ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 52, 524–543 (1992)

Choice-Based Assessment of Utility Functions

RICHARD L. DANIELS

Fuqua School of Business, Duke University

AND

L. ROBIN KELLER

Graduate School of Management, University of California, Irvine

An approach for calibrating utility functions in which consistent paired-comparison response modes are used for both elicitation and descriptive validation is proposed and experimentally evaluated. The choice-based procedure presented has the potential to avoid the systematic biases often observed in functions assessed via indifference judgments such as probability or certainty equivalents. Our results indicate that the choice-based assessment procedure outperforms calibration approaches based on indifference judgments in predicting subjects' choices among risky options for the two types of utility models studied, von Neumann–Morgenstern expected utility and lottery dependent expected utility. © 1992 Academic Press, Inc.

1. INTRODUCTION

Utility functions are conventionally calibrated by eliciting indifference judgments, *e.g.*, probability or certainty equivalents, from an individual decision maker. One problem with such an approach is that an assessed utility function can vary systematically with the type of indifference information elicited. Utility models calibrated from indifference judgments have, in addition, exhibited relatively poor performance in predicting individuals' preferences. As an alternative to indifference-based assessment, this paper presents a calibration approach which utilizes a decision maker's choices among risky options to determine a utility model's appropriate representation for that individual. This distinction between choice and indifference is based on the premise that fundamentally different processes are used to formulate these judgments, and thus while indifference can theoretically be achieved through a converging series of choices (or, alternatively, indifference judgments can be used to infer choice), in practice these judgments are not interchangeable. Two utility models, von Neumann–Morgenstern (1947) expected utility (EU) and Becker–Sarin (1987) lottery dependent expected utility (LDEU), are used to illustrate the application of choice-based assessment. Experimental results comparing the predictive performance of the choice-based models with that of their indifference-based counterparts are also provided.

We express our gratitude to Julie Edell and Michael Moore for their assistance in the statistical analysis of our results, to William Goldstein, John Hershey, John Miyamoto, Paul Schoemaker, and Mary Kay Stevenson for their valuable comments on an earlier draft of the paper, and to Michael Birnbaum for his efforts as guest action editor for this issue. Reprint requests should be addressed to L. Robin Keller, Graduate School of Management, University of California at Irvine, Irvine, CA 92717.

Although the information content contained in a single choice is clearly less than that obtained from an indifference judgment, a utility model calibrated from choices may still provide a more accurate representation of an individual's preferences. This is because the quality of the elicited choice information may be significantly higher, in terms of replicability, confidence in judgment, and consistency across judgments, than the corresponding indifference judgments. While an indifference-based assessment procedure requires relatively few judgments from a decision maker, these judgments may be difficult to provide reliably. In contrast, a choice-based assessment approach requires a larger number of judgments, each of which may be more easily elicited from decision makers. All of these factors, information content, information quality, and effort involved in the elicitation process, must be weighed in selecting an appropriate assessment method.

Conclusions regarding the descriptive validity of utility models can also be influenced by the assessment method chosen. Currim and Sarin (1989, 1992) and Daniels and Keller (1990) evaluated utility models assessed via indifference judgments by counting the number of correct predictions over a holdout sample of choice scenarios. The indifference-based response mode used to calibrate the models in these studies differed fundamentally from the choices used to test the models, confounding the evaluation process. Choice-based assessment avoids this problem by utilizing the same response mode for both model assessment and evaluation. In a related area, Schoemaker and Waid (1982) have compared multiattribute value function assessment methods under certainty using holdout samples involving both choices and direct strength of preference ratings.

The paper is organized as follows. Section 2 provides some background on problems commonly encountered in utility assessment. The expected utility and lottery dependent expected utility models are briefly reviewed in Section 3. Section 4 outlines the design of an experiment that compares the performance of the utility models calibrated via indifference vs. choice judgments in predicting subjects' choices among risky options. Details on the assessment approaches used to calibrate the models are provided in Section 5. Experimental results are discussed in Section 6, and Section 7 concludes with a summary and suggestions for further research.

2. BACKGROUND

One utility assessment procedure using indifference judgments requires a decision maker to adjust the level of a sure outcome until indifference is attained between this *certainty equivalent* and a given lottery. The utility function determined in this manner often differs from that obtained when the decision maker is asked to adjust the probability distribution of a lottery until indifference is attained between this *probability equivalent* and a given sure outcome (see, e.g., Hershey, Kunreuther, & Schoemaker, 1982; Hershey & Schoemaker, 1985; Byrd, de Neufville, & Delquie, 1987; Johnson & Schkade, 1989; and Schoemaker & Hershey, 1992). Utility functions calibrated via certainty equivalents may also vary with the assessment approach, e.g., more risk averse expected utility functions are obtained when certainty equivalents are elicited from a series of two-outcome lotteries of the form $(x,p;\ 0,1-p)$ (outcome x received with probability p, 0 otherwise, with x varied) for high values of p than for low values.

Also relevant to the selection of an appropriate assessment procedure is the discrepancy often observed between inferred preferences derived from paired comparison choices and direct rating methods. This preference reversal phenomenon (see, e.g., Lichtenstein & Slovic, 1971; and Grether & Plott, 1979) could be explained (see, e.g., Tversky, Slovic, & Kahneman, 1990) by (i) scale incompatibility (outcomes are weighted more heavily in pricing than in choice), (ii) the prominence effect (more prominent attributes are weighted more heavily when making choices), or (iii) expression effects (Goldstein & Einhorn, 1987)].

Several approaches are available to counteract the observed problems with assessment methods based on indifference information. The first is to refine elicitation methods to adjust for assessment biases. For example, since certainty equivalent methods involve one sure outcome, McCord and de Neufville (1986) and de Neufville and Delquie (1988) proposed a structure whereby the level of an outcome within a lottery is adjusted so that indifference is achieved between two gambles to offset the certainty effect (Kahneman & Tversky, 1979). Based on a series of experiments eliciting indifference between pairs of lotteries, Delquie and de Neufville (1988) revised their emphasis on the certainty effect explanation for assessment biases in favor of overweighting of the response dimension.

A second approach is to recognize that vagueness or response errors are incurred when an indifference-based assessment procedure is used, and develop a framework for adjusting for these errors. For example, Bostic, Herrnstein, and Luce (1990) found that a choice-based sequential procedure for developing tight bounds on the value of certainty equivalents holds promise for eliminating the preference reversal phenomenon of overstating the value for lotteries with a moderate probability of a large gain. In a similar vein, Weber (1987) and Nau (1990) investigated methods for guiding decision making when there is incomplete or vague information on probabilities or utilities. Weber (1985) also considered models that incorporate incomplete information for riskless decision situations, while Barron and Schmidt (1988) have examined bounds on the weights in multi-attribute measurable value functions. Chu, Moskowitz, and Wong (1988) and Moskowitz, Wong, and Chu (1989) have developed a software package that allows a person to be vague in specifying preference information (i.e., giving a range rather than a point estimate). Eliashberg and Hauser (1985) proposed a utility function estimation approach which incorporates the possibility of measurement errors in choice (or indifference) data. Schoemaker and Hershey (1992) have developed a model with random noise to clarify differences between probability and certainty equivalent judgments.

A third approach is to develop a unifying model to relate responses elicited via different response modes. Tversky, Sattath, and Slovic (1988) introduced a contingent weighting model to represent the variation in inferred preferences resulting from indifference judgments and choices. Mellers, Ordóñez, and Birnbaum (1992) have investigated whether an individual attaches an underlying value, common to multiple assessment procedures, on achieving a specific level of the relevant attribute. Busemeyer and Goldstein (1992) have developed a model unifying choice and indifference judgments, and Goldstein and Busemeyer (1992) have considered how to distinguish whether differences among paired-comparison response modes reflect changes in preference to criterion shifts. Goldstein and Einhorn's (1987) expression theory also takes this approach.

The approach adopted in this paper avoids the problems associated with assessment from indifference judgments by requiring only choice information from decision makers. To the extent that utility models represent theories of choice as opposed to theories explaining how decision makers achieve indifference, it seems reasonable that utility models should be assessed and validated using choice data. Our motivation to investigate a choice-based approach stems from work on a previous paper (Daniels & Keller, 1990) in which individual subjects' EU and LDEU models were calibrated using certainty and probability equivalents. The predictive performance of the models was then evaluated by counting the number of correct predictions over a set of 18 choice scenarios involving pairs of risky options. Using the choice data collected for that paper, we conducted preliminary simulations to estimate the potential of a choice-based assessment procedure by partitioning the set of choice scenarios into an assessment sample (from which a model would be calibrated) and a holdout sample (on which a model would be evaluated). Our preliminary results indicated that, by including some paradoxical scenarios (e.g., those linked by common ratio or common consequence transformations) in the assessment sample, the resulting EU and LDEU models performed quite well in predicting choices among similarly related holdout scenarios.

Even if choices are used to assess an expected utility function, problems still remain to be overcome. For example, subjects often prefer option $A \equiv (\$3200,1.0)$ to option $B \equiv (\$4000,.8;\$0,.2)$. The expected utility function calibrated from choices made over similar scenarios would tend to be concave, reflecting risk aversion. Alternatively, if scenarios involving options like $A' \equiv (\$3200,.1;\$0,.9)$ and $B' \equiv (\$4000,.08;\$0,.92)$ are used to calibrate an expected utility model, the assessed function would likely be convex, reflecting risk proneness. Individuals who prefer both A over B and B' over A' violate the substitution (common ratio) principle of expected utility. The problem for prescriptive decision analysis is not that expected utility is violated, since the analysis process of applying an assessed utility function to a specific decision problem will guarantee that the substitution principle and expected utility are obeyed. Rather, the problem is that assessment questions, which by expected utility standards should yield identical utility functions, can produce widely varying utility models, as demonstrated in the results of McCord and de Neufville (1984).

This problem suggests that a generalized utility theory model, such as prospect theory (Kahneman & Tversky, 1979) or lottery dependent expected utility theory (Becker & Sarin, 1987), will be needed to provide a descriptively valid model. As discussed by Keller (1989), a recurring issue in prescriptive decision analysis is determining how far to aid decision makers in restructuring the problem and the relevant preferences. Since many people violate assumptions (such as the substitution principle) of expected utility, analysts must determine the extent to which conformance with expected utility is required, and when violations of these principles should be allowed and a generalized utility model used for guiding choice.

In this paper, we have chosen to investigate a choice-based approach for both the dominant prescriptive theory, expected utility theory, and for a generalized utility theory, lottery dependent utility theory. If LDEU were to be used in predictive or prescriptive settings, it would be very easy to implement, since it has a well-specified functional form as a special case.

3. 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 lottery dependent expected utility model (see Becker, 1986; and Becker & Sarin, 1987, 1989), is more general than expected utility, allowing the utility of an outcome to depend 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 that there exists a real-valued function $h(x)$, specific to a decision maker, such that

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

Section 5 describes a cubic form for the $h(x)$ function that is useful in calibrating LDEU models for individual decision makers.

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

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

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

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 $c_F > 0$, then the LDEU 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 all available lotteries using the same exponential utility function. In this case, the exponential forms of the EU and LDEU models are equivalent.

4. EXPERIMENTAL STUDY

A total of 82 Duke University MBA students voluntarily participated in an experiment to evaluate the predictive performance of EU and LDEU models assessed with indifference and choice judgments. Subjects were initially given a set of 30 hypothetical choice scenarios consisting of pairs of risky investment options and asked to indicate the most-preferred alternative in each case. These 30 choices then formed the assessment sample from which the utility models were calibrated using a choice-based approach.

4.1. Structure of Assessment Sample

The scenarios comprising the assessment sample were linked by common ratio transformations to provide a significant predictive challenge for the models. The basic construction of scenarios in the assessment sample can be represented as follows:

$$A \equiv (\$(a + \Delta), p; \$0, 1 - p) \text{ vs.}$$
$$B \equiv \left(\$4000, p\left(\frac{\$a}{\$4000}\right); \$0, 1 - p\left(\frac{\$a}{\$4000}\right)\right).$$

Subjects were thus presented with two options in each scenario, one option involving a p chance of receiving a base amount a adjusted by factor Δ and a $1 - p$ chance of receiving \$0, and the second option involving a proportional chance ($p\$a/\$4000$) of receiving \$4000 (the largest possible outcome) or \$0 otherwise.

Three base amount values (a = \$1000, \$2000, \$3000) were included in the experimental design to represent low, moderate, and large outcomes within the range [\$0,\$4000] considered. Similarly, three adjustment factors ($\Delta = -\$500, \$0, \$500$) were selected in an effort to determine the degree of risk aversion/proneness exhibited by individual subjects. Finally, three base probability adjustments (p = 1.0, .20, .04) were included to represent no, moderate, and extreme common ratio transformations that may lead to paradoxical choice behavior inconsistent with the expected utility model. One scenario was constructed for each combination of a, Δ, and p, for a total of $3^3 = 27$ scenarios. Three additional scenarios were generated to determine bounds on subjects' $h(\$4000)$ values. These scenarios are shown below:

$$A \equiv (\$2500, 1.0) \text{ vs. } B \equiv (\$4000, .25; \$2000, .75)$$
$$A \equiv (\$3000, 1.0) \text{ vs. } B \equiv (\$4000, .50; \$2000, .50)$$
$$A \equiv (\$3500, 1.0) \text{ vs. } B \equiv (\$4000, .75; \$2000, .25).$$

Subjects were also asked to provide probability and certainty equivalence judgments for direct assessment of the models via indifference information. These questions are shown in Table 1.

4.2. Structure of Holdout Sample

In a subsequent session, subjects were provided with a holdout sample of scenarios consisting of 21 pairs of risky options and asked to indicate the most-preferred option in each scenario. As shown in Table 2, this sample consisted of seven basic scenarios whose options had an expected value ranging from \$500 to \$3500 in increments of \$500. From each of the seven original scenarios, two additional scenarios were constructed by taking moderate and extreme common ratio transformations, yielding the 21 holdout sample questions used to test the predictive performance of the models.

5. MODEL ASSESSMENT

5.1. Choice-Based Assessment of the EU Model

The strategy adopted for calibrating an expected utility model for an individual subject is quite simple. The value of $u(x)$ was varied system-

TABLE 1
ASSESSMENT QUESTIONS

Probability equivalents	Certainty equivalents
To assess $u(x)$ and $h(x)$ for x = \$500 to \$3500:	To assess $u(x)$ and $h(x)$:
(\$ 500,1.0) ~ (\$4000,p; \$0,1 − p)	(\$$x$,1.0) ~ (\$4000,.75; \$0,.25)
(\$1000,1.0) ~ (\$4000,p; \$0,1 − p)	(\$$x$,1.0) ~ (\$4000,.5; \$0,.5)
(\$1500,1.0) ~ (\$4000,p; \$0,1 − p)	(\$$x$,1.0) ~ (\$4000,.25; \$0,.75)
(\$2000,1.0) ~ (\$4000,p; \$0,1 − p)	
(\$2500,1.0) ~ (\$4000,p; \$0,1 − p)	
(\$3000,1.0) ~ (\$4000,p; \$0,1 − p)	
(\$3500,1.0) ~ (\$4000,p; \$0,1 − p)	
To assess $h(\$4000)$ and $h(\$0)$:	
(\$4000,.5; \$500,.5) ~ (\$4000,p; \$0,1 − p)	
(\$2000,.5; \$0,.5) ~ (\$4000,p; \$0,1 − p)	

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

TABLE 2
HOLDOUT SAMPLE OF SCENARIOS

Scenario (i)	Scenario classification	Option A_i	Option B_i
1	$S1$	($500,1.0)	($1000,.50;$0,.50)
2	($S1$,.50;$0,.50)	($500,.50;$0,.50)	($1000,.25;$0,.75)
3	($S1$,.10;$0,.90)	($500,.10;$0,.90)	($1000,.05;$0,.95)
4	$S4$	($1000,1.0)	($2000,.50;$0,.50)
5	($S4$,.50;$0,.50)	($1000,.50;$0,.50)	($2000,.25;$0,.75)
6	($S4$,.10;$0,.90)	($1000,.10;$0,.90)	($2000,.05;$0,.95)
7	$S7$	($1500,1.0)	($3000,.50;$0,.50)
8	($S7$,.50;$0,.50)	($1500,.50;$0,.50)	($3000,.25;$0,.75)
9	($S7$,.10;$0,.90)	($1500,.10;$0,.90)	($3000,.05;$0,.95)
10	$S10$	($2000,1.0)	($3000,.50;$1000,.50)
11	($S10$,.50;$1000,.50)	($2000,.50;$1000,.50)	($3000,.25;$1000,.75)
12	($S10$,.10;$1000,.90)	($2000,.10;$1000,.90)	($3000,.05;$1000,.95)
13	$S13$	($2500,1.0)	($4000,.50;$1000,.50)
14	($S13$,.50;$1000,.50)	($2500,.50;$1000,.50)	($4000,.25;$1000,.75)
15	($S13$,.10;$1000,.90)	($2500,.10;$1000,.90)	($4000,.05;$1000,.95)
16	$S16$	($3000,1.0)	($4000,.50;$2000,.50)
17	($S16$,.50;$2000,.50)	($3000,.50;$2000,.50)	($4000,.25;$2000,.75)
18	($S16$,.10;$2000,.90)	($3000,.10;$2000,.90)	($4000,.05;$2000,.95)
19	$S19$	($3500,1.0)	($4000,.50;$3000,.50)
20	($S19$,.50;$3000,.50)	($3500,.50;$3000,.50)	($4000,.25;$3000,.75)
21	($S19$,.10;$3000,.90)	($3500,.10;$3000,.90)	($4000,.05;$3000,.95)

Note. A total of 82 subjects chose either Option A_i or Option B_i for each scenario i.

atically over an appropriate range for x = $500, $1000, $1500, $2000, $2500, $3000, $3500, with $u(\$0) = 0$ and $u(\$4000) = 1$. To retain a tractable search procedure, each outcome was allowed to take on only seven possible utility values, shown below, with the additional constraint that the resulting combination of utilities had to be monotonically nondecreasing in x. Even with these restrictions, a wide range of risk attitudes, from substantially risk averse to substantially risk prone, can be generated by the various combinations of outcome utilities.

Outcome (x)	Possible Values of $u(x)$						
$500	.01	.05	.10	.15	.20	.30	.40
$1000	.05	.15	.20	.25	.30	.35	.50
$1500	.10	.20	.30	.35	.40	.45	.60
$2000	.25	.40	.45	.50	.55	.60	.75
$2500	.40	.55	.60	.65	.70	.80	.90
$3000	.50	.65	.70	.75	.80	.85	.95
$3500	.60	.70	.80	.85	.90	.95	.99

A total of 24,334 combinations of utility values were thus considered in the search. For each combination, predictions were formulated over the 30 scenarios in the assessment sample. These predictions were then compared with the actual choices made by the subject, and the number of correct predictions tallied. The combinations yielding the largest number of correct predictions represented the best-fitting EU functions for the subject. Multiple best-fitting models were typically identified for each subject by this process, since several utility combinations could yield an identical (maximum) number of correct predictions.

A constrained version of the expected utility model was also fitted by forcing the model to take on the following exponential form:

$$u(x) = \frac{1 - e^{-c\left(\frac{x}{\$4000}\right)}}{1 - e^{-c}}, \text{ if } c \neq 0$$

$$u(x) = \frac{x}{\$4000}, \text{ if } c = 0.$$

For this exponential model, a search was conducted over 1000 possible values for the exponential parameter c ranging from -5 to 5 in increments of .01. Negative (zero, positive) values of c indicate a risk prone (neutral, averse) utility function; therefore, the outer envelope of utility values ranged from a substantially risk prone exponential utility function to a substantially risk averse utility function. Again, predictions were generated for each combination of utility values, and the best-fitting exponential functions identified by comparing the predictions with the subject's actual choices.

5.2. *Choice-Based Assessment of the LDEU Model*

The choice-based assessment approach adopted for the LDEU model is similar to that described for the EU model. The value of $h(x)$ was varied systematically over an appropriate range for x = $0, $500, $1000, $1500, $2000, $2500, $3000, $3500, $4000. With 9 $h(x)$ values allowed to vary, it was necessary to limit each outcome to only six possible $h(x)$ values, as shown below. In addition, only combinations of $h(x)$ that were monotonically nondecreasing in x were allowed, conforming to the notion that the existence of larger outcome values contributes to risk averse choice behavior (e.g., risk averse choices are more likely for scenarios involving a sure $3500 than for scenarios with a sure $500). Consistent with the choice-based EU assessment approach, these restrictions still allowed for a wide range of risk attitudes.

Outcome (x)	Possible Values of $h(x)$					
\$0	−5.00	−3.00	−1.50	−0.75	−0.25	0.25
\$500	−4.00	−2.00	−1.00	−0.50	0.00	0.50
\$1000	−3.00	−1.50	−0.75	−0.25	0.25	0.75
\$1500	−2.00	−1.00	−0.50	0.00	0.50	1.00
\$2000	−1.50	−0.75	−0.25	0.25	0.75	1.50
\$2500	−1.00	−0.50	0.00	0.50	1.00	2.00
\$3000	−0.75	−0.25	0.25	0.75	1.50	3.00
\$3500	−0.50	0.00	0.50	1.00	2.00	4.00
\$4000	−0.25	0.25	0.75	1.50	3.00	5.00

For each combination, predictions were formulated over the 30 scenarios in the assessment sample. These predictions were then compared with the actual choices made by the subject, and the number of correct predictions calculated. The combinations yielding the largest number of correct predictions represented the best-fitting LDEU functions for the subject. Like the choice-based EU models, multiple best-fitting $h(x)$ combinations were typically identified by this process for each subject.

A constrained version of the lottery dependent expected utility model was also fitted by restricting the $h(x)$ function to take on the cubic form proposed in the study by Daniels and Keller (1990), $h(x) = r + s(x - t)^3$, with x in thousands of dollars. This form was selected based on an examination of the $h(x)$ functions that best fit subjects' actual choices. The parameter t can be interpreted as an individual's target or reference level of the outcome variable, expressed in thousands of dollars. The parameter s sets the scale of $h(x)$ over the range \$0 to \$4000 and thus controls the variability of $h(x)$ values. The parameter r then specifies the value of $h(t)$, indicating the risk attitude for a sure outcome of the neutral target amount t. A search was conducted over the following values of the parameters r, s, and t: r ranged from −5 to 5, in increments of 0.5; s ranged from 0 to 0.1, in increments of .005; and t ranged from −2 to 2, in increments of 0.2. Note that by constraining the scale parameter s to be positive, only monotonically nondecreasing values of $h(x)$ are generated.

5.3. *Model Assessment Using Indifference Judgments*

As shown in Table 1, assessment by probability equivalents required subjects to indicate the indifference probability p that satisfies $A \equiv (x, 1.0) \sim B \equiv (\$4000, p; \$0, 1 - p)$ for a given value of x. The utility of outcome x can then be expressed as

$$u(x) = pu(\$4000) + (1 - p)u(\$0) = p,$$

since $u(\$4000) = 1$ and $u(\$0) = 0$.

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

$$u(x,c_A) = \frac{1 - e^{-h(x)\left(\frac{x}{\$4000}\right)}}{1 - e^{-h(x)}} = pu(\$4000,c_B) + (1 - p)u(\$0,c_B) = p,$$

since $u(\$4000, c_B) = 1$, $u(\$0,c_B) = 0$, and $c_A = h(x)$ (assume $h(x) \neq 0$). $h(x)$ is thus set equal to the parameter of the exponential utility function that includes the points $(\$0, u(\$0) = 0)$, $(\$4000, u(\$4000) = 1)$, and $(x, u(x) = p)$. The value of $h(x)$ that satisfies the above 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 = \$500, \$1000, \$1500, \$2000, \$2500, \$3000, and \$3500. These computed values, along with the final two probability indifference judgments in Table 1, can then be used to determine $h(x)$ values for $x = \$4000$ and $x = \$0$.

Assessment by certainty equivalents required subjects to indicate the certain outcome x that satisfies $A \equiv (x, 1.0) \sim B \equiv (\$4000, p; \$0, 1 - p)$ for a given value of p. An expected utility function $u(x)$ was estimated by fixing $u(\$4000) = 1$ and $u(\$0) = 0$ and computing, as above, $u(x) = p$. As shown in Table 1, three indifference questions were required to determine the certain outcomes associated with p = .75, .50, and .25. The best-fitting exponential utility function from the specified indifference judgments over the range $\$0 \leq x \leq \$4000$ was then derived from these judgments (see Keller, 1985 for details of the fitting process).

The elicited certainty equivalents (call them CE_1, CE_2, and CE_3) can also be used to determine a lottery dependent model for the subject:

$$u(CE_1,h(CE_1)) = \frac{1 - e^{-h(CE_1)\left(\frac{CE_1}{\$4000}\right)}}{1 - e^{-h(CE_1)}} = .75u(\$4000,c_B) + .25u(\$0,c_B) = .75$$

$$u(CE_2,h(CE_2)) = \frac{1 - e^{-h(CE_2)\left(\frac{CE_2}{\$4000}\right)}}{1 - e^{-h(CE_2)}} = .50$$

$$u(CE_3,h(CE_3)) = \frac{1 - e^{-h(CE_3)\left(\frac{CE_3}{\$4000}\right)}}{1 - e^{-h(CE_3)}} = .25.$$

Solving these equations for $h(CE_1)$, $h(CE_2)$, and $h(CE_3)$ and assuming a cubic form $h(x) = r + s(x - t)^3$, a best-fitting $h(x)$ function can be derived from the specified indifference judgments. This function can then be used to calculate $h(x)$ values for any $x \in [\$0, \$4000]$.

6. RESULTS

Predictions for each of the 21 scenarios comprising the holdout sample were generated for each of the models described above. These predictions were then compared with subjects' actual choices to determine relative predictive performance. The results are contained in Table 3.

6.1. *Indifference-Based Models*

As shown in Table 3, the EU model assessed by probability equivalents generated correct predictions over 59.93% of the holdout scenarios tested, while the model calibrated from certainty equivalents predicted 59.52% of the holdout choices made by subjects. The LDEU model assessed by probability equivalents correctly predicted 52.44% of the holdout choices, while the model derived from certainty equivalents matched subjects' choices in 57.65% of the holdout scenarios.

6.2. *Choice-Based Models*

For each subject, all the EU (unconstrained and exponential) and LDEU (unconstrained and cubic $h(x)$) models that correctly predicted the largest number of actual choices over the assessment sample of scenarios were retained for model evaluation. For each best-fitting model, predictions over the holdout sample of scenarios were generated and compared with the subject's actual choices. Since the predictive performance of the best-fitting models of a given type over the holdout sample of scenarios can vary, Table 3 provides information on the average, maximum, and

TABLE 3
A COMPARISON OF PREDICTIVE PERFORMANCE

	Assessment sample		Holdout sample		
	Average No. of best-fitting models	% Correct predictions	% Correct predictions	Average maximum % correct	Average minimum % correct
Models assessed by choice judgments					
EU (unconstrained)	153	83.27	66.32	77.18	55.28
EU (exponential)	74	76.59	57.55	59.23	57.14
LDEU (unconstrained)	66	82.89	69.69	78.66	59.49
LDEU (cubic)	348	81.99	68.11	77.83	58.12
Models assessed by indifference judgments					
EU (probability equivalents)			59.93		
EU (certainty equivalents)			59.52		
LDEU (probability equivalents)			52.44		
LDEU (certainty equivalents)			57.65		

minimum number of correct predictions observed among all the best-fitting models of each type, with all values averaged over the 82 subjects.

Table 3 shows that an average of 153 best-fitting EU models were obtained per subject from the unconstrained search process, and that 83.27% of the assessment scenarios were correctly predicted by these best-fitting models. Similarly, an average of 74 best-fitting EU models constrained to an exponential form were found to correctly predict 76.59% of the assessment scenarios. The unconstrained EU models successfully forecast 66.32% of the choices made over the holdout sample of scenarios, with a maximum of 77.18% correct and a minimum of 55.28% correct (averaged over all subjects). Thus, if for each subject one best-fitting unconstrained EU model was randomly selected to predict that subject's holdout choices, we would expect 66.32% of the predictions to be correct, and the predictive performance could be as high as 77.18% or as low as 55.28%. The exponential EU models correctly predicted 57.55% of the actual choices, with a maximum of 59.23% correct and a minimum of 57.14% correct, again averaged over all subjects.

Table 3 also provides results on the LDEU models assessed from choice information. Note that the more general structure of the LDEU model does not guarantee higher prediction rates over either the assessment or holdout samples, since (i) unlike unconstrained expected utility, LDEU models must adhere to an exponential form, and (ii) computational considerations limited the unconstrained searches so that the set of EU models evaluated was not a proper subset of LDEU. For the unconstrained version of the model, an average of 66 best-fitting combinations of $h(x)$ per subject correctly predicted 82.89% of the assessment scenarios, while an average of 348 best-fitting cubic models per subject successfully forecast 81.99% of the assessment scenarios. The unconstrained LDEU model correctly predicted 69.69% of the choices made over the holdout sample of scenarios, with an average maximum of 78.66% correct and an average minimum of 59.49% correct over all subjects. The cubic LDEU model achieved similar performance, predicting 68.11% of the actual choices, with an average maximum of 77.83% and an average minimum of 58.12% over all subjects.

A within-subjects analysis of variance was performed to test for significant differences in predictive performance among the models. For the purpose of this analysis, exactly one set of holdout predictions was generated per subject for each choice-based model. This was accomplished by identifying the modal prediction among all the best-fitting models of a given type for each scenario (e.g., if 100 best-fitting cubic LDEU models were generated for subject 1 and 60 of these predicted preference for option *A* in the first holdout scenario, then the cubic LDEU model would predict that subject 1 would choose option *A* in scenario 1). Note that the set of holdout predictions obtained from this consensus decision rule may not be consistent with any of the individual best-fitting models from which the consensus was formed. As shown in Table 4, better than average predictive performance was observed for three of the four choice-based models using the consensus rule.

The results of the analysis of variance indicate that the percentage of correct predictions varies significantly with subject and model factors. Table 4 also presents a summary of the paired contrasts performed among the eight models, from which several observations can be made. First, the unconstrained LDEU model clearly outperformed all of the other models, predicting a significantly larger percentage of the actual choices made by subjects. In addition, with the exception of the exponential EU model, all of the choice-based models significantly outperformed the models calibrated from indifference judgments. Finally, while the indifference-based EU models tend to outperform their LDEU counterparts (EU_{PE} or EU_{CE} vs. $LDEU_{PE}$), assessment via choice information appears to favor the LDEU model ($LDEU_{unc}$ or $LDEU_{cub}$ vs. EU_{exp}; $LDEU_{unc}$ vs. EU_{unc}).

7. DISCUSSION AND SUMMARY

This paper has explored the potential for a choice-based mechanism to accurately represent preferences in decision making under risk. The results demonstrate that the average predictive performance of both the expected utility and the lottery dependent expected utility models improves when choice data are used to calibrate the models. That the performance of the LDEU model realized a greater improvement through

TABLE 4
SUMMARY OF WITHIN-SUBJECTS ANALYSIS OF VARIANCE

Model (i).	% Correct predictions	Model (j) $LDEU_{unc}$	$LDEU_{cub}$	EU_{unc}	EU_{PE}	EU_{CE}	$LDEU_{CE}$	EU_{exp}	$LDEU_{PE}$
$LDEU^a_{unc}$	76.52		**	**	**	**	**	**	**
$LDEU^a_{cub}$	69.10				**	**	**	**	**
EU^a_{unc}	67.48				**	**	**	**	**
EU_{PE}	59.93								**
EU_{CE}	59.52								**
$LDEU_{CE}$	57.65								*
EU^a_{exp}	57.48								*
$LDEU_{PE}$	52.44								

Note. Entries in the table above indicate that the percentage of correct predictions associated with model *i* is significantly greater than the percentage of correct predictions associated with model *j*. unc, choice-based model derived from unconstrained search; cub, choice-based LDEU model with cubic $h(x)$ function; exp, choice-based EU model constrained to an exponential form; PE, model calibrated from probability equivalents; CE, model calibrated from certainty equivalents.

[a] Predictions for the choice-based models were generated using a consensus decision rule.

*, significance at the 95% confidence level.

**, significance at the 99% confidence level.

choice-based assessment than the EU model suggests that the generalized utility model may be more sensitive to either response mode inconsistencies across model assessment and evaluation or to errors made when providing indifference information.

These results also enhance understanding about the relative variability in predictive performance of the expected utility and lottery dependent expected utility models as assessed in different response modes. In our previous study (Daniels & Keller, 1990), both models were assessed via indifference judgments, leading to predictive performance in the range of 50 to 60% correct, as replicated in this study. Indifference-based EU models tended to exhibit slightly superior performance over their LDEU counterparts in predicting choices over both paradoxical and unrelated scenarios. The results of this study indicate that choice-based assessment may better exploit the more general structure of the LDEU model in capturing subjects' choice behavior.

Choice-based assessment offers several advantages when a utility function is to be used prescriptively. First, a decision maker need only make choices, rather than the arguably more difficult or unreliable indifference judgments which are common in current assessment procedures. Second, assessment by choices is appropriate when a decision situation involves choice. Third, so long as a sufficiently general set of choice questions is used, an analyst need not specify the type of utility model to be used prior to assessment. If the choice judgments appear to conform with expected utility, that model can be used; however, other generalized utility models can possibly be fit to the choice data, without requiring the decision maker to make indifference judgments specific to the model used.

Several potential problems associated with choice-based assessment must also be acknowledged. Choices are themselves subject to framing effects (see, e.g., Tversky & Kahneman, 1981; and Slovic, Fischhoff, & Lichtenstein, 1982) and have been shown to violate the independence of irrelevant alternatives principle (Huber, Payne, & Puto, 1982). In addition, different paired-comparison response modes can lead to inconsistent judgments (Goldstein & Einhorn (1987)). The impact of these problems on the validity and predictive performance of choice-based utility models is an issue that must be considered in determining an appropriate assessment approach.

In measuring the predictive performance of a utility model, care must be taken to distinguish between systematic variability (due to response mode differences, differences in validation questions, etc.) and unsystematic error (e.g., variability in responses to the same questions). By selecting a consistent response mode for assessment and validation, a possible source of systematic error is eliminated. A gap between observed predictive performance and 100% predictive performance will still be observed due to baseline inconsistencies or unsystematic error. Such inconsistencies will not be captured by a non-stochastic model like expected utility without modifications to include random errors. Eliashberg and Hauser (1985) and Laskey and Fischer (1987) address the issues of response errors and models to incorporate random error, and Luce (1959), Luce and Suppes (1965), and Busemeyer (1965) consider the related topic of probabilistic choice models.

Our results indicate that choice-based assessment exhibits potential as an alternative method for calibrating utility functions for individual decision makers. Refinements to the choice-based assessment process require that several research questions be investigated further. When multiple best-fitting models are obtained in the calibration phase of the process, how to select a single model to guide choice remains a key issue. The expected predictive performance of the various utility models if this selection is made randomly is represented by the average percentage of correct predictions in Table 3. Table 3 also suggests that if an intelligent decision rule for identifying a single best-fitting model for prediction can be designed, a further improvement in predictive performance can be realized; alternatively, naive selection can result in poorer than average predictive performance. The consensus decision rule utilized in Table 4 is one approach that combines preference information from all best-fitting models to determine an appropriate set of predictions; other translation rules are possible and merit further investigation.

Computational issues associated with the search process in choice-based assessment also need to be addressed. When the number of monetary outcomes is large, computational tractability dictates a fundamental trade-off between search breadth (the range of utility values allowed in the search) and search intensity (the increment in utility values allowed in the search). Also, while a broad search with small utility increments can capture a wide range of possible utility models, many of these models' predictions over both the assessment and holdout samples of scenarios would be identical. In determining the proper utility range and increment governing the assessment search, care must be taken to generate as many utility models with unique choice characteristics as possible while retaining a manageable search process. This suggests that the parameters of the search must be tailored to the specific set of choice scenarios to be used in model assessment and evaluation.

Further work is also required to determine an appropriate design for the choice scenarios used to calibrate and validate utility models. Assessment samples should consist of scenarios which are both challenging and offer subjects a reasonable opportunity to exhibit a wide range of choice behavior. Interactive design, in which choice scenarios are generated based on previous choices in a manner similar to the choice-based procedures

for estimating parameters of psychological functions (see, e.g., Wasan, 1969; and Levitt, 1970), represents one promising approach. In any design, the costs (e.g., increased effort requirements or computational burden) of gathering additional choice information should be balanced with the associated benefits. Issues related to the consistency of scenarios across assessment and holdout samples (to avoid, e.g., biases incurred when models are calibrated by comparing gambles with sure rewards, but validated by comparing gambles with gambles) must also be considered.

REFERENCES

Barron, H., & Schmidt, C. P. (1988). Sensitivity analysis of additive multiattribute value models. *Operations Research*, **36**, 122–127.

Becker J. (1986). *A new model of decisions under risk using the concept of lottery dependent utility function*. Unpublished doctoral dissertation, UCLA Graduate School of Management.

Becker, J., & Sarin, R. (1987). Lottery dependent utility. *Management Science*, **33**, 1367–1382.

Becker, J., & Sarin, R. (1989). Decision analysis using lottery dependent utility. *Journal of Risk and Uncertainty*, **2**, 105–117.

Bostic, R., Herrnstein, R. J., & Luce, R. D. (1990). The effect on the preference-reversal phenomenon of using choice indifferences. *Journal of Economic Behavior and Organization*, **13**, 193–212.

Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed sample, and sequential-sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **11**, 538–564.

Busemeyer, J. R., & Goldstein, W. M. (1992). Linking together different measures of preference: A dynamic model of matching derived from choice theory. *Organizational Behavior and Human Decision Processes*, **52**, 370–396.

Byrd, J. G., de Neufville, R., & Delquie, P. (1987). *The difference between probability and certainty equivalent methods of utility assessment*. Working paper, Technology and Policy Program, Massachusetts Institute of Technology, Cambridge, MA.

Chu, P., Moskowitz, H., & Wong, R. T. (1988). Robust interactive decision analysis (RID): An Overview. *Acta Psychologica*, **68**, 255–269.

Currim, I., & Sarin, R. (1989). Prospect versus utility. *Management Science*, **35**, 22–41.

Currim, I., & Sarin, R. (1992). Robustness of expected utility model in predicting individual choices. *Organizational Behavior and Human Decision Processes*, **52**, 544–568.

Daniels, R. L., & Keller, L. R. (1990). An experimental evaluation of the descriptive validity of lottery dependent utility theory. *Journal of Risk and Uncertainty*, **3**, 115–134.

Delquie, P., & de Neufville, R. (1988). *Response-modes and inconsistencies in preference assessments*. Working paper, Technology and Policy Program, Massachusetts Institute of Technology, Cambridge, MA.

de Neufville, R., & Delquie, P. (1988). A model of the influence of certainty and probability effects in the measurement of utility. In B. Munier (Ed.), *Risk, decision, and rationality* (pp. 189–205). Dordrecht: Reidel.

Eliashberg, J., & Hauser, J. R. (1985). A measurement error approach for modeling consumer risk performance. *Management Science*, **31**, 1–25.

Goldstein, W. M., & Busemeyer, J. D. (1992). The effect of "irrelevant" variables on decision making: Criterion shifts in preferential choice? *Organizational Behavior and Human Decision Processes*, **52**, 425–454.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preferene reversal phenomena. *Psychological Review*, 94, 236–254.

Grether, D. M., & Plott, C. P. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69, 623–638.

Hershey, J. C., Kunreuther, H., & Schoemaker, P. (1982). Sources of bias in assessment procedures for utility functions. *Management Science*, 28, 936–954.

Hershey, J. C., & Schoemaker, P. (1985). Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Science*, 31, 1213–1231.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding assymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9, 90–98.

Johnson, E., & Schkade, D. (1989). Bias in utility assessments: Further evidence and explanations. *Management Science*, 35, 406–424.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.

Keller, L. R. (1985). An empirical investigation of relative risk aversion. *IEEE Transactions on Systems, Man, and Cybernetics*, 15, 475–482.

Keller, L. R. (1989). The role of generalized utility theories in descriptive, prescriptive, and normative decision analysis. *Information and Decision Technologies*, 15, 259–271.

Laskey, K. B., & Fischer, G. W. (1987). Estimating utility functions in the presence of response error. *Management Science*, 33, 965–980.

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467–476.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89, 46–55.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Luce, R. D., & Suppes, P. (1965). Preference, utility, and risk. In R. D. Luce, R. R. Bush, & R. E. Galanter (Eds.), *Handbook of mathematical psychology*. New York: Wiley.

McCord, M., & de Neufville, R. (1984). Utility dependence on probability: An empirical demonstration. *Large-Scale Systems*, 6, 91–103.

McCord, M., & de Neufville, R. (1986). Lottery equivalents: Reduction of the certainty effect in utility assessment. *Management Science*, 32, 56–60.

Mellers, B. A., Ordóñez, L. D., & Birnhaum, M. H. (1992). A change of process theory for contextual effects and preference reversals in risky decision making. *Organizational Behavior and Human Decision Processes*, 52, 331–369.

Moskowitz, H., Wong, R. T., & Chu, P. (1989). Robust interactive decision analysis (RID): Behavioral results and implications. *Journal of Behavioral Decision Making*, 2, 95–111.

Nau, R. F. (1990). *Indeterminate probabilities and utilities on finite sets*. Fuqua School of Business Working Paper 8929, Duke University.

Schoemaker, P. J. H., & Hershey, J. C. (1992). Utility measurement: Signal, noise, and bias. *Organizational Behavior and Human Decision Processes* 52, 397–424.

Schoemaker, P. J. H., & Waid, C. (1982). An experimental comparison of different approaches to determining weights in additive utility models. *Management Science*, 28, 182–196.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Response mode, framing, and information processing effects in risk assessment. In R. Hogarth (Ed.), *New directions for methodology of social and behavioral science: Question framing and response consistency*. San Francisco: Jossey-Bass.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, 95, 371–384.

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *American Economic Review*. 80, 204–217.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and linear programming*. New York: Wiley.

Wasan, M. T. (1969). *Cambridge tracts in mathematics and mathematical physics (58): Stochastic approximation*. Cambridge: Cambridge Univ. Press.

Weber, M. (1985). Method of multi-attribute decision-making with incomplete information. *Management Science*, 31, 1365–1371.

Weber, M. (1987). Decision making with incomplete information. *European Journal of Operational Research*, 28, 44–57.

RECEIVED: June 18, 1991