Journal of Risk and Uncertainty, 3:115-134 (1990)

An Experimental Evaluation of the Descriptive Validity of Lottery-Dependent Utility Theory

RICHARD L. DANIELS
Fuqua School of Business, Duke University, Durham, NC 27706

L. ROBIN KELLER
Graduate School of Management, University of California, Irvine

Key words: expected utility, lottery-dependent utility, preference assessment

Abstract

This article compares the performance of the expected utility (EU) and lottery-dependent expected utility (LDEU) models in predicting the actual choices of experimental subjects among risky options. In the process, we present two approaches for calibrating the LDEU model for an individual decision maker. The results indicate that while LDEU exhibits a higher potential for correctly predicting choice, the version of the model calibrated by indifference judgments does not outperform EU. We suggest a functional form for the parametric functions that defines the LDEU model, and discuss ways in which this function can be incorporated into choice-based assessment approaches to improve predictions.

Experiments involving choice under risk have consistently demonstrated that subjects exhibit patterns of preference that violate principles of von Neumann-Morgenstern (1947) expected utility (EU) theory. To accommodate these inconsistencies, a number of more general models of risky decision making have been proposed. Weber and Camerer (1987) reviewed several of these alternative models, including prospect theory (Kahneman and Tversky, 1979), weighted utility theory (Chew and MacCrimmon, 1979a, 1979b), and generalized utility theory (Machina, 1982).

Generalized utility models serve several potential purposes, and the need for model specificity is determined in part by how a model is to be used. If the objective is to characterize the general behavior of decision makers who conform with the assumptions of a model (to be used, e.g., in theoretical models of consumer behavior or economic agents), the model need only be specific enough to provide the behavioral patterns of interest. On the other hand, if the objective is to represent the preferences of an individual decision maker (to be used, e.g., to predict an individual's choice among several options), then the model must be precise enough for a preference function to be assessed.

The authors would like to thank Kevin Hayes and Dorlisa King of Duke University and Stuart Eriksen and Albert Lee of the University of California, Irvine for valuable research assistance, and Amihai Glazer, Howard Kunreuther, Henry McMillan, Michael Moore, Rakesh Sarin, and Robert Winkler for helpful comments on a previous draft. This research was supported in part by the Business Associates Fund at the Fuqua School of Business, Duke University.

Becker and Sarin (1987) recently introduced a generalized utility model that allows the utility of an outcome to vary with the lottery in which it is contained. Unlike some other models, this lottery-dependent expected utility (LDEU) model is specific enough to be precisely assessed, and thus can be used to both characterize and predict preferences. This article reports the results of an experiment designed to test the ability of the LDEU model to predict the actual choices of subjects among risky options.

Models of decision making under risk can be experimentally investigated in several ways. First, alternative models can be compared and evaluated by examining the properties of preference implied by the structure of the models; an example is provided by Camerer (1989), who examined several generalized utility models by comparing implied properties with the actual choice patterns exhibited by experimental subjects. Similarly, tests of the descriptive validity of the EU model have focused primarily on its structural inability to model specific patterns of preference. Predictive accuracy is a second criterion by which models may be evaluated, appropriate for the objective of representing an individual's preferences. Currim and Sarin (1989) provided an example of this approach, comparing the predictive accuracies of the prospect and EU models by assessing both for each experimental subject and comparing actual choices with those predicted by the models. In this article we adopt the latter approach, evaluating the predictive accuracy of the LDEU model using EU as a standard for comparison.

The article is organized as follows. The expected utility and lottery-dependent expected utility models are described in section 1. The exponential form proposed by Becker and Sarin (1987) as an operation version of the LDEU model is the focus of this study. The potential of EU and LDEU to predict actual choices is examined in section 2. The predictive capacity of each model was determined by conducting an exhaustive search among possible utility values for each outcome to identify the model that best conformed with the actual choices of each subject. Since this search becomes impractical for choice situations involving many outcomes, subjects' EU and LDEU models were assessed using indifference information. Section 3 presents approaches for calibrating the LDEU model from probability equivalence and certainty equivalence judgments. Predictions from the assessed models were then compared with the actual choices of subjects over two sets of choice scenarios, one involving dependent lotteries designed to illustrate several of the choice paradoxes reported in the literature, and another involving nonparadoxical, unrelated lotteries. The experimental results for the assessed models are reported in section 4. While the assessed LDEU model often predicted patterns of preference inconsistent with expected utility, neither model approached its predictive capacity as determined from the search process. Focusing on those subjects for whom LDEU most accurately predicted choices, we examined the parametric function $h(x)$ that defines an LDEU model and found that this function is well approximated by a cubic form. Section 5 presents this function and discusses alternative assessment approaches in which this form is exploited. The predictive capacity of the LDEU model with $h(x)$ constrained to this cubic form was found to be nearly as high as the unconstrained model. Section 6 concludes with a summary and suggestions for further research.

1. The models

Consider risky option F consisting of n discrete outcomes. Let x_i denote the i^{th} outcome of F, occurring with probability p_i. The expected utility of F can then be expressed as

$$U(F) = E_F[u(x)] = \sum_{i=1}^{n} p_i u(x_i),$$

where E_F denotes expectation with respect to F, and u represents a real-valued utility function defined over outcomes x_i. EU ranks risky options by their respective expected utilities, with more-preferred options having higher expected utilities.

The LDEU model is more general that expected utility, reflecting the dependence of the utility of an outcome on the lottery in which the outcome occurs. Let $u_F(x_i) = u(x_i, c_F)$ denote the utility of outcome x_i in lottery F, where c_F is a constant that depends on F. Then the lottery-dependent expected utility of F can be expressed as

$$U(F) = E_F[u_F(x)] = \sum_{i=1}^{n} p_i u(x_i, c_F).$$

Assumptions about the parameter c_F and the form of the utility function are needed to make the model operational. The parameter c_F is assumed to be linear in probabilities, so there exists a real-valued function $h(x)$, specific to a decision maker, such that

$$c_F = E_F[h(x)] = \sum_{i=1}^{n} p_i h(x_i).$$

A special case of $u(x, c_F)$ adopted throughout this article is the exponential model suggested by Becker and Sarin:

$$u(x, c_F) = \frac{1 - e^{-c_F\left(\frac{x - x_0}{x^* - x_0}\right)}}{1 - e^{-c_F}} \qquad \text{if } c_F \neq 0,$$

$$u(x, c_F) = \frac{x - x_0}{x^* - x_0} \qquad \text{if } c_F = 0,$$

where x^* and x_0 represent the best and worst attainable outcomes, respectively, in the set of available lotteries and $u(x^*, c_F) = 1$ and $u(x_0, c_F) = 0$ for any c_F. For the exponential model with c_F linear in probabilities, if $c_F = E_F[h(x)] > 0$, then the lottery-dependent utility function associated with lottery F is concave, reflecting risk aversion, while $c_F < 0$ implies a convex LDEU function and risk proneness. Note that if c_F is constant for all lotteries F, $h(x)$ must also be constant, and the corresponding decision maker will evaluate available lotteries using the same exponential utility function. In this case, the exponential forms of the EU and LDEU models are equivalent.

2. Predictive capacity

Experiments designed to characterize the properties of a given utility model generally consist of sets of risky options, over which the pattern of choices predicted by the model can be compared with the pattern of actual choices made by subjects. A model may be structurally incapable of representing some patterns of preference that are experimentally observed, (e.g., the EU model is incapable of predicting the modal response pattern in the Allais (1953) paradox) while performing quite well on other, non-paradoxical patterns. Conversely, a model capable of representing each of a given set of preference patterns may fail to match all choices when a subject exhibits two or more of these patterns simultaneously. Therefore, predictive capacity is an important measure of the ability to model choice.

2.1. Experimental design

Graduate management students at the University of California, Irvine participated voluntarily in an experiment to evaluate the predictive capacities of EU and LDEU. Subjects were given a set of 18 hypothetical choice scenarios consisting of pairs of risky investment options and asked to indicate the more-preferred alternative in each case. Two sets of 18 scenarios were used in a between-subject design to compare the models' performance. Experimental questionnaires included consistency checks to ensure that subjects were providing reliable preference information. One group of subjects was assigned independently generated scenarios, while the other group received a set of dependently constructed scenarios designed to elicit preference patterns inconsistent with expected utility. The independent scenarios were expected to be neutral to the models, while the dependent scenarios were expected to favor LDEU, since LDEU is designed to capture choice patterns which EU is structurally incapable of modeling.

Previous experiments have shown differences in risk-taking behavior in different outcome domains. For this reason, scenarios were equally partitioned according to the characteristics of the outcomes available in the lotteries. Gain-domain scenarios involved lotteries with only nonnegative outcomes, loss-domain scenarios had no positive outcomes available, and mixed-domain scenarios spanned both positive and negative outcomes. The 18 scenarios were presented to individual subjects in one of three random orders. The largest positive outcome was set at $x^* = \$4000$ and the largest possible loss at $x_0 = -\$4000$, with the expected values in each pair of options set equal, except for small differences due to rounding of probabilities. The range of monetary outcomes was selected to represent a significant portion of a graduate student's annual budget. Tables 1 and 2 present detailed descriptions of the paradoxical and independent scenarios.

A group of 44 subjects was randomly selected to provide choices over the set of independent scenarios. Independent scenarios consisted of pairs of lotteries involving two of the seven possible outcomes each, with probabilities ranging from .20 to .80 (to avoid possible overweighting of likely or unlikely outcomes). A second group of 54 subjects responded to the set of dependent, or paradoxical, scenarios. While the independent

Table 1. Paradoxical scenarios

Scenario (i)	Scenario classification	Option A_i	Option B_i
1	$S1$	(3000, 1.0) [91%]	(4000, .75; 0, .25) [9%]
2	(4000, .80; $S1$, .2)	(4000, .8; 3000, .2) [78%]	(4000, .95; 0, .05) [22%]
3	(3000, .8; $S1$, .2)	(300, 1.0) [39%]	(4000, .15; 3000, .8; 0, .05) [61%]
4	(0, .8; $S1$, .2)	(3000, .2; 0, .8) [37%]	(4000, .15; 0, .85) [63%]
5	(0, .96; $S1$, .04)	(3000, .04; 0, .96) [13%]	(4000, .03; 0, .97) [87%]
6	B_1 vs. B_3	(4000, .15; 3000, .8; 0, 0.5) [13%]	(4000, .75; 0, .25) [87%]
7	$S7$	(− 3000, 1.0) [13%]	(0, .25; − 4000, .75) [87%]
8	(0, .8; $S7$, .2)	(0, .8; − 3000, .2) [72%]	(0, .85; − 4000, .15) [28%]
9	(− 3000, .8; $S7$, .2)	(− 3000, 1.0) [33%]	(0, .05; − 3000, .8; − 4000, .15) [67%]
10	(4000, .8; $S7$, .2)	(− 3000, .2; − 4000, .8) [39%]	(0, .05; − 4000, .95) [61%]
11	(− 4000, .96; $S7$, .04)	(− 3000, .04; − 4000, .96) [43%]	(0, .01; − 4000, .99) [57%]
12	B_7 vs. B_9	(0, .05; − 3000, .8; − 4000, .15) [72%]	(0, .25; − 4000, .75) [28%]
13	$S13$	(2000, 1.0) [85%]	(4000, .75; − 4000, .25) [15%]
14	(4000, .8; $S13$, .2)	(4000, .8; 2000, .2) [83%]	(4000, .95; − 4000, .05) [17%]
15	(2000, .8; $S13$, .2)	(2000, 1.0) [69%]	(4000, .15; 2000, .8; − 4000, .05) [31%]
16	(− 4000, .8; $S13$, .2)	(2000, .2; − 4000, .8) [26%]	(4000, .15; − 4000, .85) [74%]
17	(− 4000, .96; $S13$, .04)	(2000, .04; − 4000, .96) [15%]	(4000, .03; − 4000, .97) [85%]
18	B_{13} vs. B_{15}	(4000, .15; 2000, .8; − 4000, .05) [11%]	(4000, .75; − 4000, .25) [89%]

Fifty-four subjects chose either option A_i or option B_i for each scenario i. The percentage of subjects who chose each option is given in brackets.

Table 2. Independent scenarios

Scenario (i)	Option A_i	Option B_i
1	(4000, .5; 3000, .5) [30%]	(4000, .75; 2000, .25) [70%]
2	(4000, .2; 2000, .8) [68%]	(3000, .8; 0, .2) [32%]
3	(3000, .4; 2000, .6) [82%]	(4000, .6; 0, .4) [18%]
4	(3000, .4; 0, .6) [48%]	(4000, .3; 0, .7) [52%]
5	(2000, .4; 0, .6) [80%]	(4000, .2; 0, .8) [20%]
6	(3000, .25; 2000, .75) [82%]	(3000, .75; 0, .25) [18%]
7	(− 4000, .5; − 3000, .5) [57%]	(− 4000, .75; − 2000, .25) [43%]
8	(− 4000, .2; − 2000, .8) [39%]	(− 3000, .8; 0, .2) [61%]
9	(− 3000, .4; − 2000, .6) [20%]	(− 4000, .6; 0, .4) [80%]
10	(− 3000, .4; 0, .6) [39%]	(− 4000, .3; 0, .7) [61%]
11	(− 2000, .4; 0, .6) [52%]	(− 4000, .2; 0, .8) [48%]
12	(− 3000, .25; − 2000, .75) [30%]	(− 3000, .75; 0, .25) [70%]
13	(4000, .77; − 3000, .23) [66%]	(4000, .8; − 4000, .2) [34%]
14	(3000, .5; − 3000, .5) [30%]	(4000, .43; − 4000, .57) [70%]
15	(4000, .66; − 3000, .34) [39%]	(3000, .8; − 4000, .2) [61%]
16	(2000, .28; − 3000, .72) [20%]	(4000, .2; − 3000, .8) [80%]
17	(3000, .5; − 4000, .5) [30%]	(2000, .5; − 3000, .5) [70%]
18	(2000, .63; − 4000, .37) [61%]	(4000, .4; − 3000, .6) [39%]

Forty-four subjects chose either option A_i or option B_i for each scenario i. The percentage of subjects who chose each option is given in brackets.

scenarios were generated to provide a good cross section of nonrelated scenarios, the paradoxical scenarios were constructed (over the same range of outcomes) to elicit preference patterns that illustrate violations of the common ratio, common consequence, and betweenness principles.

2.2. *Paradoxical scenarios*

As an example of the structural dependence among pairs of paradoxical scenarios, consider the following options from scenarios 1 and 4 in table 1.

	Option A_1		Option B_1
Scenario 1:	100% chance of \$3000	*vs.*	75% chance of \$4000
			25% chance of \$0
	Option A_4		Option B_4
Scenario 4:	20% chance of \$3000	*vs.*	15% chance of \$4000
	80% chance of \$0		85% chance of \$0

Observe that options A_4 and B_4 consist of an 80% chance of receiving \$0 and a 20% chance of receiving options A_1 or B_1, respectively. Many subjects prefer option A_1 over option B_1, but prefer the transformed option B_4 over the transformed option A_4 (see, e.g., Kahneman and Tversky, 1979; Keller, 1985b). This pattern of preference violates the *substitution* (or *common ratio*) principle of expected utility, which states that if alternative A_1 is preferred to B_1, then the compound alternative $[C, p; A_1, 1 - p]$ should also be preferred to $[C, p; B_1, 1 - p]$ (see Marschak, 1950). Unlike EU, the LDEU model is able to simultaneously predict preference for options A_1 and B_4 (or B_1 and A_4) by allowing the utility of \$3000 to vary with the lottery. Thus LDEU might be expected to exhibit superior predictive performance for subjects responding to paradoxical scenarios. In the gain domain, subjects' responses to scenarios 1, 4, and 5 can be examined to check conformance with the common ratio principle. Scenarios 7, 10, and 11 in the loss domain and Scenarios 13, 16, and 17 in the mixed domain are similarly related.

Conformance with the *common consequence* principle can be examined by considering the first four scenarios in each outcome domain. The last three of these scenarios are related to the first via a common ratio transformation, combining a 20% chance of receiving the original (O) scenario with an 80% chance of receiving the high (H), intermediate (I), or low (L) outcome in the given domain. The four scenarios together form the HILO structure of Chew and Waller (1986), with H, I, and L values of (\$4000, \$3000, and \$0), (\$0, −\$3000, and −\$4000), and (\$4000, \$2000, and −\$4000) in the gain, loss, and mixed domains, respectively. Expected utility requires that a subject preferring the risk-averse option A (risk-prone option B) in the original scenario should also prefer A (B) in any scenario related by the common consequence construction; further, subjects' responses to the set of transformed scenarios should be identical. In contrast, LDEU is capable of modeling 14 of the $2^4 = 16$ possible choice patterns (see Becker, 1986).

Coombs (1969, 1975) and Coombs and Huang (1970) suggested that when the expected value of a set of lotteries is held constant, subjects will exhibit different preference orderings based on their ideal risk levels. Those subjects who prefer an intermediate level of risk are said to exhibit a folded preference ordering. Subjects' responses to scenarios (1, 3 and 6), (7, 9 and 12), and (13, 15, and 18) can be examined to investigate the prevalence of folded orderings in the gain, loss, and mixed domains, respectively. Each triplet of scenarios consists of the three possible paired comparisons between a high variance option (a lottery involving only the highest and lowest outcomes in the given outcome domain), a zero variance option (a sure payoff of the intermediate outcome), and a moderate variance option (constructed as a 20% chance of receiving the high variance option and an 80% chance of receiving the zero variance option). According to expected utility, the moderate variance option can be neither the most preferred nor the least preferred of the three options. LDEU is capable of modeling violations of this *betweenness* property, as Becker and Sarin (1989) demonstrated with an example problem.

2.3. Results

The choices made by each subject over the 18 (paradoxical or independent) scenarios were used to estimate the predictive capacity of the EU and LDEU models. The best EU model was found by setting $u(\$4000) = 1$ and $u(-\$4000) = 0$ and simultaneously varying the values of $u(x)$ for $x = \$3000, \$2000, \$0, -\$2000$, and $-\$3000$. The combination of $u(x)$ values that yielded the largest number of correct predictions for an individual subject represented the best-predicting EU model for that subject. The best-predicting LDEU model was similarly obtained by varying the values of $h(x)$ for $x = \$4000, \$3000, \$2000, \$0, -\$2000, -\$3000$, and $-\$4000$.

The EU model was found to be capable of correctly predicting subjects' choices in 71.91% of the paradoxical and 83.96% of the independently generated scenarios in which the model did not predict indifference. Inspection of the best-predicting $u(x)$ values associated with individual subjects revealed a dominant pattern of risk aversion in the gain domain and risk proneness in the loss domain. Subjects were also more risk-prone in the loss domain than risk-averse in the gain domain.

A search was conducted among possible combinations of the $h(x)$ values to identify the best LDEU model for each subject. The value of $h(x)$ was varied by increments of 1 between 0 and 5 for $x = \$2000, \$3000$, and $\$4000$, between -5 and 0 for $x = -\$4000, -\$3000$, and $-\$2000$, and between -2 and 2 for $x = \$0$. The search yielded correct predictions for 93.52% of the subjects' responses to the paradoxical scenarios and 91.92% of the independent scenarios. Examination of the best-predicting LDEU models also revealed a dominant pattern, to be discussed in detail in section 5.

The results thus indicate that LDEU exhibited a higher potential for predicting choice, as evidenced by the additional 19.87% of the paradoxical responses and 8.10% of the independent responses matched by the model. However, the search strategy required to identify the best LDEU model becomes intractable as the number of possible outcomes becomes large. The next section outlines two approaches for assessing the models using indifference information.

3. Model calibration using indifference judgments

Each subject's responses to a series of indifference judgments were used to calibrate both the EU and LDEU models. Since experimental results can depend on the method by which the models were calibrated (see, e.g., Hershey, Kunreuther, and Schoemaker, 1982), two assessment approaches were adopted to allow a more complete comparison. *Direct assessment* involved eliciting *probability equivalence* (*PE*) indifference judgments to determine $u(x)$ and $h(x)$ values for seven monetary outcomes ranging from $-\$4000$ to \$4000. *Curve-fitting assessment* involved eliciting *certainty equivalence* (*CE*) indifference judgments independently for the gain domain (with outcomes ranging from \$0 to \$4000) and the loss domain (with outcomes from $-\$4000$ to \$0). The $u(x)$ and $h(x)$ values of interest were then derived from the two-piece exponential utility function that best fit the indifference judgments. Therefore, four calibrated models were formulated for each subject: EU calibrated by the direct and curve-fitting procedures and LDEU assessed by the two methods.

The two procedures used in this study were chosen to represent opposite ends of the spectrum of possible indifference judgment assessment methods. Calibration by probability equivalents provides a straightforward means of directly assessing $u(x)$ and $h(x)$ values for each model without requiring specification of functional forms. Calibration by certainty equivalents requires a different type of indifference judgment that many subjects find easier to make. However, by allowing subjects to supply certainty equivalents, the ability to directly compute $u(x)$ and $h(x)$ values is lost, since the *CE* values given by subjects rarely correspond to the specific monetary outcomes needed to predict choices. The curve-fitting approach adopted also allows separate assessment in the gain and loss domains (where subjects often exhibit different risk attitudes) and could lead to superior predictions if the selected functional forms smooth out random response errors. Table 3 shows the specific assessment questions used to calibrate the EU and LDEU models.

Table 3. Assessment questions

Probability equivalents	Certainty equivalents
To assess $u(x)$ and $h(x)$ for x = \$3000 to $-\$3000$:	*To assess $u_g(x)$ in the gain domain:*
$(3000, 1.0) \sim (4000, p; -4000, 1-p)$	$(x, 1.0) \sim (4000, .75; 0, .25)$
$(2000, 1.0) \sim (4000, p; -4000, 1-)p$	$(x, 1.0) \sim (4000, .5; 0, .5)$
$(0, 1.0) \sim (4000, p; -4000, 1-p)$	$(x, 1.0) \sim (4000, .25; 0, .75)$
$(-2000, 1.0) \sim (4000, p; -4000, 1-p)$	*To assess $u_\ell(x)$ in loss domain:*
$(-3000, 1.0) \sim (4000, p; -4000, 1-p)$	$(x, 1.0) \sim (0, .75; -4000, .25)$
	$(x, 1.0) \sim (0, .5; -4000, .5)$
	$(x, 1.0) \sim (0, .25; -4000, .75)$
	To link gain and loss domains:
	$(x, 1.0) \sim (4000, .25; 0, .5; -4000, .25)$
To assess $h(\$4000)$ and $h(-\$4000)$:	*To assess $h(\$4000)$ and $h(-\$4000)$:*
$(4000, .5; -3000, .5) \sim (4000, p; -4000, 1-p)$	$(4000, .5; x, .5) \sim (4000, .75; -4000, .25)$
$(2000, .5; -4000, .5) \sim (4000, p; -4000, 1-p)$	$(x, .5; -4000, .5) \sim (4000, .25; -4000, .75)$

Subjects were asked to supply value p to the probability equivalents questions and value x to the certainty equivalents questions that provide indifference between the two options.

3.1. Assessment by probability equivalents

Assessment by probability equivalents required subjects to indicate the indifference probability p that satisfies $A \equiv (x, 1.0) \sim B \equiv (x^*, p; x_0, 1 - p)$ for a given value of x. The utility of outcome x can then be expressed as

$$u(x) = pu(x^*) + (1 - p)u(x_0) = p,$$

since $u(x^*) = 1$ and $u(x_0) = 0$. Probability equivalents for a hypothetical subject are given below; the corresponding utilities are graphed as points in figure 1a.

Monetary amount	x	−4000	−3000	−2000	0	2000	3000	4000
Probability equivalent	PE	0	0.5	1.25	.5	.875	.95	1

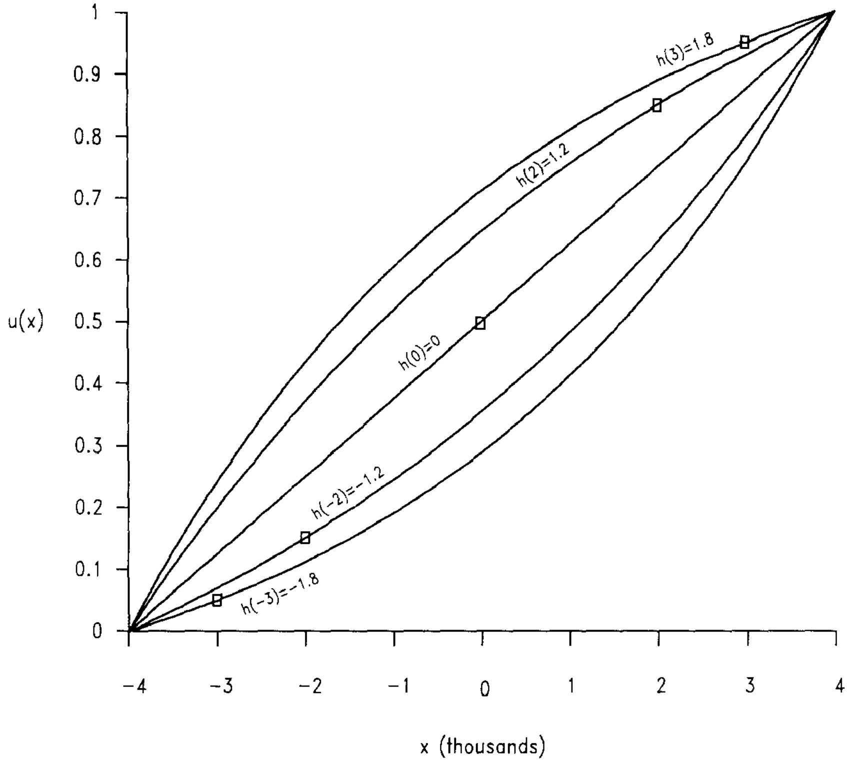

Figure 1a. Determination of hypothetical subject's $h(x)$ values from probability equivalence information. $h(x)$ values are determined as the parameter that specifies the exponential utility function that contains the points $(-4, 0)$, $(4, 1)$, and $(x, u(x))$.

The indifference probability can also be used to calculate the lottery-dependent utility outcome of x:

$$u(x, c_A) = \frac{1-e^{-h(x)\left(\frac{x-x_0}{x^*-x_0}\right)}}{1-e^{-h}(x)} = pu(x^*, c_B) + (1-p)u(x_0, c_B) = p,$$

since $u(x^*, c_B) = 1, u(x_0, c_B) = 0$, and $c_A = h(x)$ (assume $h(x) \neq 0$). As shown in figure 1a, $h(x)$ is set equal to the parameter of the exponential utility function that includes the points $(-\$4000, u(-\$4000) = 0)$, $(\$4000, u(\$4000) = 1)$ and $(x, u(x))$. The value of $h(x)$ that satisfies this equation can be used to calculate the lottery-dependent utility of any option that includes outcome x. Since the experimental scenarios constructed for this study involve few outcomes, the process described above is only required to find $h(x)$ values for $x =$ \$3000, \$2000, \$0, $-$\$2000, and $-$\$3000. These computed values, along with the final two probability indifference judgments in table 3, were then used to determine $h(x)$ values for $x = \$4000$ and $x = -\$4000$.

3.2. *Assessment by certainty equivalents*

Assessment by certainty equivalents required subjects to indicate the certain outcome x that satisfies $A \equiv (x, 1.0) \sim B \equiv (x^*, p; x_0, 1-p)$ for a given value of p. One utility function $u_g(x)$ was estimated by setting $x^* = \$4000$ and $x_0 = \$0$, fixing $u_g(x^*) = 1$ and $u_g(x_0) = 0$. The expected utility of the specified certain outcome in the gain domain is then given, as above, by $u_g(x) = p$. As shown in table 3, three indifference questions were required to determine the certain outcomes associated with $p = .75, .50$, and $.25$. The estimate $\hat{u}_g(x)$ was then derived over the range $\$0 \leq x \leq \$4000$ as the best-fitting exponential function from the specified indifference judgments (see Keller, 1985a, for details of the fitting process). The utility function derived in this manner for the gain domain is completely specified by the best-fitting exponential parameter, c_g.

A separate utility function $u_\ell(x)$ was estimated for the loss domain by setting $x^* = \$0$ and $x_0 = -\$4000$, fixing $u_\ell(x^*) = 1$ and $u_\ell(x_0) = 0$. The utility in the loss domain of the specified certain loss is then given by $u_\ell(x) = p$. Three indifference questions, corresponding to $p = .75, .50$, and $.25$, were constructed for this stage of assessment and the estimated $\hat{u}_\ell(x)$ derived over the range $-\$4000 \leq x \leq \$0$ as the best-fitting exponential function, with exponential parameter c_ℓ, from this set of indifference judgments.

The next step in the assessment process determines the linked scale for the gain and loss exponential functions shown above. This is achieved by eliciting the certainty equivalent CE that satisfies $(CE, 1.0) \sim (\$4000, .25; \$0, .50; -\$4000, .25)$. Since $u(\$4000) = 1$ and $u(-\$4000) = 0$, the expected utility of the outcome CE is given by

$$u(CE) = .25 + .50u(\$0).$$

The appropriate scale is set by determining $q = u(\$0)$. If $CE = \$0$, then the equation above becomes

$$q = .25 + .50q \rightarrow q = .50.$$

If $CE > \$0$, the utility of this outcome can be derived from the gain-domain utility function with the scaling factor q:

$$u(CE) = q + (1 - q)\,\hat{u}_g(CE).$$

Substitution yields the following:

$$q + (1 - q)\hat{u}_g(CE) = .25 + .50q,$$

or

$$q = \frac{.25 - \hat{u}_g(CE)}{.50 - \hat{u}_g(CE)}.$$

Since $q \geq 0$, the logical limit on $\hat{u}_g(CE)$ is $\hat{u}_g(CE) < .25$. If $CE < \$0$, a similar process yields the following expression for q:

$$q = \frac{.25}{\hat{u}_\ell(CE) - .50},$$

with logical limit $\hat{u}_\ell(CE) > .75$. The two-piece exponential utility function defined over the range $-\$4000 \leq x \leq \$4000$ then takes the following form:

$$u(x) = \begin{cases} q + (1 - q)\hat{u}_g(x), & \text{if } \$0 < x \leq \$4000 \\ q, & \text{if } x = \$0 \\ q\hat{u}_\ell(x), & \text{if } -\$4000 \leq x < \$0 \end{cases}$$

Therefore, the utility function as assessed by the curve-fitting method is completely specified by the exponential parameters c_g and c_ℓ, and the scale factor q. This fitted utility function is likely to be smoother than that derived from probability equivalents, as shown in figure 1b for a hypothetical subject providing the following certainty equivalents.

	$(4000, p; 0, 1 - p)$			$(0, p; -4000, 1 - p)$			$(4000, p; 0, .5; -4000, 1 - p)$
p	.75	.5	.25	.75	.5	.25	.25
CE	2000	1250	500	−250	−1000	−2000	150

The LDEU $h(x)$ values of interest can be derived from the utility function identified above. As in the direct approach, $h(x)$ for a given value of x is set equal to the unique parameter of the exponential utility function that includes the points $(-\$4000, u(-\$4000) = 0)$, $(\$4000, u(\$4000) = 1)$, and $(x, u(x))$. This can be expressed as

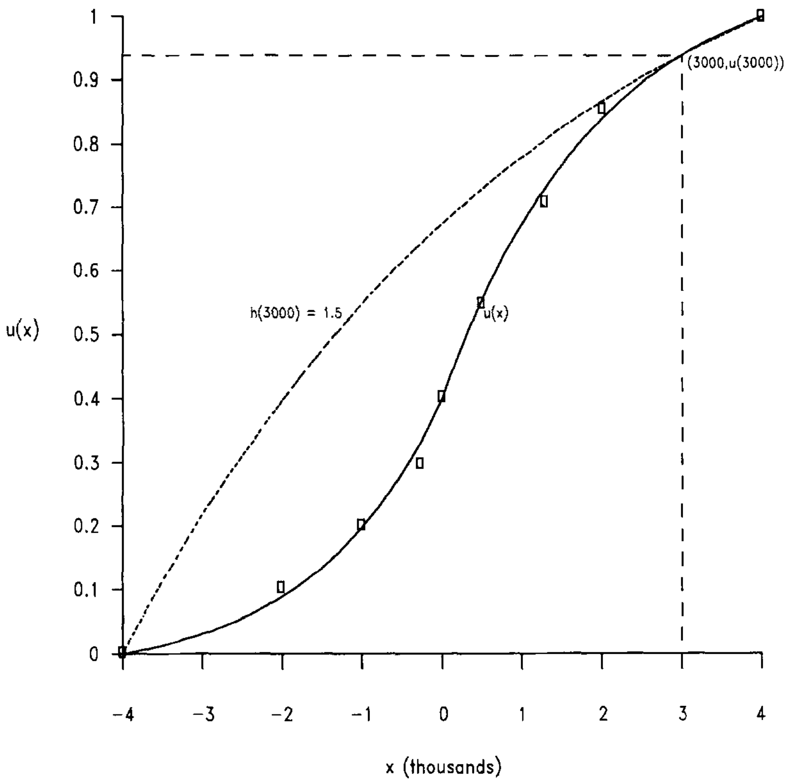

Figure 1b. Determination of hypothetical subject's $h(x)$ values from certainty equivalence information. $h(3000)$ is determined as the parameter that specifies the exponential utility function that contains the points $(-4, 0)$, $(4, 1)$, and $(3000, u(3000))$.

$$u(x,h(x)) = \frac{1-e^{-h(x)\left(\frac{x-x_0}{x^*-x_0}\right)}}{1-e^{-h(x)}} = u(x).$$

Once again, this equation can be solved for $h(x)$ and the associated value later utilized in part to determine the lottery-dependent utility of any option that includes outcome x. An example is depicted graphically in figure 1b. This approach is followed to compute $h(x)$ values for outcomes ranging from $-\$3000$ to $\$3000$; these values are then used to determine $h(x)$ values for $x = \$4000$ and $x = -\$4000$ using the final two certainty equivalence judgments in table 3.

Note that a one-piece exponential utility function is a special case of the exponential LDEU model in which $h(x)$ is a constant. Since the curve-fitting method described above links two exponential functions across the gain and loss domains, the resulting LDEU model will only yield a constant $h(x)$ if both component utility functions are linear.

3.3. *Experimental design*

All experimental subjects responded to the probability equivalence assessment questions after responding to the 18 choice scenarios. Responses to the certainty equivalence questions were collected in a similar manner one week later. A total of 117 students completed the choice and assessment parts of the experiment. Of these, 19 subjects provided certainty equivalents consistent with the logical bounds required by stochastic dominance ($u_g(CE) < .25$ and $u_\ell(CE) > .75$) but in violation of the corresponding logical limit on the fitted functions, *i.e.*, violating $\hat{u}_g(CE) < .25$ or $\hat{u}_\ell(CE) > .75$. We were unable to recontact these subjects for new responses based on the fitted functions; therefore, they were omitted, reducing our sample to $117 - 19 = 98$ subjects.

4. Results on predictive accuracy of assessed models

The main experimental research question in this section is whether or not the assessed LDEU model outperforms the assessed EU model in predicting the actual choices of subjects. Table 4 outlines our experimental design and shows that over all treatments, LDEU correctly predicted 58.42% of the responses, compared to 61.88% for the EU model. (Equal utility calculated for the two options in a scenario was counted as a correct prediction.) In addition, the relative performance of the two models depended on neither the type of scenario over which choices were predicted nor the assessment method used to calibrate the models (shown by the percentage of correct predictions for the EU and LDEU models over the set of paradoxical and independent scenarios, and over assessment by the direct *PE* and the curve-fitting *CE* methods). Details on the predictive performance of the two models are provided by further analysis.

Table 4. Experimental design and percentage of correct predictions by assessed models

	Expected utility: Percent correctly predicted		
	Probability equivalence method	Certainty equivalence method	Average
Paradoxical scenarios (N = 54)	56.28	67.80	62.04
Independent scenarios (N = 44)	53.28	70.08	61.68
Total (N = 98)	54.93	68.83	61.88

	Lottery-dependent expected utility: Percent correctly predicted		
	Probability equivalence method	Certainty equivalence method	Average
Paradoxical scenarios (N = 54)	58.02	64.09	61.06
Independent scenarios (N = 44)	50.88	59.47	55.17
Total (N = 98)	54.82	62.02	58.42

Since pairs of options in each holdout scenario were designed to have approximately equal expected value, the utilities of the options in a given pair, as computed by an assessed EU or LDEU model, often differed by very little. In these situations, a decision maker could adopt the choice of the associated model without suffering a significant utility loss. For this reason, it is important to measure performance not only as the instances in which a model correctly predicts a subject's choice, but also as the level of distinction a model draws between scenario options. The following classification scheme refines the definitions of correct and incorrect predictions to incorporate the utility difference between options in a scenario.

$CH \equiv$ *Clear Hit*—an instance in which a model correctly predicts the subject's choice *and* the utility difference between the pair of options is greater than or equal to some tolerance level α.
$NH \equiv$ *Near Hit*—correct model prediction; utility difference $< \alpha$.
$CM \equiv$ *Clear Miss*—a model incorrectly predicts a subject's choice *and* the utility difference $\leq \alpha$.
$NH \equiv$ *Near Miss*—incorrect model prediction; utility difference $< \alpha$.
$T \equiv$ *Tie*—a model predicts equal utility for both options.

Distinguishing by the utility difference between options provides a number of different ways in which a model's predictive performance can be evaluated. The frequency of accurate predictions, or hit ratio, can be represented by the quantity $\frac{CH+NH+T}{CH+NH+T+NM+CM}$. Alternatively, those instances where a model indicates no clear preference between options can be ignored to discount for the sensitivity of both predictions and choices to small response errors or inconsistencies; the appropriate performance measure in this case would be the clear hit ratio, $\frac{CH}{CH+CM}$. Finally, the percentage of clear observations, $\frac{CH+CM}{CH+NH+T+NM+CM}$, can be used to measure the tendency of a model to strongly discriminate between scenario options.

The data in table 5 show how the distribution of clear and near hits and clear and near misses varied with α, comparing the performance of the EU and LDEU models in predicting the choices of all 98 subjects. By definition, the hit ratio of each model was constant (with respect to α), and the percentage of clear observations decreased with larger values of α. The clear hit ratio of the EU model increased modestly with increasing α, while the corresponding measure for the LDEU model remain essential unchanged.

Table 6 provides further detail, comparing the predictive performance of the EU and LDEU models over paradoxical and independent scenarios, assessment by probability and certainty equivalents, and the domain of option outcomes. The percentage of clear hits and misses, near hits and misses, and ties were computed for a tolerance level $\alpha = 0.01$. The hit ratio, clear hit ratio, and percentage of clear observations are also presented to summarize the performance of each model.

Comparing across models, the results indicate that the LDEU model distinguished more strongly between scenario options, as seen by the consistently higher percentage of

Table 5. Overall model performance

	Percentages for different tolerance levels (α)				
EU model	.001	.005	.010	.030	.050
Clear Hit	42.16	36.48	31.69	18.57	11.65
Near Hit	0.96	6.64	11.43	24.55	31.47
Tie	18.76	18.76	18.76	18.76	18.76
Near Miss	1.30	7.88	12.47	24.03	30.01
Clear Miss	36.82	30.24	25.65	14.09	8.11
Hit ratio	61.88	61.88	61.88	61.88	61.88
Clear Hit ratio	53.38	54.68	55.27	56.86	58.96
% Clear observations	78.98	66.72	57.34	32.66	19.76
LDEU model					
Clear Hit	46.40	41.01	37.33	27.01	20.10
Near Hit	3.20	8.59	12.27	22.59	29.50
Tie	8.82	8.82	8.82	8.82	8.82
Near Miss	2.49	7.23	10.63	19.73	24.74
Clear Miss	39.09	34.35	30.95	21.85	16.84
Hit ratio	58.42	58.42	58.42	58.42	58.42
Clear Hit ratio	54.28	54.42	54.67	55.28	54.41
% Clear observations	85.49	75.36	68.28	48.86	36.94

clear observations. However, the generalized model did not appear to exhibit superior performance in predicting subjects' choices. This is apparent from the hit and clear hit ratios, which were notably higher for the LDEU model only in the mixed domain. While there was no significant difference in the hit ratios recorded by EU (62.04%) and LDEU (61.06%) over the set of paradoxical scenarios, the EU hit ratio (61.68%) was significantly higher than the LDEU hit ratio (55.17%) over the set of independent scenarios (difference significant at the $p = .0002$ level). Essentially identical performance was observed for the two models directly assessed by probability equivalents (EU hit ratio = 54.93% versus LDEU hit ratio = 54.82%); however, the EU hit ratio (68.83%) was significantly higher than the LDEU hit ratio (62.02%) when the models were calibrated using the curve-fitting approach (difference significant at the $p = .0001$ level).

On average, the predictive performance of LDEU favored paradoxical over independently structured scenarios (difference in hit ratios significant at the $p = .0109$ level), assessment by certainty over probability equivalents (significant at the $p = .0443$ and $p = .0086$ level for paradoxical and independent subjects, respectively), and options with mixed outcomes over those in the gain or loss domains. The model was most accurate on mixed-domain, paradoxical scenarios when assessed by certainty equivalents, achieving a hit ratio of 75.31% and a clear hit ratio of 72.66%. By contrast, EU attained its highest success on gain-domain, independent scenarios when assessed by certainty equivalents, attaining an

Table 6. Breakdown of experimental results ($\alpha = 0.01$)

	Percentage of observations					Performance measures		
	Clear Hit	Near Hit	Tie	Near Miss	Clear Miss	Hit Ratio	Clear Hit Ratio	% Clear Observations
Paradoxical scenarios								
EU	31.53	13.53	16.98	16.05	21.91	62.04	59.00	53.44
LDEU	37.35	14.71	9.00	11.47	27.47	61.06	57.62	64.82
Independent scenarios								
EU	31.88	8.84	20.96	8.08	30.24	61.68	51.32	62.12
LDEU	37.31	9.28	8.58	9.60	35.23	55.17	51.43	72.54
Probability equivalents								
EU	36.45	9.07	9.41	11.51	33.56	54.93	52.06	70.01
LDEU	39.85	11.17	3.80	9.86	35.32	54.82	53.01	75.17
Certainty equivalents								
EU	26.93	13.78	28.12	13.44	17.73	68.83	60.30	44.66
LDEU	34.81	13.38	13.83	11.39	26.59	62.02	56.69	61.40
Gain domain								
EU	30.19	13.27	21.00	13.94	21.60	64.46	58.29	51.79
LDEU	34.86	15.22	9.69	10.04	31.72	58.24	52.36	66.58
Loss domain								
EU	26.45	11.82	24.40	11.05	26.28	62.67	50.16	52.73
LDEU	27.13	15.22	10.12	14.20	33.33	52.47	44.87	60.46
Mixed domain								
EU	38.44	9.18	10.88	12.41	29.09	58.50	56.92	67.53
LDEU	50.00	7.91	6.63	7.65	27.81	64.54	64.26	77.81

82.58% hit ratio and a 72.55% clear hit ratio. The difference in hit ratios for EU with the two elicitation methods is significant at the $p = .0016$ and $p = .00007$ level for paradoxical and independent subjects, respectively.

A detailed examination of the subjects' responses and the assessed models' predictions for the paradoxical scenarios demonstrated both the existence of the paradoxical choice patterns associated with common ratio, common consequence, and folded ordering scenarios, and the capability of the assessed LDEU model to predict these patterns. Paradoxical choices were observed frequently in the data on subjects' preferences, placing severe limitations on the predictive ability of the EU model. However, neither EU nor LDEU managed to correctly predict the exact choice pattern for a large percentage of the subjects over these scenarios. Details of the results are contained in a companion working paper available from the authors. The next section provides some suggestions for reducing the disparity between the predictive capacity of the lottery-dependent model and its performance.

5. A functional form for $h(x)$

The $h(x)$ functions associated with the LDEU models that best fit the actual choices of each subject were examined in detail. A consistent pattern among these models emerged in which $h(x)$ values remained essentially constant across the range of x, with the exceptions of the upper and lower endpoints, which tended to take on significantly higher and lower $h(x)$ values, respectively. This predominant $h(x)$ pattern, shown in figure 2, indicates a relatively constant risk attitude at moderate outcome levels, with increasing risk aversion at the upper end and increasing risk proneness at the lower end of the outcome range.

The $h(x)$ pattern shown in figure 2 can be represented by the following functional form:

$$h(x) = r + s(x - t)^3,$$

where r, s, and t are constants specific to the individual decision maker. The parameter t can be interpreted as an individual's target or reference level of the outcome variable. The parameter r then specifies the value of $h(t)$, indicating the risk attitude for a sure outcome of the neutral target amount t. The parameter s sets the scale of $h(x)$ over the range $-\$4000 \leq x \leq \$4000$ and thus controls the variability of $h(x)$ values.

An advantage of the above functional form for $h(x)$ is that only three parameters need to be varied in searching for an LDEU model that best predicts a subject's choice pattern. Constraining $h(x)$ values to adhere to the cubic form shown above, a best-predicting LDEU model can be fitted to subjects' responses to choice scenarios. Table 7, which

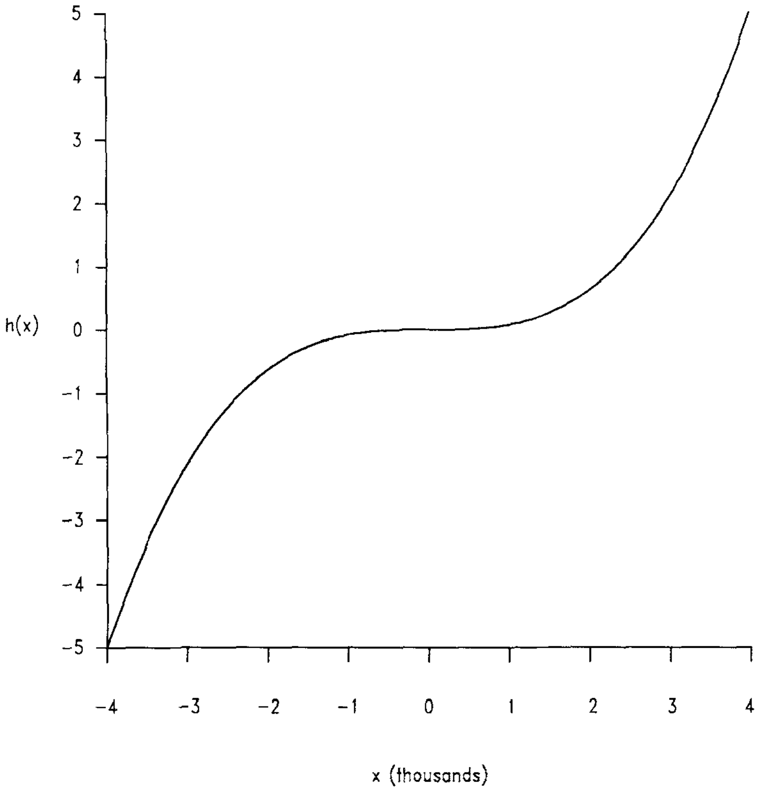

Figure 2. Predominant $h(x)$ pattern.

presents a summary of the predictive performance of the models evaluated in this study (ties are omitted in this table), shows that 92.06% of the paradoxical scenarios and 84.44% of the independent scenarios are correctly predicted by the cubic model (with the values of r and t varied between -5 and 5 in increments of 0.25 and s varied between -0.25 and 0.25 by 0.025). This indicates that only a moderate reduction in predictive capacity is associated with the introduction of the cubic constraint on $h(x)$.

Two approaches for calibrating the LDEU model that exploit the cubic form of the underlying $h(x)$ function are suggested. First, the indifference information provided by suitably chosen probability or certainty equivalence judgments can be fitted to the cubic $h(x)$ function given above, and the resulting values of r, s, and t used to generate predictions over any holdout sample of interest. A second approach would be to construct a series of choice scenarios designed to span the entire outcome range and to record a decision maker's responses to these scenarios. A search could then be conducted to find combinations of r, s, and t that yield $h(x) = r + s(x - t)^3$ values that best predict the response pattern of the subject. The feasible combinations of r, s, and t identified by this process then each correspond to one LDEU model that may be appropriate for the individual decision maker. By specifying a decision rule for selecting among the feasible combinations of r, s, and t, the associated values of $h(x)$ can be used to predict that subject's response to any additional choice scenario of interest. Such a choice-based calibration procedure requires further research on the appropriate design of assessment choice scenarios and the formulation of decision rules for selecting among the feasible values of r, s, and t. Preliminary work in this area indicates that both alternative approaches hold promise for substantially improving the predictive performance of the LDEU model.

6. Conclusions and suggestions for further research

Tversky (1967) hinted at the potential formulation of a lottery-dependent theory: "[One could] redefine the consequences so that winning a certain amount in a gamble is regarded as a different consequence from receiving the same amount as a sure-thing," but

Table 7. A comparison of predictive performance

	Percentage correct	
	Paradoxical scenarios	Independent scenarios
Best-predicting models		
EU	71.91	83.96
LDEU	93.52	91.92
LDEU ($h(x) = r + s(x - t)^3$)	92.06	84.44
Models assessed by indifference judgments		
EU (Probability Equivalents)	51.83	48.33
EU (Certainty Equivalents)	57.22	55.78
LDEU (Probability Equivalents)	55.89	49.61
LDEU (Certainty Equivalents)	58.67	52.50

was not optimistic about the prospects of implementing such a model: "In spite of its apparent plausibility, this approach does not yield testable predictions because consequences cannot be identified independently of gambles . . . utility has to be defined not on monetary outcomes but on abstract consequences which depend on subjective probabilities as well." Overcoming this modeling challenge, Becker and Sarin (1987) successfully developed a lottery-dependent utility model with a form specific enough to obtain testable predictions.

In this article, we have implemented a lottery-dependent utility model and compared its predictive capacity and predictive performance with expected utility's capacity and performance. LDEU was shown to exhibit a higher capacity for predicting the actual choices of experimental subjects among risky options. However, when the models were calibrated using probability or certainty equivalence indifference judgments, LDEU did not significantly outperform EU, and neither model approached its predictive capacity. A functional form was presented for the parametric $h(x)$ function that defines an LDEU model, and assessment approaches in which this form is exploited were proposed.

Several research directions have been identified during the course of this work. To find a more valid description of preferences under risk, other lottery-dependent models should be developed and experimentally tested. Potential extensions include relaxation of the assumption that requires the exponential parameter c_F to be linear in probabilities (e.g., weighting high probability outcomes disproportionately) and consideration of functional forms of utility other than the exponential (e.g., a power function form for $u(x, c_F)$).

Questions regarding the proper approach for assessing utility models require additional attention. The relative lack of success experienced by both the EU and LDEU models calibrated in an indifference response mode may stem from the incongruent processes involved with making ordinal versus matching judgments (see, e.g., Tversky, Sattath, and Slovic, 1988). Further research might first consider the performance of expected utility when both the assessment and test questions utilize identical response modes. Extending these results to generalized utility models may then provide an accurate and useful representation of preferences under risk.

References

Allais, Maurice. (1953). "Le Comportement de l'Homme Rational Devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine," *Econometrica* 21, 503-546.

Becker, Joao. (1986). "A New Model of Decisions under Risk using the Concept of Lottery Dependent Utility Function," unpublished Ph.D. thesis, Graduate School of Management, UCLA.

Becker, Joao, and Rakesh Sarin. (1987). "Lottery Dependent Utility," *Management Science* 33, 1367-1382.

Becker, Joao, and Rakesh Sarin. (1989). "Decision Analysis using Lottery Dependent Utility," *Journal of Risk and Uncertainty* 2, 105-117.

Camerer, Colin. (1989). "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty* 2, 61-104.

Chew, Soo Hong, and Kenneth MacCrimmon. (1979a). "Alpha-Nu Choice Theory: A Generalization of Expected Utility Theory," Working Paper No. 669, University of British Columbia Faculty of Commerce and Business Administration, Vancouver.

Chew, Soo Hong, and Kenneth MacCrimmon. (1979b). "Alpha-Nu Theory, Lottery Composition and the Allais Paradox," Working Paper No. 686, University of British Columbia Faculty of Commerce and Business Administration, Vancouver.

Chew, Soo Hong, and William Waller. (1986). "Empirical Tests of Weighted Utility," *Journal of Mathematical Psychology* 30, 55-72.

Coombs, Clyde. (1969). "Portfolio Theory: A Theory of Risky Decision Making," *La Decision*, Paris: Centre National de la Recherche Scientifique.

Coombs, Clyde. (1975). "Portfolio Theory and the Measurement of Risk." In M. Kaplan and S. Schartz (eds.), *Human Judgment and Decision Processes*. New York: Academic Press.

Coombs, Clyde, and Lily Huang. (1970). "Tests of a Portfolio Theory of Risk Preference," *Journal of Experimental Psychology* 85, 23-29.

Currim, Imran, and Rakesh Sarin. (1989). "Prospect Versus Utility," *Management Science* 35, 22-41.

Hershey, John, Howard Kunreuther, and Paul Schoemaker. (1982). "Sources of Bias in Assessment Procedures for Utility Functions," *Management Science* 28, 936-954.

Kahneman, Daniel, and Amos Tversky. (1979). "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47, 263-291.

Keller, L. Robin. (1985a). "An Empirical Investigation of Relative Risk Aversion," *IEEE Transactions on Systems, Man, and Cybernetics* 15, 475-482.

Keller, L. Robin. (1985b). "The Effects of Problem Representation on the Sure-Thing and Substitution Principles," *Management Science* 31, 738-751.

Machina, Mark. (1982). "Expected Utility Analysis Without the Independence Axiom," *Econometrica* 50, 277-323.

Marschak, Jacob. (1950). "Rational Behavior, Uncertain Prospects, and Measurable Utility," *Econometrica* 18, 111-141.

Tversky, Amos. (1967). "Additivity, Utility, and Subjective Probability," *Journal of Mathematical Psychology* 4, 175-201.

Tversky, Amos, Shmuel Sattath, and Paul Slovic. (1988). "Contingent Weighting in Judgment and Choice," *Psychological Review* 95, 371-384.

Weber, Martin, and Colin Camerer. (1987). "Recent Developments in Modeling Preferences under Risk," *OR Spektrum* 9, 129-151.

von Neumann, John, and Oskar Morgenstern. (1947). *Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press, Second Edition.