Pre-Analysis Plan: Robust Estimation of Risk Preferences

Shunto Kobayashi[1], Aldo Lucia[1]
[1]Caltech

July 19, 2023

1 Motivation

The motivation for this project stems from an extensive body of research focusing on estimating models of decision-making under risk using incentivized choices in experimental settings. A prevalent practice for obtaining quantitative measures of risk preferences is structural estimation, which DellaVigna (2018) defines in the field of behavioral economics as the "estimation of a model on data that recover estimates (and confidence intervals) for some key behavioral parameters." Within this framework, researchers need to make various assumptions about the decision model to consider, the functional assumptions within the selected model, and how to incorporate noise into the model. For instance, in the classic empirical framework proposed by Hey and Orme (1994), a decision-maker selects lottery p over lottery q if

$$V(p,q)+\varepsilon \geq 0,$$

where $V(p,q)$ is a quantity greater or equal to zero if the decision-maker prefers lottery p over lottery q, and ε is an error term normally distributed with a mean of zero and a variance of one. The specific functional form of $V(p,q)$ depends on the assumptions made regarding the decision model. For example, under Expected Utility (EU), $V(p,q)$ represents the difference in expected utilities between lottery p and lottery q.

Quantitative measures of risk preferences are important because they enable researchers to formulate predictions. However, these predictions often depend heavily on the assumptions associated with the chosen decision model and the error term. The large number of potential decision models, each bearing its own set of assumptions, adds another layer of complexity to this process. Experimental evidence displaying behavior inconsistent with EU has led to the development of numerous alternative models. While some of these have been acknowledged as leading descriptive models of behavior under risk, their ability to rationalize behavior can vary greatly across studies. To address these challenges, we propose a novel parametric approach to obtain quantitative measures of risk preferences that does not rely on specific decision models. We subsequently implement this approach in an experiment to assess its effectiveness.

This project introduces a novel methodological framework to investigate preferences under risk that differ from EU theory by relaxing its fundamental and most disputed assumption: the independence axiom. Our goal is to identify all choices between risky lotteries that can be accurately rationalized and predicted by EU. To achieve this, we develop an empirical approach that estimates the EU-core of a preference relation, which represents the largest subrelation that satisfies the independence axiom (Cerreia-Vioglio, 2009).

Estimates derived from studying the EU-core hold the potential to enhance the accuracy of the predictions, as they are not tied to specific decision models. However, the EU-core generally constitutes an incomplete binary relation, which means that predictions derived from it might be weaker than those derived from estimating a specific decision model. In particular, our estimates may not always allow us to predict a choice between two lotteries. Consequently, there exists a trade-off between the stringency of the assumptions that researchers are willing to make to estimate risk preferences and the granularity achievable in the resulting predictions. The main objective of this project is to illustrate that an empirical analysis of the EU-core can generate valuable predictions that, although less detailed, are more accurate in-sample and out-of-sample.

The rest of the pre-analysis plan is organized as follows. Section 2 introduces our methodology. Section 3 describes the experimental design. Section 4 outlines our research questions and how we plan to analyze the data.

2 Methodology

Given any reflexive, transitive, and continuous preference relation $\succsim$, its EU-core is the subrelation $\succsim^*$ such that for all lotteries p, q, r and for all $\lambda \in (0,1]$,[1]

$$p \succsim^* q \Leftrightarrow \lambda p + (1-\lambda) r \succsim \lambda q + (1-\lambda) r.$$

That is, $p \succsim^* q$ whenever the decision-maker prefers p to q and mixing both lotteries p and q with a third common lottery r does not affect the relative preferences of the decision-maker between p and q.[2] Cerreia-Vioglio (2009) proves that $\succsim^*$ is the greatest subrelation of $\succsim$ that satisfies the independence axiom.[3] If a preference $\succsim$ violates the independence axiom, then its EU-core is an incomplete preference relation and admits a multi-utility representation. In particular, there exists a set of utilities $\mathscr{W}$ such that for all lotteries p and q, we have $p \succsim^* q$ if and only if the difference in expected utilities between p and q is non-negative for all utilities within the set $\mathscr{W}$.

Our empirical approach consists in obtaining information about a preference by estimating the set of utilities that represents its EU-core. In this way, we can obtain estimates and generate predictions that do not rely on specific decision models.

[1] $\succsim^*$ is a subrelation of $\succsim$ if for all lotteries p and q, $p \succsim^* q$ implies $p \succsim q$.

[2] We study the expected utility core by considering only "one-stage" lottery mixtures, rather than two-stage compound lotteries. In other words, we focus on mixture independence, rather than compound independence, as defined in Segal (1990).

[3] That is, if $\succsim^{**}$ is another subrelation of $\succsim$ that satisfies the independence axiom, then $\succsim^{**}$ is a subrelation of $\succsim^*$.

2.1 Econometric Specification

We detail our empirical framework in the context of our experimental design. Our study involves lotteries over a finite set of K monetary prizes $X = \{x_1, \ldots, x_K\}$, with $x_1 < x_2 < \cdots < x_K$. We consider a set of L utility functions $\mathscr{W} = \{v_1, \ldots, v_L\}$, each utility $v_l \colon X \to \mathbb{R}$ is representable as a vector with its k-th component, v_{lk}, being equal to $v_l(x_k)$. We restrict our attention to normalized sets of utilities, setting $v_{11} = \cdots = v_{L1} = 0$ and $v_{1K} = \cdots = v_{LK} = 1$. This means all utilities assign zero to the worst outcome x_1 and one to the best outcome x_K. Moreover, we assume all utilities are weakly increasing.[4] In our estimation procedure, the number of utilities L is kept constant. We aim to estimate the set of utilities for different values of L and compare them by performing likelihood ratio tests for nested models. For instance, we can test the assumption that preferences satisfy the independence axiom by setting $L = 1$ as the null hypothesis.

We define $I = \{1, \ldots, N\}$ as a set of subjects in our experiment, $\triangle(X)$ as the set of lotteries over X, and by $\mathscr{D} \subseteq \triangle(X)^2$ as a subset of pairs of lotteries where the subjects express their preferences. An empirical analysis of the expected utility (EU) core requires evaluating whether it holds that $p \succsim_i^* q$ or $q \succsim_i^* p$ for each subject $i \in I$ and each pair of lotteries $(p, q) \in \mathscr{D}$. In the traditional estimation framework, where the objective is to estimate a subject's preferences, the choices made by the subjects can be used directly as inputs for the estimation. However, when the focus of the estimation shifts from a preference relation to its EU-core, additional information becomes necessary. Specifically, we need to assess whether the choices made by the subjects indicate a violation of the independence axiom.

By observing the subjects' choices in experimental settings that test the independence axiom, we construct an index, $Core_i$, for each subject i as follows: for each pair of lotteries $(p, q) \in \mathscr{D}$,

$$Core_i(p,q) \coloneqq \begin{cases} 3 & \text{if there is no experimental evidence against } p \succsim_i^* q \\ 2 & \text{if there is no experimental evidence against } q \succsim_i^* p \\ 1 & \text{otherwise.} \end{cases}$$

Section 4 describes the different experimental strategies we implement to search for evidence against $p \succsim_i^* q$ and $q \succsim_i^* p$. In this section, we consider the index $Core_i$ as a given and discuss how we utilize it in our estimation procedure.

We define $V(p, q; v_l)$ as the difference in expected utilities between lottery p and lottery q, given Bernoulli utility function v_l. For each subject $i \in I$, utility $v_l \in \mathscr{W}$, and comparison $(p, q) \in \mathscr{D}$, we associate an error term $\varepsilon_{i,l,(p,q)}$. We assume that the vector of error terms $[\varepsilon_{i,1,(p,q)}, \ldots, \varepsilon_{i,L,(p,q)}]$ across utilities follows a multivariate normal distribution with mean $[0, \ldots, 0] \in \mathbb{R}^L$ and covariance matrix $\Sigma \in \mathbb{R}^{L \times L}$. For any two lotteries p and q, and for any subject i, our empirical framework postulates that

$$Core_i(p,q) = 3 \Leftrightarrow V(p,q,v_l) - \varepsilon_{i,l,(p,q)} \geq 0, \text{ for all } l \in \{1, \ldots, L\},$$

[4] That is, we assume $v_{l1} \leq v_{l2} \leq \ldots \leq v_{lK}$ for all utilities $v_l \in \mathscr{W}$.

and

$$Core_i(p,q) = 2 \Leftrightarrow V(p,q,v_l) - \varepsilon_{i,l,(p,q)} < 0, \text{ for all } l \in \{1,\dots,L\}.$$

In other words, we postulate to find no evidence against $p \succsim_i q$ whenever the difference in expected utilities between lotteries p and q, minus an error term, is non-negative for all utilities. Similarly, we expect to find no evidence against $q \succsim_i p$ whenever the opposite condition holds.

Our flexible formulation of the error structure extends the normality assumption of the unique error term in Hey and Orme (1994), allowing us to account for potential noise in the $Core_i$ index that might arise from several sources. First, we construct this index by observing the choices of subject i in experimental settings that test the independence axiom. If these choices are noisy, then the resulting $Core_i$ index will also be noisy. Additionally, even in the absence of noise in the choices, the $Core_i$ index might still be noisy due to issues with missing data. For example, we might find no evidence against $p \succsim_i^* q$ simply because we could not observe enough choices involving lotteries p and q.

To account for variation in preferences across subjects, we employ a mixture model and postulate that each subject i belongs to one of C possible different groups. We denote by v_l^c the l-th utility in group c and by Σ_c the covariance matrix in group c, with $c \in 1,\dots,C$. Within this framework, the probability that we find no experimental evidence against $p \succsim_i^* q$ if subject i belongs to group c is:

$$\Pr(Core_i(p,q) = 3 \mid v_1^c,\dots,v_L^c,\Sigma_c) = \Phi\left(V(p,q;v_1^c),\dots,V(p,q;v_L^c);[0,\dots,0],\Sigma_c\right),$$

where Φ represents the cumulative distribution function of the multivariate normal distribution. Similarly, the probability that we find no experimental evidence against $q \succsim_i^* p$ if subject i belongs to group c is:

$$\Pr(Core_i(p,q) = 2 \mid v_1^c,\dots,v_L^c,\Sigma_c) = \Phi\left(-V(p,q;v_1^c),\dots,-V(p,q;v_L^c);[0,\dots,0],\Sigma_c\right).$$

Therefore, given the observed index $Core_i(p,q)$ for all pairs of lotteries $(p,q) \in \mathscr{D}$, the likelihood function for subject i belonging to group c is:

$$\begin{aligned} f(Core_i;v_1^c,\dots,v_L^c,\Sigma_c) = \prod_{(p,q)\in\mathscr{D}} \Big(& \mathbb{1}\left(Core_i(p,q)=3\right)\cdot \Pr(Core_i(p,q)=3 \mid v_1^c,\dots,v_L^c,\Sigma_c) \\ & + \mathbb{1}\left(Core_i(p,q)=2\right)\cdot \Pr(Core_i(p,q)=2 \mid v_1^c,\dots,v_L^c,\Sigma_c) \\ & + \mathbb{1}\left(Core_i(p,q)=1\right)\cdot \left(1-\Pr(Core_i(p,q)=3)-\Pr(Core_i(p,q)=2)\right)\Big). \end{aligned}$$

Let π_c represent the probability of a subject belonging to group type c. The log-likelihood of the finite mixture model is given by:

$$\sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c f(Core_i;v_1^c,\dots,v_L^c,\Sigma_c),$$

where the first sum is over subjects and the second sum is over groups.

Our plan is to estimate the utility functions, the parameters of the covariance matrices, and the probabilities of group membership through maximum likelihood estimation. Specifically, we will incorporate demographic information about subjects to estimate their probabilities of group membership.[5] Following the approach in Bruhin et al. (2010), we will employ model-fit measures such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), normalized entropy criterion (NEC), and integrated completed likelihood criterion (ICL) to determine the optimal number of groups.

3 Experimental Design

Our experimental design is guided by two main objectives. First, we aim to investigate the EU-core extensively by conducting tests of the independence axiom. Second, we aim to compare the accuracy of predictions derived from the empirical analysis of the EU-core with those obtained by estimating specific decision models. To pursue these objectives, we design an experiment with binary choice tasks between monetary lotteries.

3.1 Binary Choice Tasks

We elicit choices between lotteries over three monetary prizes $L < M < H$. We represent the three-outcome lottery that gives \$$L$ with probability p_L, \$$M$ with probability p_M, and \$$H$ with probability p_H as $(\$L, p_L; \$M, p_M; \$H, p_H)$. The independence axiom imposes consistency requirements on choices across two or more binary choice tasks. We first assess the independence axiom via the common ratio version of the Allais paradox, which involves two types of binary choice tasks that we call CR-tasks:

CR1: Lottery $\delta_M = (\$M, 1)$ vs. Lottery $r = (\$L, 1 - p_H; \$H, p_H)$.
CR2: Lottery $0.3\delta_M + 0.7\delta_L = (\$L, 0.7; \$M, 0.3)$ vs. Lottery $0.3r + 0.7\delta_L = (\$L, 1 - 0.3p_H; \$H, 0.3p_H)$.

In the following sections, we utilize the Marschak-Machina (MM) triangle to describe the lotteries in the experiment (Marschak, 1950; Machina, 1982). The left graph in Figure 1 illustrates the CR-tasks in the MM triangle. In the MM triangle, the probability of receiving the highest prize H is on the vertical axis, and the probability of receiving the lowest prize L is on the horizontal axis. Therefore, the generic point (p_L, p_H) in the MM triangle represents the lottery $(\$L, p_L; \$M, 1 - p_L - p_H; \$H, p_H)$. Each dashed segment connecting two lotteries indicate that there is a choice task that involves these lotteries. For instance, the black dashed segments in the left MM triangle of Figure 1 represent CR1 choice tasks, while the green dashed segments represent CR2 choice tasks.

There are two possible scenarios in which subjects' choices in CR-tasks are incompatible with the independence axiom. The Common Ratio Effect (CRE) refers to the violation of the independence axiom in which subjects in the experiment choose lottery δ_M in CR1, and lottery $0.3r + 0.7\delta_L$ in CR2. The opposite choices in CR1 and CR2 constitute the other possible violation of the independence axiom, known as the Reverse Common Ratio Effect (RCRE).

[5] Given a vector of demographic information z_i and a vector of coefficients $\theta \in \mathbb{R}^{C-1}$, we parameterize π_c as $\pi_c(z_i; \theta) = \frac{1}{1+\sum_{c=1}^{C-1} \exp(-z_i' \theta_c)}$.

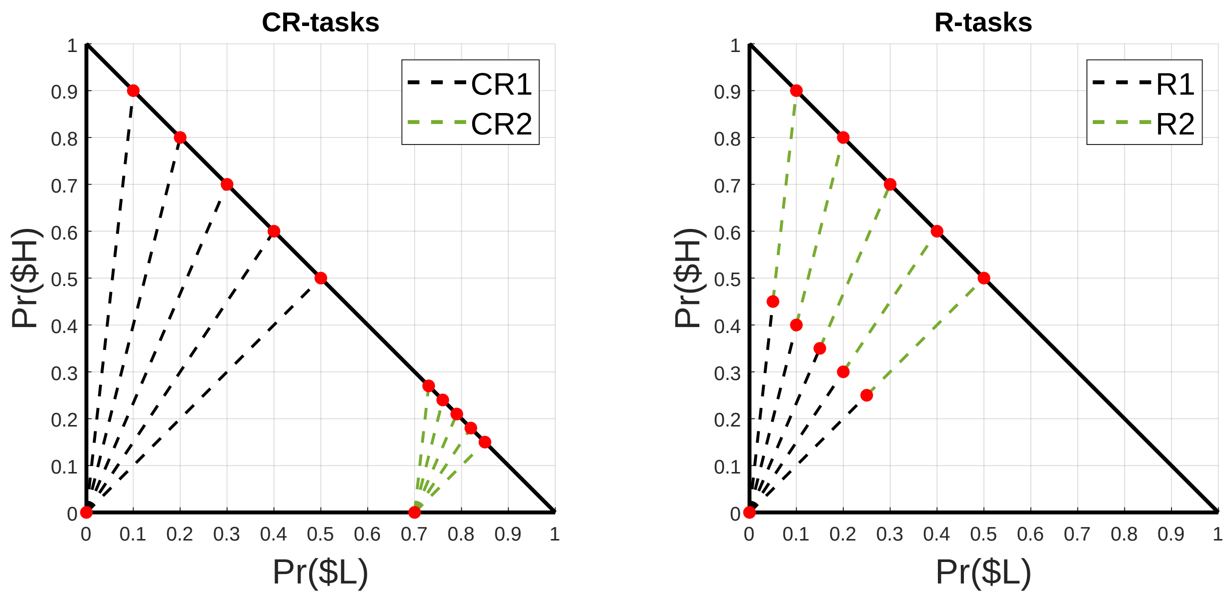

Figure 1: Choice Tasks.

As an additional assessment of the independence axiom, we also study subjects' attitudes toward randomization. To this end, we consider the following two types of binary choice tasks that we refer to as R-tasks:

R1: Lottery $\delta_M = (\$M, 1)$ vs. Lottery $0.5\delta_M + 0.5r = (\$L, 0.5(1-p_H); \$M, 0.5; \$H, 0.5p_H)$.
R2: Lottery $r = (\$L, 1-p_H; \$H, p_H)$ vs. Lottery $0.5\delta_M + 0.5r = (\$L, 0.5(1-p_H); \$M, 0.5; \$H, 0.5p_H)$.

The right MM triangle in Figure 1 represents the R1 choice tasks (depicted by black dashed segments) and the R2 choice tasks (depicted by green dashed segments). In studies exploring preferences for randomization, it is common to combine R1 and R2 into a single choice task in which subjects can select either lottery δ_M, lottery r, or a combination of the two. Choosing a mixture of lotteries δ_M and r is typically interpreted as a preference for randomization. However, this approach has a limitation: it does not allow us to observe whether subjects exhibit aversion to randomization, meaning they prefer either of the lotteries δ_M and r over the mixture. By treating R1 and R2 as separate choice tasks, we can observe both preferences for and aversion to randomization. Specifically, subjects in the experiment display a preference for randomization when they consistently choose the lottery $0.5\delta_M + 0.5r$, and aversion to randomization when they consistently reject the lottery $0.5\delta_M + 0.5r$. Both a preference for randomization and an aversion to it are behaviors that contradict the independence axiom.

All subjects engage in CR1, CR2, R1, and R2 choice tasks involving five different prize triplets (L, M, H): (0, 15, 30), (5, 15, 25), (10, 20, 30), (15, 20, 25), and (0, 10, 20). For each triplet, subjects undertake all types of tasks with five different probability values for the high prize p_H: 0.5, 0.6, 0.7, 0.8, and 0.9. In addition to these 100 choice tasks, the experiment includes two choice tasks in which one lottery stochastically dominates the other (referred to as FOSD choice tasks), and three additional types

Table 1: Experimental design.

	Block 1					Block 2		
	CR1	CR2	R1	R2	FOSD	MPL1	MPL2	MPL3
# Tasks	25	25	25	25	2	11	11	11
Order Tasks	Randomized					MPL1, MPL2, MPL3		
Order Blocks	Always First					Always Second		

of choice tasks used to elicit certainty equivalents:[6]

MPL1: Lottery $(\$X, 1)$ vs. Lottery $(\$0, 0.5; \$20, 0.5)$ for $X \in \{3, \ldots, 13\}$.
MPL2: Lottery $(\$X, 1)$ vs. Lottery $(\$5, 0.5; \$25, 0.5)$ for $X \in \{8, \ldots, 18\}$.
MPL3: Lottery $(\$X, 1)$ vs. Lottery $(\$10, 0.5; \$30, 0.5)$ for $X \in \{13, \ldots, 23\}$.

We choose not to incorporate choice tasks between certain amounts and a given lottery into a list, as is typically done using the multiple price list (MPL) method. This design decision is made to minimize the amount of instruction that subjects need to understand, retaining binary choice tasks as the sole method for expressing their preferences.[7] The certainty equivalents elicited from MPL1, MPL2, and MPL3 choice tasks will serve to further assess the out-of-sample accuracy of the predictions derived from the empirical analysis of the EU-core.

Table 1 provides a summary of our experimental design. The choice tasks in the experiment are divided into two blocks: Block 1 and Block 2. Block 1 comprises the choice tasks used to test the independence axiom (CR1, CR2, R1, and R2), along with the FOSD choice tasks. The 102 choice tasks within Block 1 are presented to subjects in a randomized order at the beginning of the experiment.

Upon completing Block 1, subjects then proceed to complete the remaining choice tasks in Block 2 (MPL1, MPL2, and MPL3), specifically designed to elicit certainty equivalents. In Block 2, subjects first encounter MPL1 tasks, followed by MPL2 tasks, and ultimately MPL3 tasks. Within each task type in Block 2, the monetary amounts are presented in ascending order.

3.2 Recruitment and Experimental Payments

We plan to recruit 500 subjects from Prolific and will conduct the experiment using Otree. All subjects must be United States citizens, possess at least a high school education, and maintain a high approval rate on Prolific. We will collect data for each subject, such as gender, age, income, insurance, and investment behavior, through Prolific.

Each subject will receive $4 upon completing the experiment. Additionally, every subject will have

[6] We plan to exclude from the analysis any subjects who violate first-order stochastic dominance more than once.

[7] Different procedures to elicit risk preferences may result in different observed behavior. Freeman et al. (2019) find that embedding a pairwise choice between a certain monetary amount and a risky lottery in a choice list increases the proportion of subjects choosing the risky lottery.

Table 2: Simulation Results.

Prize	$v_1 - \hat{v}_1$	$v_2 - \hat{v}_2$
\$5	-0.0170	0.0125
	(0.0358)	(0.0325)
\$10	-0.0120	0.0132
	(0.0253)	(0.0254)
\$15	-0.0091	0.0083
	(0.0181)	(0.0162)
\$20	0.0179	-0.0165
	(0.0318)	(0.0266)
\$25	0.0182	-0.0154
	(0.0333)	(0.0287)

Notes: Mean difference between true and estimated utility values, with standard errors in parentheses.

a one-in-six chance of being selected to receive an additional bonus payment based on their decisions during the study. Out of the 135 choice tasks, each carries an equal probability of determining the bonus payment amount. Specifically, subjects will receive the realized amount from the lottery they chose in the randomly selected choice task.

4 Sample Size: Monte Carlo Simulation

We conduct a simulation exercise to assess the reliability of the empirical approach detailed in Section 2.1, given our target sample size of 500 subjects and the lotteries that we consider in our experiment. Specifically, we posit the existence of a representative decision-maker whose EU-core is represented by a set of two utilities $\mathscr{W} = \{v_1, v_2\}$.[8] Next, we demonstrate our approach's effectiveness in recovering the true values of these utilities. Specifically, we follow a three-step procedure, repeating it 100 times:

Step 1: Draw 500 numbers from a normal distribution with mean zero and standard deviation 0.1 for each pair of lotteries $\delta_M = (\$M, 1)$ and $r = (\$L, 1 - p_H; \$H, p_H)$ in our experiment.

Step 2: Determine the value of $Core(\delta_M, r)$ for each of the 500 random numbers $\tilde{\varepsilon}$ as follows:[9]

$$Core(\delta_M, r) = 2 \times \mathbb{1}\left(\min_{v_l \in \mathscr{W}} V(\delta_M, r, v_l) \geq \tilde{\varepsilon}\right) + \mathbb{1}\left(\max_{v_l \in \mathscr{W}} V(\delta_M, r, v_l) < \tilde{\varepsilon}\right).$$

Step 3: Given the simulated dataset, we implement our maximum likelihood procedure fixing $L = 2$ to recover the utility values. We denote by $\hat{v}_l$ the estimate of utility v_l, for $l \in \{1, 2\}$.

Table 2 summarizes the average differences between the true and estimated utility values for each prize in the experiment, with standard errors displayed in parentheses. The simulation results suggest

[8] We set $v_1(\$5) = 0.25$, $v_1(\$10) = 0.4$, $v_1(\$15) = 0.5$, $v_1(\$20) = 0.65$, $v_1(\$25) = 0.9$, $v_2(\$5) = 0.35$, $v_2(\$10) = 0.45$, $v_2(\$15) = 0.55$, $v_2(\$20) = 0.6$, $v_2(\$25) = 0.85$. The simulation results do not depend on this specific choice of the parameters.

[9] The simulated values of the index *Core* differ on average from the "true" values that we would have observed without noise in the 29.35% of the observations.

Table 3: Choice patterns in CR-tasks and R-tasks.

		CR2 choice				R2 choice	
		$0.3\delta_M + 0.7\delta_L$	$0.3r + 0.7\delta_L$			$0.5\delta_M + 0.5r$	r
CR1	δ_M	EU: $(\delta_M m_1)$	non-EU: $(\delta_M m_2)$	R1	δ_M	EU: $(\delta_M m)$	non-EU: $(\delta_M r)$
choice	r	non-EU: (rm_1)	EU: (rm_2)	choice	$0.5\delta_M + 0.5r$	non-EU: (mm)	EU: (mr)

Notes: We denote choice patterns by string of chosen lotteries. To ease notation, for each pair of lotteries (δ_M, r) in our experiment, we denote lottery $0.3\delta_M + 0.7\delta_L$ by m_1, lottery $0.3r + 0.7\delta_L$ by m_2, and lottery $0.5\delta_M + 0.5r$ by m.

that our estimation approach performs effectively within our experimental design context, as the Monte Carlo sampling distributions of the estimates are centered around the true values. While this exercise assumes the existence of a representative agent, the use of mixture models allows us to accommodate heterogeneity in risk preferences, achieving similar outcomes.

5 Analysis

Table 3 summarizes the possible choice patterns we can observe for each pair of lotteries (δ_M, r) in CR-tasks and R-tasks. The two non-EU choice patterns in CR-tasks are the CRE $(\delta_M m_2)$ and the RCRE (rm_1). Similarly, subjects violate EU in R-tasks if they consistently choose the mixture (mm), indicating a strict preference for randomization, or if they never opt for the mixture (sr), demonstrating an aversion to randomization.

We will begin our analysis by summarizing EU and non-EU behavior in CR-tasks and R-tasks. The investigation of R-tasks is especially interesting due to the opportunity to observe and evaluate the relative significance of a preference for randomization and an aversion to it. Additionally, we will scrutinize the correlation between the CRE and a preference for randomization. Most earlier experimental studies have examined the CRE and preferences for randomization either separately or with a predominant emphasis on one of the two behaviors. By studying these two phenomena concurrently, we can assess whether economic models are capable of jointly rationalizing their potential emergence in our experiment.

What kind of correlation should we expect to observe between CRE and preferences for randomization under popular non-EU models? We address this question by deriving predictions under Cumulative Prospect Theory (CPT). The CPT value of a lottery $p = (\$L, p_L; \$M, p_M; \$H, p_H)$ is

$$U_{CPT}(p) = \pi(p_H)u(H) + [\pi(p_H + p_M) - \pi(p_M)]\,u(M) + [1 - \pi(p_H + p_M)]\,u(L),$$

where $v(\cdot)$ is a utility function and $\pi(\cdot)$ is a probability weighting function. We consider the functional form and parameter values of Tversky and Kahneman (1992):

$$u(x) = x^{\alpha};\ \alpha = 0.88$$

$$\pi(p) = \frac{p^{\gamma}}{[p^{\gamma} + (1-p)^{\gamma}]^{\frac{1}{\gamma}}};\ \gamma = 0.61.$$

Within the empirical framework proposed by Hey and Orme (1994), a CPT decision-maker chooses

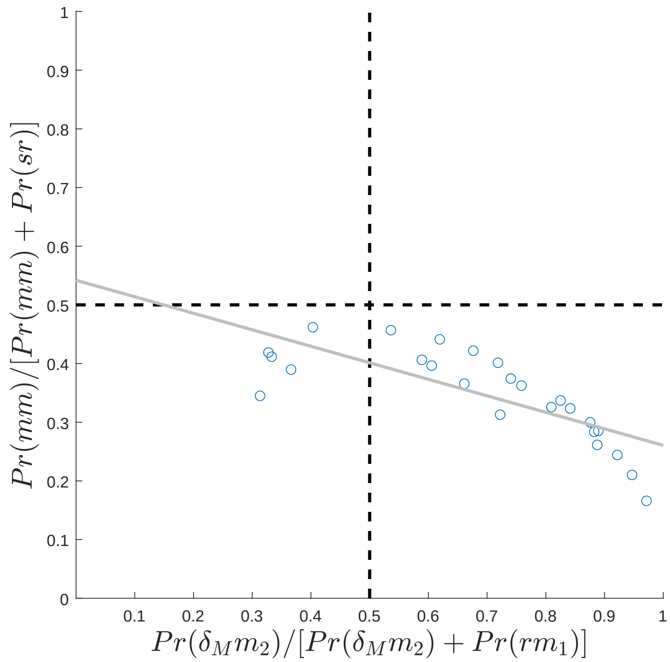

Figure 2: Predicted correlation between non-EU behaviors under CPT.

lottery p over lottery q if

$$U_{CPT}(p) - U_{CPT}(q) \geq \varepsilon,$$

where ε is an error term normally distributed with a mean of zero and a variance of one.

Assuming that errors are independent, we can compute the probability of the different choice patterns in our experiment. For instance, the probability of observing the CRE in CR-tasks is

$$Pr[\delta_M m_2] = Pr[U_{CPT}(\delta_M) - U_{CPT}(r) \geq \varepsilon] \times Pr[U_{CPT}(m_2) - U_{CPT}(m_1) \geq \varepsilon].$$

Figure 2 summarizes the predicted correlation between non-EU behaviors under CPT for all pairs of lotteries δ_M and r in our experiment. The x-axis measures the probability of CRE relative to the total probability of either CRE or RCRE. Similarly, the y-axis measures the probability of a preference for randomization relative to the combined probability of either a preference for or an aversion to randomization. The first prediction under CPT that arises from Figure 2 is that CRE and an aversion to randomization should be the prevalent non-EU behaviors in our experiment. Moreover, the emergence of CRE for one pair of lotteries should exhibit a strong negative correlation with the emergence of a preference for randomization.[10]

After evaluating the ability of CPT to explain the observed behavior in Block 1, we will proceed with the analysis of the EU-core. The first step in this analysis involves constructing the index *Core* for every pair of lotteries (δ_M, r) in our experiment. We will develop three different versions of this index:

[10]Correlation coefficient:-0.7623.

1. Construct the index $Core_{CR}$ using data solely from CR-tasks:

$$Core_{CR}(\delta_M, r) = \begin{cases} 3 & \text{if } \delta_M m_1 \\ 2 & \text{if } r m_2 \\ 1 & \text{otherwise.} \end{cases}$$

   There is no evidence against $\delta_M \succsim^* r$ in CR-tasks if lottery δ_M is chosen over lottery r and lottery m_1 is chosen over lottery m_2. When this happens, we assign to index $Core_{CR}(\delta_M, r)$ the value of three. If instead lottery r is chosen over lotteries δ_M and lottery m_2 is chosen over lottery m_1, we have no evidence against $r \succsim^* \delta_M$ and we assign to index $Core_{CR}(\delta_M, r)$ the value of two. In all the remaining cases, we assign to the index $Core_{CR}(\delta_M, r)$ the value of one because there is evidence disputing $\delta_M \succsim^* r$ and $r \succsim^* \delta_M$.

2. Construct the index $Core_R$ using data exclusively from R-tasks:

$$Core_R(\delta_M, r) = \begin{cases} 3 & \text{if } \delta_M m \\ 2 & \text{if } m r \\ 1 & \text{otherwise.} \end{cases}$$

   When we focus on R-tasks, no evidence against $\delta_M \succsim^* r$ implies that δ_M is chosen over lottery m and lottery m is chosen over lottery r. In this case, we assign to index $Core_R(\delta_M, r)$ the value of three. When instead lottery m is chosen over lotteries δ_M and lottery r is chosen over lottery m, we have no evidence against $r \succsim^* \delta_M$ and we assign to index $Core_R(\delta_M, r)$ the value of two. In all the remaining cases, we assign to the index $Core_R(\delta_M, r)$ the value of one because there is evidence disputing $\delta_M \succsim^* r$ and $r \succsim^* \delta_M$.

3. Construct the index $Core_F$ using all available data:

$$Core_F(\delta_M, r) = \begin{cases} 3 & \text{if } Core_{CR}(\delta_M, r) = 3 \text{ and } Core_R(\delta_M, r) = 3 \\ 2 & \text{if } Core_{CR}(\delta_M, r) = 2 \text{ and } Core_R(\delta_M, r) = 2 \\ 1 & \text{otherwise.} \end{cases}$$

   We assign to index $Core_F(\delta_M, r)$ the value of three if we observe no choice in the data contradicting the hypothesis that $\delta_M \succsim^* r$. Similarly, if the hypothesis that $r \succsim^* \delta_M$ is never in contrast with the data, we assign to the index $Core_F(\delta_M, r)$ the value of two. In all remaining cases, we assign to the index $Core_F(\delta_M, r)$ the value of one.

The three variants of the index *Core* differ in the information that we exploit to compute them and will be used to estimate the set of utility functions that represent the EU-core. We will account for heterogeneity in preferences by estimating multiple sets of utility functions with finite-mixture models. This procedure enables us to establish a probability distribution over the possible values of the index *Core* for each pair of lotteries (δ_M, r) and each group of subjects. Consequently, we can predict whether subjects' behavior aligns with EU or not, but we cannot differentiate between specific non-EU choice patterns. For example, in CR-tasks, we cannot distinguish between CRE and RCRE, or between a preference for randomization and an aversion to it in R-tasks.

We intend to use both CPT and EU estimates within mixture models as benchmarks to assess the accuracy of our predictions. In these models, we calculate the probability of non-EU choice patterns as the sum of the probabilities of certain non-EU behaviors. For example, in CR-tasks, we add up the probabilities of CRE and RCRE to get the probability of non-EU choice patterns. This allows us to compare the predictive accuracy of these models with those based on the EU-core, ensuring a consistent level of detail in our predictions.

Moreover, we plan to consider two machine learning algorithms: gradient boosting trees and neural networks. We aim to use these algorithms in two ways to make predictions. Like with EU and CPT, we first aim to predict choices using the machine learning algorithms, and then derive EU and non-EU choice pattern probabilities. Additionally, we plan to use machine learning algorithms to directly predict the three variants of the *Core* index. We will use as covariates information about outcomes and probabilities for each pair of lotteries (δ_M, r), and include an indicator for each subject in the experiment.

For gradient boosting, we plan to employ the LogitBoost algorithm for predicting choices and AdaBoostM2 (a multi-class problem variant of AdaBoost) for predicting EU and non-EU choice patterns. As hyperparameters, we will fine-tune the number of boosting iterations, learning rate, and the minimum number of observations required at a decision tree node to conclude decision-making. We will use 10-fold cross-validation to measure the algorithms' loss during the hyperparameter optimization. For neural networks, we will use feedforward neural network classifiers for predicting both choices and EU/non-EU choice patterns. We will tune the regularization parameter and network architecture as hyperparameters to attempt to minimize the classifier's 10-fold cross-validation loss.

How do we evaluate the predictive accuracy of the different approaches? To answer this question, we consider two generic variants of the index *Core*, denoted as $Core_1$ and $Core_2$, along with a generic model, ρ, that predicts these indices. We denote by $\hat{\rho}(\delta_M, r; Core_1) \in \mathbb{R}^3$ the vector whose j-th component equals the estimated probability under model ρ that the index $Core_1(\delta_M, r)$ takes the value of j. Similarly, we denote by $O_i(\delta_M, r; Core_2) \in \mathbb{R}^3$ the vector whose j-th component is equal to one if the observed index $Core_2(\delta_M, r)$ for subject i equals j and zero otherwise. The probabilistic error of model ρ when predicting index $Core_2$ using information from index $Core_1$ is denoted as

$$e(\rho, Core_1, Core_2, l) = \sum_{(\delta_M, r) \in \mathscr{D}} \sum_{i=1}^{N} l\left(\hat{\rho}(\delta_M, r; Core_1), O_i(\delta_M, r; Core_2)\right),$$

where $l \colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ is a loss function. We will consider two different loss functions: the L1 norm and the L2 norm. The lower the value of the loss, the greater the probabilistic accuracy of a model.

Moreover, we will analyze the ability of different models to provide accurate deterministic predictions. To do this, we denote by $\tilde{\rho}(\delta_M, r; Core_1) \in \mathbb{R}^3$ the vector whose j-th component is equal to one if the predicted value for the index $Core_1(\delta_M, r)$ under model ρ is equal to j, and zero otherwise. We compute deterministic predictions in the multi-utility model by setting the estimated variances of all the error terms to zero. For EU, CPT and machine learning algorithms, we plan to use as deterministic prediction the the value of the index *Core* with the highest predicted probability. The deterministic error of model ρ when predicting index $Core_2$ using information from index $Core_1$ is denoted as

$$d(\rho, Core_1, Core_2) = \frac{1}{|\mathscr{D}| \times N} \sum_{(\delta_M, r) \in \mathscr{D}} \sum_{i=1}^{N} \mathbb{1}\left(\tilde{\rho}(\delta_M, r; Core_1) = O_i(\delta_M, r; Core_2)\right),$$

where $|\mathscr{D}|$ denotes the cardinality of the set $\mathscr{D}$.[11]

We aim to compute both probabilistic and deterministic errors for the multi-utility model, EU, CPT and machine learning algorithms in the following prediction exercises:

- In-sample predictions and cross-validation analyses focusing solely on CR-tasks.
- In-sample predictions and cross-validation analyses focusing exclusively on R-tasks.
- Predicting behavior in R-tasks based on estimates derived from CR-tasks.
- Predicting behavior in CR-tasks based on estimates derived from R-tasks.

Moreover, the estimated sets of utility functions in the mixture model will allow us to differentiate groups of subjects along two dimensions. The first dimension pertains to how far a group of subjects deviates from EU. The second dimension pertains to risk aversion. Specifically, given an estimated set of utilities $\hat{\mathscr{W}}$ and two lotteries $\delta_M = (\$M, 1)$ and $r = (\$L, 1 - p_H; \$H, p_H)$, we can compute the range of probabilities $[\underline{p}_H, \overline{p}_H]$, where

$$\underline{p}_H := \max\left\{p_H \in [0,1] \,:\, \hat{v}(M) \geq p_H \hat{v}(H) + (1 - p_H)\,\hat{v}(L) \text{ for all } \hat{v} \in \hat{\mathscr{W}}\right\},$$

and

$$\overline{p}_H := \min\left\{p_H \in [0,1] \,:\, p_H \hat{v}(H) + (1 - p_H)\,\hat{v}(L) \geq \hat{v}(M) \text{ for all } \hat{v} \in \hat{\mathscr{W}}\right\}.$$

We plan to compute the range of probabilities $[\underline{p}_H, \overline{p}_H]$ for all triplets of prizes in our experiment. These ranges can be used as proxies for the extent to which a subject adheres to EU. A wider range of probabilities suggests that a smaller proportion of the data is consistent with behavior predicted by EU. Consequently, given a triplet of prizes in the experiment, we can classify one group of subjects as more non-EU than another if the associated range of probabilities for the first group is larger, in terms of set inclusion. In the extreme case where behavior always complies with EU, both $\underline{p}_H$ and $\overline{p}_H$ would be equal to a fixed probability p_{EU}, which characterizes a subject's risk attitude.

Specifically, in our experiment, the mid-value prize M is always set as the mean of the high prize H and the low prize L. Consequently, given a triplet of prizes, we can categorize a subject as risk averse if $p_{EU} > 0.5$, risk neutral if $p_{EU} = 0.5$, and risk-seeking if $p_{EU} < 0.5$. Moving beyond the EU framework, the range of probabilities $[\underline{p}_H, \overline{p}_H]$ enables us to explore risk attitudes in situations where preferences

[11]In our experiment, $|\mathscr{D}|$ is equal to 25.

violate the independence axiom. In this context, we categorize a subject as risk averse if $\underline{p}_H > 0.5$, risk-seeking if $\overline{p}_H < 0.5$, and as neither risk averse nor risk-seeking in all other cases. Furthermore, we can categorize one group of subjects as more risk-averse than another if both the lower and upper bounds of the probability range are weakly higher, with at least one of the two bounds being strictly greater.

We plan to use the characterization of risk attitudes for the different groups of subjects, established using data from Block 1, to make predictions in Block 2. Specifically, we plan to use the estimates arising from all the three versions of index *Core* to execute the following prediction exercises. The first hypothesis that we aim to test is that if one group is consistently classified as more risk-averse than another for the majority or all of the prize triplets in Block 1, then the empirical distribution of certainty equivalents in the more risk-averse group should stochastically dominate that of the less risk-averse group. Additionally, given a lottery and a set of estimated utility functions, we can predict a range of potential values for the lottery's certainty equivalent. As a result, we can assess how often the observed certainty equivalents fall within these predicted ranges.

A common observation in studies that elicit certainty equivalents using the MPL method is that subjects sometimes switch multiple times between preferring a certain amount and favoring a fixed risky lottery.[12] Recent experimental evidence by Chew et al. (2022) indicates that this multiple switching behavior is positively correlated with deliberate randomization in successive tasks, suggesting that this behavior might result from randomization rather than noise. In Block 2, we plan to use the predicted range of certainty equivalents to assess this hypothesis. Specifically, multiple switching behavior can only be rationalized as deliberate randomization if it occurs between choice tasks where the certain amounts fall within the range of certainty equivalents. Furthermore, we intend to test whether our range of certainty equivalents can validate the findings of Agranov and Ortoleva (2021), which demonstrate that subjects are willing to randomize between the certain amount and the risky lottery in the MPL method.[13]

Lastly, we intend to explore the correlation between subjects' risk attitudes, their deviations from EU, and their demographic information, along with their insurance and investment behaviors outside of the lab environment. To achieve this, we will examine the correlation between the membership probability in the finite mixture model and this information, gathered from Prolific. As detailed in Section 2.1, we also plan to incorporate demographic information and proxies for behavior outside the lab into the estimation of the mixture model. The factors under consideration include subjects' gender, age, and income. Additionally, we will incorporate information about subjects' insurance and investment behaviors into the estimation.[14] For insurance, we will focus on whether they have purchased insurance for products such as mobile phones. Regarding investments, we will examine whether they have participated in stock trading and whether they possess any cryptocurrencies. We carry out this analysis by estimating the finite mixture model separately on the index cores $Core_{CR}$ observed in CR-tasks and $Core_R$ observed in R-tasks. In addition, we also estimate the model using $Core_F$ defined above to analyze the overall pattern

[12] Crosetto and Filippin (2016) reported an average frequency of 14.3 percent for this multiple switching behavior across 41 studies.

[13] Specifically, we aim to predict the average lower and upper bounds for the range of certain amounts for which subjects chose to randomize between a certain amount within the range and the lottery that pays either \$20 or \$0 with equal chance ("Q1r task" in their paper).

[14] This information is supplied by the subjects when they register on the Prolific platform.

across these tasks.

References

Agranov, M. and Ortoleva, P. (2021). Ranges of randomization. Technical report.

Bruhin, A., Fehr-Duda, H., and Epper, T. (2010). Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. *Econometrica*, 78(4):1375–1412.

Cerreia-Vioglio, S. (2009). Maxmin expected utility on a subjective state space: Convex preferences under risk. Technical report, Mimeo, Bocconi University.

Chew, S. H., Miao, B., Shen, Q., and Zhong, S. (2022). Multiple-switching behavior in choice-list elicitation of risk preference. *Journal of Economic Theory*, 204:105510.

Crosetto, P. and Filippin, A. (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics*, 19(3):613–641.

DellaVigna, S. (2018). Structural behavioral economics. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 1, pages 613–723. Elsevier.

Freeman, D. J., Halevy, Y., and Kneeland, T. (2019). Eliciting risk preferences using choice lists. *Quantitative Economics*, 10(1):217–237.

Hey, J. D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica: Journal of the Econometric Society*, pages 1291–1326.

Machina, M. J. (1982). " expected utility" analysis without the independence axiom. *Econometrica: Journal of the Econometric Society*, pages 277–323.

Marschak, J. (1950). Rational behavior, uncertain prospects, and measurable utility. *Econometrica: Journal of the Econometric Society*, pages 111–141.

Segal, U. (1990). Two-stage lotteries without the reduction axiom. *Econometrica: Journal of the Econometric Society*, pages 349–377.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323.