), Choi et al. (2014), Carvalho et al. (2016), Carvalho and Silverman (2019), and Halevy et al. (2018). All of these experimental studies involve subjects making decisions under risk or uncertainty, and

CCEI was proposed as a measure of consistency with general utility maximization, not EU, the most commonly-used theory to explain choices under risk or uncertainty.[1]

Of course, there is nothing wrong with studying general utility maximization in environments with risk and uncertainty, but the data is ideally suited to studying theories of choice under risk and uncertainty, and it should be of great interest to evaluate EU using this data. We shall argue (on both theoretical and empirical grounds) that our method provides a more accurate and intuitive measure of consistency with EU than using CCEI.

Our main contribution is to propose a measure of how far a dataset is from being consistent with EU, under the assumption of risk aversion. The measure is different from CCEI: we explain theoretically why our measure, and not CCEI, best captures the distance of a dataset to EU theory. The key insight is that our measure captures perturbations, or errors, that are reflected in deviations from the first-order conditions for optimizing behavior. In this sense, our paper is about "approximate" EU rationalizations. Such perturbations can be (equivalently) interpreted as random utility, miss-perceived or miss-measured prices, as well as incorrect beliefs. Indeed our measure captures the size of the perturbations that would be needed to reconcile the observed choice data with EU.

We also argue on empirical grounds that our measure passes "smell tests" that CCEI fails. For example, CCEI ignores the manifest violations of EU where subjects make first-order stochastically dominated choices. And CCEI does not correlate well with the property of downward-sloping demand, a property that is implied by EU maximization. Roughly speaking, prices and quantities must be inversely related, subject to certain qualifications. Such a property can be seen to characterize consistency with EU in the risk-averse case. We also provide a revealed preference axiomatization of the measure based on observed prices and consumption.

In this paper, we first lay out the implications of EU that cannot be captured by CCEI, and give an overview of our approach. After a theoretical discussion of our measure of consistency (with objective EU discussed in Section 3 and subjective EU in Online Appendix B), we present an empirical application using data from experiments on choices under risk (Section 4).

Our empirical application has two purposes. The first is to illustrate how our method can be applied and to argue that our measure of distance to EU is useful and sensible. The second is to offer new insights into the empirical validity of EU. We use data from three large-scale experiments (Choi et al., 2014; Carvalho et al., 2016; Carvalho and Silverman, 2019), each with over 1,000 subjects, that involve choices under risk. Consistency with general utility maximization is well understood in these studies using CCEI. We test for EU theory using our methodology and

[1] See Allen and Rehbeck (2021), Echenique (2021), and Polisson and Quah (2022) for a general discussion of the CCEI, irrespective of the question of testing for EU that is the focus of the present paper.

obtain some important findings.

There are three main conclusions from our empirical application. The first is that there is a sizable gap between consistency with general utility maximization and consistency with expected utility. Many subjects are utility maximizers, or display a CCEI that is close to one, but are far from consistent with EU as measured by our proposed distance to EU rationality.[2] Indeed, subjects with a CCEI that is close to one exhibit varying degrees of consistency with EU. If one were to estimate, for example, a risk-averse utility from subjects with a high CCEI, then the exercise would be misspecified. We show that subjects who are close to being consistent with EU according to our measure, tend to have smaller estimated coefficients of risk aversion than subjects who are far from being consistent with EU. This means that if one estimates risk aversion from a (potentially misspecified) EU model, the subjects who are estimated to be less risk averse are closer (according to our measure) to EU than those who are estimated to be more risk averse. No such correlation is found with CCEI.

The second main conclusion is that a simple and intuitive "downward-sloping demand" property is strongly related to distance to EU. When consumption in each state reacts negatively to state prices (actually risk-neutral prices), then subjects are typically much closer to being consistent with EU than when this property is violated. Another, perhaps more expected, feature of the data is that subjects who exhibit some kind of violation of monotonicity with respect to first-order stochastic dominance are far from being consistent with EU.

The third conclusion emerges from looking at the correlation between demographic characteristics and consistency with EU. We find that younger subjects, those who have high cognitive abilities, and those who are working, are closer to EU behavior than older, low cognitive ability, or non-working, subjects. For some of the three experiments, we also find that highly educated, high-income, and male subjects, are closer to EU. These observations suggest that our measure complements CCEI as an empirical toolkit and provides additional insights on datasets that had been analyzed primarily with CCEI.

1.1 How to Measure Deviations from EU

The CCEI is meant to test deviations from general utility maximization. If an agent's behavior is not consistent with utility maximization, then it cannot possibly be consistent with EU maximization. Thus it stands to reason that if an agent's behavior is far from being rationalizable as measured by CCEI, then it is also far from being rationalizable with an EU function. The problem is, of course, that an agent's behavior may be rationalizable with a general utility function but not with EU.

[2] Such a comparison is subjective but seems often rather clear-cut.

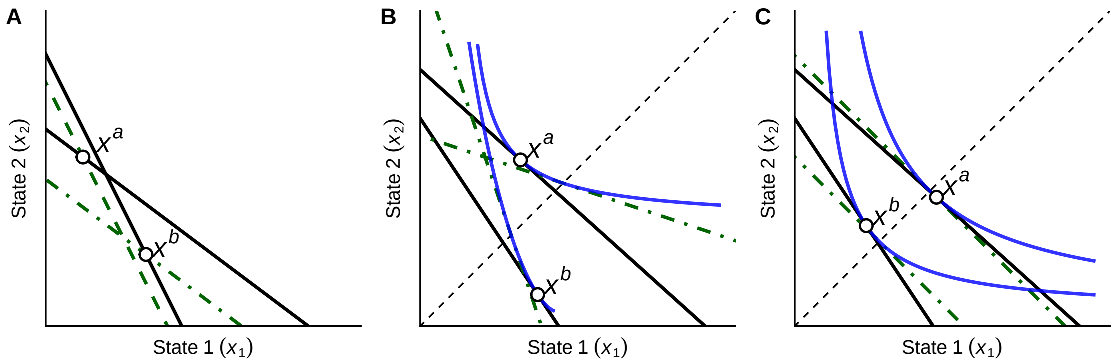

FIGURE 1: (A) A violation of WARP. (B) A violation of EU: $x_2^a > x_1^a$, $x_1^b > x_2^b$, and $p_1^b/p_2^b < p_1^a/p_2^a$. (C) A choice pattern consistent with EU.

Broadly speaking, the CCEI proceeds by "amending" inconsistent choices through the device of changing income. This works for general utility maximization, but it is the wrong way to amend choices that are inconsistent with EU. Since EU is about getting marginal rates of substitution right, prices, not incomes, need to be changed. The problem is illustrated with a simple example in Figure 1.

Suppose that there are two states of the world, labeled 1 and 2. An agent purchases a state-contingent asset $x = (x_1, x_2)$ given Arrow-Debreu prices $p = (p_1, p_2)$ and her income. Prices and income define a budget set. In Figure 1A, we are given two choices for the agent, x^a and x^b, for two different budgets. The choices in Figure 1A are inconsistent with utility maximization: they violate the weak axiom of revealed preference (WARP). When x^b (x^a) was chosen, x^a (x^b, respectively) was strictly inside of the budget set. This violation of WARP can be resolved by shifting down the budget line associated with choice x^b to the dashed line passing through x^a. Alternatively, the violation can be resolved by shifting down the budget line associated with choice x^a to the dash-dotted line passing through x^b. CCEI is the smallest of the two shifts that are needed: the smallest proportion of shifting down a budget line to resolve WARP violation. Therefore, the CCEI of this dataset is given by the dashed line passing through x^a. That is, the CCEI is $(p^b \cdot x^a)/(p^b \cdot x^b)$.

Now consider the example in Figure 1B. There are again two choices, x^a and x^b, for two different budgets. These choices do not violate WARP and comply with the theory of utility maximization with CCEI = 1. The choices in the panel are *not*, however, compatible with EU. To see why, assume that the dataset were rationalized by an expected utility: $\mu_1 u(x_1^k) + \mu_2 u(x_2^k)$, where (μ_1, μ_2) are the probabilities of the two states, and u is a (smooth) concave utility function over money. Note that the slope of a tangent line to the indifference curve at a point x^k is equal to the marginal rate of substitution (MRS): $\mu_1 u'(x_1^k)/\mu_2 u'(x_2^k)$. Moreover, at the 45-degree line (i.e., when $x_1^k = x_2^k$), the slope must be equal to $\mu_1 \cancel{u'(x_1^k)}/\mu_2 \cancel{u'(x_2^k)} = \mu_1/\mu_2$. This is a contradiction because in Figure 1B, the two tangent lines (dash-dotted lines) associated with x^a and x^b cross each other. The source of the problem lies in which observations are below, and which are above the 45-degree line. Note that x^a is above the diagonal, which means that the MRS at the 45-degree line must be flatter than the budget line; x^b is below the diagonal, which further means that the MRS at the 45-degree line must be steeper than the budget line. The latter budget line is steeper than the former, which is inconsistent with MRS being the same on any point on the 45-degree line. Finally, Figure 1C shows an example of choices that are consistent with EU. Note that tangent lines at the 45-degree line are parallel in this case. Here the choice of x^a below the diagonal is made at the flatter budget set, while x^b above the diagonal is at the steeper budget line. In consequence, the choices in Figure 1C are consistent with the EU.

Importantly, the violation in Figure 1B cannot be resolved by shifting budget lines up or down, or more generally by adjusting the agent's expenditures. The reason is that *the empirical content of expected utility is captured by the relation between prices and marginal rates of substitution. The slope, not the level, of the budget line, is what matters.* The basic insight comes from the equality of marginal rates of substitution and relative prices:

$$\frac{\mu_1 u'(x_1^k)}{\mu_2 u'(x_2^k)} = \frac{p_1^k}{p_2^k}. \tag{1}$$

Since marginal utility is decreasing, equation (1) imposes a negative relation between prices and quantities. The distance to EU is directly related to how far the data is to complying with such a negative relation between prices and quantities. The formal connection is established in Theorem 2. Empirically, as we shall see, the degree of compliance of a subject's choices with this "downward-sloping demand" property, goes a long way to capturing the degree of compliance of the subject's choices with EU.

We propose a measure of how close the data is to being consistent with EU maximization. Our measure is based on the idea that marginal rates of substitution have to conform to EU maximization, i.e., whether data conform to equation (1). If one "perturbs" marginal utility enough, then a dataset is always consistent with expected utility. Our measure is simply a measure of how large of a perturbation is needed to rationalize the data. Perturbations of marginal utility can be interpreted in three different, but equivalent, ways: as measurement error on prices, as random shocks to marginal utility in the fashion of random utility theory (McFadden, 1974), or as perturbations to agents' beliefs. For example, if the data in Figure 1B is "e away" from being consistent with expected utility given a positive number e, then one can find beliefs μ^a and μ^b, one for each observation so that EU is maximized for these observation-specific beliefs, and the degree of perturbation of beliefs is bounded by e.

Our measure can be applied in settings where probabilities are known and objective, for which we develop a theory in Section 3, and an application to experimental data in Section 4. It can also be applied to settings where probabilities are not known, and therefore subjective (Online Appendix B).

Finally, we propose a statistical methodology for testing the null hypothesis of consistency with EU (Section 4.3). Our test relies on a set of auxiliary assumptions. The test indicates moderate levels of rejection of the EU hypothesis.

1.2 Related Literature

Revealed preference theory has developed tests for consistency with general utility maximization. A test, in this literature, provides a yes or no answer to the question of whether the data is consistent with utility maximization. The seminal papers include Samuelson (1938), Afriat (1967), and Varian (1982). See Chambers and Echenique (2016) for an exposition of the basic theory.

More recent work has explored the testable implications of EU theory. This work includes Green and Srivastava (1986), Chambers et al. (2016), Kübler et al. (2014), Echenique and Saito (2015), and Polisson et al. (2020). The first four papers focus, as we do here, on rationalizability for risk-averse agents. Green and Srivastava (1986) and Chambers et al. (2016) allow for many goods in each state, which our methodology cannot accommodate. Polisson et al. (2020) present a general approach to testing that allows for a test of EU in isolation, not jointly with risk aversion. Our assumptions are the same as in Kübler et al. (2014) and Echenique and Saito (2015). In fact, the techniques used to obtain an axiomatization of the data that is consistent with EU, and some perturbation of the model, is very much borrowed from Echenique and Saito (2015). This paper turns the first-order conditions into a linear system, which is first perturbed for technical reasons, so as to have a system with rational coefficients, and then analyzed using duality to obtain the axioms. The technical perturbation has to, in the end, be undone. In the present paper, the perturbation has a totally different meaning. It is introduced to relax the conditions needed for consistency with EU, and when we work out the dual, the relaxation is summarized by a single number (e), which is incorporated into the axiom. In sum, the two papers rely on the same broad techniques but differ substantially in how they are used, and the nature of the results.

Compared to most of the existing revealed preference literature on EU, our focus is on measuring consistency with EU, not on providing a test. Thus our paper is inscribed in the literature that seeks to measure the degree of compliance of a dataset with a theory. Our assumption of monetary payoffs and risk aversion is restrictive but consistent with how EU theory is used in economics. Many economic models assume EU together with risk aversion. Our results speak directly to the empirical relevance of such models. A further motivation for focusing on risk aversion is empirical: in the data we have looked at, corner choices are very rare. This would rule out risk-seeking behavior in the context of EU. Thus, arguably, EU and risk-loving behavior would not be a candidate explanation of the experimental data we examine in this paper.

As mentioned, the CCEI was proposed by Afriat (1972). Varian (1990) proposes a modification, and Echenique et al. (2011) and Dean and Martin (2016) propose alternative measures. Dziewulski (2020) provides a foundation for CCEI based on the model in Dziewulski (2016), which seeks to rationalize violations of utility-maximizing behavior with a model of just-noticeable differences. Compared to the literature based on the CCEI, we present explicit model specifications of the errors that would explain the deviation from EU. Each of these specifications introduces errors in beliefs, prices, or utilities. As a consequence, our measure of consistency with EU is based on three possible "stories" for why choices are inconsistent with EU, even though we consider one story at a time and do not study them jointly. And, as we have explained above, the nature of EU-consistent choices is poorly reflected in the CCEI's budget adjustments.

Apesteguia and Ballester (2015) propose a general method to measure the distance between theory and data in revealed preference settings. For each possible preference relation, they calculate the *swaps index*, which counts the number of alternatives that must be swapped with the chosen alternative in order for the preference relation to rationalize the data. Then, Apesteguia and Ballester (2015) consider the preference relation that minimizes the total number of swaps in all the observations, weighted by their relative occurrence in the data. Apesteguia and Ballester (2015) assume that there is a finite number of alternatives, and thus a finite number of preference relations over the set of alternatives. Because of the finiteness, they can calculate the swaps index for each preference relation and find the preference relation that minimizes the swaps index. This method by Apesteguia and Ballester (2015) is not directly applicable to our setup because in our setup, a set of alternatives is a budget set and contains infinitely many elements; moreover, the number of expected utility preferences relation is infinite.[3]

There are many other studies of revealed preference that are based on a notion of distance between the theory and the data. For example, Halevy et al. (2018) use such distances as a guide in estimating parametric functional forms for the utility function. Echenique et al. (2020) focus on discounted utility, and use a statistical formulation that is connected to how we use e in the present paper (this work does not develop the measure methodologically as we do here, and does not include an axiomatization that captures the degree of consistency). In terms of testing con-

[3]In Appendix D.1 of Apesteguia and Ballester (2015), they consider the swaps index for expected utility preferences while assuming the finiteness of the set of alternatives. In their Appendix D.3, without axiomatization, they consider the swaps index for an infinite set of alternatives using the Lebesgue measure to "count" the number of swaps. However, they do not study the case where the number of alternatives is infinite and the preference relations are expected utility.

sistency with EU, Friedman et al. (2022) use the correlation of changes in quantities with changes in prices as an intuitive notion of distance to the theory. They do not explore the methodological aspects of this measure, but we find in our work that it is closely related to the measure we propose. Finally, Aguiar and Kashaev (2021) considers a general approach toward testing models that are formulated through first-order conditions. They focus on measurement error as a source of perturbations, which is one of the sources considered in our paper, but they focus for the most part on problems of intertemporal choice. Their methodology is particularly useful for identification and estimation.

Polisson et al. (2020) develop a general method called the Generalized Restriction of Infinite Domain (GRID) for testing consistency with models of choice under risk and uncertainty. Using GRID, they provide a way to calculate CCEI for departures from EU. Importantly, and in contrast with our measure, their approach does not rely on risk aversion. They present measures of departure from EU and risk-averse EU. We compare empirically our measure to theirs in Section 4.2 (the Online Appendix has additional details). Suffice it to say here that the measures are similar, but distinct, when applied to the data, and that the differences cannot be attributed to risk aversion. Theoretically, our approach has the advantage of modeling a specific source of deviations from EU, and our results connect the measure to certain observable behavioral patterns. These include exact behavioral patterns described by the theorems, but also an empirically motivated observation that our measure captures compliance with downward-sloping demand.

Finally, perhaps the most closely related paper is de Clippel and Rozen (forthcoming), who measure consistency with utility maximization by way of departures from first-order conditions, an approach similar to ours. Their FOC-Departure Index (FDI) can be computed for different classes of utility functions. In particular, their FDI measure for risk-averse EU is equivalent to our measure, except for the use of different scaling (their measure $\varepsilon \in [0, 1]$ is the same as a transformation of our measure $e \geq 0$, with $\varepsilon = e/(1 + e)$). While de Clippel and Rozen propose the same measure as we do, there are some significant differences. First, the axiomatic exercises in both papers are very different. Their axioms are on measures of compliance with first-order conditions, while our axiomatization is standard revealed preference characterizations of the dataset that are a particular distance from being consistent with EU. They take as primitives the weak orderings on pairs of price and utility gradient (derivatives of utility function), and then propose certain axioms on such orderings, which turn out to pin down the FDI.[4] In contrast, our paper proposes an axiom, which takes the form of a test (much like GARP is a test), that describes the dataset that are a distance e from being EU rational. A second difference with de Clippel and

[4]In their paper, $(p, g) \geq (p_0, g_0)$ means that "the utility gradient g is farther apart from the price vector p than g_0 is from p_0."

Rozen is that we show how our measure can be used in a statistical test, an avenue that they do not pursue. Finally, there are also some key differences in the empirical exercise in both papers. We put significant emphasis on the connection with downward-sloping demand, again an aspect that they do not investigate. The result in their Proposition 8 is perhaps closest in spirit to our exercise, where they show that computing the measure reduces to checking a set of inequalities. See Remarks C.1 and C.2 in Online Appendix C of our paper. Finally, we should emphasize that de Clippel and Rozen's work is independent and contemporaneous to ours.

2 Model

Let S be a finite set of *states*. Let $\Delta_{++}(S) = \{\mu \in \mathbf{R}^{|S|}_{++} \mid \sum_{s\in S} \mu_s = 1\}$ denote the set of strictly positive probability measures on S. In our model, the objects of choice are state-contingent monetary payoffs, or *monetary acts*. A monetary act is a vector in $\mathbf{R}^{|S|}_{+}$.

Definition 1. *A* dataset *is a finite collection of pairs* $(x, p) \in \mathbf{R}^{|S|}_{+} \times \mathbf{R}^{|S|}_{++}$.

The interpretation of a dataset $(x^k, p^k)_{k=1}^K$ is that it describes K purchases of a state-contingent payoff x^k at some given vector of prices p^k, and income $p^k \cdot x^k = \sum_{s\in S} p^k_s x^k_s$. Let $\mathcal{K}$ denote the set $\{1, \dots, K\}$. For any prices $p \in \mathbf{R}^{|S|}_{++}$ and positive number $I > 0$, the set $B(p, I) = \{y \in \mathbf{R}^{|S|}_{+} \mid p\cdot y \leq I\}$ is the *budget set* defined by p and I.

Expected utility theory requires a decision maker to solve the problem

$$\max_{x\in B(p,I)} \sum_{s\in S} \mu_s u(x_s), \tag{2}$$

when faced with prices $p \in \mathbf{R}^{|S|}_{++}$ and income $I > 0$, where $\mu \in \Delta_{++}(S)$ is a belief and u is a concave utility function over money. We are interested in concave u; an assumption that corresponds to risk aversion.

The belief μ will have two interpretations in our model. First, in Section 3, we shall focus on decisions taken under *risk*. The belief μ will be a known "objective" probability measure $\mu^* \in \Delta_{++}(S)$. Then, in Online Appendix B, we study choice under *uncertainty*. Consequently, The belief μ will be a subjective belief, which is unobservable to us as outside observers.

The following definition formalizes the concept of as-if choices (Echenique and Saito, 2015).

Definition 2. *A dataset* $(x^k, p^k)_{k=1}^K$ *is* Objective Expected Utility (OEU) rational *if there exists a concave and strictly increasing function* $u : \mathbf{R}_+ \to \mathbf{R}$ *such that, for all* $k \in \mathcal{K}$,

$$y \in B(p^k, p^k \cdot x^k) \implies \sum_{s\in S} \mu^*_s u(y_s) \leq \sum_{s\in S} \mu^*_s u(x^k_s),$$

where $\mu^* \in \Delta_{++}(S)$ *is an objective probability.*

When imposed on a dataset, expected utility maximization (2) may be too demanding. We are interested in situations where the model in (2) holds *approximately.* As a result, we shall relax (2) by "perturbing" some elements of the model. The exercise will be to see if a dataset is consistent with the model in which some elements have been perturbed. Specifically, we shall perturb beliefs, prices, or utilities.

First, consider a perturbation of beliefs. We allow μ to be different for each choice problem k. That is, given price $p \in \mathbf{R}^{|S|}_{++}$ and income $I > 0$ in choice problem k, a decision maker solves the problem

$$\max_{x \in B(p,I)} \sum_{s \in S} \mu^k_s u(x_s), \tag{3}$$

where $\{\mu^k\}_{k \in \mathcal{K}} \subset \Delta_{++}(S)$ is a set of beliefs and u is a concave utility function over money.

In the second place, consider a perturbation of prices. Our consumer faces perturbed prices $\tilde{p}^k_s = \varepsilon^k_s p^k_s$, with a perturbation ε^k_s that depends on the choice problem k and the state s. A decision maker solves the problem

$$\max_{x \in B(\tilde{p},I)} \sum_{s \in S} \mu^*_s u(x_s),$$

when faced with income $I > 0$ and the perturbed prices $\tilde{p}^k_s = \varepsilon^k_s p^k_s$ for each $k \in \mathcal{K}$ and $s \in S$. Here $\{\varepsilon^k_s\}_{s \in S, k \in \mathcal{K}}$ is a set of perturbations, and u is a concave utility function over money.

Finally, consider a perturbation of utility u. We allow u to depend on the choice problem k and the realization of the state s. We suppose that the utility of consumption x_s in state s is given by $\varepsilon^k_s u(x_s)$, with ε^k_s being a multiplicative perturbation in utility. To sum up, a decision maker solves the problem

$$\max_{x \in B(p,I)} \sum_{s \in S} \mu^*_s \varepsilon^k_s u(x_s),$$

when faced with prices $p \in \mathbf{R}^{|S|}_{++}$ and income $I > 0$. As before, $\{\varepsilon^k_s\}_{s \in S, k \in \mathcal{K}}$ is a set of perturbations and u is a concave utility function over money.

Observe that our three sources of perturbations have different interpretations, each can be traced back to a long-standing tradition for how errors are introduced in economic models. Perturbed prices can be thought of as prices subject to measurement error, measurement error being a very common source of perturbations in econometrics (Griliches, 1986). Perturbed utility is an instance of random utility models (McFadden, 1974), at least for the kind of perturbations that are the focus of our paper. Finally, perturbations of beliefs can be thought of as a kind of random utility, or as an inability to use probabilities exactly. Note that we perturb one source at a time and do not consider combinations of perturbations.

3 Perturbed Objective Expected Utility

In this section, we discuss choice under risk: there exists a known "objective" belief $\mu^* \in \Delta_{++}(S)$ that determines the realization of states. The experiments we discuss in Section 4 are all on choice under risk.

We go through each of the sources of perturbation: beliefs, utility, and prices. We seek to understand how large a perturbation has to be in order to rationalize a dataset. It turns out that, for this purpose, all sources of perturbations are equivalent.

3.1 Belief Perturbation

Deviations from EU are accommodated by allowing a different belief at each observation. So we assume a belief μ^k for each choice k, and allow μ^k to differ from the objective μ^*. We seek to understand how much the belief μ^k deviates from the objective belief μ^* by evaluating how far the ratio,

$$\frac{\mu_s^k/\mu_t^k}{\mu_s^*/\mu_t^*},$$

where $s \neq t$, differs from 1. If the ratio is larger (smaller) than one, then it means that in choice k, the decision maker believes the relative likelihood of state s with respect to state t is larger (smaller, respectively) than what he should believe, given the objective belief μ^*.[5]

Given a non-negative number e, we say that a dataset is e-belief-perturbed objective expected utility (OEU) rational, if it can be rationalized using expected utility with perturbed beliefs for which the relative likelihood ratios do not differ by more than e from their objective equivalents.

Definition 3. *Let $e \in \mathbf{R}_+$. A dataset $(x^k, p^k)_{k=1}^K$ is* e-belief-perturbed OEU rational *if there exist $\mu^k \in \Delta_{++}(S)$ for each $k \in \mathcal{K}$, and a concave and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$, such that, for all $k \in \mathcal{K}$,*

$$y \in B(p^k, p^k \cdot x^k) \implies \sum_{s \in S} \mu_s^k u(y_s) \leq \sum_{s \in S} \mu_s^k u(x_s^k),$$

and for each $k \in \mathcal{K}$ and $s, t \in S$,

$$\frac{1}{1+e} \leq \frac{\mu_s^k/\mu_t^k}{\mu_s^*/\mu_t^*} \leq 1 + e. \tag{4}$$

[5] Alternatively, we can introduce multiplicative noise terms ε_s^k (as we do below) and perturb beliefs by $\mu_s^k = \varepsilon_s^k \mu_s^* / \sum_{s' \in S} \varepsilon_{s'}^k \mu_{s'}^*$. Then, the ratio can be rewritten as

$$\frac{\mu_s^k/\mu_t^k}{\mu_s^*/\mu_t^*} = \frac{\varepsilon_s^k}{\varepsilon_t^k}.$$

We are interested in how much the ratio differs from 1.

When $e = 0$, e-belief-perturbed OEU rationality requires that $\mu_s^k = \mu_s^*$ for all s and k, so the case of exact consistency with expected utility is obtained with a zero bound of belief perturbations. Moreover, it is easy to see that by taking e to be large enough, any dataset can be e-belief-perturbed rationalizable.

We should note that e bounds belief perturbations for all states and observations. As such, it can be sensitive to extreme observations and outliers (the CCEI is also subject to this critique: see Echenique et al., 2011). In our empirical application, we carry out a robustness analysis to account for such sensitivity (see Online Appendix F.4).

Finally, we mention a potential relationship with models of non-expected utility. One could think of rank-dependent utility, for example, as a way of allowing an agent's beliefs to adapt to his observed choices. However, unlike e-belief-perturbed OEU, the non-expected utility theory requires some consistency on the dependency. For example, in the case of rank-dependent utility, the agent's beliefs over the states is affected by the ranking of the outcomes across states.

3.2 Price Perturbation

We now turn to perturbed prices: think of them as prices measured with error. The perturbation is a multiplicative noise term ε_s^k to the Arrow-Debreu state price p_s^k. Thus, perturbed state prices are $\varepsilon_s^k p_s^k$. Note that if $\varepsilon_s^k = \varepsilon_t^k$ for all s, t, then introducing the noise does not affect anything because it only changes the scale of prices. In other words, what matters is how perturbations affect relative prices, that is $\varepsilon_s^k / \varepsilon_t^k$.

We can measure how much the noise ε^k perturbs relative prices by evaluating how much the ratio,

$$\frac{\varepsilon_s^k}{\varepsilon_t^k},$$

where $s \neq t$, differs from 1.

Definition 4. *Let $e \in \mathbf{R}_+$. A dataset $(x^k, p^k)_{k=1}^K$ is* e-price-perturbed OEU rational *if there exists a concave and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$, and $\varepsilon^k \in \mathbf{R}_+^{|S|}$ for each $k \in \mathcal{K}$ such that, for all $k \in \mathcal{K}$,*

$$y \in B(\tilde{p}^k, \tilde{p}^k \cdot x^k) \implies \sum_{s \in S} \mu_s^* u(y_s) \leq \sum_{s \in S} \mu_s^* u(x_s^k),$$

where for each $k \in \mathcal{K}$ and $s \in S$

$$\tilde{p}_s^k = p_s^k \varepsilon_s^k$$

and for each $k \in \mathcal{K}$ and $s, t \in S$

$$\frac{1}{1+e} \leq \frac{\varepsilon_s^k}{\varepsilon_t^k} \leq 1 + e. \tag{5}$$

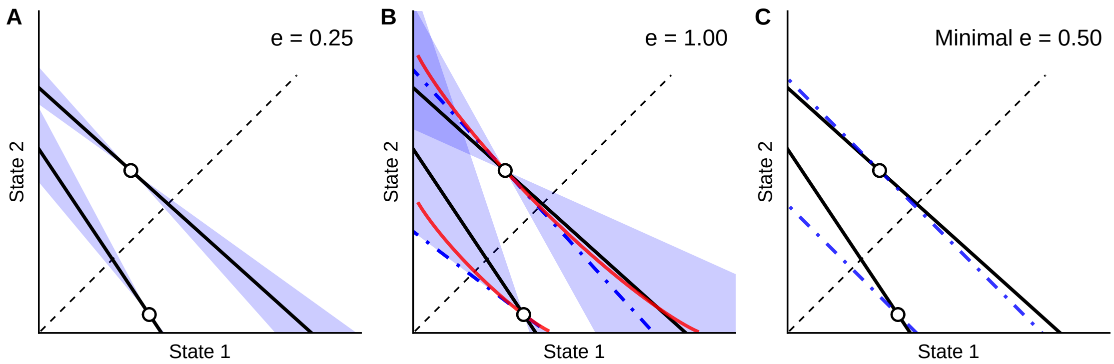

FIGURE 2: Illustration of e-price-perturbation. *Notes*: Panels A and B show the set of possible e-perturbed budget sets with $e \in \{0.25, 1\}$. Panel B presents an example of price-perturbed OEU rationalization. The dash-dotted lines represent the perturbed budget sets and the curves represent the indifference curves of an agent with the CRRA utility function. Panel C presents the minimal price perturbation for OEU rationalization. In panels B and C, we present e-price-perturbation under the assumption that two states are equally likely ($\mu_1^* = \mu_2^* = 0.5$).

It is without loss of generality to add an additional restriction that $\tilde{p}^k \cdot x^k = p^k \cdot x^k$ for each $k \in \mathcal{K}$ because what matters are the relative prices.

The idea is illustrated in Figure 2. The figure shows how the perturbations to relative prices affect budget lines, under the assumption that $|S| = 2$. For each value of $e \in \{0.25, 1\}$ and $k \in \{1, 2\}$, the blue area represents the set

$$\left\{ x \in \mathbf{R}_+^2 \;\middle|\; \tilde{p} \cdot x = \tilde{p} \cdot x^k \text{ for some } \tilde{p} \in \mathbf{R}_{++}^2 \text{ such that } \forall s, t \in S, \frac{1}{1+e} \leq \frac{\tilde{p}_s / p_s^k}{\tilde{p}_t / p_t^k} \leq 1 + e \right\}$$

of perturbed budget lines. The dataset in the figure is the same as in Figure 1B, which is not rationalizable with any expected utility function as we discussed.

Figure 2B illustrates how we rationalize the dataset in Figure 1B using price perturbation. The dash-dotted lines are perturbed budget lines and the curves are indifference curves of an agent with CRRA utility function that are tangent to perturbed budget lines at each of the x^k in the data. Since the perturbed budget lines are inside the bound given by $e = 1$ (shaded areas), we say that the dataset is price-perturbed OEU rational with $e = 1$. There are other perturbations that "work," and in particular, the one in Panel C corresponds to the smallest amount of perturbation we need to rationalize this dataset under $\mu_1^* = \mu_2^* = 0.5$.

3.3 Utility Perturbation

Finally, we turn to perturbed utility. As explained above, perturbations are multiplicative and take the form $\varepsilon_s^k u(x_s^k)$.[6] As for price perturbations, we seek to measure how much the ε^k perturbs utilities at choice problem k by evaluating how much the ratio,

$$\frac{\varepsilon_s^k}{\varepsilon_t^k},$$

where $s \neq t$, differs from 1.

Definition 5. *Let $e \in \mathbf{R}_+$. A dataset $(x^k, p^k)_{k=1}^K$ is* e-utility-perturbed OEU rational *if there exists a concave and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$ and $\varepsilon^k \in \mathbf{R}_+^{|S|}$ for each $k \in \mathcal{K}$ such that, for all $k \in \mathcal{K}$,*

$$y \in B(p^k, p^k \cdot x^k) \implies \sum_{s \in S} \mu_s^* \varepsilon_s^k u(y_s) \leq \sum_{s \in S} \mu_s^* \varepsilon_s^k u(x_s^k),$$

and for each $k \in \mathcal{K}$ and $s, t \in S$

$$\frac{1}{1+e} \leq \frac{\varepsilon_s^k}{\varepsilon_t^k} \leq 1 + e. \tag{6}$$

3.4 Equivalence of Belief, Price, and Utility Perturbations

The first observation we make is that the three sources of perturbations are equivalent, in the sense that for any e, a dataset is e-perturbed rationalizable according to one of the sources if and only if it is also rationalizable according to any of the other sources with the same e. By virtue of this result, we can interpret our measure of deviations from OEU in any of the ways we have introduced.

Theorem 1. *Let $e \in \mathbf{R}_+$, and D be a dataset. The following are equivalent:*

- *D is e-belief-perturbed OEU rational;*
- *D is e-price-perturbed OEU rational;*
- *D is e-utility-perturbed OEU rational.*

The proof appears in Appendix A. To see the intuition behind the equivalence, let us assume for simplicity that u is differentiable. The first-order condition for the maximization of perturbed utility for choice problem k is

$$\lambda^k p_s^k = \mu_s^* \varepsilon_s^k u'(x_s^k),$$

[6] We consider state-contingent perturbations. As such, perturbed utilities fall outside of the domain of EU theory. We thank Jose Apesteguía and Miguel Ballester for pointing this out to us.

for each $s \in S$, where $\lambda^k > 0$ is a Lagrange multiplier. By rearranging the multiplicative noise terms ε_s^k, we can obtain the first-order conditions for belief-perturbed OEU and price-perturbed OEU.

In light of Theorem 1, we shall simply say that a dataset is *e-perturbed OEU rational* if it is e-belief-perturbed OEU rational, and this will be equivalent to being e-price-perturbed OEU rational, and e-utility-perturbed OEU rational.

3.5 Characterizations

We proceed to give a characterization of the dataset that is e-perturbed OEU rational. Specifically, given $e \in \mathbf{R}_+$, we propose a revealed preference axiom and prove that a dataset satisfies the axiom if and only if it is e-perturbed OEU rational.

Before we state the axiom, we need to introduce some additional notation. In the current model, where μ^* is known and objective, what matters to an expected utility maximizer is not the state price itself, but instead the *risk-neutral* price.

Definition 6. *For any dataset $(p^k, x^k)_{k=1}^K$, the* risk-neutral price $\rho_s^k \in \mathbf{R}_{++}^{|S|}$ *in choice problem k at state s is defined by*

$$\rho_s^k = \frac{p_s^k}{\mu_s^*}.$$

As in Echenique and Saito (2015), the axiom we propose involves a sequence $(x_{s_i}^{k_i}, x_{s_i'}^{k_i'})_{i=1}^n$ of pairs satisfying certain conditions. Note that $k_i, k_i' \in \mathcal{K}$ and $s_i, s_i' \in S$ in each pair $(x_{s_i}^{k_i}, x_{s_i'}^{k_i'}) \in \mathbf{R}_+^2$, so each object $x_{s_i}^{k_i}$ in the sequence is the payoff in state s_i purchased in observation k_i under price $p_{s_i}^{k_i}$.

Definition 7. *A sequence of pairs $(x_{s_i}^{k_i}, x_{s_i'}^{k_i'})_{i=1}^n$ is called a* test sequence *if*

(i) $x_{s_i}^{k_i} > x_{s_i'}^{k_i'}$ for all $i = 1, \dots, n$;

(ii) each $k \in \mathcal{K}$ appears as k_i (on the left of the pair) the same number of times it appears as k_i' (on the right of the pair).

Echenique and Saito (2015) provide an axiom for OEU rationalization, termed the Strong Axiom for Revealed Objective Expected Utility (SAROEU), which states that for any test sequence $(x_{s_i}^{k_i}, x_{s_i'}^{k_i'})_{i=1}^n$, we have

$$\prod_{i=1}^n \frac{\rho_{s_i}^{k_i}}{\rho_{s_i'}^{k_i'}} \leq 1. \tag{7}$$

SAROEU is equivalent to the axiom provided by Kübler et al. (2014).

It is easy to see why SAROEU is necessary for OEU rationalization. Assuming (for simplicity of exposition) that u is differentiable, the first-order condition of the maximization problem (2) for choice problem k is

$$\lambda^k p_s^k = \mu_s^* u'(x_s^k), \text{ or equivalently, } \rho_s^k = \frac{u'(x_s^k)}{\lambda^k},$$

where $\lambda^k > 0$ is a Lagrange multiplier.

By substituting this equation on the left-hand side of (7), we have

$$\prod_{i=1}^n \frac{\rho_{s_i}^{k_i}}{\rho_{s_i'}^{k_i'}} = \prod_{i=1}^n \frac{\lambda^{k_i'}}{\lambda^{k_i}} \cdot \prod_{i=1}^n \frac{u'(x_{s_i}^{k_i})}{u'(x_{s_i'}^{k_i'})} \leq 1.$$

To see that this term is smaller than 1, note that the first term of the product of the λ-ratios is equal to one because of the condition (ii) of the test sequence: all λ^k must cancel out. The second term of the product of u'-ratio is less than one because of the concavity of u, and the condition (i) of the test sequence (i.e., $u'(x_{s_i}^{k_i})/u'(x_{s_i'}^{k_i'}) \leq 1$). Thus, SAROEU is implied. It is more complicated to show that SAROEU is sufficient (see Echenique and Saito, 2015).

Now, e-perturbed OEU rationality allows the decision maker to use different beliefs $\mu^k \in \Delta_{++}(S)$ for each choice problem k. Consequently, SAROEU is not necessary for e-perturbed OEU rationality. To see that SAROEU can be violated, note that the first-order condition of the maximization (3) for choice k is as follows: there exists a positive number (Lagrange multiplier) λ^k such that for each $s \in S$,

$$\lambda^k p_s^k = \mu_s^k u'(x_s^k), \text{ or equivalently, } \rho_s^k = \frac{\mu_s^k}{\mu_s^*} \frac{u'(x_s^k)}{\lambda^k}.$$

Suppose that $x_s^k > x_t^k$. Then (x_s^k, x_t^k) is a test sequence (of length one) according to Definition 7. We have

$$\frac{\rho_s^k}{\rho_t^k} = \left(\frac{\mu_s^k}{\mu_s^*} \frac{u'(x_s^k)}{\lambda^k} \right) \Big/ \left(\frac{\mu_t^k}{\mu_t^*} \frac{u'(x_t^k)}{\lambda^k} \right) = \frac{u'(x_s^k)}{u'(x_t^k)} \frac{\mu_s^k/\mu_t^k}{\mu_s^*/\mu_t^*}.$$

Even though $x_s^k > x_t^k$ implies the first term of the ratio of u' is less than one, the second term can be strictly larger than one. When x_s^k is close enough to x_t^k, the first term is almost one while the second term can be strictly larger than one. Consequently, SAROEU can be violated.

However, by (4), we know that the second term is bounded by $1 + e$. So we must have

$$\frac{\rho_s^k}{\rho_t^k} \leq 1 + e.$$

In general, for a sequence $(x^{k_i}_{s_i}, x^{k'_i}_{s'_i})^n_{i=1}$ of pairs, one may suspect that the bound is calculated as $(1+e)^n$. This is not true because if x^k_s appears both as $x^{k_i}_{s_i}$ for some i (on the left of the pair) and as $x^{k'_j}_{s'_j}$ for some j (on the right of the pair), then all μ^k_s can be canceled out. What matters is the number of times x^k_s appears without being canceled out. This number can be defined as follows.

Definition 8. *Consider any sequence $(x^{k_i}_{s_i}, x^{k'_i}_{s'_i})^n_{i=1}$ of pairs. Let $\sigma \equiv (x^{k_i}_{s_i}, x^{k'_i}_{s'_i})^n_{i=1}$. For any $k \in \mathcal{K}$ and $s \in S$,*

$$d(\sigma, k, s) = \#\{i \mid x^k_s = x^{k_i}_{s_i}\} - \#\{i \mid x^k_s = x^{k'_i}_{s'_i}\},$$

and

$$m(\sigma) = \sum_{s\in S} \sum_{k\in\mathcal{K}:d(\sigma,k,s)>0} d(\sigma, k, s).$$

Note that, if $d(\sigma, k, s)$ is positive, then $d(\sigma, k, s)$ is the number of times μ^k_s appears as a numerator without being canceled out. If it is negative, then $d(\sigma, k, s)$ is the number of times μ^k_s appears as a denominator without being canceled out. So $m(\sigma)$ is the "net" number of terms such as μ^k_s/μ^k_t that are present in the numerator. Thus the relevant bound is $(1+e)^{m(\sigma)}$.

Given the discussion above, it is easy to see that the following axiom is necessary for e-perturbed OEU rationality.

Axiom 1 (e-Perturbed Strong Axiom for Revealed Objective Expected Utility (e-PSAROEU))**.** *For any test sequence of pairs $(x^{k_i}_{s_i}, x^{k'_i}_{s'_i})^n_{i=1} \equiv \sigma$, we have*

$$\prod_{i=1}^{n} \frac{\rho^{k_i}_{s_i}}{\rho^{k'_i}_{s'_i}} \le (1+e)^{m(\sigma)}.$$

The main result of this section is to show that the axiom is also sufficient.

Theorem 2. *Given $e \in \mathbf{R}_+$, and let D be a dataset. The following are equivalent:*

- *D is e-perturbed OEU rational.*
- *D satisfies e-PSAROEU.*

The proof appears in Appendix A.[7]

[7] We should mention that Theorem 2 is similar in spirit to some of the results in Allen and Rehbeck (2020), who consider approximate rationalizability of quasilinear utility. They present a revealed preference characterization with a measure of error "built in" to the axiom, similar to ours, which they then use as an input to a statistical test. The two papers were developed independently, and since the models in question are very different, the results are unrelated.

Axioms like e-PSAROEU can be interpreted as a statement about downward-sloping demand (see Echenique et al., 2020). For example, $(x_s^k, x_{s'}^k)$ with $x_s^k > x_{s'}^k$ is a test sequence. If risk-neutral prices satisfy $\rho_s^k > \rho_{s'}^k$, then the dataset violates downward-sloping demand. Now e-PSAROEU measures the extent of the violation by controlling the size of $\rho_s^k/\rho_{s'}^k$. Specifically, it says that $\rho_s^k/\rho_{s'}^k \leq (1+e)$, as it is easy to see that we have $m(\sigma) = 1$ in this case. So e-PSAROEU describes how large a violation of downward-sloping demand is allowed: when quantities are ordered so that more is consumed in state s than in state s', despite consumption in state s than in state s', when evaluated at risk-neutral prices, our axiom uses the number e to restrict the magnitude of the extent to which price ratio is "incorrect." Indeed, as we have already mentioned, 0-PSAROEU is equivalent to SAROEU. And, when $e = \infty$, the e-PSAROEU always holds because $(1+e)^{m(\sigma)} = \infty$. Now, most of the time, we shall turn the axiom on its head. We will calculate the smallest value of e for which the data satisfies e-PSAROEU, and use this as our measure of deviations from OEU.

In its connection to downward-sloping demand, Theorem 2 formalizes the idea of testing OEU through the correlation of risk-neutral prices and quantities: see Friedman et al. (2022) and our discussion in Section 4.2. Theorem 2 and the axiom e-PSAROEU give the precise form that the downward-sloping demand property takes in order to characterize OEU, and provide a non-parametric justification to the practice of analyzing the correlation of prices and quantities. Here it is important to account for imbalances in a test sequence. These are manifested in the value of $m(\sigma)$. A test sequence may involve the same states appearing on the left and on the right of a pair, for the same observation. When this occurs, any potential "violation" of OEU is inconsequential, as it does not affect marginal rates of substitution. By including the number $m(\sigma)$ we account for the degree of imbalance, which amplifies a given error e and provides the correct slack in satisfying downward-sloping demand with error e.

Given a dataset, we shall calculate the *smallest* e for which the dataset satisfies e-PSAROEU. It is easy to see that such a minimal level of e exists.[8] We explain in Online Appendix C how it is calculated in practice.

Definition 9. Minimal e*, denoted* e_**, is the smallest* $e' \geq 0$ *for which the data satisfies* e'*-PSAROEU.*

The number e_* is a crucial component of our empirical analysis. Importantly, it is the basis of a statistical procedure for testing the null hypothesis of OEU rationality.

As mentioned above, e_* is a bound that has to hold across all observations, and therefore may be sensitive to extreme outliers. It is, however, easy to check the sensitivity of the calculated

[8]In Online Appendix C, we show that e_* can be obtained as a solution of minimization of a continuous function on a compact space. Hence, the minimum exists.

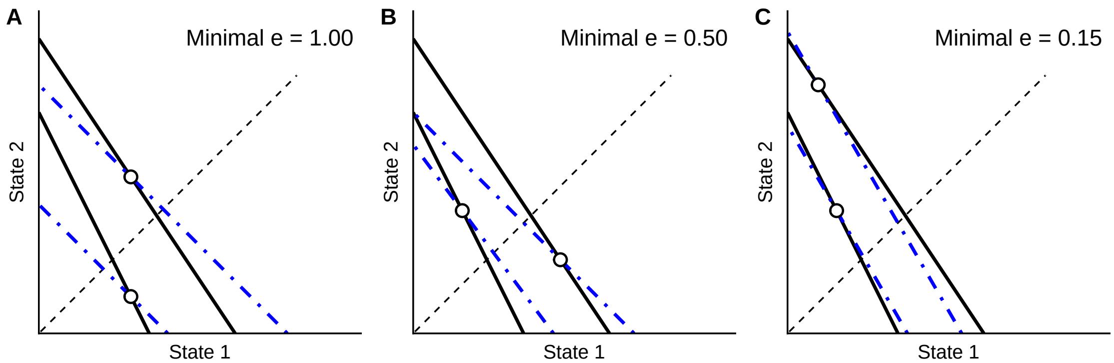

FIGURE 3: Minimal price perturbation for OEU rationalization. *Notes*: We assume $\mu_1^* = \mu_2^* = 0.5$. See Online Appendix D.1 for details.

e_* to an extreme observation. One can, for example, re-calculate e_* after dropping one or two observations, and look for large changes.

Finally, e_* depends on the prices and the objective probability that a decision maker faces. In particular, it is clear from e-PSAROEU that $1+e$ is bounded by the maximum ratio of risk-neutral prices (i.e., $\max_{k,k'\in\mathcal{K},s,s'\in S} \rho_s^k/\rho_{s'}^{k'}$).

Remark. It is easy to visualize perturbation corresponding to e_* if we take the price perturbation interpretation of e. Solving the constrained minimization problem described in Online Appendix C gives us ratios of perturbations $\varepsilon_s^k/\varepsilon_t^k$, for all $k \in \mathcal{K}$ and $s, t \in S$, that corresponds to e_*. We can compute perturbed relative prices

$$\frac{\tilde{p}_s^k}{\tilde{p}_t^k} = \frac{p_s^k}{p_t^k}\frac{\varepsilon_s^k}{\varepsilon_t^k}.$$

Note that perturbed budgets must pass through the chosen bundles. Figure 3 illustrates these "minimally-perturbed" budgets (dash-dotted lines) under which observed choices are e_*-perturbed OEU rational assuming that two states are equally likely.

4 Measuring the Deviation from Objective Expected Utility

We apply our methodology to data from three large-scale online experiments. The experiments were implemented through representative surveys, and the task involved objective risk, not uncertainty. The data are taken from Choi et al. (2014, hereafter CKMS), Carvalho et al. (2016, hereafter CMW), and Carvalho and Silverman (2019, hereafter CS). All three experiments share a common experimental structure, the portfolio allocation task introduced by Loomes (1991) and Choi et al. (2007).

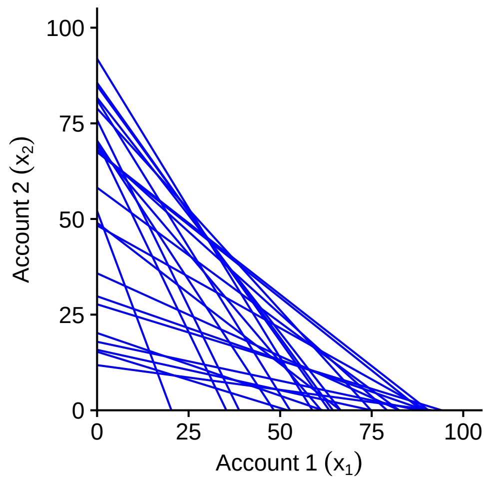

FIGURE 4: Sample budget lines. A set of 25 budgets presented to one of the subjects in Choi et al. (2014).

It is worth mentioning again that the three studies focus on CCEI as a measure of violation of basic rationality. We shall instead look at OEU, and use e_* as our measure of violations of OEU. The procedure for calculating e_* is explained in Online Appendix C.

4.1 Datasets

In the experiments, subjects were presented with a sequence of decision problems under risk in a graphical illustration of a two-dimensional budget line. They were asked to select a point (x_1, x_2), an "allocation," by clicking on the budget line (subjects were therefore forced to exhaust the income). The coordinates of the selected point represent an allocation of points between "accounts" 1 and 2. They received the points allocated to one of the accounts, determined at random with an equal chance ($\mu_1^* = \mu_2^* = 0.5$). Subjects faced 25 budgets, as illustrated in Figure 4.

We note some interpretations of the design that matter for our posterior discussion. First, points on the 45-degree line correspond to equal allocations between the two accounts and therefore involve no risk. The 45-degree line is the "full insurance" line. Second, we can interpret the slope of a budget line as a price in the usual sense: if the x_2-intercept is larger than the x_1-intercept, points in the account 2 are "cheaper" than those in the account 1.

Choi et al. (2014) implemented the task using the instrument of the CentERpanel, randomly recruiting subjects from the entire panel sample in the Netherlands. Carvalho et al. (2016) administered the task using the GfK KnowledgePanel, a representative panel of the adult U.S. population. Carvalho and Silverman (2019) used the Understanding America Study panel. The number of subjects who completed the task in each study is 1,182 in CKMS, 1,119 in CMW, and 1,423 in

CS.

The survey instruments in these studies allowed the researchers to collect a wide variety of individual demographic and economic information from the respondents. The main demographic information they obtained include gender, age, education level, household income, occupation, and household composition.

The selection of 25 budget lines was independent across subjects in CKMS (i.e., the subjects were given different sets of budget lines), fixed in CMW (i.e., all subjects saw the same set of budgets), and semi-randomized across subjects in CS (i.e., each subject drew one of the prepared sets of 25 budgets).

4.2 Results

There are 3,724 subjects in three experiments, including five subjects who are "exactly" OEU rational. About 76% of the subjects never chose corners of the budget lines, and there is only two percent of the entire sample who chose corners in more than half of the 25 questions. Finally, no subjects chose corners in all 25 questions. Given these observations, our focus on risk aversion does not seem to be too restrictive in these datasets.

We calculate e_* for each individual subject.[9,10] The distributions of e_* are displayed in Figure 5A.[11] The CKMS sample has a mean e_* of 3.034, and a median of 2.729. The CMW subjects have a mean of 2.480 and a median of 2.533. The CS sample has a mean of 2.490 and a median of 2.081.[12] Recall that the smaller a subject's e_* is, the closer are her choices to OEU rationality. It is, however, hard to exactly interpret the magnitude of e_*. We turn to this issue in Section 4.3.

Downward-sloping demand and e_*. Perturbations in beliefs, prices, or utility, seek to accommodate a dataset so that it is OEU rationalizable. The accommodation can be seen as correcting a mismatch of relative prices and marginal rates of substitution: recall our discussion in the introduction. Another way to see the accommodation is through the relation between prices and quantities. Our revealed preference axiom, e-PSAROEU, bounds certain deviations from downward-sloping demand. The minimal e is therefore a measure of the kinds of deviations from downward-sloping demand that are crucial to OEU rationality.

[9] By definition, five OEU rational subjects have $e_* = 0$.

[10] Earlier drafts of the paper (posted before summer 2019) reported $\log(1 + e_*)$, not e_* itself.

[11] The empirical CDF for the CMW data has several "steps" since all subjects saw the same set of 25 budget lines. For example, there are 172 subjects with $e_* = 3.5925$. The maximum adjustment required to make their data e-perturbed OEU rational is on the budget line with prices $(p_1, p_2) = (1, 0.2177)$.

[12] Since e_* depends on the design of the set(s) of budgets, comparing the values of e_* across studies requires caution.

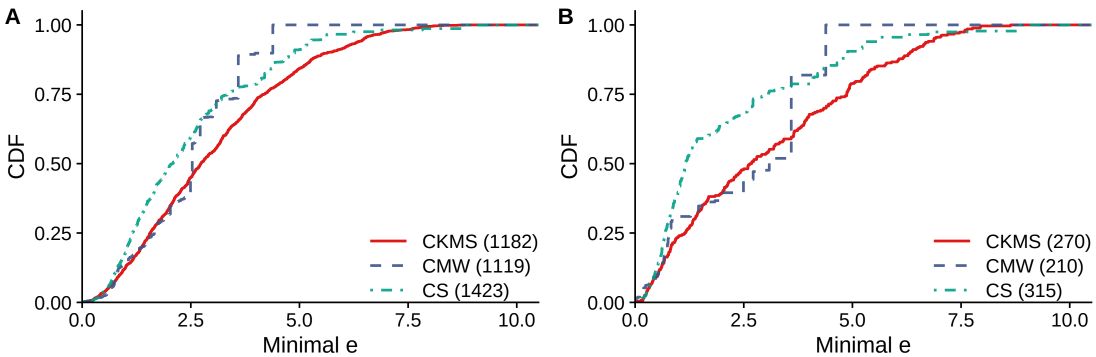

FIGURE 5: Empirical CDFs of e_*. (A) All subjects. (B) The subsample of subjects with CCEI = 1. *Notes*: The number of observations in each dataset is presented in parentheses. Table F.1 in the Online Appendix presents summary statistics.

Figure 6 illustrates this idea. We calculate the correlation coefficient between $\log(x_2/x_1)$ and $\log(p_2/p_1)$ for each subject in the datasets as a measure for the degree of compliance with downward-sloping demand.[13] Roughly speaking, downward-sloping demand is captured by the negative correlation between changes in quantities $\log(x_2/x_1)$ and changes in prices $\log(p_2/p_1)$. The idea is that if a subject properly responds to price changes, then as $\log(x_2/x_1)$ becomes larger, $\log(x_2/x_1)$ should become smaller. The correlation is close to zero if subjects do not respond to price changes.

The top panels (A1-C1) of Figure 6 confirm that e_* and the correlation between prices and quantities are closely related. This means that subjects with smaller e_* tend to exhibit downward-sloping demand, while those with larger e_* are insensitive to price changes. Across all three datasets, e_* and downward-sloping demand are strongly and positively related.

The CCEI, on the other hand, is not clearly related to downward-sloping demand. As illustrated in the bottom panels (A2-C2) of Figure 6, the relation between CCEI and the degree of downward-sloping demand is not monotonic, as indicated by U-shaped LOESS curves. Agents who are closer to complying with utility maximization do not necessarily display a stronger negative correlation between prices and quantities.[14]

The finding is consistent with our comment about CCEI, e_*, and OEU rationality: CCEI measures the distance from general utility maximization, which is related to parallel shifts in budget lines, while e_* and OEU are about the slopes of the budget lines, and about a negative relation

[13]Note that $\log(x_2/x_1)$ is not defined at the corners. We thus adjust corner choices (less than 5% of all choices) by a small constant, 0.1% of the budget in each choice, in the calculation of the correlation coefficient.

[14]Note that CCEI and the degree of downward-sloping demand exhibit a monotonic relation if we restrict attention to subjects who comply with downward-sloping demand in the sense of significant negative correlation between $\log(x_2/x_1)$ and $\log(p_2/p_1)$ (i.e., observations that locate left of the vertical dashed line in Figure 6).

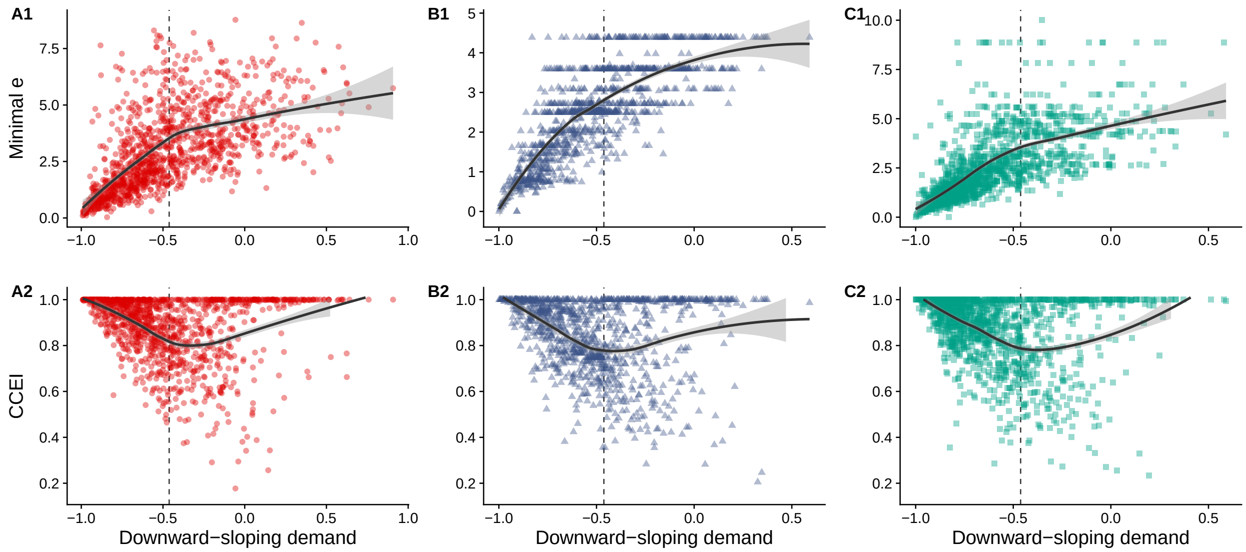

FIGURE 6: Downward-sloping demand and measures of rationality. Panels: (A) CKMS, (B) CMW, (C) CS. *Notes*: The x-axis shows Spearman's correlation coefficient between $\log(x_2/x_1)$ and $\log(p_2/p_1)$ as a proxy for the degree of downward-sloping demand. The vertical dashed line indicates the threshold below which correlation is significantly negative (one-sided, at the 1% level). Black curves represent LOESS smoothing with 95% confidence bands.

between quantities and prices.

We should mention the practice by some authors, notably, Friedman et al. (2022), to evaluate compliance with OEU by looking at the correlation between risk-neutral prices and quantities. Our e_* is related to that idea, and the empirical results presented in this section can be read as a validation of the correlational approach. Friedman et al. (2022) use their approach to estimate a parametric functional form, using experimental data in which they vary objective probabilities, not just prices. Our approach is non-parametric, and focused on testing OEU itself, not estimating any particular utility specification.

First-order stochastic dominance and e_*. In the experiments we consider, choosing (x_1, x_2) at prices (p_1, p_2) violates *monotonicity with respect to first-order stochastic dominance* (hereafter *FOSD-monotonicity*) when either (i) $p_1 > p_2$ and $x_1 > x_2$ or (ii) $p_2 > p_1$ and $x_2 > x_1$. Since the two states have the same objective probability in our datasets, choosing a greater payoff in the more expensive state violates FOSD-monotonicity. Violations of FOSD-monotonicity are related to downward-sloping demand, as they involve consuming more in the more expensive state. Choices that violate FOSD-monotonicity are not uncommon in the data (on average, subjects violated FOSD-monotonicity in six to nine budgets out of 25; see Online Appendix F.2).

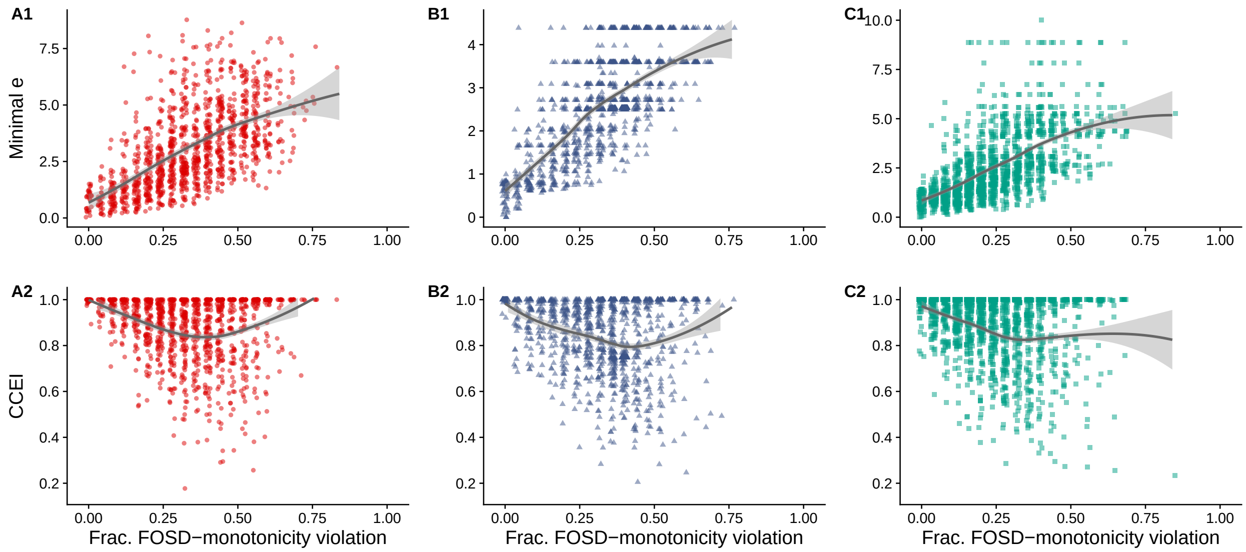

FIGURE 7: Violation of FOSD-monotonicity and measures of rationality. Black curves represent LOESS smoothing with 95% confidence bands. Panels: (A) CKMS, (B) CMW, (C) CS.

Since OEU-rational choices must satisfy FOSD-monotonicity, $e_* = 0$ implies no violations of FOSD-monotonicity. Moreover, the value of e_* is a good indicator of FOSD-monotonicity violations. See the positive relationship between the fraction of FOSD-monotonicity violations and e_* in the top panels (A1-C1) of Figure 7: subjects who frequently made choices violating FOSD-monotonicity tend to have larger e_* compared to those with fewer such violations.

The relation between e_* and violations of FOSD-monotonicity stands in sharp contrast with CCEI. First, choices that violate FOSD-monotonicity *can* be consistent with GARP. Our data exhibits subjects who pass GARP while making choices that violate FOSD-monotonicity (an empirical fact that was first pointed out by Choi et al., 2014). The bottom panels (A2-C2) of Figure 7 show U-shaped relationship between the fraction of FOSD-monotonicity violations and CCEI. Subjects who made frequent violations of FOSD-monotonicity do not necessarily have lower values of CCEI.

Typical patterns of choices. We can gain some insights into the data by considering "typical" patterns of choice. Figure 8 presents choice patterns of six selected subjects with CCEI = 1 and varying degrees of e_*.[15] Panels A-F plot observed choices and panels a-f plot the relationship between $\log(x_2/x_1)$ and $\log(p_2/p_1)$ associated with each choice pattern. As discussed above, panels a-f would exhibit a strongly negative correlation (downward-sloping demand) for the subject to be OEU rational.

[15]The patterns in Figure 8 are not an exhaustive list by any means. See Online Appendix F.8 for more examples.

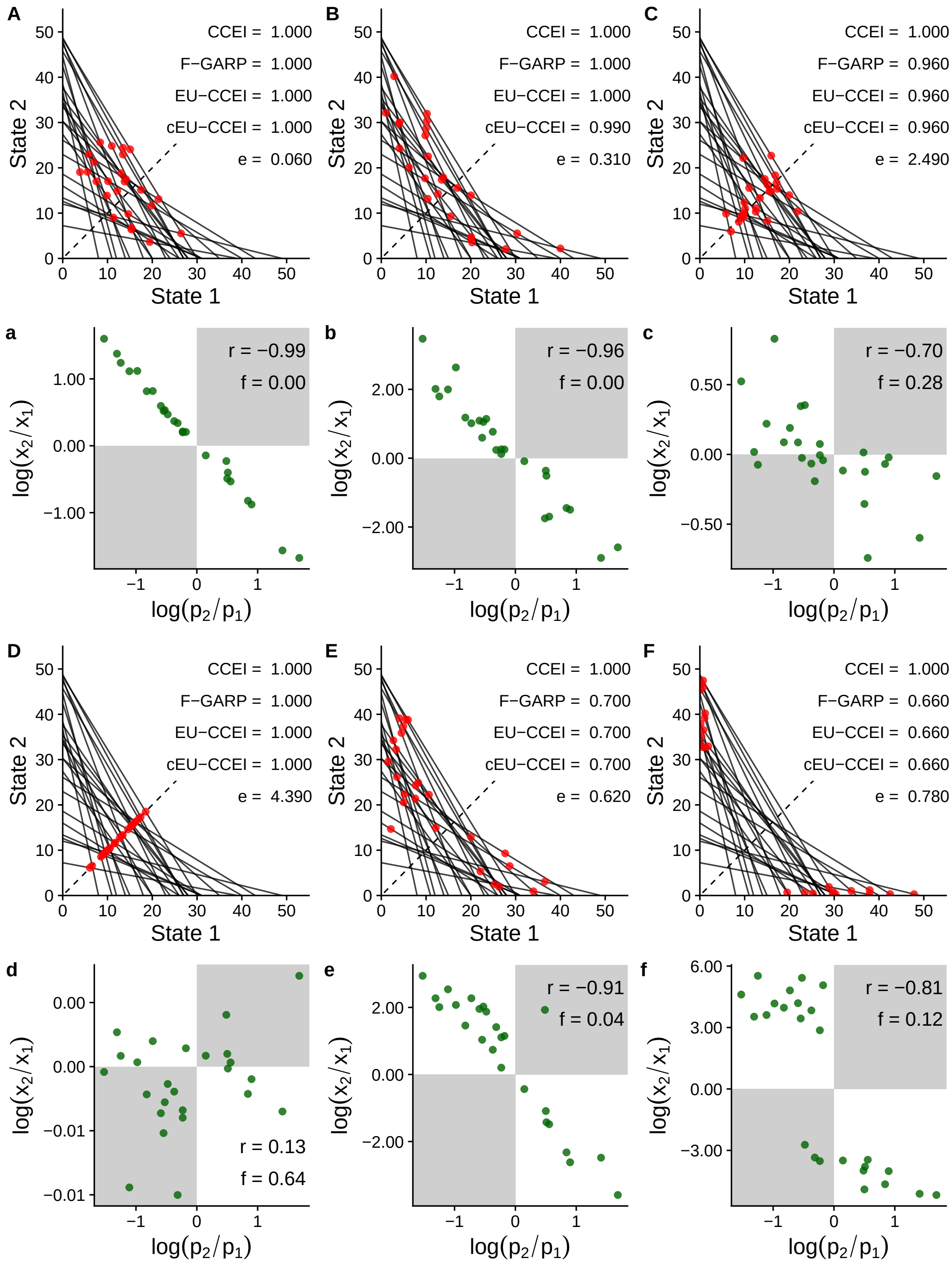

FIGURE 8: Choice patterns from six subjects in the CMW data with CCEI = 1 and varying e_*. (A-F) Observed choices. (a-f) The relation between $\log(x_2/x_1)$ and $\log(p_2/p_1)$. *Notes*: Choices that appear in shaded areas violate FOSD-monotonicity. r indicates the correlation coefficient and f indicates the fraction of choices violating FOSD-monotonicity. In this data, the median CCEI is 0.889, the median EU-CCEI is 0.730, and the median e_* is 2.533. F-GARP, EU-CCEI, cEU-CCEI are calculated with the GRID method of Polisson et al. (2020).

Panel A presents a choice pattern that is "almost" consistent with OEU. The relation between $\log(x_2/x_1)$ and $\log(p_2/p_1)$ fits close to a line with negative slope. Panel B also shows a pattern that does not involve any FOSD-monotonicity violations but is not OEU rational due to small deviations from the downward-sloping demand (see panel b). The pattern in panel C exhibits larger deviations from the downward-sloping demand (panel c), which push its e_* higher than the previous two subjects.

The subject's choices in panel D are close to the 45-degree line. At first glance, such choices might seem to be rationalizable by a very risk-averse expected utility function. However, as panel d shows, the subject's choices deviate from the downward-sloping demand property, and hence cannot be rationalized by any risk-averse expected utility function. Note that the "size" of the deviation from the downward-sloping demand is small (see the scale of the y-axis in panel d). One might be able to rationalize the choices made in panel D with some models of errors in choices, but not with the types of errors captured by our model.[16] We will discuss the other two subjects (panels E and F) below.

Figure 8 also illustrates a simplified way of assessing the size of e_* when there are two states. Under the price-perturbation interpretation, it measures how big of an adjustment of prices would be needed to satisfy downward-sloping demand. Such adjustments will be represented as "horizontal shifts" of points in the bottom panels of the figure (since we fix the chosen bundle and rotate the budget line), and the largest adjustment corresponds to e_*. A scatterplot of $\log(x_2/x_1)$ versus $\log(p_2/p_1)$, as in panels a-f of Figure 8, works as a graphical tool to get a sense of whether a subject's e_* is big or small. Online Appendix F.6 discusses this idea in more detail and illustrates e-price-perturbed OEU rationalization using the choice data presented in Figure 8.

Relationship between e_*, CCEI, and EU-CCEI. CCEI serves a different purpose than e_*; it is meant to capture deviations from general utility maximization, and not OEU. Nevertheless, it is informative to understand the relationship between these measures in the data. We also comment on the recent proposal by Polisson et al. (2020) of an adaptation of CCEI to test for OEU.

We observe, in Figure 5, that the distribution of e_* among subjects whose CCEI is equal to one (panel B) varies as much as in the whole population (panel A). Many subjects have CCEI equal to one, but their e_*'s can be far from zero. This means that consistency with general utility maximization is not necessarily a good indication of consistency with OEU.

[16]This is, in our opinion, a strength of our approach. We do not ex-post seek to invent a model of errors that might rescue EU. Instead, we have written down what we think are natural sources of errors and perturbation (random utility, beliefs, and measurement errors). Our results deal with what can be rationalized when these sources of errors, and only those, are used to explain the data. A general enough model of errors will, of course, render the theory untestable.

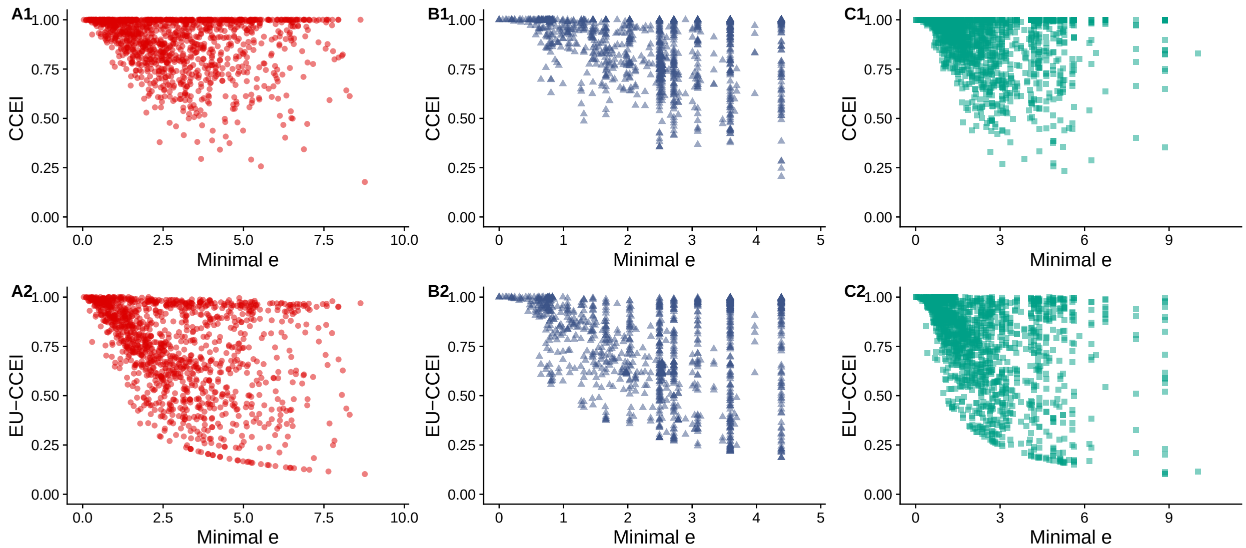

Figure 9: Correlation between e_* and CCEI (top panels) and EU-CCEI of Polisson et al. (2020) (bottom panels). Panels: (A) CKMS, (B) CMW, (C) CS.

That said, the measures are clearly correlated. Figure 9, top panels (A1-C1), plot the relation between CCEI and e_*. As we expect from their definitions (*larger* CCEI and *smaller* e_* correspond to higher consistency), there is a negative and significant relationship between them (correlation coefficient: $r = -0.18$ for CKMS, $r = -0.12$ for CMW, $r = -0.35$ for CS, all $p < 0.001$). Of course, subjects that are consistent with OEU as measured by e_* (they have $e_* = 0$) must exhibit CCEI = 1.

Notice that the variability of the CCEI widens as e_* becomes larger. Obviously, subjects with a small e_* are close to being consistent with general utility maximization, and therefore have a CCEI that is close to one. However, subjects with large e_* seem to have dispersed values of CCEI.

Polisson et al. (2020) propose a version of CCEI meant to measure departures from EU using their GRID method. We term this measure EU-CCEI. In contrast with our measure e_*, which assumes risk aversion and is based on rotating budget lines, EU-CCEI does not impose risk aversion and uses the same idea of shrinking budget lines as in standard CCEI. The bottom panels (A2-C2) of Figure 9 exhibit the relationship between e_* and EU-CCEI. It is clear that the relation between e_* and EU-CCEI is similar to that between e_* and CCEI. The two measures are strongly correlated, but they also provide different conclusions for many subjects.

There are many subjects that EU-CCEI deems consistent with OEU, but have high levels of e_*. This could be attributed to the more restrictive theory being tested by e_*. Subjects with EU-CCEI close to one and large e_* could simply be non-risk-averse OEU maximizers. Perhaps more interesting is the existence of subjects that e_* sees as relatively closer to OEU than others while EU-CCEI does not, in the sense that their e_* and EU-CCEI are below the first quartiles of the empirical distributions (CKMS: three subjects; CMW: 28 subjects; CS: four subjects).

The same conclusions hold true for modified CCEI indices for two additional models considered in Polisson et al. (2020): stochastically monotone utility maximization and risk-averse EU. We call these indices F-GARP and cEU-CCEI, respectively. Values of these indices are also reported for the patterns in Figure 8. See Figures F.15-F.17 in the Online Appendix for pairwise scatter plots of five indices (CCEI, F-GARP, EU-CCEI, cEU-CCEI, and e_*). The modified CCEI measures provide a more refined index for consistency for EU than CCEI, but differences with e_* persist.

It is hard to investigate the differences between EU-CCEI and e_* methodologically. EU-CCEI does not specify a source of deviations from OEU, so we cannot say that one measure emphasizes one source of errors and the other a different source. Instead, we look at some of the patterns in the data that give rise to differences. An example of a choice pattern in which e_* and EU-CCEI differ is provided by Figure 8, panel D. The subject in question exhibits CCEI = EU-CCEI = 1, while e_* is large and indicates a violation of OEU. (The pattern involves choices close to the 45-degree line, but with a clear violation of downward-sloping demand, see panel d.) Panels E and F exhibit subjects that e_* says are close to (risk-averse) OEU, but EU-CCEI deems far from OEU. We see in panels e and f that the conclusion using e_* can be understood by the subjects' compliance with downward-sloping demand. The subjects in panels E and F make a few FOSD-monotonicity violations, which might explain the behavior of EU-CCEI, but that cannot be the end of the story because the subject in panel D makes substantial FOSD-monotonicity violations and exhibits the opposite behavior of e_* and EU-CCEI. Finally, we should say that there are many other patterns for which the conclusions of e_* and EU-CCEI differ (see Online Appendices F.3 and F.8).

With that said, we can still gain some understanding of the similarities and differences between e_*, EU-CCEI, and cEU-CCEI by focusing on a simple environment with two equally-likely states, and two budget sets. The Online Appendix D.2 has a detailed development of this exercise. In brief, while the relative location of the two choices plays a central role in determining the value of e_*, EU-CCEI and cEU-CCEI also depend to some extent on how far apart they are to each other. The way these measures treat and penalize violations of FOSD-monotonicity also differs.

Correlation with demographic characteristics. We investigate the correlation between our measure of consistency with OEU, e_*, and various demographic variables available in the data. The exercise is analogous to findings in Choi et al. (2014) that use CCEI.[17]

We find that younger subjects, those who have high cognitive abilities, and those who are working, are closer to being consistent with OEU than older, low cognitive ability, or non-working,

[17] We used Welch's t-test to evaluate differences between demographic groups.

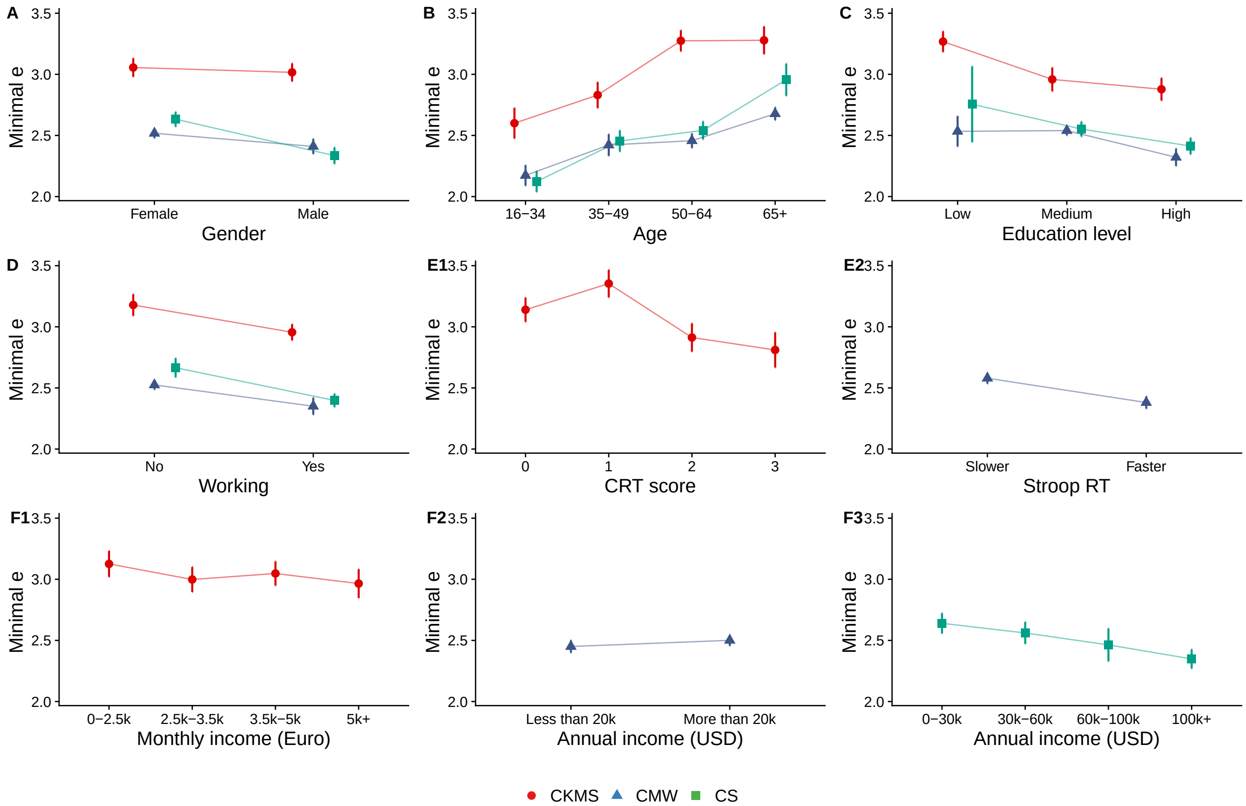

Figure 10: Correlation between e_* and demographic variables. *Notes*: Dots represent mean e_* and bars represent standard errors of means.

subjects. For some of the three experiments we also find that highly educated, high-income subjects, and males, are closer to OEU. Figure 10 summarizes the mean e_* (along with the standard error of the mean) across several socioeconomic categories. We use the same categorization as in Choi et al. (2014) to compare our results with their Figure 3.

We observe statistically significant (at the 5% level) gender differences in CS ($t(1385) = -3.28$, $p = 0.001$) but not in CKMS ($t(1163) = -0.37$, $p = 0.708$) and CMW ($t(717) = -1.46$, $p = 0.144$). Male subjects were on average closer to OEU rationality than female subjects in the CS sample (panel A).

We find significant effects of age in all three datasets. Panel B shows that younger subjects are on average closer to OEU rationality than older subjects (the comparison between age groups 16-34 and 65+ reveals a statistically significant difference in all three datasets; all t-tests give $p < 0.001$).

We observe weak effects of education on e_* (panel C).[18] Subjects with higher education are

[18]The low, medium, and high education levels correspond to primary or pre-vocational secondary education,

on average closer to OEU than those with lower education in CKMS ($t(827) = 3.11$, $p = 0.002$), but the difference is not significant in the CMW and CS ($t(122) = 1.48$, $p = 0.140$ in CMW; $t(47) = 1.08$, $p = 0.286$ in CS).

Panel D shows that subjects who were working at the time of the survey are on average closer to OEU than those who were not ($t(865) = 2.03$, $p = 0.043$ in CKMS; $t(471) = 2.19$, $p = 0.029$ in CMW; $t(974) = 2.79$, $p = 0.005$ in CS).

In panels E1 and E2, we classify subjects according to their Cognitive Reflection Test score (CRT; Frederick, 2005) or average reaction times in the numerical Stroop task.[19] The average e_* for those who correctly answered two questions or more of the CRT is lower than the average for those who answered at most one question ($t(929) = -3.16$, $p = 0.002$). Subjects with lower response times in the numerical Stroop task have significantly lower e_* ($t(1104) = -2.96$, $p = 0.003$).

One of the key findings in Choi et al. (2014) is that consistency with utility maximization as measured by CCEI correlates with household wealth. When we look at the relation between e_* and household income, there is a negative trend but the differences across income brackets are not statistically significant (bracket "0-2.5k" vs. "5k+", $t(528) = 1.02$, $p = 0.309$; panel F1). Panel F2 presents a similar result between subjects who earned more than 20 thousand USD annually or not in the CMW sample ($t(1012) = 0.75$, $p = 0.455$). When we compare poor households (annual income less than 20 thousand USD) and wealthy households (annual income more than 100 thousand USD) from the CS sample, average e_* is significantly smaller for the latter sample ($t(854) = 2.55$, $p = 0.011$; panel F3).

Robustness of the results. The measure e_* is a bound that has to hold across all observations and states (see conditions (4), (5), and (6) in the definitions of e-perturbed OEU). One may wonder how sensitive e_* is to a small number of "bad" choices. Online Appendix F.4 presents two robustness checks. In the first robustness check, we recalculate e_* using subsets of observed choices after dropping one or two "critical mistakes". More precisely, for each subject, we calculate e_* for all combinations of $25 - m$ ($m = 1, 2$) choices and pick the smallest e_* among them. In the second robustness check, we calculate the "average" perturbation necessary to rationalize the data to

pre-university secondary education or senior vocational training, and vocational college or university education, respectively.

[19]CRT consists of three questions, all of which have intuitive and spontaneous, but incorrect, answers, and a deliberative and correct answer. In the numerical Stroop task, subjects are presented with a number, such as 888, and are asked to identify the number of times the digit is repeated (in this example the answer is "3", while an "intuitive" response is "8"). It has been shown that response times in this task capture the subject's cognitive control ability.

mitigate the effect of extreme mistakes. These alternative ways of calculating e_* do not change the general pattern of correlation between e_* and CCEI or e_* and demographic variables. The main empirical results are robust to the presence of a small number of bad choices.

4.3 Minimum Perturbation Test

Our discussion so far has sidestepped one issue: How are we to interpret the absolute magnitude of e_*? *When can we say that e_* is large enough to "reject" consistency with OEU rationality allowing perturbations?* To answer this question, we present a statistical test of the hypothesis that an agent is OEU rational. Of course, with enough noise, any data may be accommodated as OEU rational. We need to quantify the amount of noise needed, which translates into a test based on e_*. The test needs some assumptions, but it gives us a threshold level (a critical value) for e_*. Any value of e_* that exceeds the threshold indicates inconsistency with perturbed OEU at some given statistical significance level.[20]

Our approach is inspired by the methodology laid out in Varian (1985), Echenique et al. (2011), and Echenique et al. (2016). The overall idea is to adopt as the null hypothesis that the data was produced by an agent who is OEU rational, but who may misperceive prices (or, alternatively, who perceives prices correctly, but the data includes prices with measurement error). In other words, we pursue the price-perturbation interpretation of e in Section 3.2. Then we seek to calculate the probability of observing the magnitude of noise that would be needed to reconcile the observed data with the null hypothesis. This probability is the basis for a statistical test: it allows us to formulate critical regions and p-values for the null hypothesis that the agent's behavior is OEU rational.

Now we should explain two related issues. The first is that the whole exercise rests on making an assumption regarding the distribution of noise ε, and in particular its variance. The second issue is why we need to adopt the price perturbation interpretation of e. We work with price perturbations because it allows us to use the induced variability in price to get a handle on the assumptions we need to make on perturbed prices. In particular, it allows us to make an informed decision regarding which variance to assume for the distribution of noise.

More formally, let $D_{\text{obs}} = (p^k, x^k)_{k=1}^K$ denote an observed dataset and $D_{\text{true}} = (\tilde{p}^k, x^k)_{k=1}^K$ denote the "true" (but unobserved) dataset. Let us suppose that observed prices and the "true" prices are

[20] Another way to assess the magnitude of observed e_* is to simulate choice data assuming some behavioral model and calculate e_* on the simulated dataset. In Online Appendix F.5, we compare e_* from real choice data against two sets of simulated choice data. We observe that choices made by subjects in the experiments are closer to OEU than synthetic subjects who choose uniformly randomly on budget lines. We also observe that synthetic subjects who choose randomly while respecting FOSD-monotonicity have e_* that are substantially smaller than real subjects in the experiments.

related in the following way: $\tilde{p}^k_s = p^k_s \tilde{\varepsilon}^k_s$, where $\tilde{\varepsilon}^k_s > 0$ for all $s \in S$ and $k \in \mathcal{K}$. We call $(\tilde{p}^k, x^k)_{k=1}^K$ the "true" dataset and $\tilde{p}^k$ the "true" prices because we can interpret $\tilde{p}^k$ as the (misperceived) prices the agent had in mind when she made a decision, or alternatively, we can interpret p^k as prices measured with error.

The null hypothesis we consider is

$$\text{H}_0: \text{ The "true" dataset } D_{\text{true}} = (\tilde{p}^k, x^k)_{k=1}^K \text{ is OEU rational.}$$

If we could (somehow) observe D_{true}, we could compute

$$\mathcal{E} = \max_{k \in \mathcal{K}, s, t \in S} \frac{\tilde{\varepsilon}^k_s}{\tilde{\varepsilon}^k_t}$$

and reject the null hypothesis if $\mathcal{E}$ is "too large" in the sense that it exceeds a certain threshold. Now, we do not know D_{true}, but we can show that e_*, which we can compute from D_{obs}, provides a lower bound on $\mathcal{E}$ (see Online Appendix E). In consequence, we use e_* as our test statistic.

Now, we need to compare e_* against the distribution of $\mathcal{E}$ under the null hypothesis. What do we assume regarding the variance of ε? To address this question, we take seriously the idea that the agent mistakes perturbed prices for actual prices, and we would like the misperception to be plausible. So it should not be obvious to the agent that one distribution of prices is significantly different from the other. To operationalize this idea, we imagine an agent who conducts a statistical test for the variance of prices. If the true variance of p is σ_0^2 and the variance of $\tilde{p}$ is $\sigma_1^2 > \sigma_0^2$, then the agent could conduct a test for the null of $\sigma^2 = \sigma_0^2$ against the alternative of $\sigma^2 = \sigma_1^2$. We want the variances to be close enough that the agent might reasonably get inconclusive results from such a test (i.e., the agent may reasonably mistake true prices p with perturbed prices $\tilde{p}$, as we assumed). Specifically, *we assume the sum of probabilities of type I and type II errors in this test, $\eta^I + \eta^{II}$, is relatively large.*[21]

Altogether, our statistical test for OEU rationality works as follows: (i) we compute e_*, the relevant test statistic; (ii) we calculate σ_0^2 from observed prices; (iii) we set the pair (η^I, η^{II}) to obtain σ_1^2, which in turn gives the variance of ε; (iv) we simulate the distribution of $\mathcal{E}$, and find the critical value C_α given the significance level α; and (v) we reject the null hypothesis that D_{true} is OEU rational if $e_* > C_\alpha$. Additional details are presented in Online Appendix E.

The results are summarized in Figure 11. Consider, for example, our results for CKMS. The outermost numbers assume that $\eta^I + \eta^{II} = 0.7$. For such numbers, the rejection rates range from 5% to 30%. This means that if prices p and $\tilde{p}$ are close enough so that the agent may misperceive the

[21]The problem of variance is pervasive in statistical implementations of revealed preference tests. See Varian (1985), Echenique et al. (2011), and Echenique et al. (2016), for example. The use of the sum of type I and type II errors to calibrate a variance is new to the present paper.

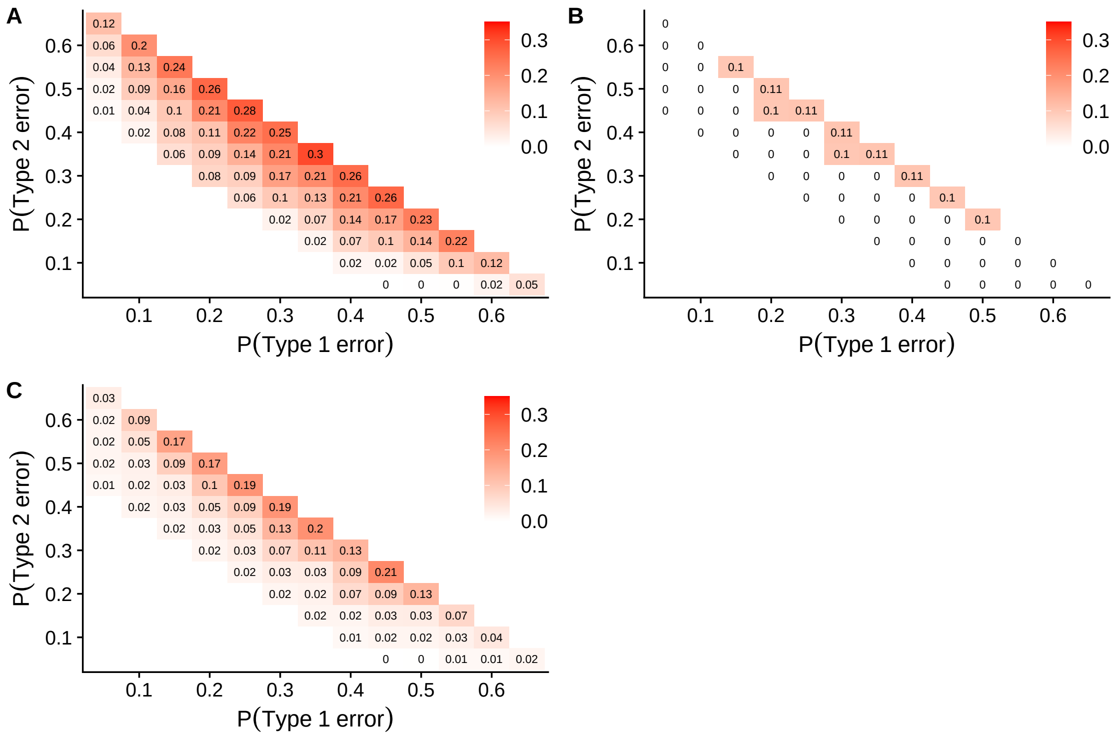

FIGURE 11: Rejection rates under each combination of type I and type II error probabilities (η^I, η^{II}). Panels: (A) CKMS, (B) CMW, (C) CS.

prices and make type I and type II errors with probability 70%, then we can reject the hypothesis that the agent is an OEU maximizer at most 30% of the cases.

Overall, it is fair to say that rejection rates of the hypothesis that the decision maker is an OEU maximizer are modest. Notice also that smaller values of $\eta^I + \eta^{II}$ corresponds to smaller rejection rates. This is because when values of $\eta^I + \eta^{II}$ are smaller (i.e., the decision maker does not misperceive prices much), the difference between p and $\tilde{p}$ should be large, which corresponds to larger variances of ε. Larger variance, in turn, leads to smaller rejection rates. The figure also illustrates that the conclusions of the test are very sensitive to what one assumes about variances, through the assumptions about η^I and η^{II}. But if we look at the largest rejection rates, for the largest values of $\eta^I + \eta^{II}$, we get 30% for CKMS, 11% for CMW, and 21% for CS. Hence, while many subjects in the experiments are inconsistent with OEU, for most of these subjects, our statistical tests would attribute such inconsistency to misperception of prices and do not reject that the subjects are OEU maximizers.

5 Conclusion

We present a measure of deviations from expected utility theory, called minimal e (or e_*), that is based on a revealed preference characterization of the "perturbed" version of the model.

We start with an observation that the empirical content of EU is captured by the relation between prices and marginal rates of substitution. We measure the deviations from EU by the smallest amount of perturbations one needs to add in order to get the "right" relation between prices and marginal rates of substitution. There are three components of the EU model, beliefs, prices, and utilities, which we can perturb, but we can interpret the measure in any of the ways (Theorem 1).

We apply our method to data from three large-scale experiments and find that the measure delivers additional insights on datasets that had been analyzed with CCEI, a measure of consistency with general utility maximization. Our measure can be used as an additional toolkit for data analysis in empirical studies employing choices from linear budgets.

Appendix A

A.1 Proof of Theorem 1

First, we prove a lemma that implies Theorem 1, and is useful for the sufficiency part of Theorem 2. The lemma provides "Afriat inequalities" for the problem at hand.

Lemma 1. *Given $e \in \mathbf{R}_+$, and let $(x^k, p^k)_{k=1}^K$ be a dataset. The following statements are equivalent.*

(a) $(x^k, p^k)_{k=1}^K$ *is e-belief-perturbed OEU rational.*

(b) There are strictly positive numbers v_s^k, λ^k, μ_s^k, for $s \in S$ and $k \in \mathcal{K}$, such that

$$\mu_s^k v_s^k = \lambda^k p_s^k, \quad \text{and} \quad x_s^k > x_{s'}^{k'} \implies v_s^k \leq v_{s'}^{k'}, \tag{8}$$

and for all $k \in \mathcal{K}$ and $s, t \in S$,

$$\frac{1}{1+e} \leq \frac{\mu_s^k / \mu_t^k}{\mu_s^* / \mu_t^*} \leq 1 + e. \tag{9}$$

(c) $(x^k, p^k)_{k=1}^K$ *is e-price-perturbed OEU rational.*

(d) There are strictly positive numbers $\hat{v}_s^k$, $\hat{\lambda}^k$, and ε_s^k for $s \in S$ and $k \in \mathcal{K}$, such that

$$\mu_s^* \hat{v}_s^k = \hat{\lambda}^k \varepsilon_s^k p_s^k, \quad \text{and} \quad x_s^k > x_{s'}^{k'} \implies \hat{v}_s^k \leq \hat{v}_{s'}^{k'},$$

and for all $k \in \mathcal{K}$ and $s, t \in S$,

$$\frac{1}{1+e} \leq \frac{\varepsilon_s^k}{\varepsilon_t^k} \leq 1+e.$$

(e) $(x^k, p^k)_{k=1}^K$ *is e-utility-perturbed OEU rational.*

(f) There are strictly positive numbers $\hat{v}_s^k$, $\hat{\lambda}^k$, *and* $\hat{\varepsilon}_s^k$ *for* $s \in S$ *and* $k \in \mathcal{K}$, *such that*

$$\mu_s^* \hat{\varepsilon}_s^k \hat{v}_s^k = \hat{\lambda}^k p_s^k, \quad \text{and} \quad x_s^k > x_{s'}^{k'} \implies \hat{v}_s^k \leq \hat{v}_{s'}^{k'},$$

and for all $k \in \mathcal{K}$ and $s, t \in S$,

$$\frac{1}{1+e} \leq \frac{\hat{\varepsilon}_s^k}{\hat{\varepsilon}_t^k} \leq 1+e.$$

Proof. The equivalence between (a) and (b), the equivalence between (c) and (d), and the equivalence between (e) and (f) follow from arguments in Echenique and Saito (2015). The equivalence between (d) and (f) with $\varepsilon_s^k = 1/\hat{\varepsilon}_s^k$ for each $k \in \mathcal{K}$ and $s \in S$ is straightforward. Thus, to show the result, it suffices to show that (b) and (d) are equivalent.

To show that (d) implies (b), define $v = \hat{v}$ and $\mu_s^k = \frac{\mu_s^*}{\varepsilon_s^k} / \left(\sum_{s \in S} \frac{\mu_s^*}{\varepsilon_s^k} \right)$ for each $k \in \mathcal{K}$ and $s \in S$ and $\lambda^k = \hat{\lambda}^k / \left(\sum_{s \in S} \frac{\mu_s^*}{\varepsilon_s^k} \right)$ for each $k \in \mathcal{K}$. Then, $\mu^k \in \Delta_{++}(S)$. Since $\mu_s^* \hat{v}_s^k = \hat{\lambda}^k \varepsilon_s^k p_s^k$, we have $\mu_s^k v_s^k = \lambda^k p_s^k$. Moreover, for each $k \in \mathcal{K}$ and $s, t \in S$, $\frac{\varepsilon_s^k}{\varepsilon_t^k} = \frac{\mu_s^k / \mu_t^k}{\mu_s^* / \mu_t^*}$. Hence, $\frac{1}{1+e} \leq \frac{\varepsilon_s^k}{\varepsilon_t^k} \leq 1+e$.

To show that (b) implies (d), for all $s \in S$ define $\hat{v} = v$ and for all $k \in \mathcal{K}$, $\hat{\lambda}^k = \lambda^k$. For all $k \in \mathcal{K}$ and $s \in S$, define $\varepsilon_s^k = \frac{\mu_s^*}{\mu_s^k}$. For each $k \in \mathcal{K}$ and $s \in S$, since $\mu_s^k u_s^k = \lambda^k p_s^k$, we have $\mu_s^* v_s^k = \hat{\lambda}^k \varepsilon_s^k p_s^k$. Finally, for each $k \in \mathcal{K}$ and $s, t \in S$, $\frac{\varepsilon_s^k}{\varepsilon_t^k} = \frac{\mu_s^* / \mu_s^k}{\mu_t^* / \mu_t^k} = \frac{\mu_t^k / \mu_s^k}{\mu_t^* / \mu_s^*}$. Therefore, we obtain $\frac{1}{1+e} \leq \frac{\varepsilon_s^k}{\varepsilon_t^k} \leq 1+e$. □

A.2 Proof of the Necessity Direction of Theorem 2

Lemma 2. *Given $e \in \mathbf{R}_+$, if a dataset is e-belief-perturbed OEU rational, then the dataset satisfies e-PSAROEU.*

Proof. Fix any sequence $(x_{s_i}^{k_i}, x_{s_i'}^{k_i'})_{i=1}^n \equiv \sigma$ of pairs that satisfies conditions (i) and (ii) in Definition 7. By Lemma 1, there exist $v_{s_i}^{k_i}, v_{s_i'}^{k_i'}, \lambda^{k_i}, \lambda^{k_i'}, \mu_{s_i}^{k_i}, \mu_{s_i'}^{k_i'}$ such that $v_{s_i'}^{k_i'} \geq v_{s_i}^{k_i}$ and $v_{s_i}^{k_i} = \frac{\mu_{s_i}^*}{\mu_{s_i}^{k_i}} \lambda^{k_i} \rho_{s_i}^{k_i}$, and $v_{s_i'}^{k_i'} = \frac{\mu_{s_i'}^*}{\mu_{s_i'}^{k_i'}} \lambda^{k_i'} \rho_{s_i'}^{k_i'}$. Thus, we have

$$1 \geq \prod_{i=1}^n \frac{\lambda^{k_i} (\mu_{s_i'}^{k_i'} / \mu_{s_i'}^*) \rho_{s_i}^{k_i}}{\lambda^{k_i'} (\mu_{s_i}^{k_i} / \mu_{s_i}^*) \rho_{s_i'}^{k_i'}} = \prod_{i=1}^n \frac{\mu_{s_i'}^{k_i'} / \mu_{s_i'}^*}{\mu_{s_i}^{k_i} / \mu_{s_i}^*} \prod_{i=1}^n \frac{\rho_{s_i}^{k_i}}{\rho_{s_i'}^{k_i'}},$$

where the second equality holds by condition (ii). Hence,

$$\prod_{i=1}^{n} \frac{\rho_{s_i}^{k_i}}{\rho_{s_i'}^{k_i'}} \leq \prod_{i=1}^{n} \frac{\mu_{s_i}^{k_i}/\mu_{s_i}^{*}}{\mu_{s_i'}^{k_i'}/\mu_{s_i'}^{*}}.$$

In the following, we evaluate the right-hand side. For each (k, s), we first cancel out all the terms μ_s^k that can be canceled out. Then, the number of μ_s^k's that remain in the numerator is $d(\sigma, k, s)$, as in Definition 8. Since the number of terms in the numerator and the denominator must be the same, the number of remaining fractions is $m(\sigma) \equiv \sum_{s\in S} \sum_{k\in\mathcal{K}:d(\sigma,k,s)>0} d(\sigma, k, s)$. So by relabeling the index i to j if necessary, we obtain

$$\prod_{i=1}^{n} \frac{\mu_{s_i}^{k_i}/\mu_{s_i}^{*}}{\mu_{s_i'}^{k_i'}/\mu_{s_i'}^{*}} = \prod_{j=1}^{m(\sigma)} \frac{\mu_{s_j}^{k_j}/\mu_{s_j}^{*}}{\mu_{s_j'}^{k_j'}/\mu_{s_j'}^{*}}.$$

Consider the corresponding sequence $(x_{s_j}^{k_j}, x_{s_j'}^{k_j'})_{j=1}^{m(\sigma)}$. Since the sequence is obtained by canceling out x_s^k from the first element and the second element of the pairs, and since the original sequence $(x_{s_i}^{k_i}, x_{s_i'}^{k_i'})_{i=1}^{n}$ satisfies condition (ii), it follows that $(x_{s_j}^{k_j}, x_{s_j'}^{k_j'})_{j=1}^{m(\sigma)}$ satisfies condition (ii).

By condition (ii), we can assume without loss of generality that $k_j = k_j'$ for each j. Therefore, by the condition on the perturbation,

$$\prod_{j=1}^{m(\sigma)} \frac{\mu_{s_j}^{k_j}/\mu_{s_j}^{*}}{\mu_{s_j'}^{k_j'}/\mu_{s_j'}^{*}} \leq (1+e)^{m(\sigma)}.$$

In conclusion, we obtain that $\prod_{i=1}^{n}(\rho_{s_i}^{k_i}/\rho_{s_i'}^{k_i'}) \leq (1+e)^{m(\sigma)}$. □

A.3 Proof of the Sufficiency Direction of Theorem 2

We need three lemmas to prove the sufficiency direction. The idea behind the argument is the same as in Echenique and Saito (2015). We know from Lemma 1 that it suffices to find a solution to the relevant system of Afriat inequalities. We take logarithms to linearize the Afriat inequalities in Lemma 1. Then we set up the problem to find a solution to the system of linear inequalities.

The first lemma, Lemma 3, shows that e-PSAROEU is sufficient for e-belief-perturbed OEU rationality under the assumption that the logarithms of the prices are rational numbers. The assumption of rational logarithms comes from our use of a version of the theorem of the alternative (see Lemma 12 in Online Appendix B.4): when there is no solution to the linearized Afriat inequalities, a rational solution to the dual system of inequalities exists. Then we construct a violation of e-PSAROEU from the given solution to the dual.

The second lemma, Lemma 4, establishes that we can approximate any dataset satisfying e-PSAROEU with a dataset for which the logarithms of prices are rational, and for which e-PSAROEU is satisfied.

The last lemma, Lemma 5, establishes the result by using another version of the theorem of the alternative, stated as Lemma 11 in Online Appendix B.4.

The rest of the section is devoted to the statement of these lemmas.

Lemma 3. *Given $e \in \mathbf{R}_+$, let a dataset $(x^k, p^k)_{k=1}^k$ satisfy e-PSAROEU. Suppose that $\log(p_s^k) \in \mathbf{Q}$ for all $k \in \mathcal{K}$ and $s \in S$, $\log(\mu_s^*) \in \mathbf{Q}$ for all $s \in S$, and $\log(1+e) \in \mathbf{Q}$. Then there are numbers v_s^k, λ^k, μ_s^k for $s \in S$ and $k \in \mathcal{K}$ satisfying* (8) *and* (9) *in Lemma 1.*

Lemma 4. *Given $e \in \mathbf{R}_+$, let a dataset $(x^k, p^k)_{k=1}^k$ satisfy e-PSAROEU with respect to μ^*. Then for all positive numbers $\overline{\varepsilon}$, there exist a positive real numbers $e' \in [e, e+\overline{\varepsilon}]$, $\mu'_s \in [\mu_s^* - \overline{\varepsilon}, \mu_s^* + \overline{\varepsilon}]$, and $q_s^k \in [p_s^k - \overline{\varepsilon}, p_s^k]$ for all $s \in S$ and $k \in \mathcal{K}$ such that $\log q_s^k \in \mathbf{Q}$ for all $s \in S$ and $k \in \mathcal{K}$, $\log(\mu'_s) \in \mathbf{Q}$ for all $s \in S$, and $\log(1+e') \in \mathbf{Q}$, $\mu' \in \Delta_{++}(S)$, and the dataset $(x^k, q^k)_{k=1}^k$ satisfy e'-PSAROEU with respect to μ'.*

Lemma 5. *Given $e \in \mathbf{R}_+$, let a dataset $(x^k, p^k)_{k=1}^k$ satisfy e-PSAROEU with respect to μ. Then there are numbers v_s^k, λ^k, μ_s^k for $s \in S$ and $k \in \mathcal{K}$ satisfying* (8) *and* (9) *in Lemma 1.*

A.3.1 Proof of Lemma 3

The proof is similar to the proof of the main result in Echenique and Saito (2015), which corresponds to the case $e = 0$. By log-linearizing the equation in system (8) and the inequality (9) in Lemma 1, we have for all $s \in S$ and $k \in \mathcal{K}$, such that

$$\log \mu_s^k + \log v_s^k = \log \lambda^k + \log p_s^k, \tag{10}$$

$$x_s^k > x_{s'}^{k'} \implies \log v_s^k \leq \log v_{s'}^{k'}, \tag{11}$$

and for all $k \in \mathcal{K}$ and $s, t \in S$,

$$-\log(1+e) + \log \mu_s^* - \log \mu_t^* \leq \log \mu_s^k - \log \mu_t^k \leq \log(1+e) + \log \mu_s^* - \log \mu_t^*. \tag{12}$$

We are going to write the system of inequalities (10)-(12) in matrix form, following Echenique and Saito (2015) with some modifications.

Let A be a matrix with $K \times |S|$ rows and $2(K \times |S|) + K + 1$ columns, defined as follows: We have one row for every pair (k, s), two columns for every pair (k, s), one column for each k, and one last column. In the row corresponding to (k, s), the matrix has zeroes everywhere with the following exceptions: it has 1's in columns for (k, s); it has a -1 in the column for k; it has $-\log p_s^k$ in the very last column. The matrix A looks as follows:

$$
\begin{array}{c}
 \\ \\ (k,s) \\ (k,t) \\ (l,s) \\ (l,t) \\ \\
\end{array}
\begin{array}{c}
\begin{array}{ccccccc|cccccc|cccc|c}
\cdots & v_s^k & v_t^k & v_s^l & v_t^l & \cdots & \cdots & \mu_s^k & \mu_t^k & \mu_s^l & \mu_t^l & \cdots & \cdots & \lambda^k & \lambda^l & \cdots & p
\end{array}
\\
\left[\begin{array}{cccccc|cccccc|cccc|c}
 & \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots & & \vdots \\
\cdots & 1 & 0 & 0 & 0 & \cdots & \cdots & 1 & 0 & 0 & 0 & \cdots & \cdots & -1 & 0 & \cdots & -\log p_s^k \\
\cdots & 0 & 1 & 0 & 0 & \cdots & \cdots & 0 & 1 & 0 & 0 & \cdots & \cdots & -1 & 0 & \cdots & -\log p_t^k \\
\cdots & 0 & 0 & 1 & 0 & \cdots & \cdots & 0 & 0 & 1 & 0 & \cdots & \cdots & 0 & -1 & \cdots & -\log p_s^l \\
\cdots & 0 & 0 & 0 & 1 & \cdots & \cdots & 0 & 0 & 0 & 1 & \cdots & \cdots & 0 & -1 & \cdots & -\log p_t^l \\
 & \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots & & \vdots
\end{array}\right]
\end{array}.
$$

Next, we write the system of inequalities (11) and (12) in a matrix form. There is one row in matrix B for each pair (k, s) and (k', s') for which $x_s^k > x_{s'}^{k'}$. In the row corresponding to $x_s^k > x_{s'}^{k'}$, we have zeroes everywhere with the exception of a -1 in the column for (k, s) and a 1 in the column for (k', s'). Matrix B has additional rows, that capture the system of inequalities (12), as follows:

$$
\begin{array}{c}
\begin{array}{ccccccc|cccccc|cccc|c}
\cdots & v_s^k & v_t^k & v_s^l & v_t^l & \cdots & \cdots & \mu_s^k & \mu_t^k & \mu_s^l & \mu_t^l & \cdots & \cdots & \lambda^k & \lambda^l & \cdots & p
\end{array}
\\
\left[\begin{array}{cccccc|cccccc|cccc|c}
 & \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots & & \vdots \\
\cdots & 0 & 0 & 0 & 0 & \cdots & \cdots & 1 & -1 & 0 & 0 & \cdots & \cdots & 0 & 0 & \cdots & \log(1+e) - \log \mu_s^* + \log \mu_t^* \\
\cdots & 0 & 0 & 0 & 0 & \cdots & \cdots & -1 & 1 & 0 & 0 & \cdots & \cdots & 0 & 0 & \cdots & \log(1+e) + \log \mu_s^* - \log \mu_t^* \\
\cdots & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & -1 & 1 & \cdots & \cdots & 0 & 0 & \cdots & \log(1+e) + \log \mu_s^* - \log \mu_t^* \\
\cdots & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 1 & -1 & \cdots & \cdots & 0 & 0 & \cdots & \log(1+e) - \log \mu_s^* + \log \mu_t^* \\
 & \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots & & \vdots
\end{array}\right]
\end{array}.
$$

Finally, we have a matrix E which has a single row and has zeroes everywhere except for 1 in the last column.

To sum up, there is a solution to the system (10)-(12) if and only if there is a vector $u \in \mathbf{R}^{2(K\times|S|)+K+1}$ that solves the system of equations and linear inequalities

$$
S1 : \begin{cases} Au = 0, \\ Bu \geq 0, \\ Eu > 0. \end{cases}
$$

The entries of A, B, and E are either 0, 1 or -1, with the exception of the last column of A and B. Under the hypotheses of the lemma we are proving, the last column consists of rational numbers. By Motzkin's theorem, then, there is such a solution u to $S1$ if and only if there is no rational vector (θ, η, π) that solves the system of equations and linear inequalities

$$S2: \begin{cases} \theta \cdot A + \eta \cdot B + \pi \cdot E = 0, \\ \eta \geq 0, \\ \pi > 0. \end{cases}$$

In the following, we shall prove that the non-existence of a solution u implies that the dataset must violate e-PSAROEU. Suppose then that there is no solution u and let (θ, η, π) be a rational vector as above, solving system $S2$.

The outline of the rest of the proof is similar to the proof of Echenique and Saito (2015). Since (θ, η, π) are rational vectors, by multiplying a large enough integer, we can make the vectors integers. Then we transform the matrices A and B using θ and η. (i) If $\theta_r > 0$, then create θ_r copies of the rth row; (ii) omitting row r when $\theta_r = 0$; and (iii) if $\theta_r < 0$, then θ_r copies of the rth row multiplied by -1.

Similarly, we create a new matrix by including the same columns as B and η_r copies of each row (and thus omitting row r when $\eta_r = 0$; recall that $\eta_r \geq 0$ for all r).

By using the transformed matrices and the fact that $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$ and $\eta \geq 0$, we can prove the following claims:

Claim. *There exists a sequence $(x^{k_i}_{s_i}, x^{k'_i}_{s'_i})^{n^*}_{i=1}$ of pairs that satisfies conditions (i) and (ii) in Definition 7.*

Proof. We can construct a sequence $(x^{k_i}_{s_i}, x^{k'_i}_{s'_i})^{n^*}_{i=1}$ in a similar way to the proof of Lemma 11 of Echenique and Saito (2015). By construction, the sequence satisfies condition (i) that $x^{k_i}_{s_i} > x^{k'_i}_{s'_i}$ for all i.

In the following, we show that the sequence satisfies condition (ii) that each k appears as k_i the same number of times it appears as k'_i. Let $n(x^k_s) \equiv \#\{i \mid x^k_s = x^{k_i}_{s_i}\}$ and $n'(x^k_s) \equiv \#\{i \mid x^k_s = x^{k'_i}_{s'_i}\}$. It suffices to show that for each $k \in \mathcal{K}$, $\sum_{s \in S} \left[n(x^k_s) - n'(x^k_s) \right] = 0$.

Recall our construction of the matrix B. We have a constraint for each triple (k, s, t) with $s < t$. Denote the weight on the rows capturing $\frac{\mu^k_s/\mu^k_t}{\mu^*_s/\mu^*_t} \leq 1 + e$ by $\eta(k, s, t)$ and $1 + e \leq \frac{\mu^k_s/\mu^k_t}{\mu^*_s/\mu^*_t}$ by $\eta(k, t, s)$.

For each $k \in \mathcal{K}$ and $s \in S$, in the column corresponding to μ^k_s in matrix A, remember that we have 1 if we have $x^k_s = x^{k_i}_{s_i}$ for some i and -1 if we have $x^k_s = x^{k'_i}_{s'_i}$ for some i. This is because a row in A must have 1 (-1) in the column corresponding to v^k_s if and only if it has 1 (-1, respectively) in the column corresponding to μ^k_s. By summing over the column corresponding to μ^k_s, we have $n(x^k_s) - n'(x^k_s)$.

Now we consider matrix B. In the column corresponding to μ_s^k, we have 1 in the row multiplied by $\eta(k,t,s)$ and -1 in the row multiplied by $\eta(k,s,t)$. By summing over the column corresponding to μ_s^k, we also have $-\sum_{t\neq s}\eta(k,s,t)+\sum_{t\neq s}\eta(k,t,s)$.

For each $k\in\mathcal{K}$ and $s\in S$, the column corresponding to μ_s^k of matrices A and B must sum up to zero; so we have

$$n(x_s^k)-n'(x_s^k)+\sum_{t\neq s}\left[-\eta(k,s,t)+\eta(k,t,s)\right]=0. \tag{13}$$

Hence for each $k\in\mathcal{K}$, $\sum_{s\in S}\left[n(x_s^k)-n'(x_s^k)\right]=0$. □

Claim. $\prod_{i=1}^{n^*}(\rho_{s_i}^{k_i}/\rho_{s_i'}^{k_i'})>(1+e)^{m(\sigma^*)}$.

Proof. By (13), for each $s\in S$,

$$\sum_{k\in\mathcal{K}}\sum_{s\in S}\sum_{t\neq s}\left[\eta(k,s,t)-\eta(k,t,s)\right]\log\mu_s^*=\sum_{k\in\mathcal{K}}\sum_{s\in S}\left[n(x_s^k)-n'(x_s^k)\right]\log\mu_s^*=\sum_{i=1}^{n^*}\log\frac{\mu_{s_i}^*}{\mu_{s_i'}^*},$$

where the last equality holds by the definition of n and n'. Moreover, since $d(\sigma^*,k,s)=n(x_s^k)-n'(x_s^k)=\sum_{t\neq s}\left[\eta(k,s,t)-\eta(k,t,s)\right]\leq\sum_{t\neq s}\eta(k,s,t)$, we have

$$m(\sigma^*)\equiv\sum_{s\in S}\sum_{k\in\mathcal{K}:d(\sigma^*,k,s)>0}d(\sigma^*,k,s)=\sum_{s\in S}\sum_{k\in\mathcal{K}}\min\{n(x_s^k)-n'(x_s^k),0\}\leq\sum_{s\in S}\sum_{k\in\mathcal{K}}\sum_{t\neq s}\eta(k,s,t).$$

By the equality and the inequality above and by the fact that the last column must sum up to zero and E has one at the last column, we have

$$\begin{aligned}0&>\sum_{i=1}^{n^*}\log\frac{p_{s_i'}^{k_i'}}{p_{s_i}^{k_i}}+\log(1+e)\sum_{k\in\mathcal{K}}\sum_{s\in S}\sum_{t\neq s}\eta(k,s,t)+\sum_{k\in\mathcal{K}}\sum_{s\in S}\sum_{t\neq s}(\eta(k,s,t)-\eta(k,t,s))\log\mu_s^*\\&=\sum_{i=1}^{n^*}\log\frac{p_{s_i'}^{k_i'}}{p_{s_i}^{k_i}}-\sum_{i=1}^{n^*}\log\frac{\mu_{s_i}^*}{\mu_{s_i'}^*}+\log(1+e)\sum_{k\in\mathcal{K}}\sum_{s\in S}\sum_{t\neq s}\eta(k,s,t)\\&=\sum_{i=1}^{n^*}\log\frac{\rho_{s_i'}^{k_i'}}{\rho_{s_i}^{k_i}}+\log(1+e)\sum_{k\in\mathcal{K}}\sum_{s\in S}\sum_{t\neq s}\eta(k,s,t)\geq\sum_{i=1}^{n^*}\log\frac{\rho_{s_i'}^{k_i'}}{\rho_{s_i}^{k_i}}+\log(1+e)m(\sigma^*).\end{aligned}$$

That is, $\sum_{i=1}^{n^*}\log(\rho_{s_i}^{k_i}/\rho_{s_i'}^{k_i'})>m(\sigma^*)\log(1+e)$. This is a contradiction. □

A.3.2 Proof of Lemma 4

Let $\mathcal{X}=\{x_s^k\mid k\in\mathcal{K},s\in S\}$. Consider the set of sequences that satisfy conditions (i) and (ii) in Definition 7:

$$\Sigma=\left\{(x_{s_i}^{k_i},x_{s_i'}^{k_i'})_{i=1}^n\subseteq\mathcal{X}^2\;\middle|\;\begin{array}{l}(x_{s_i}^{k_i},x_{s_i'}^{k_i'})_{i=1}^n\text{ satisfies conditions (i) and (ii)}\\\text{in Definition 7 for some }n\end{array}\right\}.$$

For each sequence $\sigma \in \Sigma$, we define a vector $t_\sigma \in \mathbf{N}^{K^2|S|^2}$. For each pair $(x^{k_i}_{s_i}, x^{k'_i}_{s'_i})$, we shall identify the pair with $((k_i, s_i), (k'_i, s'_i))$. Let $t_\sigma((k,s),(k',s'))$ be the number of times that the pair $(x^k_s, x^{k'}_{s'})$ appears in the sequence σ. One can then describe the satisfaction of e-PSAROEU by means of the vectors t_σ. Observe that t depends only on $(x^k)^K_{k=1}$ in the dataset $(x^k, p^k)^K_{k=1}$. It does not depend on prices.

For each $((k,s),(k',s'))$ such that $x^k_s > x^{k'}_{s'}$, define $\delta((k,s),(k',s')) = \log(p^k_s / p^{k'}_{s'})$. And define $\delta((k,s),(k',s')) = 0$ when $x^k_s \leq x^{k'}_{s'}$. Then, δ is a $K^2|S|^2$-dimensional real-valued vector. If $\sigma = (x^{k_i}_{s_i}, x^{k'_i}_{s'_i})^n_{i=1}$, then

$$\delta \cdot t_\sigma = \sum_{((k,s),(k',s')) \in (KS)^2} \delta((k,s),(k',s')) t_\sigma((k,s),(k',s')) = \log\left(\prod_{i=1}^n \frac{\rho^{k_i}_{s_i}}{\rho^{k'_i}_{s'_i}}\right).$$

So the dataset satisfies e-PSAROEU with respect to μ if and only if $\delta \cdot t_\sigma \leq m(\sigma)\log(1+e)$ for all $\sigma \in \Sigma$.

Enumerate the elements in $\mathcal{X}$ in increasing order: $y_1 < y_2 < \cdots < y_N$, and fix an arbitrary $\underline{\xi} \in (0,1)$. We shall construct by induction a sequence $(\varepsilon^k_s(n))^N_{n=1}$, where $\varepsilon^k_s(n)$ is defined for all (k,s) with $x^k_s = y_n$.

By the denseness of the rational numbers, and the continuity of the exponential function, for each (k,s) such that $x^k_s = y_1$, there exists a positive number $\varepsilon^k_s(1)$ such that $\log(\rho^k_s \varepsilon^k_s(1)) \in \mathbf{Q}$ and $\underline{\xi} < \varepsilon^k_s(1) < 1$. Let $\varepsilon(1) = \min\{\varepsilon^k_s(1) \mid x^k_s = y_1\}$.

In second place, for each (k,s) such that $x^k_s = y_2$, there exists a positive $\varepsilon^k_s(2)$ such that $\log(\rho^k_s \varepsilon^k_s(2)) \in \mathbf{Q}$ and $\underline{\xi} < \varepsilon^k_s(2) < \varepsilon(1)$. Let $\varepsilon(2) = \min\{\varepsilon^k_s(2) \mid x^k_s = y_2\}$.

In third place, and reasoning by induction, suppose that $\varepsilon(n)$ has been defined and that $\underline{\xi} < \varepsilon(n)$. For each (k,s) such that $x^k_s = y_{n+1}$, let $\varepsilon^k_s(n+1) > 0$ be such that $\log(\rho^k_s \varepsilon^k_s(n+1)) \in \mathbf{Q}$, and $\underline{\xi} < \varepsilon^k_s(n+1) < \varepsilon(n)$. Let $\varepsilon(n+1) = \min\{\varepsilon^k_s(n+1) \mid x^k_s = y_n\}$.

This defines the sequence $(\varepsilon^k_s(n))^N_{n=1}$ by induction. Note that $\varepsilon^k_s(n+1)/\varepsilon(n) < 1$ for all n. Let $\bar{\xi} < 1$ be such that $\varepsilon^k_s(n+1)/\varepsilon(n) < \bar{\xi}$.

For each $k \in \mathcal{K}$ and $s \in S$, let $\hat{\rho}^k_s = \rho^k_s \varepsilon^k_s(n)$, where n is such that $x^k_s = y_n$. Choose $\mu' \in \Delta_{++}(S)$ such that for all $s \in S$ $\log \mu'_s \in \mathbf{Q}$ and $\mu'_s \in [\bar{\xi}\mu_s, \mu_s/\bar{\xi}]$ for all $s \in S$. Such μ' exists by the denseness of the rational numbers. Now for each $k \in \mathcal{K}$ and $s \in S$, define

$$q^k_s = \frac{\hat{\rho}^k_s}{\mu'_s}. \tag{14}$$

Then, $\log q^k_s = \log \hat{\rho}^k_s - \log \mu'_s \in \mathbf{Q}$.

We claim that the dataset $(x^k, q^k)^K_{k=1}$ satisfies e'-PSAROEU with respect to μ'. Let δ^* be defined from $(q^k)^K_{k=1}$ in the same manner as δ was defined from $(\rho^k)^K_{k=1}$.

For each pair $((k,s),(k',s'))$ with $x_s^k > x_{s'}^{k'}$, if n and m are such that $x_s^k = y_n$ and $x_{s'}^{k'} = y_m$, then $n > m$. By definition of ε,

$$\frac{\varepsilon_s^k(n)}{\varepsilon_{s'}^{k'}(m)} < \frac{\varepsilon_s^k(n)}{\varepsilon(m)} < \bar{\xi} < 1.$$

Hence,

$$\delta^*((k,s),(k',s')) = \log \frac{\rho_s^k \varepsilon_s^k(n)}{\rho_{s'}^{k'} \varepsilon_{s'}^{k'}(m)} < \log \frac{\rho_s^k}{\rho_{s'}^{k'}} + \log \bar{\xi} < \log \frac{\rho_s^k}{\rho_{s'}^{k'}} = \delta((k,s),(k',s')).$$

Now, we choose e' such that $e' \geq e$ and $\log(1+e') \in \mathbf{Q}$.

Thus, for all $\sigma \in \Sigma$, $\delta^* \cdot t_\sigma \leq \delta \cdot t_\sigma \leq m(\sigma)\log(1+e) \leq m(\sigma)\log(1+e')$ as $t_\cdot \geq 0$ and the dataset $(x^k, p^k)_{k=1}^K$ satisfies e-PSAROEU with respect to μ.

Thus the dataset $(x^k, q^k)_{k=1}^K$ satisfies e'-PSAROEU with respect to μ'. Finally, note that $\underline{\xi} < \varepsilon_s^k(n) < 1$ for all n and each $k \in \mathcal{K}, s \in S$. So that by choosing $\underline{\xi}$ close enough to 1, we can take $\hat{\rho}$ to be as close to ρ as desired. By the definition, we also can take μ' to be as close to μ as desired. Consequently, by (14), we can take $(q^k)_{k=1}^K$ to be as close to $(p^k)_{k=1}^K$ as desired. We also can take e' to be as close to e as desired.

A.3.3 Proof of Lemma 5

We use the following notational convention: For a matrix D with $2(K \times |S|) + K + 1$ columns, write D_1 for the submatrix of D corresponding to the first $K \times |S|$ columns; let D_2 be the submatrix corresponding to the following $K \times |S|$ columns; D_3 correspond to the next K columns; and D_4 to the last column. Thus, $D = [D_1 | D_2 | D_3 | D_4]$.

Consider the system comprised by (10), (11), and (12) in the proof of Lemma 3. Let A, B, and E be constructed from the dataset as in the proof of Lemma 3. The difference with respect to Lemma 3 is that now the entries of A_4 and B_4 may not be rational. Note that the entries of E, B, and A_i, $i = 1,2,3$ are rational.

Suppose, towards a contradiction, that there is no solution to the system comprised by (10), (11), and (12). Then, by the argument in the proof of Lemma 3 there is no solution to system $S1$. Lemma 11 (in Online Appendix B.4) with $\mathbf{F} = \mathbf{R}$ implies that there is a real vector (θ, η, π) such that $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$ and $\eta \geq 0, \pi > 0$. Recall that $E_4 = 1$, so we obtain that $\theta \cdot A_4 + \eta \cdot B_4 + \pi = 0$.

Consider $(q^k)_{k=1}^K$, μ', and e' be such that the dataset $(x^k, q^k)_{k=1}^K$ satisfies e'-PSAROEU with respect to μ', and $\log q_s^k \in \mathbf{Q}$ for all k and s, $\log \mu'_s \in \mathbf{Q}$ for all $s \in S$, and $\log(1+e') \in \mathbf{Q}$. (Such $(q^k)_{k=1}^K$, μ', and e' exist by Lemma 4.) Construct matrices A', B', and E' from this dataset in the same way as A, B, and E is constructed in the proof of Lemma 3. Note that only the prices, the objective probabilities, and the bounds are different. So $E' = E$ and $A'_i = A_i$ and $B'_i = B_i$ for $i = 1,2,3$. Only A'_4 and B'_4 may be different from A_4 and B_4, respectively.

By Lemma 4, we can choose q^k, μ', and e' such that $|(\theta \cdot A'_4 + \eta \cdot B'_4) - (\theta \cdot A_4 + \eta \cdot B_4)| < \pi/2$. We have shown that $\theta \cdot A_4 + \eta \cdot B_4 = -\pi$, so the choice of q^k, μ', and e' guarantees that $\theta \cdot A'_4 + \eta \cdot B'_4 < 0$. Let $\pi' = -\theta \cdot A'_4 - \eta \cdot B'_4 > 0$.

Note that $\theta \cdot A'_i + \eta \cdot B'_i + \pi' E_i = 0$ for $i = 1, 2, 3$, as (θ, η, π) solves system $S2$ for matrices A, B and E, and $A'_i = A_i$, $B'_i = B_i$ and $E_i = 0$ for $i = 1, 2, 3$. Finally, $\theta \cdot A'_4 + \eta \cdot B'_4 + \pi' E_4 = \theta \cdot A'_4 + \eta \cdot B'_4 + \pi' = 0$. We also have that $\eta \geq 0$ and $\pi' > 0$. Therefore θ, η, and π' constitute a solution to $S2$ for matrices A', B', and E'.

Lemma 11 then implies that there is no solution to system $S1$ for matrices A', B', and E'. So there is no solution to the system comprised by (10), (11), and (12) in the proof of Lemma 3. However, this contradicts Lemma 3 because the dataset (x^k, q^k) satisfies e'-PSAROEU with μ', $\log(1 + e') \in \mathbf{Q}$, $\log \mu'_s \in \mathbf{Q}$ for all $s \in S$, and $\log q^k_s \in \mathbf{Q}$ for all $k \in \mathcal{K}$ and $s \in S$.

References

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

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

Aguiar, V. H. and N. Kashaev (2021): "Stochastic Revealed Preferences with Measurement Error," *Review of Economic Studies*, 88, 2042–2093.

Ahn, D. S., S. Choi, D. Gale, and S. Kariv (2014): "Estimating Ambiguity Aversion in a Portfolio Choice Experiment," *Quantitative Economics*, 5, 195–223.

Allen, R. and J. Rehbeck (2020): "Satisficing, Aggregation, and Quasilinear Utility," SSRN: `https://doi.org/10.2139/ssrn.3180302`.

——— (2021): "Measuring Rationality: Percentages vs Expenditures," *Theory and Decision*, 91, 265–277.

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

Carvalho, L., S. Meier, and S. W. Wang (2016): "Poverty and Economic Decision Making: Evidence from Changes in Financial Resources at Payday," *American Economic Review*, 106, 260–284.

Carvalho, L. and D. Silverman (2019): "Complexity and Sophistication," NBER Working Paper No. 26036.

Chambers, C. P. and F. Echenique (2016): *Revealed Preference Theory*, Cambridge: Cambridge

University Press.

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

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

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

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

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

Dziewulski, P. (2016): "Eliciting the Just-Noticeable Difference," Unpublished manuscript.

——— (2020): "Just-Noticeable Difference as a Behavioural Foundation of the Critical Cost-Efficiency Index," *Journal of Economic Theory*, 105071.

Echenique, F. (2021): "On the Meaning of the Critical Cost Efficiency Index," arXiv:2109.06354.

Echenique, F., T. Imai, and K. Saito (2016): "Testable Implications of Models of Intertemporal Choice: Exponential Discounting and Its Generalizations," Caltech HSS Working Paper 1388.

——— (2020): "Testable Implications of Models of Intertemporal Choice: Exponential Discounting and Its Generalizations," *American Economic Journal: Microeconomics*, 12, 114–143.

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

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

Frederick, S. (2005): "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19, 25–42.

Friedman, D., S. Habib, D. James, and S. Crockett (2022): "Varieties of Risk Preference Elicitation," *Games and Economic Behavior*, 133, 58–76.

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

Griliches, Z. (1986): "Economic Data Issues," in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, Elsevier, vol. 3, 1465–1514.

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

Kübler, F., L. Selden, and X. Wei (2014): "Asset Demand Based Tests of Expected Utility Maximization," *American Economic Review*, 104, 3459–3480.

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

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.

Polisson, M. and J. K.-H. Quah (2022): "Rationalizability, Cost-Rationalizability, and Afriat's Efficiency Index," Bristol Economics Discussion Papers 22/754.

Polisson, M., J. K.-H. Quah, and L. Renou (2020): "Revealed Preferences over Risk and Uncertainty," *American Economic Review*, 110, 1782–1820.

Samuelson, P. A. (1938): "A Note on the Pure theory of Consumer's Behaviour," *Economica*, 5, 61–71.

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

——— (1985): "Non-Parametric Analysis of Optimizing Behavior with Measurement Error," *Journal of Econometrics*, 30, 445–458.

——— (1990): "Goodness-of-Fit in Optimizing Models," *Journal of Econometrics*, 46, 125–140.