Games and Economic Behavior 132 (2022) 22–40

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Do people maximize quantiles? ☆

Luciano de Castro [a], Antonio F. Galvao [b], Charles N. Noussair [c,*], Liang Qiao [c]

[a] Department of Economics, University of Iowa, Iowa City, USA
[b] Department of Economics, Michigan State University, East Lansing, USA
[c] Department of Economics, University of Arizona, Tucson, USA

ARTICLE INFO

Article history:
Received 3 July 2020
Available online xxxx

JEL classification:
D81
C91

Keywords:
Quantile preference
Risk attitude
Experiment

ABSTRACT

Quantiles are used for decision making in investment analysis and in the mining, oil and gas industries. However, it is unknown how common quantile-based decision making actually is among typical individual decision makers. This paper describes an experiment that aims to (1) compare how common is decision making based on quantiles relative to expected utility maximization, and (2) estimate risk attitude parameters under the assumption of quantile preferences. The experiment has two parts. In the first part, individuals make pairwise choices between risky lotteries, and the competing models are fitted to the choice data. In the second part, we directly elicit a decision rule from a menu of alternatives. The results show that a quantile preference model outperforms expected utility for 32%–55%, of participants, depending on the metric. The majority of individuals are risk averse, and women are more risk averse than men, under both models.

1. Introduction

Expected Utility (EU) maximization, first axiomatized by von Neumann and Morgenstern (1944) and Savage (1954), is the standard decision model assumed in economic and financial theory for the analysis of choices under risk. The model is elegant, general, and amenable to theoretical modeling. The assumption of maximization of average utility, the average being a simple measure of central tendency, has intuitive appeal as a behavioral postulate. Nevertheless, the EU framework has been subjected to a number of criticisms, mostly arising from experimental evidence that has not been supportive. Early studies suggesting that individuals did not always employ objective probabilities, as well as the robustness of the Allais (1953) paradox, resulted in Prospect Theory (Kahneman and Tversky, 1979), Rank-Dependent Expected Utility Theory (Quiggin, 1982), and Cumulative Prospect Theory (Tversky and Kahneman, 1992).[1] Rabin (2000) criticized EU theory on the basis of some implausible predictions it generates, arguing that EU would require unreasonably large levels of risk aversion

☆ The authors would like to express their appreciation to James Cox, Ernan Haruvy, Tibor Neugebauer, Julian Romero, Vjollca Sadiraj, Stefan Zeisberger and participants at Georgia State University, Michigan State University, and the Virtual Experimental Finance Workshop for helpful comments and discussions. We especially thank Glenn Harrison for providing us with a number of very valuable comments and suggestions. We would also like to thank the Economic Science Laboratory at the University of Arizona for funding. Luciano de Castro acknowledges the support of the National Council for Scientific and Technological Development – CNPq.

* Corresponding author.
E-mail addresses: decastro.luciano@gmail.com (L. de Castro), agalvao@msu.edu (A.F. Galvao), cnoussair@email.arizona.edu (C.N. Noussair), liangqiao@email.arizona.edu (L. Qiao).

[1] In particular, compelling empirical evidence has been gathered regarding the violation of the independence axiom (e.g. Allais, 1953; Starmer and Sugden, 1989; List and Haigh, 2005).

https://doi.org/10.1016/j.geb.2021.11.010

to explain the data from some small-stakes laboratory experiments.[2] In response to the critiques it has received, the EU model has been successfully generalized to accommodate a variety of behavioral phenomena, a testament to the flexibility and tractability of the EU structure.[3]

In this paper, we are concerned with a different, though also fundamental, potential source of departure from the EU Theory. This is the assumption that individuals seek to maximize an *average* of some function of their payoffs. To us, it is not obvious why a decision maker would necessarily seek to maximize a mathematical expectation of a distribution of either true or perceived payoffs. An alternative is the possibility that individuals attempt to maximize a *quantile* of their payoff distribution. In other words, they maximize the payoff that they have at most a given probability of failing to attain. The simplest example of such a rule is that of maximizing one's median payoff, the 50th percentile. Another example, familiar in risk management, is Value-at-Risk[4] targeting, in which a firm attempts to maximize a low quantile (typically the 1st or 5th percentile) of the payoff from an investment. In oil and gas extraction, executives are frequently concerned with the P90, P50 and P10 indicators of a prospective well, which correspond to the levels of production that are exceeded with 90%, 50% (median) or 10% probability, respectively.[5]

Manski (1988) was the first to study the properties of quantile preferences (QP), which were later axiomatized by Chambers (2009) and Rostek (2010). In his article, Manski (1988) also discussed a notion of risk aversion for the quantile model, which was further developed by Rostek (2010), and is related to the idea of quantile-preserving spread introduced by Mendelson (1987). The risk attitude is captured by a single-dimensional parameter, the quantile $\tau \in (0, 1)$, with lower quantiles corresponding to more aversion to risk. QP have several attractive features. An individual's decision is independent of the form of her utility function and thus an optimal choice is relatively easy to compute. The measure of risk attitude is simple and intuitive. The structure is robust and flexible, with a family of preferences indexed by quantiles. Recently, models of QP have been attracting increasing attention.[6] However, to our knowledge, no empirical work has considered the incidence of QP in any population.

While the QP model is theoretically appealing and is known to describe decision making in some specific contexts, its usefulness ultimately depends on the breadth of its empirical relevance. This paper presents the results of an experiment designed to study the incidence of QP maximization, relative to EU maximization, among individuals choosing between risky lotteries.

Our experiment has two parts. In the first, participants engage in 242 binary decisions between risky lotteries, and we test whether their pattern of decisions conforms more closely to the EU or the QP model. We also estimate the participants' risk aversion parameters under both models. Under EU, the risk aversion parameter is a measure of the curvature of the utility function, while under QP, it is the quantile being maximized. We employ structural estimation for both models and use a maximum likelihood estimator with a parametric functional form for the underlying latent choice models.

In the second portion of the experiment, we directly elicit individuals' intended choice rules. Six alternative decision rules are presented to participants in each period. Three of the rules involve the maximization of EU and three are quantile maximization rules. We ask subjects to choose one of the rules, with the understanding that a bot will employ the rule on their behalf in a series of lottery choice tasks. Our methodology has the features that it elicits the *intended* decision rule of the participant, and that it requires no statistical assumptions to estimate the rule being used, as would be necessary with data from choices between lotteries.

Two recent experiments that have taken approaches that are similar, in that they allow the choice of decision rules in risky choice tasks, are those of Nielsen and Rehbeck (2020) and Halevy and Mayraz (2020). Nielsen and Rehbeck offer individuals the choice of different fundamental axioms of choice, such as the Independence of Irrelevant Alternatives or First-Order Stochastic Dominance, that bots can employ on their behalf in risky choice tasks. The participants then make pairwise choices between lotteries in a traditional manner. They are then presented with the inconsistencies between the rules they chose and the decisions they made in the pairwise tasks and have an opportunity to change their decisions. The authors find a strong tendency for individuals to follow the choice axioms and that they view the decisions violating the axioms as mistakes. Halevy and Meyraz study how individuals solve an investment portfolio problem. Participants have an opportunity to construct their own portfolios and to choose rules for portfolio selection. In a later phase of the experiment, they then can choose whether to make their own selections or to apply an investment rule of their choosing. The authors find that a two-thirds majority of participants chooses to employ investment rules.

[2] Cox and Sadiraj (2006) argue that Rabin's critique is only valid if participants in experiments fully integrate the payoffs that they earn in the experiment with their overall wealth. That is, Rabin's critique is valid for agents maximizing the expected utility of their overall wealth, but not of their current income. The fact that experimental procedures typically are intended to achieve isolation from participants' experience outside the laboratory makes this counterargument more compelling for experimental research.

[3] Two of the more well-known generalizations are to include the presence of regret (Bell, 1982) and ambiguity in beliefs about probabilities (Gilboa and Schmeidler, 1989).

[4] See, e.g., Duffie and Pan (1997) and Jorion (2007). The VaR measure is one of the main practical tools for reporting the exposure to risk by financial institutions.

[5] See, e.g., Apiwatcharoenkul et al. (2016) and Fanchi and Christiansen (2017).

[6] Bhattacharya (2009) studies the problem of optimally dividing individuals into peer groups to maximize a quantile of social gains given heterogeneous peer effects. Giovannetti (2013) models a two-period economy with one risky and one risk-free asset, where the agent has QP instead of the standard EU preferences. de Castro and Galvao (2019) develop a dynamic model of rational behavior under uncertainty, in which the agent maximizes a stream of future quantile utilities.

Two initial remarks are in order regarding our choice of experimental design and the interpretation of our results. The first is that we view pitting the QP model against EU as the natural starting point in evaluating the QP model, since EU is the standard model employed in economics. If the QP model performs well against EU, it can be tested against other models in follow-up work. The second regards what a reasonable standard of performance is. We do not expect QP to describe every individual's decisions, or even the decisions for a majority of participants, better than EU. However, if the QP model describes even a substantial minority of individuals at least as well as EU, then it should be taken seriously as a descriptive model of behavior. On the other hand, if very few individuals can be classified as QP maximizers, it would suggest that QP maximization is confined to a few special situations, rather than a widely and commonly employed decision model.

We use statistical classification methods to compare the capacity of the two models to describe participants' behavior in the first part of the experiment. First, we classify individuals as users of the QP or EU model based on the models' ability to predict their decisions, given their own estimated parameters for each model.[7] We find that a considerable minority of participants, 32%–42% depending on the classification criteria, behave as QP maximizers rather than EU maximizers.[8] Second, we classify participants based on both a proportion test and a model selection test (Schennach and Wilhelm (2017)). The results show that 43%–55% and 48%–49% of participants are classified as QP maximizers under the two tests, respectively.

The individual risk attitude estimates suggest considerable heterogeneity among participants. The average estimated quantile that is being maximized is 0.42, and a majority of subjects are risk averse. Most individuals, about 80%, have quantile estimates between 0.3 and 0.6.[9] For the EU model, we estimate the Power Utility version of the CRRA utility function, given by $u(x) = x^{\gamma}$.[10] Our average estimate is $\gamma = 0.87$, closely in line with the existing experimental literature (see Harrison and Rutström (2008) for a review).[11]

The design of our experiment allows us to consider whether the estimates differ between genders. We find that women are significantly more risk averse than men on average under the EU model, a result in agreement with much of the previous literature (see Eckel and Grossman, 2008). We also obtain an analogous result for the QP model, in that women maximize significantly lower quantiles than men on average, a pattern consistent with greater risk aversion on the part of women.[12]

The data from the second part of the experiment is also used to classify subjects as QP or EU maximizers, on the basis of their stated preferred decision rule. Here, no statistical method is required for classification, and a mere count of their choices can be used with varying levels of tolerance for deviations from the models. The results concur with the estimates from the first part of the experiment, and show that a significant number of subjects, in the range of 40%–50% depending on how it is measured, are quantile maximizers rather than EU maximizers.

The results illustrate the empirical relevance of QP. A very substantial minority of our participants exhibits behavior that is more consistent with the QP than the EU model. The strength of the empirical support for QP stands in contrast with the scant attention that quantile models have received in decision science. In our view, the incidence of quantile maximization that we have observed justifies the continued exploration of the properties of QP and their implications for economic and financial decision making.

The remainder of the paper is structured as follows. Section 2 reviews the QP model. Section 3 discusses the design and implementation of the experiment. In Section 4, we describe and discuss the main results for the first part of the experiment, and Section 5 further studies how the results correlate with gender. Section 6 presents results pertaining to the decision rules chosen in the second portion of the experiment. Finally, Section 7 concludes.

2. Quantile model

This section describes the quantile model. In Section 2.1, we review the definition of quantile preferences (QP) and the model's notion of risk attitude. In Section 2.2, we discuss pairwise lottery decisions under QP.

[7] Harrison and Rutström (2009) discuss the possibility of a situation in which several latent behavioral rules coexist in a population. Specifically, they estimate a mixture model in which individuals may be either Expected Utility Maximizers or Prospect Theory adherents. They estimate that 55% of participants are better described by EUT and 45% by Prospect Theory.

[8] In Section 3 of the Online Appendix, we report results using out-of-sample accuracy rate predictions, which suggest a QP maximizer proportion of 39%–44%.

[9] Given that a parametric specification is used in the estimation, it is important to consider alternative specifications and functional forms, even if the ones we have chosen are standard. Results reported in Section 2 of the Online Appendix show that varying the scale parameter underlying the logistic likelihood function does not qualitatively affect the results.

[10] See Wakker (2008) for a discussion of the properties of the CRRA utility function.

[11] Tversky and Kahneman (1992) and Abdellaoui et al. (2008) estimate the Power utility specification and obtain median estimates of $\gamma = 0.88$ and $\gamma = 0.86$, respectively. These are very close to ours. Other studies estimate a CRRA utility function of the form $u(x) = \frac{x^{1-\rho}}{1-\rho}$. Some of the resulting estimates are $\rho = .61$ (Hey and Orme (1994), sample median), $0.15 - 0.68$ depending on the stakes (Holt and Laury (2002), sample median), $0.26 - 0.54$ (Harrison et al. (2005), average estimate), and 0.89 (Noussair et al. (2014), pooled estimate for representative individual). These are similar to ours in that they reflect a moderate degree of risk aversion, with estimates lying between risk neutrality $\rho = 0$ and $u(x) = \ln(x)$, (the limit as $\rho \to 1$).

[12] We also find that results are affected by a modest change in stakes; see Section 4 of the Online Appendix. As the stakes increase, estimates of the CRRA model under EU show greater risk aversion, which is also consistent with previous work (Holt and Laury, 2002; Harrison et al., 2005). In contrast, the estimated quantile maximized is not affected by the modest changes in monetary stakes that are present in our experiment.

2.1. Quantile preferences

Consider a random variable X, and let F_X denote its cumulative distribution function (CDF), that is, $F_X(\alpha) \equiv \Pr[X \leqslant \alpha]$. The quantile function $Q : [0,1] \to \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ is the generalized inverse of F:

$$Q_\tau[X] \equiv \begin{cases} \inf\{\alpha \in \mathbb{R} : F_X(\alpha) \geqslant \tau\}, & \text{if } \tau \in (0,1] \\ \sup\{\alpha \in \mathbb{R} : F_X(\alpha) = 0\}, & \text{if } \tau = 0. \end{cases}$$

A well-known and useful property of quantiles is "invariance" with respect to monotonic transformations, that is, if $g : \mathbb{R} \to \mathbb{R}$ is a continuous and strictly increasing function, then $Q_\tau[g(X)] = g(Q_\tau[X])$.

Quantile preferences (QP), which have been axiomatized by Chambers (2009), Rostek (2010), and de Castro and Galvao (2021), are described by the following relation:

$$X \succeq Y \iff Q_\tau[u(X)] \geqslant Q_\tau[u(Y)].$$

Since quantiles are not affected by monotonic transformations, QP are invariant with respect to the utility function. Let $u(X)$, where $u : \mathbb{R} \to \mathbb{R}$, be a continuous and increasing utility function describing an individual's preferences. Then, for a given quantile $\tau \in (0,1)$, the optimization problem is

$$X^* \in \arg\max_{X \in \mathcal{R}^*} Q_\tau[u(X)] \iff X^* \in \arg\max_{X \in \mathcal{R}^*} Q_\tau[X], \tag{1}$$

where $\mathcal{R}^* \subset \mathcal{R}$ is the subset of random variables (lotteries) available. Equation (1) shows that the quantile optimization problem using a given utility function is equivalent to maximizing the quantile obtained directly from the distribution of the outcome variable. Hence, the optimal choice under QP does *not* depend on any particular specification of the utility function.

The risk attitude under QP is determined by τ. It can be shown that an agent with a quantile given by τ is more risk-preferring than another agent with quantile τ' if $\tau > \tau'$, independently of his or her utility function. In other words, a decision maker that maximizes a lower quantile is more risk averse than one who maximizes a higher quantile. Thus, under QP, risk attitude is defined by the quantile that one maximizes rather than by the concavity of the utility function. This is formalized in Section 1 of the Online Appendix using the notion of quantile-preserving spread introduced by Mendelson (1987). The Online Appendix also provides a formal definition and further details regarding risk attitude under QP.

2.2. Decision between pairwise risky lotteries under quantile preferences

We now describe the decision of an agent maximizing a given quantile τ when making a pairwise choice between two risky lotteries. Consider a choice between two lotteries A and B. The decision of a QP maximizer is based on evaluating the difference between the two lotteries at a given quantile τ. If we understand A and B as random variables, this difference is:

$$\Delta Q(\tau) = Q_\tau[B] - Q_\tau[A].$$

The agent maximizing quantile τ prefers B to A when, for that τ, the payoff of lottery B is larger than A. That is,

$$\Delta Q(\tau) = Q_\tau[B] - Q_\tau[A] > 0.$$

To illustrate this, we consider an example of a decision taken when employing QP. There are two lotteries, A and B, that each have two possible payoffs, $\{a_1, a_2, b_1, b_2\}$ and corresponding outcome probabilities p and q, such that:

- Lottery A yields $a_1 = 6$ with probability p, and $a_2 = 10$ with probability $1 - p$.
- Lottery B yields $b_1 = 2$ with probability q, and $b_2 = 16$ with probability $1 - q$.

This lottery can be represented graphically, as illustrated in Fig. 1.

The calculation of $\Delta Q(\tau) = Q_\tau[B] - Q_\tau[A]$ depends on the quantile τ, the payoffs $\{a_1, a_2, b_1, b_2\}$, and the probabilities (p, q). Specifically,

$$\Delta Q(\tau) = \begin{cases} b_1 - a_1, & \text{if } \tau \leqslant \min\{p, q\} \\ b_1 - a_2, & \text{if } p < \tau \leqslant q \\ b_2 - a_1, & \text{if } q < \tau \leqslant p \\ b_2 - a_2, & \text{if } \tau > \max\{p, q\}. \end{cases} \tag{2}$$

To make the example more specific, fix the quantile at the median, $\tau = 0.5$, and the outcome probabilities at $p = 0.3$ and $q = 0.1$. The quantile functions of lotteries A and B are plotted in the left panel of Fig. 2. Lottery A (solid line) pays $a_1 = 6$ with probability 0.3 and $a_2 = 10$ with probability 0.7, and lottery B (dashed line) pays $b_1 = 2$ with probability

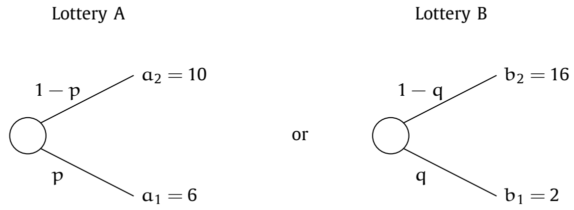

Fig. 1. Lottery Choice in Task L1.

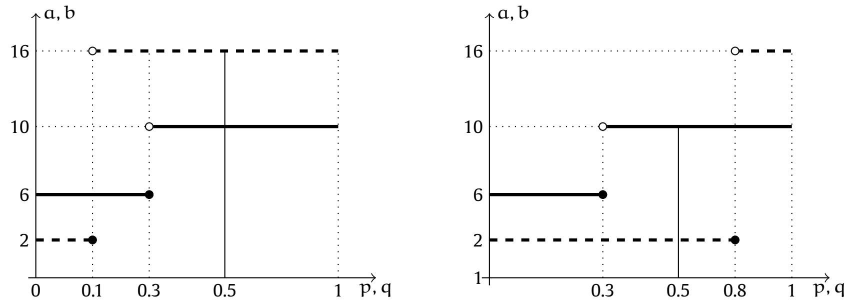

Fig. 2. Quantile function of lotteries A (solid line) and B (dashed line). The left plot considers $p = 0.3$ and $q = 0.1$. The right plot considers $p = 0.3$ and $q = 0.8$.

0.1, and $b_2 = 16$ with probability 0.9. The solid vertical line at 0.5 represents the quantile of interest. To complete the example of an agent maximizing the median, $\tau = 0.5$, and choosing between lotteries A and B, we compute $\Delta Q(\tau)$. The calculation is simple and only requires one to subtract the quantile of A from that of B. From the left panel in Fig. 2, we can see that $\Delta Q(0.5) = Q_{0.5}[B] - Q_{0.5}[A] = 16 - 10 = 6$. Therefore, the agent chooses lottery B.

Suppose now that we slightly modify the lotteries by changing the probability q in lottery B, so that we have:

- Lottery A: $a_1 = 6$ with probability $p = 0.3$, and $a_2 = 10$ with probability $1 - p = 0.7$;
- Lottery B: $b_1 = 2$ with probability $q = 0.8$, and $b_2 = 16$ with probability $1 - q = 0.2$.

The quantile functions of the new A and B lotteries are displayed in the right panel of Fig. 2. Lottery A (solid line) pays $a_1 = 6$ with probability 0.3 and $a_2 = 10$ with probability 0.7, and lottery B (dashed line) pays $b_1 = 2$ with probability 0.8, and $b_2 = 16$ with probability 0.2. In this case, we can see that the calculation of $\Delta Q(0.5)$ for the median (solid vertical line) is $\Delta Q(0.5) = Q_{0.5}[B] - Q_{0.5}[A] = 2 - 10 = -8$, and hence, the agent chooses lottery A.

3. The experiment

The experiment consisted of 8 sessions conducted at the Economic Science Laboratory at the University of Arizona, in February and March, 2020. All procedures received prior approval from the Institutional Review Board at the university. The 61 participants in the experiment were all undergraduate students at the university, who were recruited from a subject pool maintained by the Laboratory. 52.5% of the subjects were male, and 50.8% of them were majoring in Business or Economics.[13] The average cognitive reflection test (CRT) score (see Frederick, 2005) was 1.1 out of a maximum of 3. The experiment was computerized and used the ZTree platform.

A session consisted of ten periods. Periods 1 and 2 were referred to as Part 1 of the session. At the outset of the session, the instructions for these two periods were read and the two periods took place. Afterward, the instructions for Part 2 of the session, consisting of Periods 3–10 were read. Participants did not receive any feedback regarding the outcome of their choices until the session ended. The subjects then completed Periods 3–10. The instructions are available in Section 6 of the Online Appendix.

[13] The number of participants in our experiment gives us the power to detect a medium-sized difference (0.5 standard deviations) in risk aversion parameters between men and women, given a one-sided hypothesized difference, of 0.61. One-sided hypotheses are appropriate here for a gender comparison in light of a body of previous work finding that women are more risk averse than men, and the lack of evidence of the opposite effect. As we show later, we do observe highly significant gender differences in risk aversion estimates under both models with the sample size we have. The power to detect a medium-sized difference in risk aversion estimates between two different tasks within our experiment, where the comparison can be made within-subject, is 0.97 for a two-sided hypothesis.

Table 1
Lottery Payoffs For All Periods.

Task (payoff vector)	Lottery A		Lottery B		Description
	a_1 ($)	a_2 ($)	b_1 ($)	b_2 ($)	
L1	6	10	2	16	Used in Period 1 or 2
L2	6	10	2	24	Used in Period 1 or 2
L3	6	10	2	16	Same as L1
L4	6	10	2	24	Same as L2, L3's $b_2 + 8$
L5	3	5	1	12	50% of L4
L6	8	10	6	17	L5+5
L7	8	10	0.5	19.25	HL lotteries
L8	8	10	0.5	9.25	L7's $b_2 - 10$
L9	12	15	0.7	28.75	150% of L7
L10	13	15	5.5	24.25	L7+5

3.1. Periods 1 and 2

In each one of Periods 1 and 2, subjects made 121 separate decisions between two lotteries: called A and B. For ease of notation, we denote the tasks used in the two periods of Part 1 of the session as L1 and L2. The payoffs in the two tasks are given in Table 1. In the table, L3–L10 refer to the payoffs in Period 3–10, which will be described in Section 3.2. The payoffs from lottery A were $a_1 = 6$ and $a_2 = 10$ in both L1 and L2. However, those from lottery B were $b_1 = 2$ and $b_2 = 16$ in L1, and $b_1 = 2$ and $b_2 = 24$ in L2. In each of the 121 decisions of a period, the participant made a choice between Option A, which yielded a_1 with probability p and a_2 with probability $1 - p$, and Option B, which yielded b_1 with probability q and b_2 with probability $1 - q$. The probabilities of the relatively low payoffs under each lottery, p and q, took on values from the set {0,0.1,...,0.9,1}. Each of the 11 possible values of p was paired with each possible value of q in exactly one decision, generating $11 \times 11 = 121$ decisions in each of the first two periods.

To control for any order effect, the sequence of the pairwise choice lottery was counterbalanced. That is, we assigned the order of the lottery pairs L1 and L2 in the first two periods to participants on a random basis, with about one half of subjects receiving each order.[14] The 121 decisions within a period were presented in random order on 11 separate computer screens, with 11 decisions displayed on each screen.

As described in Section 2.2, we are able to identify the quantile of interest by keeping the payoffs $\{a_1, a_2, b_1, b_2\}$ constant and varying the associated probabilities (p, q). In addition, the difference in the high payment b_2 in lottery B between the two periods allows us to look at a specific comparative static. This is the effect of changing exactly one of the possible outcomes. If individuals use the same quantile independently of the magnitude of payoffs involved, decisions would be identical in the two periods. However, it is known that on average individuals tend to become more risk averse as the monetary stakes of the gambles they are confronted with increase, when risk aversion parameters are estimated under the assumption of expected utility. An analogous pattern, less risk taking behavior as stakes increase, for quantile decision makers, would mean that lower quantiles would be employed at higher stakes.

3.2. Periods 3–10

In Periods 3–10, participants were presented with the payoffs $\{a_1, a_2, b_1, b_2\}$, indicated as L3–L10 in Table 1. In each period, they were offered a menu of six decision rules, and were required to specify a rule from the menu for a computer robot to use on their behalf. The computer then made 121 decisions of the form that participants made in Periods 1 and 2, based on the rule participants specified. The six decision rules were the following:

- Maximize average earnings (EN)
- Maximize the average square root of earnings (EA)
- Maximize the average square of earnings (ES)
- Maximize the amount of money that you have at least a 50% chance of earning (Q50)
- Maximize the amount of money that you have at least a 75% chance of earning (Q25)
- Maximize the amount of money that you have at least a 25% chance of earning (Q75)

We denote the six decision rules as EN, EA, ES, Q50, Q25, Q75, respectively. EN, EA and ES apply the expected utility (EU) model with varying degrees of risk aversion. Specifically, EN orders the robot to make those choices that maximize the expected value of a risk-neutral utility function $u(x) = x$. EA commands the bot to maximize the expectation of a risk-averse utility function $u(x) = \sqrt{x}$, and ES does so for a risk-seeking utility function $u(x) = x^2$. Q50, Q25 and Q75

[14] As a result, 33 out of 61 participants received L1 in Period 1 and L2 in Period 2. The remaining 28 participants saw L2 in Period 1 and L1 in Period 2.

specify three different quantile preferences (QP) rules. Q50 orders the robot to maximize the participant's median payoff. Q25 and Q75 instruct the bot to maximize the 25th and 75th percentiles, respectively.[15]

The description of the choice rules is longer and arguably more complex for the EU rules, and this is because EU is more complicated than QP. The EU rules require an explanation of how payoffs are transformed by the utility function. This is the first attempt, to the best of our knowledge, to describe the decision rules of a bot applying expected utility theory, or indeed any completely specified rule of choice under risk, to experimental subjects. Future studies could try to improve on this initial attempt. However, because the data are comparable between designs 1 and 2 in terms of how many individuals are classified as EU and QP, it appears that the complexity of our descriptions did not create a bias in favor of or against the EU or QP models.

To overcome possible order effects in the way the rules were presented, we used two different versions of the instructions. In the instructions that one-half of the participants received, the EU rules were described first, and for the other half, the QP rules were described first. The position of the six rules on participants' computer screens was randomly reshuffled in each period. This was done in an independent random sequence for each individual.[16]

The potential lottery payoffs in Periods 3–10 are indicated as L3–L10 in Table 1. The order in which each payoff vector appeared to subjects was generated randomly and independently, without replacement, for each participant. L3 and L4 are the same lotteries used in Periods 1 and 2. The payoffs in L5 are equal to 50% of those in L4. L6 is derived by adding a constant of $5 to all payoffs in L5. L7 is proportional, in terms of payoffs, to those in the lotteries employed in the classic study of Holt and Laury (2002). L8 is similar to L7 except that b_2 is $10 lower. L9 has the payoffs of L7 scaled up by 50 percent, and L10 adds a constant of $5 to all payoff levels of L7.

We illustrate the decisions taken under the EN, EA, ES, Q50, Q25 and Q75 rules in Fig. 3. The shaded area of each individual panel indicates the combinations of probabilities p and q for which lottery B is chosen in L1, according to the decision rule depicted. We can see from the figure that, as risk tolerance under either QP or EU increases, the choice of lottery B becomes more likely. Nevertheless, the choices taken under the two models do exhibit substantial differences.

To describe the rules in a manner that participants could be expected to understand, we chose our phrasing very carefully. The instructions are available in Section 6 of the Online Appendix. To describe a rule that maximized $u(x) = \sqrt{x}$, we wrote *"This rule will take each possible amount of money you could earn, and give it a point score equal to the square root of the payment. It then makes a choice to maximize the average number of points you get. In other words, it calculates the average square root of the amount of money that you would receive under each choice, and always decides for the choice that has a higher average square root of the payment."* Similar language was used to describe maximization of $u(x) = x^2$.

Describing a quantile maximization rule is also not straightforward. For example, we chose the following language to represent the rule where $\tau = 0.25$: *This rule will consider each choice and ask the question: "How much money do I have at least a 75% chance of earning?" It will then decide on the choice that has a higher payment in answer to this question.* The responses on the quiz administered following the instructions indicate that the task was quite well understood.

3.3. Timing of events and payment

The instructions for Periods 1 and 2 were given at the beginning of the experiment. After participants finished Periods 1 and 2, the instructions for Periods 3–10 were provided, and were followed by a quiz to check participants' understanding of the decision rules introduced in the new instructions. The experimenter checked the quiz and helped participants to correct wrong answers privately. Participants were not allowed to continue until their quizzes were corrected and they demonstrated an understanding of the correct responses. Then, Periods 3–10 were conducted, in which participants selected decision rules. At the end of the session, participants filled in a brief questionnaire and were paid privately.

The final payment to participants equaled the sum of the $5 show-up fee and the earnings from one of the ten periods in the session. The period that counted toward the final payment was determined randomly at the end of the experiment by the computer. In the period that counted, one of the 121 decisions in the period was randomly chosen to count toward earnings. The average payment was US$16.6.

4. Results

This section reports findings for the first part of the experiment. It is organized in the following manner. Section 4.1 reports summary statistics, the structural estimation methods, and the risk attitude parameter estimates for both the quantile preferences (QP) and expected utility (EU) models using the data from the first two periods. In Section 4.2, we use statistical classification methods to consider the percentage of individuals that adhere more closely to each model.[17]

[15] The decision made under the Q50 rule is slightly different from the one maximizing $\tau = 0.5$, because of the discontinuity of probabilities in the experiment. More precisely, Q50 can be viewed as maximizing the $0.5 + \epsilon$ quantile, or $\tau = 0.5001$. We choose this wording to describe the quantile rules to facilitate the participant's comprehension.

[16] Shaw et al. (2000) show that individuals prefer locations in the middle of the physical space. In their study, they found that most participants chose the middle item from a set of three highlighters and on survey responses. They also found that most participants selected the middle chair from a row of three chairs to sit on. Because of this preference for locational centrality, we scrambled the location of the different choice options on the screen from period to period, and from subject to subject, to ensure that centrality effects did not affect the likelihood that the different rules were chosen.

[17] The p-values reported in our analysis are not corrected for multiple comparisons.

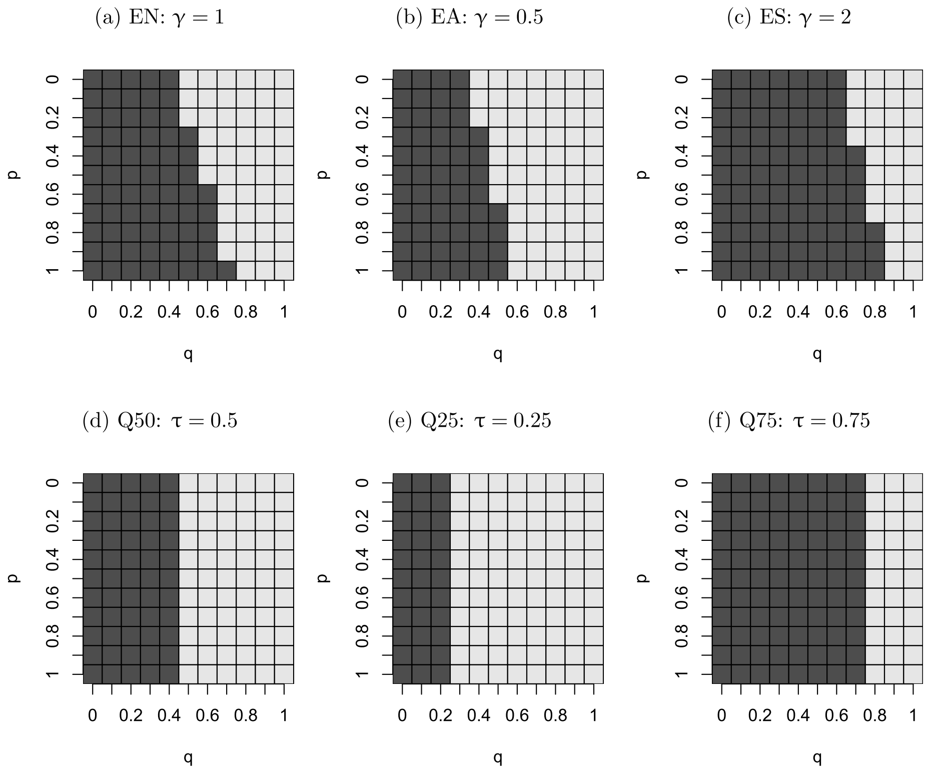

Fig. 3. Choices of lottery B over lottery A in L1 based on the decision rule specified. (a) EN with $\gamma = 1$, (b) EA with $\gamma = 0.5$, (c) ES with $\gamma = 2$, (d) Q50 with $\tau = 0.5$, (e) Q25 with $\tau = 0.25$, (f) Q75 with $\tau = 0.75$. The darker color represents choosing lottery B for the cell, while the lighter color indicates choosing lottery A. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

4.1. Risk attitude parameters

4.1.1. Summary statistics of periods 1 and 2

In each of the first two periods, every participant makes 121 decisions between lotteries A and B. Fig. 4 provides heat maps that summarize the observed choices. The figure shows the proportion of the participants choosing lottery B over A in each decision as a function of the probabilities p and q. The figure divides the data into five proportion intervals: $[0, 0.2)$, $[0.2, 0.4)$, $[0.4, 0.6)$, $[0.6, 0.8)$, $[0.8, 1]$, which are represented with five different color shades. A darker color corresponds to a higher proportion of choices of B. The left panel of the figure shows the combined results for Periods 1 and 2. The middle and right panels display separate results for tasks 1 and 2 (L1 and L2), respectively.

A comparison of the observed choice results in the left panel of Fig. 4 with the theoretical patterns in Fig. 3 gives the impression that the representative individual's decisions lie somewhere between the EU and QP models, and display a considerable aversion to risk. A contrast between the middle and right panels in Fig. 4 shows that there are more choices of lottery B in L2, where b_2 is larger, than in L1. This is consistent with substitution toward lottery B when it has a greater expected value. As we shall see later, however, the difference is smaller than would be predicted if the coefficient of risk aversion γ were the same in the two tasks, and the estimated coefficient is smaller for L2 than for L1, indicating more risk aversion in L2. The relatively low responsiveness to the change in b_2 is also consistent with a fraction of participants employing the QP model, which predicts no difference in choices between the two tasks.

4.1.2. Estimation method

We estimate the risk aversion parameters for both the QP and EU models using structural maximum likelihood estimation methods.[18] Under the assumption of EU, a utility function must be specified. We assume constant relative risk aversion (CRRA), with the specific power functional form,

$$u(x) = x^{\gamma}. \tag{3}$$

[18] An alternative would be to estimate a mixture model, allowing choices to be characterized by EU and QP. See Harrison and Rutström (2009) for a mixture approach in the context of EU and prospect theory.

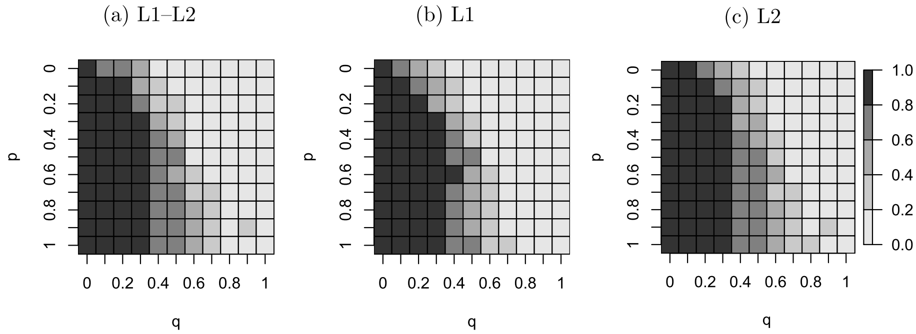

Fig. 4. Heat maps for observed choices of lottery B over lottery A. The left panel contains the pooled data for Periods 1 and 2. The middle panel shows the data for L1, where $b_2 = \$16$, and the right panel the data for L2, where $b_2 = \$24$.

The variable x is the monetary payoff and γ is the risk aversion parameter to be estimated. We choose the CRRA functional form for two reasons. The first is that it is the most widely employed in economics and this facilitates a comparison of our estimates with others in the literature. The second is that it has one parameter, and as such, can be readily compared in terms of fit to the QP model without necessitating a correction for a different number of parameters in the two models.[19] We follow the literature in estimating the parameter γ. We use a latent model with an error term together with a maximum likelihood estimator (MLE). We omit the details of the statistical model for the EU model because we use an analogous procedure for the τ parameter of the QP model, which we describe in detail below. We refer the reader to, e.g., Harrison and Rutström (2008) and Moffatt (2016) and references therein for details of the estimation procedure for the EU model.

With regard to the risk attitude parameter for the QP model, the quantile τ, recall from Section 2.1 that QP are invariant with respect to the utility function. Hence, it is not necessary to specify any particular parametric functional form of utility to estimate the quantile. We adopt the following structural estimation strategy to estimate the parameter τ using MLE.

Let $x_{l,k}$ be the monetary payoff associated with the probability $p_{l,k}$ of the outcome k for lottery l. The payoff $x_{l,k}$ and probabilities $p_{l,k}$ are induced by the experimenter, so that the cumulative distribution function for each lottery l is given by:

$$F_l(x_{l,k}) = P_l(X_l \leqslant x_{l,k}),$$

for outcomes $k = 1, ..., K$. In our experiments, $k = 2$, and there are two lotteries, so that $l = \{A, B\}$.

To compare the lotteries, we employ the framework described in Section 2.2 above. In particular, we specify two lotteries: B, the "high risk" lottery, associated with CDF F_B, and A, the "low risk" lottery, with CDF F_A. We use the quantile preserving spread concept, and assume that F_B crosses F_A from below at the quantile τ^*, making lottery B riskier than A.

Under the QP model, for each lottery pair in task i and τ^* (crossing point of the CDFs), the choice between lotteries A_i and B_i is based on the difference of quantiles that is calculated for a candidate estimate of τ by evaluating the differences index

$$\Delta Q_i(\tau) \equiv Q_\tau[B_i] - Q_\tau[A_i]. \tag{4}$$

Equation (4) indicates that for a fixed quantile τ, an agent chooses B_i over A_i when $Q_\tau[B_i] > Q_\tau[A_i]$. The latent index, based on latent preferences, is then linked to the observed choices using a specified CDF, denoted by $G(\cdot)$. In particular, we follow the literature and use a latent variable model to derive the likelihood function to estimate the parameter τ.

We model the latent (unobserved) variable y_i^* for task i as the difference in quantiles plus an error term, and assume that the observed binary response variable y_i is derived in the following manner.

$$y_i^* = \Delta Q_i(\tau) + \varepsilon_i$$
$$y_i = \begin{cases} 1, & \text{if } y_i^* > 0 \text{ (choose lottery B)} \\ 0, & \text{otherwise.} \end{cases}$$

Let the CDF of ε_i be denoted by G. The CDF G, or link function, takes any argument on the real line, in this case the difference $\Delta Q(\tau)$, and transforms it into a number between 0 and 1. Thus, the probability of choosing lottery B,

[19] One popular alternative is the expo-power utility function, $u(x) = \theta - \exp\{-\beta x^\alpha\}$, introduced by Saha (1993). This utility function has two free parameters (when θ is fixed as a constant), and allows testing of whether decreasing absolute risk aversion and increasing relative risk aversion are present. While this form can typically provide a better fit than CRRA, it would have to be compared to a quantile rule with two parameters, perhaps with one parameter indicating how the quantile being maximized varies by income.

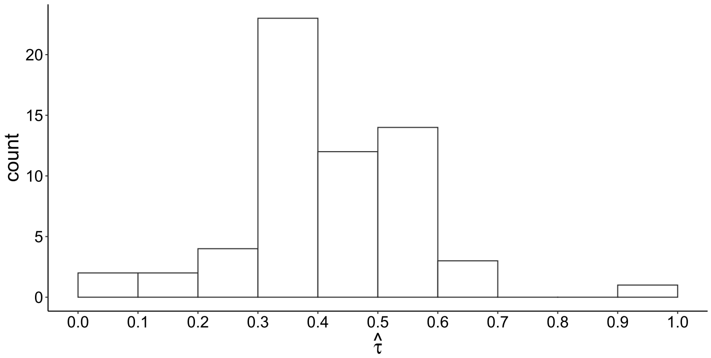

Fig. 5. Histogram of individual $\hat{\tau}$ estimates, all participants (mean=0.42; standard error=0.02).

$\text{Prob}(\text{choose lottery B}) = \text{Prob}(y_i = 1)$, which depends on the link function, can be calculated. The choice probability can be written as

$$\text{Prob}(y_i = 1) = P(\varepsilon_i > -\Delta Q_i(\tau)) = 1 - G(-\Delta Q_i(\tau))$$
$$\text{Prob}(y_i = 0) = G(-\Delta Q_i(\tau)).$$

To write the likelihood function, suppose we have a random sample (y_i, w_i) for tasks $i = 1, ..., n$, with $w_i = \Delta Q_i(\tau)$. For a symmetric link function G, the likelihood is

$$\mathcal{L}(y_i, w_i \mathcal{I} \tau) = \prod_{i:y=0} (1 - G(\Delta Q_i(\tau))) \prod_{i:y=1} G(\Delta Q_i(\tau)) = \prod_{i=1}^{n} G_i^{y_i} (1 - G_i)^{1-y_i}.$$

Then, the log-likelihood function is given by

$$\ell(y_i, w_i \mathcal{I} \tau) = \sum_{i=1}^{n} \left[y_i \log G(\Delta Q_i(\tau)) + (1 - y_i) \log(1 - G(\Delta Q_i(\tau))) \right]. \tag{5}$$

The MLE of τ is simply the value of τ that maximizes the log-likelihood function in equation (5). The CDF function $G(\cdot)$ can be specified as, for instance, the Logit (logistic distribution), the Probit (normal distribution), or another distribution.

We remark that, in contrast to the EU case, where the log-likelihood is smooth and differentiable everywhere, the log-likelihood function in equation (5) is not smooth. Nevertheless, there is existing literature in econometrics establishing the asymptotic properties – consistency, asymptotic normality, and bootstrap inference – for this class of semiparametric estimators (as the MLE), where the criterion function does not obey standard smoothness conditions. The theories allow for non-smooth objective functions of finite-dimensional unknown parameters (e.g., Pakes and Pollard (1989) and Newey and McFadden (1994, Section 7)) and both finite-dimensional and infinite-dimensional parameters (e.g., Chen et al. (2003)). In addition, Chen et al. (2003) show that bootstrapping for these methods provides asymptotically correct confidence regions for finite-dimensional parameters. Throughout this paper, we apply bootstrap procedures, with 1,000 bootstrap repetitions, to compute the standard errors of the parameters of interest.

4.1.3. Estimates of risk attitude under both models

We specify a Logit model for the estimation of the parameters in both QP and EU models, so that the link function is a logistic distribution with location parameter zero and scale two.[20] We first estimate the parameters for each participant separately. Fig. 5 presents a histogram of the distribution of quantile estimates $\hat{\tau}$ (risk attitude) from tasks 1–2 together (242 decisions). The figure shows that most subjects have an estimated risk attitude below 0.5, indicating that they are more risk averse than a maximizer of the median payoff. About 80% of subjects have quantile estimates between 0.3 and 0.6. The average, over all the quantile estimates, is 0.42, and the sample standard error is 0.02.[21] As Fig. 5 demonstrates, the distribution is skewed to the right, with relatively few estimates exhibiting risk attitude above $\hat{\tau} = 0.6$.

[20] Harrison and Rutström (2008) demonstrate, based on the work of Wilcox (2008) and Wilcox (2011), that CRRA estimates under the error structure we assume are sensitive to the scaling of the errors. We have chosen our scale so that the CRRA parameter estimate corresponds closely to that typically found in other studies, and apply the same scale to the estimation of the QP model. A scale that is too small makes the likelihood function insensitive to large values in the index. In Section 2 of the Online Appendix we present estimates for different scale parameters and discuss their effect on the results.

[21] Since the likelihood function for the QP model is not smooth, the quantile estimate used throughout our analysis is the midpoint of the interval in which the likelihood function is maximized.

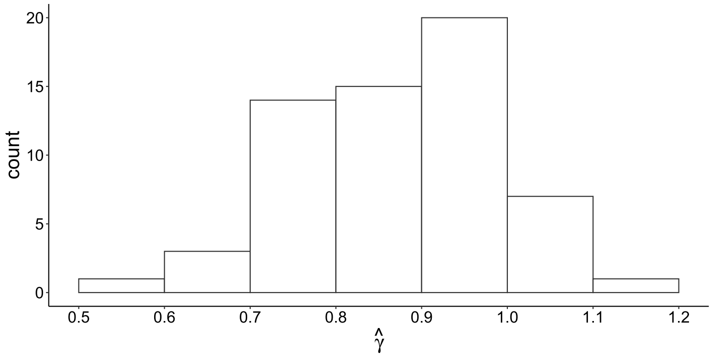

Fig. 6. Histogram of individual γ estimates, all participants (mean=0.87; standard error=0.02).

Fig. 6 shows the distribution of CRRA coefficient $\widehat{\gamma}$ estimates in (3) using the analogous Logit MLE for the data from tasks 1–2. The estimates are in line with the existing literature. The average estimate of $\widehat{\gamma}$ is 0.87, with a standard error of 0.02. Relatively few of estimates exhibit $\widehat{\gamma} > 1$. Moreover, about 79% of subjects present $\widehat{\gamma}$ smaller than 0.975, suggesting that a large majority are risk averse. About 11% of subjects have $\widehat{\gamma}$ larger than 1.025 and thus can be classified as risk seeking. The remaining 10% of $\widehat{\gamma}$ is in the interval of $[0.975, 1.025]$ which is essentially risk neutral. The proportions that are risk averse, risk seeking and risk neutral are comparable to those in previous studies. This suggests that the scale parameter in our estimation is appropriately calibrated for comparison to previous work.

We also compute pooled risk attitude parameter estimates for both the QP and EU, treating the entire sample as one "representative" participant. For risk aversion under the EU model, we obtain a point estimate of $\widehat{\gamma} = 0.85$ with a standard error of 0.02. For the QP model, the point estimate is $\widehat{\tau} = 0.35$ with a standard error of 0.01. The standard error is computed by applying the bootstrap-based cluster robust standard errors of Cameron et al. (2008), with the clustering at the level of the individual participant. In other words, assuming that individuals have CRRA utility of the power utility form, the group of 61 subjects behaves like a representative person with a risk aversion coefficient of 0.85. Analogously, under the assumption of QP maximization, the "representative" subject acts as if she is maximizing the 35th percentile of her earnings.

4.2. Comparison of QP and EU models

We now consider the relative performance of the EU and QP models. We use two methods to compare the models. First, we classify participants as QP or EU maximizers based on the corresponding statistical model that better describes their decisions using classification methodology. Second, we use a model selection test to choose among the parametric likelihoods.

For the first comparison measure, we use simple classification methods. We compute a confusion matrix (see, e.g., James et al. (2017)) using the Logit model for both QP and EU. A confusion matrix compares the predictions of the statistical model with the actual choices for all the observations in the data set. For each of the two models, the first step in computing the matrix is to calculate the predictions for all observations for each subject. This is implemented by computing the probability $P(y_i = 1)$. In our case, choice 1, $y = 1$, is choosing lottery B. The alternative choice, $y = 0$, is choosing lottery A. Once the parameters have been estimated, it is easy to compute the predicted probability of choosing lottery B for any given difference in payoffs. For example, one computes $\widehat{P}(y_i = 1)$ for every decision i. Nevertheless, since this predicted probability is a number between zero and one, to classify an observation as choice 1, we have to specify a threshold. For instance, if 0.5 is the threshold, then for $\widehat{P}(y_i = 1) \geqslant 0.5$ we assign observation i to class 1, otherwise observation i is assigned to class 0. Finally, these predictions can be compared to the actual choices, and to the percentage of correct predictions, and the accuracy rate can be computed.

By constructing the confusion matrix and computing the accuracy rate of both the QP and EU models for each subject's data, we can classify individuals into three categories: EU maximizer, QP maximizer, and tie. The methodology to classify a subject into a group simply compares the accuracy rate of the two models. The accuracy rate is defined as the rate of the correct classifications – true positive plus true negative – divided by the total number of classifications (see, e.g., Kuhn and Johnson (2013) for a discussion on different measures of performance in classification models.) Therefore, first, for each QP and EU model, we compute the confusion matrices. Second, for each QP and EU model, we compute its accuracy rate. Third, we compare the accuracy rates for both models. If the accuracy rate over all of the decisions is higher for the QP model than the EU model, the subject is classified as a QP maximizer. When the contrary occurs, the individual is an EU maximizer. A tie happens when the accuracy rates are equal for the two models. The results for the classification appear in Table 2 for three different choices of the threshold parameter: 0.4, 0.5, and 0.6. The table shows the number of subjects in each category, the percentage classified as QP or EU (excluding ties), and 95% confidence intervals of these percentages.[22]

[22] We compute the confidence intervals using binomial proportions.

Table 2
Number and Percentage of participants classified as QP Maximizers, EU Maximizers, and Ties. Classification based on the accuracy rate of models' predictions. Percentages are given in parentheses, and 95% confidence intervals of percentages are in brackets.

Threshold probability	# QP Maximizers	# EU Maximizers	# Ties
$p = 0.40$	21 (35.6%)	38 (64.4%)	2
	[23.4%, 47.8%]	[52.2%, 76.6%]	
$p = 0.50$	19 (32.2%)	40 (67.8%)	2
	[20.3%, 44.1%]	[55.9%, 79.7%]	
$p = 0.60$	25 (41.7%)	35 (58.3%)	1
	[29.2%, 54.1%]	[45.9%, 70.8%]	

Table 3
Number and Percentage of participants classified as QP Maximizers, EU Maximizers, and Ties. Classification based on the test for proportion of accuracy rate of models' predictions.

Threshold probability	# QP Maximizers	# EU Maximizers	# Ties
$p = 0.40$	10 (50.0%)	10 (50.0%)	41
	[28.1%, 71.9%]	[28.1%, 71.9%]	
$p = 0.50$	10 (43.4%)	13 (56.6%)	38
	[23.2%, 63.7%]	[36.3%, 76.8%]	
$p = 0.60$	11 (55.0%)	9 (45.0%)	41
	[33.2%, 76.8%]	[23.2%, 66.8%]	

The table indicates that a majority of participants are classified as EU under each of the three criteria. However, a significant proportion of the subjects who can be classified, between 32% (19 out of 59) and 42% (25 out of 60), adhere to QP. This result is a clear indication of the relevance of the QP model.

We also employ a validation set approach as an alternative classification method. In this case, we split the sample into training and testing parts. The former sample is used for estimating the model, and the latter for classification. All these results are reported in Section 3 of the Online Appendix. We redo Tables 2 and 3 for the validation approach and the results are qualitatively similar to those in this section.

Moreover, we compare the two decision models using both parametric and non-parametric within-subject tests for the threshold probability $p = 0.5$. We first subtract the accuracy rate of the EU from that of the QP model for each individual to compute a within-subject measure of relative performance of the two models. The sample average for the difference is -0.009 with a standard error of 0.006. We then test the null hypothesis that the average difference is equal to zero. The corresponding t-test statistic is -1.45 (p-value 0.15), and we cannot reject the null hypothesis, at standard levels of significance, that the average difference in the accuracy rates of the QP and EU models is statistically equal to zero. The sample median for the difference is -0.017. We test the null hypothesis that this median is equal to zero. A Wilcoxon signed rank test yields a test statistic equal to 666.5 (p-value 0.10). Moreover, a paired-sample sign test of the null hypothesis that the difference is equally likely to be positive or negative yields a p-value of 0.01. Hence, there is some evidence that the EU model describes decisions better than the QP model for a significant majority of participants.[23]

The classification reported in Table 2 directly compares the accuracy rate in the two models. Next we report results for classification after formally testing for the equality of the accuracy from the two observed confusion matrices. Specifically, we test for the equality of the proportion of concordant elements in both QP and EU classifications. Define the accuracy proportion as π. The null hypothesis is $H_0 : \pi_{QPL} = \pi_{EUL}$. This test is performed using the classical approach based on the comparison between two proportions at a 10% level of significance. The results are reported in Table 3. The first feature in the table is that the number of ties, that is the number of non-rejections of the null hypothesis, is relatively large. When examining the percentage of comparisons that reject the null and are classified, the results show evidence that about half of the subjects are EU maximizers and half QP maximizers.

The second method we use to compare the relative performance of EU and QP is a model selection test. We apply a test recently suggested by Schennach and Wilhelm (2017) for choosing among two parametric likelihoods. This test builds on Vuong (1989) and it is a simple method that delivers a model selection criterion based on the Kullback-Leibler (KL) discrepancy and yet only involves a test statistic that is asymptotically normally distributed in all cases (nested, nonnested, or overlapping), under the null that the two models fit the data equally well. Therefore, no pretesting is required, complicated limiting distributions are entirely avoided, and importantly, the test uniformly controls for size.[24]

[23] For the threshold probabilities of $p = 0.4$ and $p = 0.6$, the results for the average and median tests reach the same conclusions as for the $p = 0.5$ case. The results for the paired-sample sign tests have p-values of 0.04 and 0.25, for $p = 0.4$ and $p = 0.6$, respectively.

[24] Vuong (1989) established that the difference between the KL information criterion of two competing models exhibits a wide variety of limiting distributions, depending on whether the two models are overlapping or not, or whether one of the models is correctly specified or not. As a result, using the Vuong's test typically requires pretesting to establish which distribution to use for the computation of critical values for the tests.

Table 4
Number and Percentage of participants classified as QP Maximizers, EU Maximizers, and Ties. Classification based on the Shennach and Wilhelm (2017) test.

Threshold probability	# QP Maximizers	# EU Maximizers	# Ties
$\epsilon = 0$	18 (48.6%)	19 (51.4%)	24
	[32.5%, 64.8%]	[35.2%, 67.5%]	
$\epsilon = 0.25$	19 (47.5%)	21 (52.5%)	21
	[32.0%, 63.0%]	[37.0%, 68.0%]	
$\epsilon = 0.5$	19 (48.7%)	20 (51.3%)	22
	[33.0%, 64.4%]	[35.6%, 67.0%]	

A model is defined to be better if it is closer to the true distribution in the KL sense. Let $P_{\theta^*_{EU}}$ and $P_{\theta^*_{QP}}$ be the distributions in sets $\mathcal{P}_{EU}$ and $\mathcal{P}_{QP}$, which are closest to the true distribution, P_0, respectively. Formally, model EU is defined to be better than model QP if model EU's KL distance to the truth is smaller than that of model QP, that is, $K(P_0 : \mathcal{P}_{EU}) < K(P_0 : \mathcal{P}_{QP})$. If the two KL distances are equal, then we say models EU and QP are equivalent. The procedure proposed in Schennach and Wilhelm (2017) selects the better model based on performing a test of $H_0 : K(P_0 : \mathcal{P}_{EU}) = K(P_0 : \mathcal{P}_{QP})$, that is, models EU and QP are equivalent, against model EU is better, $H_a : K(P_0 : \mathcal{P}_{QP}) > K(P_0 : \mathcal{P}_{EU})$, or QP is better, $H_b : K(P_0 : \mathcal{P}_{QP}) < K(P_0 : \mathcal{P}_{EU})$. The practical implementation of the test requires the choice of a regularization parameter, ϵ.[25] The results are presented in Table 4 for three different choices of ϵ at a 10% level of significance, and show that overall about half of the subjects are QP maximizers and half EU maximizers. These results for the Vuong's type of test are, in terms of percentages, very similar to those using the proportion test discussed in Table 3.

5. Gender effects

In this section, we consider gender differences in the risk attitude estimates and the relative performance of the two models.[26] As discussed earlier, a substantial fraction of prior studies have found that on average, women are more risk averse than men (see, e.g., Eckel and Grossman (2008) for a review). If a similar pattern exists for QP maximization, one might expect women on average to have lower τ estimates than men.

Figs. 7 and 8 present histograms of the distribution of the risk attitude parameter estimates for the EU and QP models, for females and males separately. Fig. 7 reveals a gender difference in the $\widehat{\gamma}$ estimates. The top histogram shows that overall, females have smaller risk aversion coefficients, and thus are more risk averse, compared to males. The average estimate for γ is 0.83 for women and 0.91 for men. The t-statistic for the equality of average γ for males and females is $t = 2.81$, which rejects the null of equality at the 1% level. The tests are one-sided, since there is prior evidence that women are more risk-averse than men on average. In Fig. 8, we observe an analogous pattern for the QP model. The quantile estimates are more concentrated on small values for females. A larger proportion of females, relative to males, has an estimated $\widehat{\tau} < 0.3$, while the opposite is observed for $\widehat{\tau} > 0.6$. The average estimate of τ is 0.38 for women and 0.45 for men. The t-statistic for the null hypothesis of the equality of average τ for males and females is $t = 1.87$, so that we reject the null of equality at the 5% level.[27]

We now consider whether there are gender differences in the percentage of participants adhering to each model. The results are given in Table 5, for different threshold probabilities. The table shows that the QP model fits well for a greater percentage of females than males. We formally test whether the proportion of adherents that are female is the same in the QP and EU groups. The test statistics for threshold probabilities $p = 0.4$ and $p = 0.5$ are -0.76 and -1.29, respectively. Hence, we are not able to reject the null hypothesis that the gender ratio is equal in the QP and EU groups. However, the test statistic is -2.27 (p-value < 0.03) for the threshold probability $p = 0.6$. Therefore, there is some weak evidence that the percentage of females adhering to QP is greater than that of males.

We also provide classification results by gender using the proportion test for accuracy, and the Schennach and Wilhelm (2017) model selection test. The results are provided in Tables 6 and 7, respectively. As in the previous section, these results have a larger number of ties. Both tables show a larger number of females for the QP model.

In Section 4 of the Online Appendix we investigate the effect of stakes on the risk attitude estimates.

[25] The procedures in Schennach and Wilhelm (2017) require the likelihood functions to be twice continuously differentiable, which is not satisfied in the QP. Nevertheless, we include the results here in spite of this technical requirement.

[26] In Section 5 of the Online Appendix, we present the results of regressions of QP and EU estimates on gender, the program of study and CRT score.

[27] We can also conduct tests for each of the two decision problems separately. For the QP model, the t-statistic for the difference in means between males and females, when $b_2 = 16$, is $t = 1.69$. The t-statistic for the difference in means between males and females for $b_2 = 24$ is $t = 1.98$. Moreover, for the EU model, the t-statistic for the difference in means between males and females when $b_2 = 16$ is $t = 2.33$, and for $b_2 = 24$, it is $t = 2.84$. Thus, for both QP and EU models, we reject the null hypothesis that the risk aversion parameter is the same for males and females in all cases at the 5% significance level.

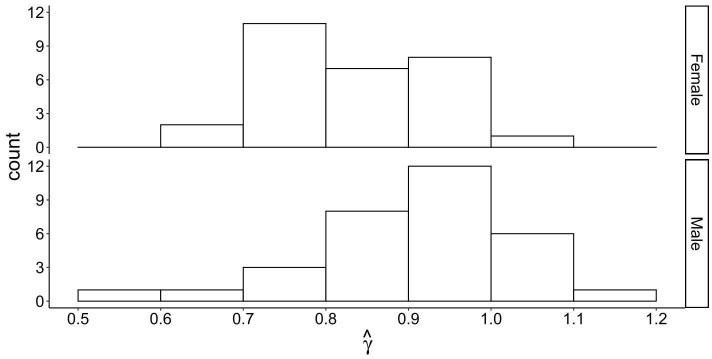

Fig. 7. Histograms of individual γ estimates by gender. Female mean= 0.83 and standard error=0.02, male mean=0.91 and standard error=0.02.

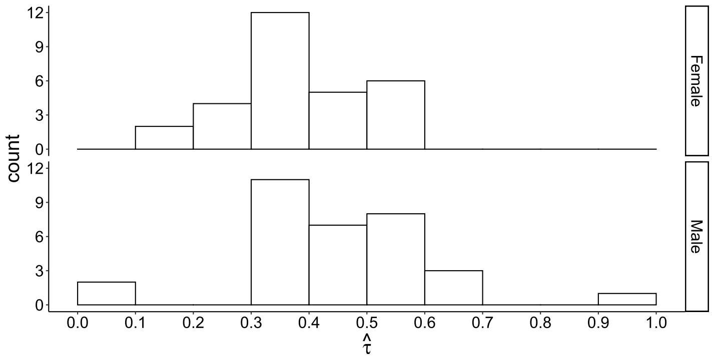

Fig. 8. Histograms of individual τ estimates by gender. Female mean=0.38 and standard error=0.02, male mean=0.45 and standard error=0.03.

Table 5
Number of participants classified as QP Maximizers, EU Maximizers, and Ties, by gender. Classification based on the accuracy rate of models' predictions.

Threshold probability	# QP Maximizers		# EU Maximizers		# Ties	
	Male	Female	Male	Female	Male	Female
$p = 0.40$	10	11	22	16	0	2
$p = 0.50$	8	11	24	16	0	2
$p = 0.60$	9	16	23	12	0	1

Table 6
Number of participants classified as QP Maximizers, EU Maximizers, and Ties, by gender. Classification based on the proportion test of model prediction accuracy.

Threshold probability	# QP Maximizers		# EU Maximizers		# Ties	
	Male	Female	Male	Female	Male	Female
$p = 0.40$	4	6	8	2	20	21
$p = 0.50$	4	6	9	4	19	19
$p = 0.60$	5	6	6	3	21	20

Table 7
Number of participants classified as QP Maximizers, EU Maximizers, and Ties, by gender. Classification based on the Shennach and Wilhelm (2017) test.

Threshold probability	# QP Maximizers		# EU Maximizers		# Ties	
	Male	Female	Male	Female	Male	Female
$\epsilon = 0$	7	11	12	7	13	11
$\epsilon = 0.25$	8	11	13	8	11	10
$\epsilon = 0.5$	8	11	13	7	11	11

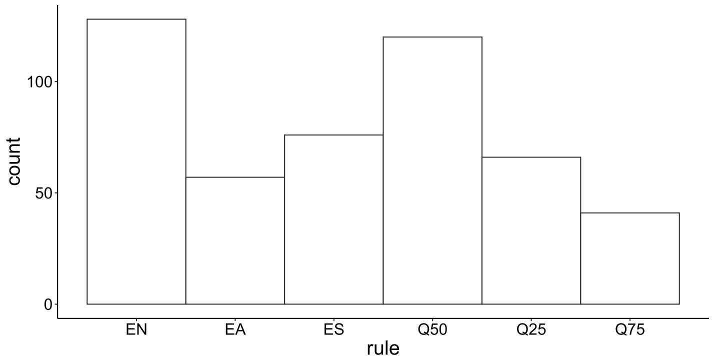

Fig. 9. Histogram of observed rule choices in L3–L10, all participants. The horizontal labels are as follows: EN risk neutral; EA risk averse; ES risk seeking; Q50 the median; Q25 the 25th percentile; Q75 the 75th percentile.

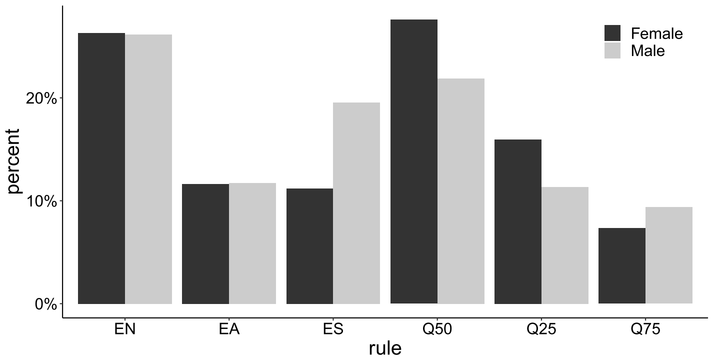

Fig. 10. Histogram of observed rule choices in L3–L10 by gender. The horizontal labels are as follows: EN risk neutral; EA risk averse; ES risk seeking; Q50 the median; Q25 the 25th percentile; Q75 the 75th percentile.

6. What decision model do individuals intend to use?

In this section we analyze the data from Periods 3–10. Recall from Section 3 that in these periods we directly ask subjects to choose the decision rule they wish to employ from six available alternatives. The goal is to directly measure the intended rule choice of participants. One very important advantage of this new methodology is that there is no requirement for a statistical model to analyze the data. We measure the decisions subjects make directly.

6.1. Summary statistics for periods 3 to 10

As discussed in Section 3.2 the rules EN, EA, and ES correspond to decisions within the EU model, whereas rules Q50, Q25, and Q75 are implementations of the QP model. Fig. 9 plots the histogram of the decision rule choices of all participants. There are two prominent patterns in the figure. The first is that there is a comparable level of choices of EU and QP rules. The second is that maximizing expected value (EN) and median payoff (Q50) are the most commonly used rules within the EU and QP classes, respectively. These two rules capture risk neutrality in their respective models.

Fig. 10 displays a histogram of the observed rule choices for females and males separately. The figure shows that more women choose QP rules than men, while men are more likely to opt for the risk-seeking EU rule than are women. We reject the hypothesis that the proportion of females in the QP group is equal to the proportion in the EU group ($t = -1.83$, p-value < 0.1). Thus, we obtain some more evidence that women tend to use quantile rules more than men. Within the subset of those individuals who choose EU rules, we reject the hypothesis that there is an equal proportion of men and women opting for the ES rule ($t = -1.98$, p-value < 0.05).

Fig. 11 depicts the decision rule choices for each of the eight Periods L3–L10, separately. Each panel concerns a different payoff vector for each period, and the relationships among L3–L10 are summarized in Table 1 in Section 3.2 above. The figure shows similar patterns as those in Fig. 9. There is a comparable incidence of choices of EU and QP rules, and the most popular rules are EN and Q50. The Q50 rule is the most commonly used in L3, L4, L6, and L9, while EN is most common in L5, L8, and L10. L5 and L8 are the lotteries with the lowest stakes. In addition, task L8 is the only one for which the quantile-preserving spread relationship does not hold between lotteries A and B. The percentages choosing QP rules in the eight tasks L3–L10 are, respectively, 48%, 46%, 38%, 51%, 46%, 38%, 51%, and 56%.

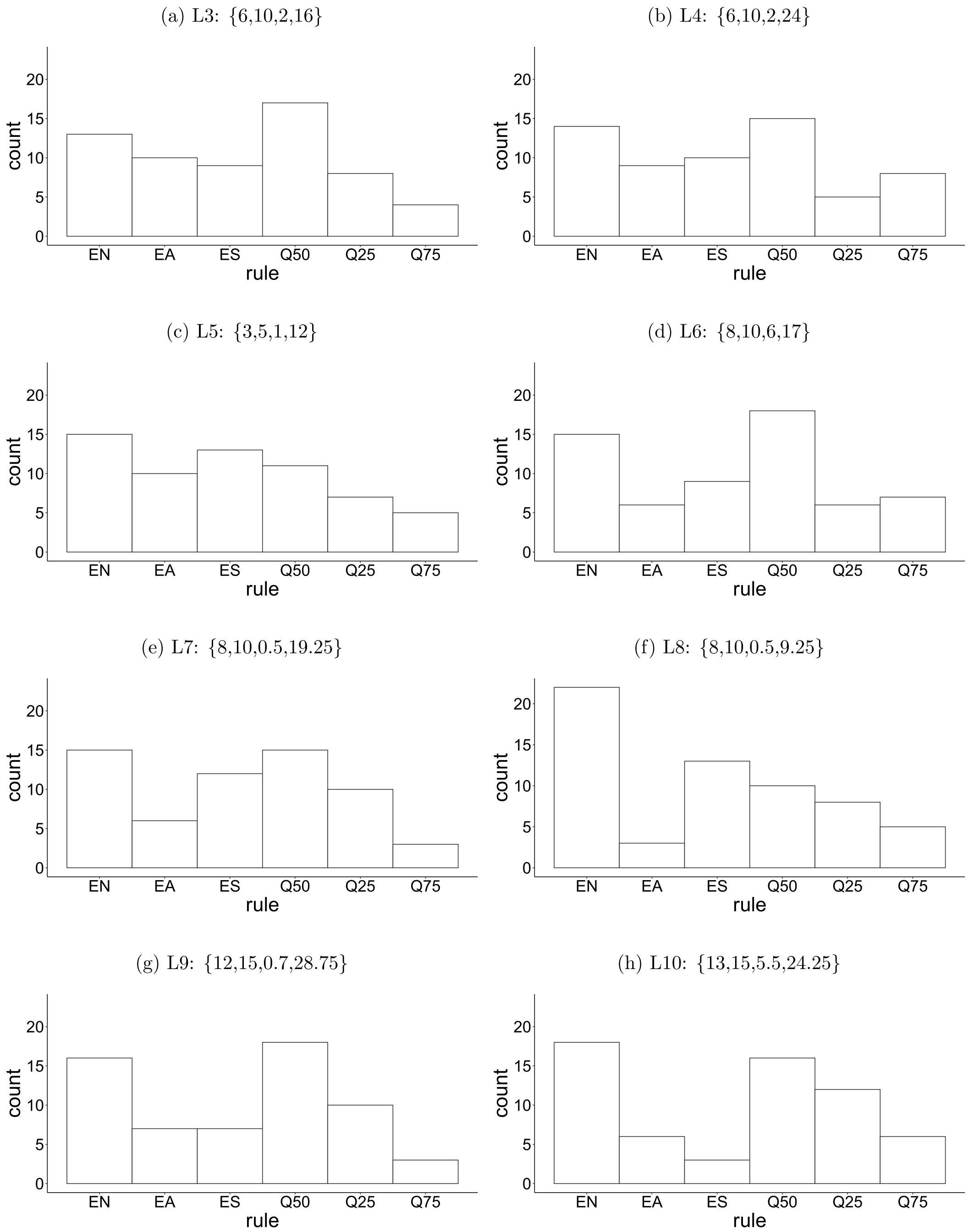

Fig. 11. Histograms of observed rule choices in L3–L10. The horizontal labels are as follows: EN risk neutral; EA risk averse; ES risk seeking; Q50 the median; Q25 the 25th percentile; Q75 the 75th percentile. The numbers in curly brackets indicate the payoff vector $\{a_1, a_2, b_1, b_2\}$.

6.2. Classification

We classify subjects as QP or EU maximizers based on the decision rules they choose. There is no need to use an underlying statistical model, and instead we employ a simple count of their choices. We adopt three different criteria to assign participants into three categories: QP maximizer, EU maximizer, and mixture. To satisfy the least strict criterion, a subject has to choose a rule from one model in at least 5 of the 8 total periods. If the criterion is not satisfied (in cases where an individual chooses EU and QP rules in four instances each), the person is classified as employing a mixture. The stricter criteria require a subject to be consistent with one specific model in at least 6 and 7 periods, respectively. Thus, we

Table 8
Number and Percentage of participants classified as QP Maximizers, EU Maximizers, and Mixture categories, all participants pooled and by gender. Classification based on rules selected in tasks L3–L10. Percentages are in parentheses, and 95% confidence intervals of percentages are in brackets.

Number Choices		# QP Maximizers		# EU Maximizers		# Mixture
⩾ 5		19 (39.6%) [25.7%, 53.5%]		29 (60.4%) [46.6%, 74.3%]		13
⩾ 6		13 (43.3%) [25.6%, 61.1%]		17 (56.7%) [38.9%, 74.4%]		31
⩾ 7		10 (45.5%) [24.6%, 66.3%]		12 (54.5%) [33.7%, 75.5%]		39
	Male	Female	Male	Female	Male	Female
⩾ 5	10	9	18	11	4	9
⩾ 6	6	7	11	6	15	16
⩾ 7	3	7	8	4	21	18

Table 9
Number and Percentage of participants classified as QP Maximizers, EU Maximizers, and Mixture categories, all participants pooled and by gender. Classification based on rules selected in tasks L3–L10 (excluding L8). Percentages are in parentheses, and 95% confidence intervals of percentages are in brackets.

Number Choices		# QP Maximizers		# EU Maximizers		# Mixture
⩾ 4		27 (44.3%) [31.8%, 56.7%]		34 (55.7%) [43.3%, 68.2%]		–
⩾ 5		16 (44.4%) [28.2%, 60.7%]		20 (55.6%) [39.3%, 71.8%]		25
⩾ 6		12 (50%) [30%, 70%]		12 (50%) [30%, 70%]		37
	Male	Female	Male	Female	Male	Female
⩾ 4	13	14	19	15	–	–
⩾ 5	8	8	12	8	12	13
⩾ 6	5	7	8	4	19	18

allow varying degrees of tolerance of error from the intended decision model under the different criteria. We assume that there is an equal likelihood of a QP agent playing EU as an EU adherent playing QP.

The number of subjects in each category, the percentage classified as QP or EU (excluding mixture), and the 95% confidence intervals for the percentages, computed using binomial proportions, are reported in Table 8. The lower part of the table displays the data for the number of subjects classified in each group for each gender separately. The table shows that while a plurality of subjects has the intention to maximize expected utility, a considerable minority chooses to maximize a quantile function. Applying the least strict threshold, 38 individuals choose a rule; of those, about 40% of participants choose a quantile rule, while about 60% intend to maximize EU. When we tighten the criterion to $\geqslant 7$, the number of individuals classified as using the QP and EU models becomes very similar, with 10 (46%) and 12 (55%) of those classified employing each model, respectively. Overall, these results are similar to those in Table 2 and corroborate the evidence from the first part of the experiment that a relatively large minority of individuals' economic decisions are more consistent with quantile, than with expected utility, maximization.

The results by gender are given in the lower part of Table 8. They show that for those adhering to the QP model, the numbers of males and females are close to equal for the two less strict categories. For the strictest category, there is a greater number of females than males. We conduct a formal test of the null hypothesis of equality of the proportion of females between the QP and EU categories. Applying the $\geqslant 5$ criterion yields a two-sample test statistic of -0.65, so that we are not able to reject the null hypothesis that the models are identical in terms of the proportion of adherents who are female. If we apply the $\geqslant 6$ criterion, the test statistic is -1.02. However, when using the $\geqslant 7$ criterion, the test statistic is equal to -1.71, sufficient to conclude, at the 10% level of significance, that males are less likely to choose QP maximization, despite the low power to reject equality under this strict criterion for what constitutes an observation. This constitutes additional weak evidence that women have a greater tendency to employ the QP model than men.

As discussed in the previous section, the quantile-preserving spread relationship does not hold between lotteries A and B in task L8. Hence, we recompute the results in Table 8 excluding task L8. In this case, we have a total of seven tasks and use 4, 5, and 6 as the number of choices consistent with a model as classification criteria. The results are given in Table 9, with the lower part separating the results by gender. When comparing the results between Tables 8 and 9, we see that the number of QP maximizers increases substantially when L8 is removed, and about 44% of the total number of subjects are classified as QP maximizers for the least stringent $\geqslant 4$ rule, while an equal number are classified as EU and QP under the strictest $\geqslant 6$ standard (12 employing each model). The patterns regarding gender in Table 9 are similar to those in Table 8, with females having a stronger tendency to employ QP rules than males. We conduct a two-sample test to assess the equality of proportions of gender between QP maximizers and EU maximizers. The two-sample test statistics are both

Table 10
Level of agreement for the classification of participants in L1 vs. L3, and L2 vs. L4. Classification based on the accuracy rate of models' predictions in L1 and L2 and the decision rule chosen in L3 and L4.

Threshold probability	L1 vs. L3	L2 vs. L4
$p = 0.40$	32/61	31/61
$p = 0.50$	30/61	31/61
$p = 0.60$	27/61	28/61

-0.60 when applying the $\geqslant 4$ and $\geqslant 5$ criteria, and -1.23 under the $\geqslant 6$ criterion. Thus, we are not able to reject the null hypothesis of no gender difference in the use of the QP and EU models in any of the three cases.

While the percentages of individuals who adhere more closely to each model are similar in both parts of the experiment, there is no evidence of consistency between the two methods of classification at the individual level. Table 10 reports the proportion of individuals for which there is an agreement between how they are classified from their choices in L1 vs. L3, and in L2 vs. L4, the two pairs of decisions which have identical parameters. The table reports the percentage classified consistency based on the accuracy rate with three different probability thresholds for the first part of the experiment.

The data show that the agreement proportion is not greater than chance. Such instability in fitting decisions in experimental individual risky decision making tasks to models is quite typical. For example, Deck et al. (2013) find little evidence of correlation of risk aversion estimates among four different tasks to measure risk aversion (only two pairwise correlations out of six are significant). Dave et al. (2010) and Crosetto and Filippin (2016) also find that different techniques to measure risk aversion yield very different estimates for the exact same individuals. Both Dave et al. (2010) and Crosetto and Filippin (2016) find that individuals are more risk averse under the Holt-Laury than the Eckel-Grossman procedures, the two most common risk aversion measurement protocols. Such instability may reflect, at least in part, intentional mixing by individuals (Agranov and Ortoleva, 2017). Here, we find that there is similar instability when different techniques are used to classify individuals by the decision model that they employ.

7. Conclusion

This paper compared the performance of the standard model of choice under risk, expected utility (EU), with a model of quantile preference (QP) maximization. As the EU model requires a choice of the utility function, we have assumed the CRRA form, since it is also standard. We considered: (1) the fit of the two models to participants' binary lottery decisions, and (2) the percentage of the time agents chose to employ each model from the description. The EU model provides a better fit for a small majority of decision makers, while the QP model outperforms for a considerable fraction, 32%–55% of our sample, depending on the classification criterion. In our view, the large fraction using quantile rules justifies continued exploration of the properties of quantile models of decision making and their application to economic contexts.[28]

Of course, this study is only a first step in the evaluation of the model. One might argue that the CRRA functional form is not appropriate to employ with EU, and that the mediocre performance of the EU model is due to the assumption of CRRA. Perhaps a utility function with more parameters that is known to provide a better fit to laboratory data, such as expo-power utility, might perform better. However, it is also not obvious how one could describe expo-power utility in terms that participants could understand, as would be necessary for Part 2 of our experiment. Expo-power utility has two parameters, and thus is less parsimonious than CRRA. Unless we correct for the different numbers of parameters, it would have to be evaluated against an extension of the quantile model with an additional parameter, perhaps one that would allow the quantile to vary depending on a measure of the stakes of the lottery payoffs. Indeed, one of the advantages of the QP model is that decisions do not depend on the utility function.

Many issues remain to be investigated. In addition to pitting the quantile model against different functional forms under EU, it would be interesting to evaluate it against cumulative prospect theory, or other models that allow probability weighting. Again, however, it is unclear how a protocol like that of our Part 2 could be used, because probability weighting would have to be explained to individuals who are choosing between decision rules. A model with probability weighting and a utility curvature parameter would presumably provide a better fit to the data than EU, which is a special case of the model. However, it would be less parsimonious, with more free parameters, than the QP model. Another direction for future research would be to offer a 'no rule' choice option in the second part of the experiment. This would provide further clarity on how well each of the two models explains choices and may allow any other favored decision rules to emerge.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.geb.2021.11.010.

[28] We do not argue that a model with a single τ parameter would apply at any level of monetary stakes. Clearly, if we were to consider a case where $b_2 = \$1,000,000$. We presumably would not observe many individuals maximizing the same quantile as they would at $b_2 = \$10$. However, the same criticism would also apply to CRRA preferences under EU, whose parameter estimates for given individuals are known to differ considerably as the stakes of gambles change.

References

Abdellaoui, M., Bleichrodt, H., L'Haridor, O., 2008. A tractable method to measure utility and loss aversion under prospect theory. J. Risk Uncertain. 36, 245–266.
Agranov, M., Ortoleva, P., 2017. Stochastic choice and preferences for randomization. J. Polit. Econ. 125, 40–68.
Allais, M., 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'ecole americaine. Econometrica 21, 503–546.
Apiwatcharoenkul, W., Lake, L., Jensen, J., 2016. Uncertainty in proved reserves estimates by decline curve analysis. In: SPE/IAEE Hydrocarbon Economics and Evaluation Symposium. Society of Petroleum Engineers.
Bell, D.E., 1982. Regret in decision making under uncertainty. Oper. Res. 30, 961–981.
Bhattacharya, D., 2009. Inferring optimal peer assignment from experimental data. J. Am. Stat. Assoc. 104, 486–500.
Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-based improvements for inference with clustered errors. Rev. Econ. Stat. 90, 414–427.
Chambers, C.P., 2009. An axiomatization of quantiles on the domain of distribution functions. Math. Finance 19, 335–342.
Chen, X., Linton, O., van Keilegom, I., 2003. Estimation of semiparametric models when the criterion function is not smooth. Econometrica 71, 1591–1608.
Cox, J.C., Sadiraj, V., 2006. Small- and large-stakes risk aversion: implications of concavity calibration for decision theory. Games Econ. Behav. 56, 45–60.
Crosetto, P., Filippin, A., 2016. A theoretical and experimental appraisal of four risk elicitation methods. Exp. Econ. 19, 613–641.
Dave, C., Eckel, C.C., Johnson, C.A., Rojas, C., 2010. Eliciting risk preferences: when is simple better? J. Risk Uncertain. 41, 219–243.
de Castro, L., Galvao, A.F., 2019. Dynamic quantile models of rational behavior. Econometrica 87, 1893–1939.
de Castro, L., Galvao, A.F., 2021. Static and dynamic quantile preferences. Econ. Theory. https://doi.org/10.1007/s00199-021-01355-8.
Deck, C., Lee, J., Reyes, J.A., Rosen, C.C., 2013. A failed attempt to explain within subject variation in risk taking behavior using domain specific risk attitudes. J. Econ. Behav. Organ. 87, 1–24.
Duffie, D., Pan, J., 1997. An overview of value at risk. J. Deriv. 4, 7–49.
Eckel, C.C., Grossman, P.J., 2008. Men, women and risk aversion: experimental evidence. In: Plott, C., Smith, V. (Eds.), Handbook of Experimental Economics Results, vol. 1. Elsevier, New York, pp. 1061–1073.
Fanchi, J.R., Christiansen, R.L., 2017. Introduction to Petroleum Engineering. Wiley Online Library.
Frederick, S., 2005. Cognitive reflection and decision making. J. Econ. Perspect. 19, 25–42.
Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with a non-unique prior. J. Math. Econ. 18, 141–153.
Giovannetti, B.C., 2013. Asset pricing under quantile utility maximization. Rev. Financ. Econ. 22, 169–179.
Halevy, Y., Mayraz, G., 2020. Identifying rule-based rationality. University of Toronto. Mimeo.
Harrison, G.W., Johnson, E., McInnes, M.M., Rutström, E.E., 2005. Risk aversion and incentive effects: comment. Am. Econ. Rev. 95, 897–901.
Harrison, G.W., Rutström, E.E., 2008. Risk aversion in the laboratory. In: Cox, J., Harrison, G. (Eds.), Risk Aversion in Experiments. In: Research in Experimental Economics, vol. 12. Emerald Group Publishing Limited, Bingley.
Harrison, G.W., Rutström, E.E., 2009. Expected utility theory and prospect theory: one wedding and a decent funeral. Exp. Econ. 12, 133–158.
Hey, J.D., Orme, C., 1994. Investigating generalizations of expected utility theory using experimental data. Econometrica 62, 1291–1326.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. Am. Econ. Rev. 92, 1644–1655.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2017. An Introduction to Statistical Learning with Applications in R. Springer, New York, NY.
Jorion, P., 2007. Value at Risk: The New Benchmark for Managing Financial Risk. McGraw-Hill.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–292.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York.
List, J.A., Haigh, M.S., 2005. A simple test of expected utility theory using professional traders. Proc. Natl. Acad. Sci. 102, 945–948.
Manski, C., 1988. Ordinal utility models of decision making under uncertainty. Theory Decis. 25, 79–104.
Mendelson, H., 1987. Quantile-preserving spread. J. Econ. Theory 42, 334–351.
Moffatt, P., 2016. Experimetrics: Econometrics for Experimental Economics. Red Globe.
Newey, W.K., McFadden, D.L., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. North Holland, Elsevier, Amsterdam.
Nielsen, K., Rehbeck, J., 2020. When choices are mistakes. California Institute of Technology. Mimeo.
Noussair, C.N., Trautmann, S., van de Kuilen, G., 2014. Higher order risk attitudes, demographics, and financial decisions. Rev. Econ. Stud. 81, 325–355.
Pakes, A., Pollard, D., 1989. Simulations and the asymptotics of optimization estimators. Econometrica 57, 1027–1057.
Quiggin, J., 1982. A theory of anticipated utility. J. Econ. Behav. Organ. 3, 323–343.
Rabin, M., 2000. Risk aversion and expected-utility theory: a calibration theorem. Econometrica 68, 1281–1292.
Rostek, M., 2010. Quantile maximization in decision theory. Rev. Econ. Stud. 77, 339–371.
Saha, A., 1993. Expo-power utility: a 'flexible' form for absolute and relative risk aversion. Am. J. Agric. Econ. 75, 905–913.
Savage, L., 1954. The Foundations of Statistics. Wiley, New York.
Schennach, S.M., Wilhelm, D., 2017. A simple parametric model selection test. J. Am. Stat. Assoc. 112, 1663–1674.
Shaw, J.I., Bergen, J.E., Brown, C.A., Gallagher, M.E., 2000. Centrality preferences in choices among similar options. J. Gen. Psychol. 127, 157–164.
Starmer, C., Sugden, R., 1989. Violations of the independence axiom in common ratio problems: an experimental test of some competing hypotheses. Ann. Oper. Res. 19, 79–102.
Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. J. Risk Uncertain. 5, 297–323.
von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior. Princeton University Press.
Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.
Wakker, P.P., 2008. Explaining the characteristics of the power (CRRA) utility family. Health Econ. 17, 1329–1344.
Wilcox, N.T., 2008. Stochastic models for binary discrete choice under risk: a critical primer and econometric comparison. In: Cox, J., Harrison, G. (Eds.), Risk Aversion in Experiments. In: Research in Experimental Economics, vol. 12, Emerald Group, Bingley.
Wilcox, N.T., 2011. 'Stochastically more risk averse:' a contextual theory of stochastic discrete choice under risk. J. Econom. 162, 89–104.