Evaluation of Similarity Models for Expected Utility Violations

David E. Buschena*
Montana State University

Joseph A. Atwood
Montana State University

Abstract

A body of work proposes a decision cost argument to explain expected utility (EU) violations based on pair similarity. These similarity models suggest various measures over the risky pairs that define decision costs and benefits. This paper assesses the empirical modeling success of these similarity measures in explaining risky choice patterns showing EU independence violations. We also compare model fit for these similarity models relative to EU and to a selected generalized EU model. Although the candidate models exhibit some degree of substitutability, our results indicate support for models that use relatively simple measures as instruments for similarity.

*David Buschena, Dept. of Agricultural Economics and Economics, Montana State University, Bozeman, MT 59717, phone: (406) 994-5623, fax: (406) 994-4838, e-mail: buschena@montana.edu

Keywords: expected utility, information criteria, risk, similarity

JEL classification: C520, C900, D810

Evaluation of Similarity Models for Expected Utility Violations

Maximization of expected utility (EU) has been the dominant explanation for risky choices. However, violations of EU first suggested by Allais (1953) and famously addressed by Kahneman and Tversky (1979) are of considerable interest in numerous fields and have led to a search for alternative models. The generalized EU models (GEU) have been the primary approach to these EU independence violations (see the survey by Starmer, 2000). These GEU models have modified choice axioms, giving rise to a reduced-form maximization that reflects both underlying preference and nonlinear treatment of the probabilities.

Another approach to EU violations considers pair similarity (Rubinstein, 1988; Leland, 1994; Buschena and Zilberman, 1995, 1999a, b; Coretto, 2002; Loomes, 2006), which assumes that decision makers may use different decision-making criteria for disparate risky choice pairs. These decision makers are modeled to apply a two-stage decision process: first, selecting a decision rule, and then making the actual risky choice. This decision algorithm selection depends on the similarity between the risky choices. A simple algorithm such as expected value maximization is chosen when the alternatives are similar, and a more complex algorithm (EU) is chosen when alternatives are not similar.

This similarity approach can be interpreted as taking into account the mental transaction costs associated with making risky choices, as well as the cost of making the wrong choice. In this way the similarity approach has much in common with bounded rationality models for risky choice such as Payne, Bettmann, and Johnson (1993) and Conlisk (1996). Harrison (1994) raised a related issue questioning the saliency of the risky pairs used to show EU violations. These similarity models also relate to a growing literature that models error specifications for risky choice in Ballinger and Wilcox (1997), Hey and Orme (1994), Hey (1995), Loomes and Sudgen (1998), Moffatt (2005), and Loomes (2005).

A primary challenge for this framework is finding appropriate similarity measures that explain observed behavior. This paper develops a methodology to resolve this problem empirically and applies it to a large data set of choices over risky pairs. We test three different similarity measures, including the Kullback-Leibler cross-entropy measure we introduce here. The data come from experimental choice sessions with more than 300 undergraduate students facing choices over gambles offering risk-return tradeoffs. Most subjects had an opportunity to play one of their selected lotteries for real payoffs. The risky pairs' similarities vary for each subject's set of decisions, resulting in several thousand observations that vary in choice pair similarity.

After first assessing previous statistical methods used to assess EU violations and the decision limitation explanations for them, we test three candidate similarity models for risky choice patterns. We also test model fit of EU, a selected GEU model, and three models of EU with heteroscedastic error. Because most of the models evaluated are nonnested, an information criteria measure is utilized to assess their empirical fit.

I. The Pioneers: Simple Means Tests across Two Risky Pairs

Von Neumann and Morgenstern's (1953) EU model for risky choice provides the EU for a probability vector $\underline{p} = (p_1, p_2, \ldots p_n)$ over a fixed vector of outcomes $\underline{x} = (x_1, x_2,$

...x_n). The probability of outcome x_i is given by p_i, and outcomes can take either positive or negative values. The probabilities satisfy $p_i \geq 0$ for all i, and $\sum_i p_i = 1$. The expected utility of gamble $\underline{p}$ is:

$$EU(\underline{p}) = \sum_{i=1}^{n} p_i \bullet u(x_i) . \tag{1}$$

EU remains a powerful and widely used tool for the analysis of choice under risk.

Shortly after von Neumann and Morgenstern's paper, Allais (1953) raised the issue of systematic departures from EU. Although Allais did not carry out rigorous tests, a substantial body of subsequent evidence showed statistically significant violations of EU related to its independence axiom. These violations, in addition to other behavior inconsistent with EU, came to the fore in economics by the publication of Kahneman and Tversky's (1979) seminal paper. The nature of these EU violations, and especially what to do about them, remains the topic of considerable interest within economics, psychology, management science, and other fields.

Kahneman and Tversky's (1979) certainty effect pairs, part of a more general family referred to as common ratio effect pairs, illustrate the nature of these violations:

Pair 1: Choose between gamble A and B:

A:	gives \$3000 with probability 1.0	B:	gives \$4000 with probability .8 gives \$ 0 with probability .2

Pair 2: Choose between gamble C and D:

C:	gives \$3000 with probability .25 gives \$ 0 with probability .75	D:	gives \$4000 with probability .2 gives \$ 0 with probability .8.

Most experimental subjects select lotteries A and D. This choice pattern violates EU, as shown clearly by rewriting the gambles comprising the second pair as linear combinations of gambles A and B plus a gamble (\$0) that denotes a gamble giving a zero payoff with certainty. This alternative presentation relies on the EU independence axiom. Under EU, if A is preferred to B, then C must be preferred to D because:

$C = 1/4 * A + 3/4 * (\$0)$ and $D = 1/4 * B + 3/4 * (\$0)$.

The empirical analysis of these original EU violations consisted of means tests of choice proportions between pair AB and CD. These tests are simple, and in a sense powerful. These tests, however, offer a limited view of the nature of these violations, and in particular how robust they are to gambles that differ in the probabilities of the alternative outcomes.

II. The Extensive Tests: Multiple Choices Per Person and Panel Data

In response to the simple EU violations above, numerous GEU models have been developed. These GEU models essentially introduced a nonlinear treatment, through a function $\pi(\cdot)$, of the probabilities for the valuation of risky alternatives, defining:

$$GEU(\underline{p}) = \sum_{i=1}^{n} \pi_i(\underline{p}) u(x_i). \tag{2}$$

It is important to note that the function $\pi(\cdot)$ is defined over the entire probability vector $\underline{p}$. The specific nature of $\pi(\cdot)$ distinguishes the competing GEU models. Virtually all of these models retain some of the EU axioms (chiefly monotonicity and transitivity), while weakening the independence axiom in order to allow for common patterns of EU violations. Particularly good summaries of these models can be found in Fishburn (1988), and in Harless and Camerer (1994).

In order to test between the many competing GEU models, Harless and Camerer (1994), Hey and Orme (1994), Hey (1995), Wilcox (1993), Ballinger and Wilcox (1997), and others evaluated choice over extensive set of risky pairs. Some of these empirical estimations apply maximum likelihood methods and information criteria to the analysis of the competing GEU and related models. Ballinger and Wilcox and also Hey additionally and gainfully explored the role of heteroscedastic error structures for explaining risky choice.

III. A Decision Cost Explanation: Similarity Models

In contrast to the "reduced form" treatment of choice patterns through the weighting function $\pi(\cdot)$ under the GEU models, an alternative approach establishes instruments for decision costs and benefits based on the importance and/or the difficulty of risky choice selection. To motivate this approach, consider again the Kahneman and Tversky pairs AB and CD above. The similarity approach (Rubinstein, 1988, 2003; Leland, 1994; Buschena and Zilberman, 1995, 1999ab; Coretto, 2002; Loomes, 2006) posits that the difference in the relative importance in choice between dissimilar pair AB and similar pair CD explains patterns of choice violating EU.

In the similarity approach, respondents are held to place less importance on the choice between pair CD relative to the choice between pair AB. Both pairs offer a choice between a higher risk/higher return alternative (B and D) and a lower risk/lower return alternative (A and C). Respondents use EU to select between pair AB but use expected value maximization to select between the less important pair CD. This choice algorithm selection results in a higher likelihood of selection of the more risky alternative D over C than for the more risky alternative B over A.

These similarity models suggest a structural approach whereby choice models are augmented with a measure that reflects the importance and difficulty of choice. Specifically, assuming that computation extracts mental cost, the decision maker selects from two algorithms. One algorithm is more effort intensive (EU), and the other less effort intensive (in this application expected value maximization). The similarity approach can be interpreted as the outcome of an optimization process where a decision maker maximizes expected utility of choice less mental cost of calculation. When outcomes are similar, and the loss for making the wrong choice is small, the decision maker may prefer the less-effort intensive approach. On the other hand, when outcomes are dissimilar, and the loss for making the wrong choice is large, the decision maker may prefer the more effort intensive and more accurate approach.

Ballinger and Wilcox (1997) found some support for similarity effects in their exploration of heteroscedastic error structures and risky choice. Buschena and Zilberman (1995, 1999a, b) tested a similarity model against GEU models, and found support for similarity effects through a structural approach.

One difficulty with the similarity approach is in defining an approximate measure because "similarity" is subjective. Rubinstein (1988) considers fairly simple gambles that offer only one nonzero outcome. In his model, the more risky gamble offers a p chance at outcome x, while the safer gamble offers a q chance of y ($p < q$ and $x > y$). Rubinstein provides an axiomatized model for choice under EU preference with similarity effects, and considers two measures for similarity on both outcomes and probabilities (defined here for probabilities).

Rubinstein's first measure is the absolute difference between the probabilities, $|p - q|$. His relative difference measure is the ratio of the probabilities, p/q. Rubinstein also includes a qualitative similarity criterion; probability q is dissimilar to p if $q = 1$ and $p <$

1; that is, degenerate probabilities that offer an outcome with certainty are viewed as dissimilar to all other probabilities.

Because most risky pairs of interest offer more than one nonzero outcome, more general similarity measures have been proposed. Three similarity models are considered in this paper. All of the models used nonlinear measures over the probability vectors to define similarity, and all of these models differ in how they handle qualitative similarity (lotteries offering certainty of a positive payoff).

Candidate Model 1: Distance-based similarity. This similarity model is relatively simple, using the Euclidian distance between the probability vectors $\underline{p}$ and $\underline{q}$ to describe differences between gambles (Buschena and Zilberman, 1995, 1996, 2000). This distance is:

$$D(\underline{p},\underline{q}) = \left[\sum_{i=1}^{n} (p_i - q_i)^2 \right]^{1/2} \tag{3}$$

Moffatt (2005) has recently used this distance measure to assess decision time for choice over the risky gambles from an experiment from Hey and Orme (1994). This distance measure is reasonably applied for the set of gambles considered here that offer nontrivial risk-return tradeoffs. A more generalized difference measure can also be considered as the absolute differences in the distributions' Cumulative Distribution Functions (Buschena and Zilberman, 2000).

In our empirical estimation, both the distance measure and its square will be considered. In addition to the distance measure, some of the models estimated also include a binary measure, quasi-certainty, that generalizes the qualitative component in Rubinstein's similarity definition. Quasi-certainty takes the value 1 if the less risky of the two gambles in the pair gives greater than a zero payoff with certainty.

Candidate Model 2: (Dis)similarity as defined through cross entropy. Entropy provides a useful alternative measure for describing the differences between probability distributions. Specifically, the cross-entropy measure described below serves to define the similarity between risky alternatives.

Shannon's (1948) entropy stems from information theory and serves as a measure of how a particular distribution differs from a uniform distribution (Golan, Judge, and Miller, 1996; Preckel, 2001). The entropy measure for a discrete distribution is:

$$Entropy(\underline{p}) = -\sum_{i=1}^{n} p_i \ln(p_i) \,. \tag{4}$$

The smaller a distribution's entropy, the more information it contains. A uniform distribution has maximum entropy and is least informative, while a degenerate distribution with p_i taking only values 0 or 1 has minimum entropy and is most informative. As pointed out by a reviewer, there is a brief but far-reaching history of discussion of entropy in economics. Marschak (1959) introduced entropy as a measure of information, and Arrow (1971) extended this discussion to relate entropy as the supply price of information. More recently for risk applications, Coretto (2002) develops a risky choice model combining EU with Shannon's entropy as an explanation of EU violations.

More useful than Shannon's entropy for evaluating risky pair similarity is the Kullback-Leibler (1951) cross entropy that measures how two distributions (our $\underline{p}$ and $\underline{q}$) differ from one another:

$$Cross\ Entropy(\underline{p},\underline{q}) = \sum_{i=1}^{n} p_i \ln(p_i / q_i) \,. \tag{5}$$

Two identical (extremely similar) distributions would have a cross entropy value of 0. As the distributions differ more from one another (become increasingly dissimilar), the cross-entropy measure increases. Note that the cross entropy functional is not symmetric, with *Cross Entropy* ($\underline{p}$,$\underline{q}$) ≠ *Cross Entropy* ($\underline{q}$,$\underline{p}$). Empirically, we will consistently define cross entropy with $\underline{p}$ defining the less risky and $\underline{q}$ the more risky (higher variance) gambles. Although there is a symmetric form for cross entropy, our use of the asymmetric form is not expected to affect our empirical results in a significant way.[1] In order to provide additional background, there are also other candidate measures for pair similarity using entropy (Read and Cressie, 1988).

Under the entropy-based candidate similarity model, the cross entropy of distributions is used to define the similarity of the pairs. For our empirical application, q_i never takes the value 0. If p_i takes the value 0, the value inside the sum in (5) is set to zero for our estimations.

In our empirical application, we consider subsets of models including the risky pairs' cross entropy, its square, and the previously defined quasi-certainty measure.

Candidate Model 3: Loomes' nonlinear similarity. Loomes (2006) recently introduced a nonlinear similarity measure (ϕ). This measure allows for both the ratio of the gamble probabilities and the difference in the probabilities to affect choice. The relative effects of the probability ratios and differences can vary across individuals in Loomes' measure, with these relative effects defined through two parameters, α and β, in the following formulation. Loomes' model allows for gambles over as many as three outcomes, with these gambles defined through their probability vectors p and q:

---

[1] Our thanks to an anonymous reviewer for pointing out the symmetric form for cross entropy:
$\sum_{i=1}^{n} p_i \ln(p_i / q_i) + \sum_{i=1}^{n} q_i \ln(q_i / p_i) = \sum_{i=1}^{n} (p_i - q_i) \ln(p_i / q_i)$.

$$
\begin{aligned}
&\phi(\alpha,\beta,\underline{p},\underline{q}) = (fgh)^{\beta}[(a_I / a_J)^{(a_I+a_J)^{\alpha}}], \text{where} \\
&f = [1-(p_1 / q_1)]; \\
&g = [1-(q_2 / p_2)]; \\
&h = [1-(p_3 / q_3)]; \\
&a_I = (q_1 - p_1); \text{and} \\
&a_J = (q_3 - p_3).
\end{aligned}
\qquad (6)
$$

Loomes development of $\phi(\cdot)$ includes a discussion of its two component parts, relating to the two coefficients α and β. The coefficient α is restricted to be nonpositive and defines the degree of divergence between the perceived (subjective) and objective (actual) probability ratios. If α = 0, there is no difference between these ratios; as α declines, this perceived vs. objective difference becomes larger.

The $(fgh)^{\beta}$ component of $\phi(\cdot)$ "scales down" the bracketed portion that relates to how close the less risky alternative is to certainty. In particular, Loomes' is working to define qualitative certainty-type effects more generally than through a certainty or quasi-certainty definition. The β coefficient is restricted to be nonnegative. A value of β = 0 in (5) indicates that the ratio of probabilities does not affect choice, while β values of larger absolute value indicate a person strongly affected by these ratios, and thus more influenced by the less risky gamble's relationship to certainty. As for α, the β coefficient likely differs across individuals in Loomes' formulation.

Alternative Models: RDEU formulations and heteroscedastic error structures. As an alternative to similarity approaches, numerous models generalizing EU have been extensively tested versus one another, and versus the similarity models and against various heteroscedastic error models (Hey and Orme, 1994; Hey, 1995; Buschena and Zilberman, 2000; Moffatt, 2005). Although the focus of our paper is on assessing the various similarity formulations, we include estimation results from GEU model that has shown empirical promise in previous research. We will assess for comparison the empirical performance of the similarity models versus two formulations of Quiggin's (1982) Rank-Dependent EU (RDEU) model and also three models of EU with heteroscedastic error.

We selected the RDEU models because (1) they have performed well empirically for these data (Buschena and Zilberman 1999a, b) and another extensive data set (Hey and Orme, 1994; Hey, 1995; Buschena and Zilberman, 2000), and (2) they are specified by a relatively small number of parameters among the GEU models. This second point is important for our relatively small sample. Readers are directed to Quiggin (1982) and these empirical papers testing RDEU for the definition of this model.

One of our RDEU model specifications does not impose any particular form for the utility function, using instead separate parameters for the outcomes (a two-parameter utility model) as in Hey and Orme (1994). This utility function is defined parametrically for a risky alternative defined through its probability vector p over three outcomes as:

$$EU(\underline{p}) = \alpha * x_2 * (p_2) + \beta * x_3 * (p_3) \tag{7}$$

In (7), no parameter is necessary for outcome $0 without any loss of generality, with the utility of a payoff of zero normalized to zero as in Hey and Orme.

The other RDEU model tested imposes the one-parameter constant absolute risk aversion (CARA) utility structure used in Moffatt (2005). Relative to the formulation in (7), this model exchanges an anticipated reduction in fit for a reduction by one in the number of parameters describing the utility function. The utility function over the outcomes is defined for each outcome x_i as:

$$U(x_i) = \frac{1 - \exp(-\alpha x_i)}{\alpha}. \tag{8}$$

In addition to the two RDEU formulations, we consider three models for EU with heteroscedastic error. One of the heteroscedastic error structures holds that the error variance decreases with the pair's Euclidian distance. Under a logistic distribution for discrete choice, this error variance is $\sigma^2_D = \exp(\alpha^* D)^2$ for D as in (3) above. A generalization of this heteroscedastic error approach was tested in Buschena and Zilberman (2000).

Another heteroscedastic error structure we test was introduced in Hey (1995) and also has empirical support in Buschena and Zilberman (2000), where the heteroscedastic variance depends on the number of outcomes supported (having positive probability) for each gamble in the risky pairs. Under the logistic distribution, this heteroscedastic error variance is $\sigma^2_N = exp(\beta^* N)^2$, where N is the average number of outcomes supported in the pair. The value of N ranges from 1.5 to 3 in our experiment.

A third heteroscedastic error structure was also tested in Hey (1995), where the variance depends on the absolute difference between the expected utilities of the gambles. We define this heteroscedastic error through $\sigma_N = exp(\gamma^* A)^2$, where A is the absolute difference in the gambles' expected utilities for utility defined as in (7).

IV. The Data: "Industrial Strength" Probability Triangle with Real Payoffs

Our experiment was designed to intensively test for similarity effects on EU violations. The risky pairs allow for clear tests of both quantitative (such as measured by Euclidian distance) and qualitative effects (such as measured by quasi-certainty) across a large set of risky pairs. The set of pairs given to each respondent was devised so that parametric estimation of the three candidate similarity models in explaining EU violations should be possible. Each respondent faced both hypothetical and "real" gambles, where for each subject one of their selected gambles from a randomly chosen pair was played for actual cash. Subjects generally completed the experiment in 20 minutes.

More than 300 undergraduate students in the experiment selected between gamble pairs offering a risk/return tradeoff. For each subject set of pairs, EU would predict that (1) either the less risky choice would always be selected for every choice pair by an individual, or (2) the more risky choice would always be selected. This design characteristic is important for our empirical analysis because it allows for a relatively small number of parameters to define the differences between our candidate models.

Only a few (27, 8.5%) of subjects had choices completely consistent with EU in that they always selected the least risky option in every pair (not one of the 313 subjects always selected the most risky option). Put another way, the majority of subjects provided a set of choices showing some violation of EU. Our empirical focus is to assess these violations using candidate models of similarity, RDEU, and heteroscedastic error. The experiment and data are additionally described in Buschena and Zilberman (1999b).

The risky pairs in the experimental design differed considerably in their quantitative and qualitative similarity, allowing an extensive test of how similarity (defined through our three candidate measures) relates to the occurrence of EU violations. Gamble $p = (p_1, p_2, p_3)$ and $q = (q_1, q_2, q_3)$ were defined for each subject over a common set of outcomes $x = (x_1, x_2, x_3)$ for all questions faced by each subject.

Each subject selected a lottery for every pair presented, with a total of 26 risky pairs drawn randomly from a larger set of 106 pairs. Respondents faced either two or three of these pairs using a compound lottery formulation that alters choice and EU violations (see Fishburn, 1988); these lotteries were omitted from the analysis for every subject.

The pattern for this set of gambles began with Kahneman and Tversky's (1979) certainty effect pairs AB and CD described at the outset of our paper. Kahneman and Tversky's four original gambles are included in the entire set of risky pairs we use, but are augmented by an additional 104 gamble pairs. In each of these pairs, one gamble ($\underline{s}$) is less risky (lower expected value and variance) than the other gamble ($\underline{r}$).

The gambles making up the pairs are illustrated in Figure 1. Lotteries on the borders of the unit triangle are listed in the table below Figure 1, with each border pair defined by the probability vectors $\mathbf{b^1}$ and $\mathbf{b^2}$. Kahneman and Tversky's pairs are *RV* and *kℓ* in this Figure. Every other lottery on the locus of points is defined using a scalar $\alpha \in (0, 1)$, as a linear combination of the border pairs: $\alpha \mathbf{b^1} + (1-\alpha)\mathbf{b}^2$. For example, gamble *B* in the figure is a combination of border pairs *A* and *C* where $\alpha = 0.5$. For interior pairs on the *DH, RV*, and *fj* loci, α takes values 0.25, 0.5, and 0.75. For interior pairs on the *IQ* and *We* loci, α takes values 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, and 0.875. A risky pair was defined for every possible combination of points on each locus—e.g., additional pairs on the *DH* locus were *DE, DF, DG, EF, EG, EH, FG, FH,* and *GH*. There were 106 gamble pairs in total to be selected from. The complete set of risky pairs varies considerably in their values for the candidate similarity measures, RDEU predictions, and measures defining the heteroscedastic error variance.

V. What Do Decision Makers View as Similar?

Our three candidate models for risky pair similarity were designed to explain choice patterns violating EU, and that explanation is our primary interest. We do not know whether subjects view these pairs as similar or not. To further explore these measures, we assess respondents' subjective similarity over these pairs. The experiment elicited subjective responses regarding the similarity of the gamble pairs, elicited using a visual scale on the computer screen over a subset of risky choice pairs (5-6 per person). This similarity elicitation has a history of use for decision analysis in psychology (see, for example, Mellers, Ordónez, and Birnbaum, 1992). Respondents moved a cursor over a computer screen to indicate their subjective similarity level on the scale below:

1____10

very dissimilar	moderately similar	very similar

The purpose of our inquiry into how subjective similarity relates to objective measures is to measure and test the effects of similarity on choice as is illustrated in Figure 2. Median similarity reported over all respondents for the subset of 1672 questions using the similarity scale is graphed against the percent of respondents selecting the more risky alternative in the pair. This relationship is in the direction hypothesized, with a correlation of .72.

Estimation. The three candidate models were used to assess the subjective responses via OLS regression, $n = 1672$. The error structure did not exhibit any apparent error assumption violations. There were no observations at the lower bound and only eight observations at the upper bound of 10. Estimations using censored models with both upper and lower bounds showed no significant improvement in fit over OLS for the two linear-in-parameters models.

The Euclidian distance model for explaining subjective similarity included a constant, distance, its square, and the quasi-certainty measure. The cross-entropy model included a constant, the pairs' cross entropy, and the quasi-certainty measure; a model including the square of the cross entropy showed no significant improvement in model fit. The Loomes' power function model included a constant and the function $\varphi(\cdot)$ defined through parameters α and β in (6).

The results of the estimation for the subjective similarity responses under the Euclidian distance model are given in Table 1. The subjects viewed a pair as increasingly less similar at a decreasing rate as the pair's distance increases. Subjects also viewed quasi-certain pairs as significantly less similar.

The results of the cross-entropy model in Table 2 show the pairs' subjective similarity decreasing as the cross entropy increases. Quasi-certain pairs are subjectively less similar. This cross-entropy model had the worst fit (judging by the Akaike information criteria) of the three models.

The results of the Loomes' model for predicting subjective similarity are reported in Table 3. The estimated alpha and beta coefficients were inside the allowable range (recall that alpha must be nonpositive and beta must be nonnegative). Model fit using the Akaike information criteria measure is virtually identical to that for the distance model in Table 1.

These subjective similarity estimations provide some support for the three objectively defined similarity measures. These measures' empirical value in predicting risk choice patterns is assessed below.

VI. Estimated Effects of Candidate Models on Choice

A. Estimation and assessment

Estimation. Variants of each of the three candidate similarity models, the two RDEU formulations, and the three models using EU with heteroscedastic error were estimated in a logit framework for discrete choice (0 = the less risky lottery selected, 1 = the more risky lottery selected). These were estimated separately for each respondent's choices, where there were generally 24 observations per respondent.[2] The degrees of freedom in these estimations ranged from 23 for EU to 19 for the models with the largest number of parameters.

The error structure for the discrete choice problem enters the individual's valuation of each pair of gambles additively as in Hey and Orme (1994):

$$y^* = V(\underline{p},\underline{q}) + \varepsilon. \tag{8}$$

In equation (8), the function $V(\cdot,\cdot)$ is defined for either EU, EU with similarity arguments, or RDEU. All of these candidate models are discussed above. The error term ε is a random draw from a logistic distribution, rather than from a normal as in Hey and Orme, to allow for the potential for our relatively small sample size to give rise to higher variance and heavier tails. The error variance is homoscedastic for some of the models we test, and heteroscedastic in others, also discussed above.

[2] The non-EU estimations were not carried out for the 24 subjects who always selected the least risky lottery, but these subjects are included in the summary statistics in Tables 5 and 6.

Estimations were carried out through a grid search using the public domain software R. This software has become a powerful statistical tool that is particularly good at data management and model diagnostics. We estimated a total of 16 different models for each subject's choices. This fairly extensive coverage of the various models was undertaken in order to give an extensive test of each models' predictive power. The 16 candidate models are listed in Table 4, with the variable set omitting the constant terms that were included for each model.

Model L2 was added to the list of estimated models after preliminary runs over model L1 gave rise to β estimates of zero for all of the first 10 subjects' considered. There were some subjects who had no risky lotteries on the border of Figure 1, so none of their observed choices had a value other than 0 for the quasi-certainty variable. For these subjects, models D3, D4, C3, and C4 were estimated omitting the quasi-certainty variable; for example, for these subjects, model D1 was estimated instead of model D3.

Our model assessment was carried out in two steps in an approach consistent with Hey and Orme, first using likelihood ratio tests for nested models and then using an information criteria measure.

Assessment Step 1. Likelihood ratio tests were used to select a model from within each family of nested models. Consider the Euclidian distance family; Model D4 was tested against D2, D1, and EU; Model D2 was subsequently tested against D1 and EU; and then D1 was tested against EU. The same procedure was used to assess D4, D1, and EU. For the nonnested three-parameter models D2 vs. D3, the model with the superior log-likelihood was selected in the event that the likelihood ratio tests did not favor D4, D1, or EU over these two models. For clarity, consider also the cross-entropy family. Likelihood ratio tests were carried out for C4 vs. C2, C1, EU, etc.

Loomes' Model L1 was assessed using the 10% likelihood ratio tests against the restricted Model L2 and EU. L2 was then assessed relative to EU. The restricted RDEU Model R2 was tested against R1 and EU, and R2 was tested against EU using the same test statistic and critical level. Each of the heteroscedastic specifications (H1, H2, and H3) were assessed vs. EU using likelihood ratio tests.

Assessment Step 2. Information criteria provide a useful method of fit comparison for the multiple nonnested models we consider. Both the Akaike (1973) and the Bayesian information criteria measure (BIC) developed in Schwartz (1978) gave virtually identical ranking results. The BIC has the added benefit of providing posterior odds ratios for the likelihood the data were drawn from a particular model (discussed below), so we report the BIC ranks here. Schwartz's BIC for model *i* with sample size *n* is:

$$BIC_i = [-LLF_i(\,.\,) + 1/2k_i * log(n)]. \tag{9}$$

This BIC, like other measures such as the Akaike information criteria, ranks model *j* based on its log-likelihood function, $LLF_j(\,.\,)$, plus a penalty for the number of parameters in the model, k_j. These ranks are from lowest (the best model) to highest BIC.

B. Results

Information from the likelihood ratio tests and information criteria rankings for each model over every individual's choices is given in Tables 5 and 6. We first report (Table 5) the number of subjects for which EU was not rejected at various significance levels in favor of the set of alternative models listed in Table 4. We then report (Table 6) the BIC ranking results and posterior odds for the selected models from each one of the seven "families" of nested models in Table 4. These families are distance (D1-D4), cross entropy (C1-C4), Loomes power (L1-L2), RDEU (R1- R2), and the three heteroscedastic error families H1, H2, and H3.

Higher Order Models vs. EU. The proportion of respondents whose choices did not support each one of the higher order models over EU are given in Table 5. The likelihood ratio tests determining these proportions are listed for the 1%, 5%, 10%, and the 25% levels. Although some of the reported results are somewhat redundant given the nested nature of these models (e.g., D1 nested inside D2, D2 nested inside D4), we report all of these proportions for completeness.

There is clearly considerable difference between the models across the columns under various significance levels. The distance family (D1-D4), Loomes (L2), and the RDEU family (R1) fared best at the 1% level. Additional support was given for every family as the test significance level was weakened to 5%, 10%, and 25% levels.

It is useful to gauge these results relative to another data set that has been extensively studied in order to evaluate our experiment and to select the significance level for the likelihood ratio tests. The proportion rejecting EU under the 10% significance level for Model R1 in our experiment with (generally) 24 observations compares with the proportion rejecting EU in favor of the same RDEU formulation in Hey and Orme's combined data set over 200 choices for 80 subjects. Hey and Orme (Table VI, page 1311) report test results for both a 1% and a 5% level, and use the 1% level for the nested tests in the remainder of their paper. We will use the 10% level of significance for our nested tests given our smaller sample.

Although it is difficult to generally assess the individual specifications within each family, the unrestricted Loomes model (L1) is clearly not supported by these data. Indeed, this model was ranked quite poorly for all subjects. The restricted Loomes model (L2) alternatively shows some promise. At least one of the heteroscedastic models, H3, also appears to have quite limited statistical support.

Model selection results. Table 6 provides significance test information and BIC ranking results for each of the seven families of nested models individually given in Tables 4 and 5. As discussed above, likelihood ratio tests were used to select between the specifications for the Euclidian distance family (D1-D4), the cross-entropy family (C1-C4), the Loomes power family (L1-L2), the RDEU family (R1-R2), and the three heteroscedastic error families (H1, H2, and H3). This selection process creates a set of nonnested models representing each of the families of models, plus EU.

The second column in Table 6 lists the number of subjects for whom EU was not rejected in favor of the selected model from a family (e.g., D1) in the likelihood ratio test. The cross-entropy family is favored under this criterion, with the heteroscedastic models generally doing poorly. Note also that there were 47 subjects for whom EU was not rejected in favor of any of the alternative models; that is, there were 266 subjects for whom EU was rejected in favor of at least one alternative model.

The third column in Table 6 provides the proportion of subjects for whom the selected model for each family was either ranked highest by the BIC criteria for nonnested models, or in the case of EU where it was not rejected in favor of any other model using likelihood ratio statistics. The cross-entropy family at 35% and the distance family at 31% of the population exhibit the strongest support using this criterion, while the three heteroscedastic error specifications have quite low support. The EU model is the best model for 15% of the respondents under this ranking criterion. Of this 15%, approximately half (24) of these subjects had choices that exhibited no EU violations; that is, they always selected the less risky lottery for every pair. An additional 23 of these respondents exhibited some EU violations, but EU was statistically the best model under the likelihood ratio test statistics.

Posterior probabilities using the BIC measures. Even though a model might not be the most favored under the likelihood ratio tests and counts of BIC #1 rankings in column 3 of Table 6, it may have merit in explaining choice under risk. A model might for instance score a large number of second-place BIC ranks and perform relatively well on average for the entire set of respondents. One way to provide this information is to proceed as in Hey (1995), who provides subject counts for all 11 ranks for the set of models he considers.

For some subjects, however, many of these seven model families did not have support vs. EU in likelihood ratio tests (column 2 in Table 6). The question then becomes how to rank these families since they all nest EU as a submodel, and they are in a sense tied with one another when more than one of the model families do not support higher order models beyond EU. Such questions arise for using either counts of rank placement (2nd, 3rd, etc.), and summary measures such as average ranks. An additional consideration is that assessing model ranks becomes increasingly difficult as the number of models increases.

We use a summary statistic that mitigates some of the ranking concerns and additionally provides useful model fit information. The Bayesian structure of the BIC statistic in equation (8) allows construction of an approximate measure of the (posterior) probability for observing the estimated BIC conditional on a uniform (uninformative) prior. We construct and estimate these estimated posterior probabilities for the families of the candidate models, where again the representative from each family was selected via the likelihood ratio tests.

Note that in the event the EU model was selected as the representative for a particular family of models, the posterior conditional probabilities for this family were calculated using the EU likelihood ratio and the models' single parameter. For the 47 subjects for whom EU was not rejected for any of the candidate model families, the posterior conditional probabilities were equal for all seven of the families with a value of .143. Because EU was selected to represent multiple families for numerous subjects, we do not report the conditional probabilities for the EU model.

The approximate posterior conditional probabilities, the BIC weights, for model *j* and respondent *i* are constructed as:

$$BIC\ Weight_{ji} = exp(\text{-}BIC_{ji}) \,/ \sum_{k} exp(-BIC_{ki})\,. \tag{10}$$

As shown in Schwartz, these BIC weights are the approximate posterior probabilities of observing the sample conditional on an uninformative prior (see also Ramsey and Schafer, 2002). We provide summary statistics for these BIC weights across the set of respondents for each of the candidate models in the last three columns in Table 6. Although omitted, minimums for these approximate posterior probabilities were uniformly zero for each model, reflecting that no one model family fits the data from every subject.

The mean posterior probabilities illustrate the relative substitutability of the candidate models, with the results favoring the Euclidian distance based models (D1-D4), and the models using cross entropy (C1-C4).

VII. Conclusion

Experimental studies suggest that the traditional frameworks for analyzing decision making under uncertainty have to be modified. The similarity approach can explain some of the behavioral paradoxes detected by experimental studies, and it also suggests that different algorithms are used for different choices under uncertainty. In particular, this similarity approach suggests that EU is used when choices are dissimilar, and as a result the stakes are higher, and a simpler rule is used when the alternatives are more similar and the stakes are correspondingly lower. The task, then, is to identify which measure of similarity is triggering the algorithm choice.

We introduce a new measure for risky pair similarity using the Kullback-Liebler cross entropy. We test this cross-entropy measure against the previously proposed Euclidian distance measure and against a nonlinear similarity specification proposed by Loomes. The empirical fit of each of these three nonnested candidate similarity measures was evaluated relative to the empirical fit of a selected GEU model and three models of EU with heteroscedastic error.

The empirical analysis suggests that the Euclidean distance measure and the cross- entropy measures are most consistent with the observed data. The distance measure is quite simple and has been empirically supported in previous work. The heretofore untested cross-entropy models use a natural measure for differences in the information content of the distributions that define the risky pairs.

The methodology presented here allows for selection between similarity measures using (1) robust grid search maximum likelihood estimation routines over the full set of choices made by a respondent, and (2) the Bayesian information criteria measures. These methods provide a general evaluation of numerous competing models such as similarity approaches, generalized EU, and heteroscedastic error formulations. In this regard, we extend work by Hey and Orme (1994), and by Hey (1995), to which risk analysts are substantially indebted.

References

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle, in: B. N. Petrov and F. Csaki (Eds.), Second international symposium on information theory. Akademiai Kiado, Budapest, pp. 267-281.

Allais, M., 1953, Le comportement de l'homme rationnel devant le risque: Critique de postulats et axiomes de l'ecole américaine. Econometrica 21, 503-546.

Arrow, K. J., 1971, The value of and demand for information, in: C.B. McGuire and J. Marshack (Eds.), Decision and Organization: A Volume in Honor of Jacob Marshack. North-Holland, Amsterdam, pp. 131-140.

Ballinger, T. P. and N. T. Wilcox, 1997, Decisions, error, and heterogeneity. The Economic Journal 107, 1090-1105.

Buschena, D. E. and D. Zilberman, 1995, Performance of the similarity hypothesis relative to existing models of risky choice. Journal of Risk and Uncertainty 11, 233-262.

Buschena, D. E. and D. Zilberman, 1999a, Testing the effects of similarity on risky choice. In M. Machina and B. Munier. Beliefs, Interactions and Preferences in Decision Making, Dordrecht, Kluwer.

Buschena, D. E. and D. Zilberman, 1999b, Testing the effects of similarity on risky choice: Implications for violations of expected utility. Theory and Decision 46, 253-280.

Buschena, D. E. and D. Zilberman, 2000, Generalized expected utility, heteroscedastic error, and path dependence in risky choice. Journal of Risk and Uncertainty 20, 67-88.

Conlisk, J., 1996, Why bounded rationality. Journal of Economic Literature 34, 669-700.

Coretto, P., 2002, A theory of decidability: Entropy and choice under uncertainty. Rivista di Politica Economica 92, 29-62.

Fishburn, P. C., 1988, Nonlinear preference and utility theory, Johns Hopkins University Press, Baltimore, MD.

Golan, A., G. Judge, and D. Miller, 1996, Maximum Entropy Econometrics: Robust Estimation with Limited Data, John Wiley and Sons, New York.

Harless, D. and C. F. Camerer, 1994, The predictive utility of generalized expected utility theory. Econometrica 62, 1251-1290.

Harrison, G., 1994, Expected utility theory and the experimentalists. Empirical Economics 19, 223-253.

Hey, J. D., 1995, Experimental investigations of errors in decision making under risk. European Economic Review 39, 633-640.

Hey, J. D. and C. Orme, 1994, Investigating generalizations of expected utility theory using experimental data. Econometrica 62, 1291-1326.

Kahneman, D. and A. Tversky, 1979, Prospect theory: An analysis of decision under risk. Econometrica 47, 263-291.

Kullback, S. and R. A. Leibler, 1951, On information and sufficiency. Annals of Mathematical Statistics 22, 79-86.

Leland, J. W., 1994, Generalized similarity judgments: An alternative explanation for choice anomalies. Journal of Risk and Uncertainty 9, 151-172.

Loomes, G. 2005, Modeling the stochastic component of behaviour in experiments: some issues for the interpretation of data. Experimental Economics 8, 301-323

Loomes, G., 2006, The improbability of a general, rational and descriptively adequate theory of decision under risk. Working Paper. School of Economics, University of East Anglia.

Loomes, G. and R. Sudgen, 1998, Testing different stochastic specifications of risky choice. Economica 65, 581-598.

Marschak, J. 1959, Remarks on the Economics of Information. Contributions to Scientific Research in Management, Los Angeles: Western Data Processing Center, University of California, 79-98.

Mellers, B. A., L. Ordónez, and M. H. Birnbaum, 1992, A change-of-process theory for contextual effects and preference reversals in risky decision making. Organizational Behavior and Human Decision Processes 52, 319-330.

Moffatt, P. G. 2005, Stochastic choice and the allocation of cognitive effort. Experimental Economics 8, 369-388.

Payne, J. W., J. R. Bettmann, and E. J. Johnson, 1993, The Adaptive Decision Maker. Cambridge University Press, Cambridge, United Kingdom.

Preckel, P., 2001, Least squares and entropy: A penalty function perspective. American Journal of Agricultural Economics 83, 366-377.

Quiggin, J. 1982, A theory of anticipated utility. Journal of Economic Behavior and Organization 3, 323-343.

Ramsey, F. L. and D. W. Schafer, 2002, The Statistical Sleuth: A Course in Methods of Data Analysis, 2nd ed., Duxbury, Thompson Learning, Belmont, CA.

Read, T. R. C. and N. A. C Cressie, 1988, Goodness-of-Fit Statistics for Discrete Multivariate Data, Springer-Verlag, New York.

Rubinstein, A., 1988, Similarity and decision making under risk: Is there a utility theory resolution to the Allais Paradox? Journal of Economic Theory 46, 145-153.

Schwartz, G., 1978, Estimating the dimension of a model. The Annals of Statistics 6, 461-464.

Shannon, C. E., 1948, A mathematical theory of communication. Bell System Technical Journal 27, 379-423, 623-59.

Starmer, C., 2000, Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. Journal of Economic Literature 38, 332-382.

von Neumann, J. and O. Morgenstern, 1953, Theory of Games and Economic Behavior, 3rd ed., Princeton University Press, Princeton, NJ.

Wilcox, N. T., 1993, Lottery choice: Incentives, complexity and decision time. Economic Journal 103, 1397-1417.

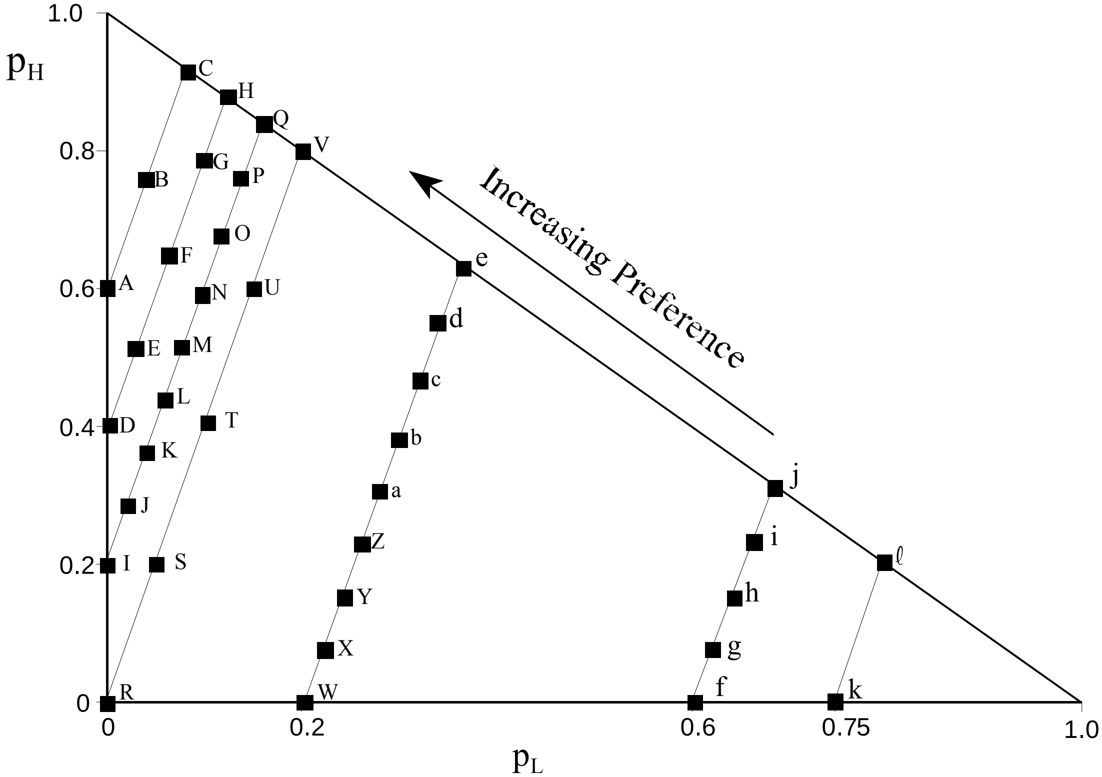

Fig. 1. Probability triangle for experimental risky choice pairs

Probabilities of the border pairs

Pair	p_L	p_M	p_H	q_L	q_M	q_H
AC	.00	.40	.60	.08	.00	.92
DH	.00	.60	.40	.12	.00	.88
IQ	.00	.80	.20	.16	.00	.84
RV	.00	1.00	.00	.20	.00	.80
We	.20	.80	.00	.36	.00	.64
fj	.60	.40	.00	.68	.00	.32
Kl	.75	.25	.00	.80	.00	.20

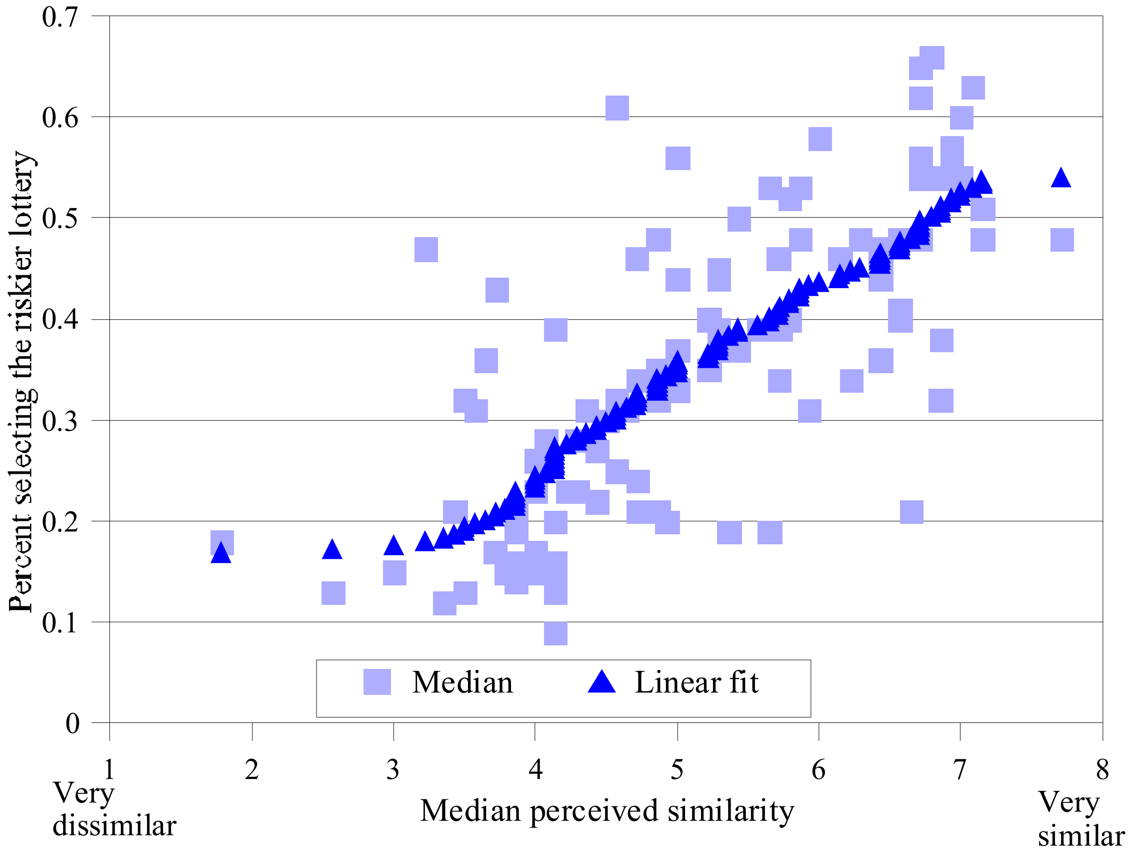

Fig. 2. Median choice and perceived similarity

Table 1. Subjective similarity and Euclidian distance

Variable	Coefficient est.	Standard error
Constant	6.42***	.065
Distance	-5.84***	.361
Distance SQ	3.06***	.291
Quasi certainty	-.674***	.151

LLF -3285.2, AIC = 3.93
N = 1672
R^2 = .201
R^2 adjusted = .200
*** indicates significance at the 1% level

Table 2. Subjective similarity and cross entropy

Variable	Coefficient est.	Standard error
Constant	4.73***	.050
Cross entropy	-.47E-02**	.26E-02
Cross entropy squared	-.13E-04	.16E-04
Quasi certainty	-1.01***	.151

LLF -3414.1, AIC = 4.09
N = 1672
R^2 = .068
R^2 adjusted = .067
*** indicates significance at the 1% level

Table 3. Subjective similarity and Loomes' power function

Variable	Coefficient est.	Standard error
Constant	3.61***	0.202
Alpha	-0.23***	0.011
Beta	2.20***	1.05

LLF -3286.0, AIC = 3.93

*** indicates significance at the 1% level

Table 4. List of candidate models

Model	Measure/Model	Variables (in addition to a constant)
EU	Expected utility	None
D1	Euclidian distance	Distance
D2	Euclidian distance	Distance, distance squared
D3	Euclidian distance	Distance, quasi-certainty
D4	Euclidian distance	Distance, distance squared, quasi certainty
C1	Cross entropy	Cross entropy
C2	Cross entropy	Cross entropy, cross entropy squared
C3	Cross entropy	Cross entropy, quasi certainty
C4	Cross entropy	Cross entropy, cross entropy squared, quasi certainty
L1	Loomes' power	Loomes power function, unrestricted model
L2	Loomes' power	Loomes power function, β restricted to zero
R1	RDEU Quiggin	Quiggin's RDEU model, general utility
R2	RDEU Quiggin	Quiggin's RDEU model, CARA utility
H1	Heteroscedasticity	EU with heteroscedastic error: Euclidian Distance
H2	Heteroscedasticity	EU with heteroscedastic error: average number of outcomes
H3	Heteroscedasticity	EU with heteroscedastic error: absolute difference in EU

Table 5. Percentage of subjects (N = 313) with models significantly different from EU

	Subjects with significant likelihood ratio tests statistics vs. EU at various significance levels			
Model	1% level	5% level	10% level	25% level
D1	19%	36%	42%	54%
D2	19%	31%	40%	60%
D3	19%	33%	41%	56%
D4	19%	31%	42%	60%
C1	10%	18%	28%	28%
C2	11%	27%	44%	66%
C3	9%	24%	35%	53%
C4	14%	33%	44%	64%
L1	0%	0%	0%	0%
L2	17%	31%	37%	51%
R1	17%	30%	36%	44%
R2	7%	20%	30%	50%
H1	1%	10%	24%	56%
H2	2%	12%	22%	56%
H3	0%	3%	5%	12%

Table 6: Results for nested significance tests and Bayesian information criteria, selecting the best model for each subject. N = 313.

Model	Number for whom EU not rejected at 10%	Favored model using nested tests or BIC	BIC posterior odds			
			Mean	Median	Max.	Min.
0. Expected utility	(47)†	15%				
D1-D4. Euclidian distance family*	150	31%	28%	16%	100%	0%
C1-C4. Cross-entropy family**	129	35%	25%	14%	100%	0%
L2. Loomes' power formulation, restricted model	195	7%	14%	10%	100%	0%
R1. Quiggin's RDEU, no assumption on utility	219	8%	13%	9%	100%	0%
H1. EU with hetero. error, distance	240	0%	7%	7%	26%	0%
H2. EU with hetero. error, number of outcomes	263	3%	8%	6%	92%	0%
H3. EU with hetero. error, absolute EU difference	298	0%	6%	5%	40%	0%

† There were 47 subjects whose choice patterns were such that no higher level model was significantly better than EU at the 10% level.