1988

A Comparison of Four Multiattribute Utility Function Elicitation Procedures for Preference Predictions.

Phillip C. Fry
Louisiana State University and Agricultural & Mechanical College

Order Number 8904537

A comparison of four multiattribute utility function elicitation procedures for preference predictions

Fry, Phillip C., Ph.D.

The Louisiana State University and Agricultural and Mechanical Col., 1988

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A COMPARISON OF FOUR MULTIATTRIBUTE UTILITY FUNCTION ELICITATION PROCEDURES FOR PREFERENCE PREDICTIONS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Interdepartmental Program in Business Administration

by
Phillip C. Fry
B.A., The University of Arkansas, 1975
M.B.A., The University of Arkansas, 1977
M.S., Louisiana State University, 1983
August 1988

This dissertation is dedicated to my family and friends.

ACKNOWLEDGEMENTS

I would like, first of all, to express my appreciation to my major professor, Dr. Dan B. Rinks, whose suggestions and guidance provided invaluable assistance and encouragement during the dissertation.

The members of my doctoral committee, Dr. Stephen W. Looney, Dr. Michael H. Peters, Dr. Kwei Tang, Dr. Robert E. Martin, and Dr. Gary Schaffer also offered helpful guidance and comments. Special thanks is extended to Dr. Robert E. Martin and Dr. Stephen W. Looney for their willingness to serve as committee members during their sabbaticals. I would also like to thank Dr. Michael H. Peters, Chairman of the Department of Quantative Business Analysis, for providing me with a teaching assistantship during my graduate study.

Although not a member of my doctoral committee, a special thanks is owed to Dr. Jeffrey L. Ringuest who first introduced me to decision analysis and provided considerable insight into my research. In addition, Dr. Ringuest provided direction and encouragement during my graduate work.

Finally, I would like to thank my loving wife, Susan, for her infinite patience, sacrifice, and support during my graduate study. Without her assistance my dream would never have been realized.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

LIST OF NOTATION

MAUA - multiattribute utility analysis

MAUF - multiattribute utility function

$u_i(x_i)$ - the single attribute utility function for attribute i

$f[u_1(x_1),u_2(x_2),\ldots,u_n(x_n)]$ - the functional form

$U(\underline{x})$ - the MAUF normalized by $U(\underline{x}^*)=1$ and $U(\underline{x}^o)=0$

$\underline{x}$ - multidimensional consequence, $\underline{x} = (x_1,x_2,\ldots,x_n)$

x_i - attribute i in $\underline{x}$

x_i^* - best level of attribute i in $\underline{x}$

x_i^o - worst level of attribute i in $\underline{x}$

$u_i(\cdot)$ - single attribute utility function normalized by $u_i(x_i^*)=1$ and $u_i(x_i^o)=0$

K, k_i, k_{ij}, k_{ijl} - scaling parameters

$(x_1^o,x_2^o,\ldots,x_{i-1}^o,x_i^*,x_{i+1}^o,\ldots,x_n^o)$ - the consequence having all attributes at their worst levels except attribute x_i which is at its best level

KR - Keeney-Raiffa procedure

HOPE - holistic orthogonal parameter estimation procedure

SMART - simple multiattribute rating technique

KR4M - underlying four attribute multiplicative utility function

KR6M - underlying six attribute multiplicative utility function

KR4A - underlying four attribute additive utility function

KR6A - underlying six attribute additive utility function

PUL - proportion of utility loss

SD - standard deviation

ABSTRACT

Multiattribute utility analysis is a set of formal procedures designed to assist a decision maker in resolving problems requiring tradeoffs among competing objectives. The approach elicits a multiattribute utility function (MAUF) intended to represent the decision maker's preferences for multivariate alternatives. While different MAUF assessment technologies are available, few studies have been conducted that compare assessment techniques. The purpose of this research was to investigate the predictive ability of four MAUF elicitation procedures.

Keeney-Raiffa (KR), holistic orthogonal parameter estimation (HOPE), simple multiattribute rating technique (SMART), and mathematical programming elicited MAUFs were compared in terms of 1) ordinal preference predictions, 2) first preference predictions, and 3) ability to capture tradeoffs between competing objectives. Since actual decision tasks do not offer a "correct" solution, it is difficult to compare the performances of alternative assessment procedures in a real-world setting. Consequently, a hypothetical decision making environment was established to determine a procedure's performance in the presence of elicitation errors of known type and magnitude. The inputs necessary to implement each procedure were provided by an artificial decision maker. While behavioral influences on a technique's performance were not considered, each experimental elicitation contained one or more of the following assessment errors: noisy respondent inputs; incorrectly specified functional forms;

and incompletely defined attribute sets. Validity was determined by comparing MAUF predictions to the artificial subject's known preference constructs.

The four procedures were not uniformly robust within or across the elicitation scenarios considered. KR encoded models typically provided the most consistently accurate predictions. HOPE first preference predictions were sensitive to the level of noise contained in the subject's responses. SMART and mathematical programming performances were generally inferior to those of KR and HOPE. SMART offered reasonable overall preference orderings, but was less effective at predicting the subject's most desired alternative. Predictions offered by mathematical programming MAUFs were relatively unstable, at times yielding high proportions of utility loss from incorrect first preferences. This research provides an initial set of guidelines to assist a decision analyst in choosing among competing MAUF elicitation techniques.

CHAPTER 1 INTRODUCTION

1.1 Introduction and Background

Decision making requires that alternative courses of action be evaluated so that the strategy which yields the most preferred outcome can be identified. For simple decision tasks, this generally requires that the decision maker optimize with respect to a single objective. For highly repetitive problems having only minor consequences, acceptable actions may be suggested by informal intuitive evaluations.

More important decisions, however, can involve several objectives. When these objectives are commensurate, decision consequences can be expressed by one variable (e.g., money). In such situations a single standard of performance such as maximum profit or minimum cost can identify an appropriate decision strategy. However, when objectives are incommensurate, and decision consequences cannot be easily modeled by a single attribute, the decision problem is better formulated using multiple criteria. For these, alternatives of choice are characterized by vectors of quantifiable attributes describing the level of each objective. At times, these objectives conflict with one another and improvements in one objective can occur only at the expense of another. Such decision making is inherently complex because tradeoffs must be made among objectives when evaluating choice alternatives. For example, a bank officer's commercial loan decisions must consider tradeoffs among projected returns, payback periods, individual loan risk characteristics, overall loan

portfolio characteristics, and banking regulations. Problems of this nature are often encountered by managerial and governmental decision makers (see Hwang and Yoon 1981, Zeleny 1982, von Winterfeldt and Edwards 1986).

While complex decisions often involve significant consequences, some decision makers may be tempted to base a choice solely on unaided innate evaluations. Unfortunately, intuitive judgments can produce unsatisfactory outcomes. Bowman (1963) revealed how selective cueing and information overload could cause unaided intuitive deliberations to yield inferior choices. A decade of research led Slovic (1972) to conclude "... that humans are quite bad at making complex unaided decisions."

To overcome the shortcomings of unaided intuitive evaluations, formal decision aids have been developed. One such decision aid is multiattribute utility analysis (MAUA). The primary purpose of MAUA is to derive a model which reveals a rational decision maker's preferences for contemplated choices. This is accomplished by eliciting the individual subject's multiattribute utility function (MAUF). The MAUF is a mathematical model which assigns a real number (called a utility value) to multivariate decision outcomes, so that the utility for consequence i is greater than the utility for consequence j, if and only if consequence i is preferred to consequence j. The assessed utility values therefore represent the final product of the decision maker's internal preference deliberations.

MAUA is not the only methodology available for modeling individual preferences. Such techniques as regression analysis, conjoint

analysis, and logit analysis have also been used for this purpose. Unlike these alternative approaches, however, MAUA can offer specific advantages to preference modeling. First, MAUA directly embodies individual decision maker risk attitudes in its axiomatic development. This enables explicit incorporation of individual risk preferences. Other techniques are limited to a single variable to capture risk considerations. Second, the functional form of the MAUF can be directly established from a set of verifiable assumptions. Finally, in addition to capturing the decision maker's underlying preferences, MAUF scaling constants depict the decision maker's willingness to make tradeoffs between competing objectives (Hauser and Urban 1979).

These advantages, combined with a solid theoretical foundation (Keeney and Raiffa; Farquhar 1977), and evidence of successful application (Keeney 1973, 1977; Keeney and Wood 1977; Keeney and Sicherman 1983; Keeney, Lathrop, and Sicherman 1986), reveal the potential of formal decision analysis for resolving multiattribute decision tasks. Unfortunately, MAUA has not yet gained widespread acceptance as a practical decision making tool. One explanation for its limited acceptance is the difficulty often encountered in discovering the decision maker's MAUF. The large information requirements and the intricacies of the concepts and constructs involved in eliciting the MAUF can prove overwhelming for any but the most adept decision maker. Consequently, the assistance of a skilled decision analyst is typically enlisted to structure and administer the MAUF elicitation process.

One responsibility of the analyst is to select a procedure for encoding the individual decision maker's MAUF. The assessment technology chosen will determine the general approach of the preference modeling exercise. Although several procedures are available, every technique employs either a decomposition or holistic assessment framework.

Instead of assessing a decision maker's MAUF directly, decomposition procedures first separate (decompose) multiattribute decision tasks into their individual components. A single attribute utility function is then encoded for each decision criterion independently of other attributes. The single attribute functions, together with an elicited set of scaling parameters, are finally aggregated to form the MAUF. In other words,

$$U(x_1,x_2,\ldots,x_n) = f[u_1(x_1),u_2(x_2),\ldots,u_n(x_n)]$$

where, $U(x_1,x_2,\ldots,x_n)$ is the MAUF, $u_i(x_i)$ is the utility function for attribute i, and $f[\cdot]$ is the functional form used to aggregate the n univariate utility functions (where n is the total number of decision objectives involved) (Keeney and Raiffa 1976, Fischer 1979, Zeleny 1892). The appropriate functional form is generally determined by verifying a set of utility independence conditions which describe the subject's underlying conditional preferences for the attributes comprising the decision task.

Alternatively, the decision analyst can employ a holistic elicitation procedure. Such techniques require the decision maker to provide direct, global evaluations of decision outcomes by assigning a utility value to each consequence. In general, no formal check is

conducted to directly verify a specific utility functional form. Rather, the analyst employs a general utility form not dependent on a set of independence conditions, or a simplified function believed to offer a reasonable approximation to the decision maker's internal preference configuration.

Having decided on a general approach to utility assessment, the analyst must then select a specific elicitation technique. The selection process is complicated by the fact that the procedures within each category vary in terms of their operational complexity. For instance, those techniques most closely linked to the theory underlying MAUA can be time-consuming and cognitively demanding to implement. Conversely, techniques which employ simplifying assumptions to facilitate their implementation are less theoretically exact. While some decision makers are hesitant to use a complicated procedure regardless of its advantages (Einhorn and McCoach 1977), others prefer a more difficult elicitation procedure (Hobbs 1980). In any event, the technique selected must be structurally capable of encoding valid preference models.

1.2 Need For A Comparative Study of MAUF Elicitation Procedures

The analyst's choice of an MAUF elicitation procedure is a critical step in the application of MAUA. When deciding among assessment procedures, the analyst must consider the technique's suitability for a decision context, its ease of use, its validity, and the sensitivity of preference evaluations to the technique employed (Hobbs 1986).

To be meaningful in an applied decision context, the assessed utility model must order decision outcomes in accordance with the decision maker's internal preference structure. However, because human decision makers are imperfect information processors, the responses they provide during the elicitation exercise will be "noisy". As a result, the encoded MAUF will not be error free. Barron (1983) described the sources of error that can potentially arise during the elicitation process. Given these assessment errors, it becomes imperative that the analyst select an MAUF encoding procedure which is structurally capable of modeling a robust utility function. The selection process would be facilitated by a rigorous study indicating the relative preference prediction capabilities of alternatively encoded MAUFs. Given the importance of choosing a suitable assessment procedure, it is surprising that little research has been conducted to determine the relative performance of elicitation procedures under different conditions.

Of the studies available, most have examined the effects of various weighting schemes and model forms on preference predictions. While these works are informative, they do not systematically compare the sensitivity of an MAUF's revealed preferences to the elicitation procedure used. Such information would allow the analyst to tailor the selection of an elicitation procedure to a specific decision task. For example, if the procedure used does not affect the quality of the assessed MAUF, then the analyst can emphasize ease of implementation when choosing an elicitation technique. However, if mean-

ingful preference predictions are technique dependent, then validity, rather than ease of use, becomes the primary choice criterion.

The works of Fisher (1977), Barron and Person (1979), Barron (1980), Eils and John (1980), and Farmer (1987) are representative of the small, but growing, body of research comparing the preference predictions of alternatively encoded models. However, the decision contexts examined in these studies were restricted by the number of attributes, the functional forms, and the elicitation procedures considered. Furthermore, the incomplete and contradictory conclusions offered by these studies provide little direction to an analyst deliberating over the choice of an assessment technology.

Research examining an assessment technique's structural ability to encode reliable MAUFs over different elicitation scenarios would represent an important step in advancing the application of MAUA to complex decision making. The results would extend existing research by establishing a set of guidelines to assist a potential user in selecting an appropriate MAUF elicitation procedure.

1.3 Research Objective and Plan

This research is concerned with the structural abilities of four MAUF elicitation procedures to capture an underlying preference structure. To determine, in a general way, each procedure's sensitivity to elicitation errors of known type and magnitude, an experimental decision making environment was created. This approach neutralized behavioral influences and provided a criterion for quantifying a procedure's performance. From the results a set of guide-

lines was established to assist an analyst in selecting an elicitation methodology.

The general plan of the study was as follows. Two MAUFs, assessed by a procedure widely regarded as being theoretically valid, were selected from the applied decision analysis literature, and modified to produce two additional composite model forms. These four MAUFs then became an experimental decision maker's "true" preference structures for different decision contexts. Experimental decision maker and decision analyst roles were assumed by a computer and the researcher, respectively. The artificial decision maker simulated the inputs needed to implement each assessment technology. The degree of conformity between the elicited models' preference predictions and the decision maker's "true" internal choices were then evaluated.

Two potential sources of assessment error were included: random error and model misspecification error. Random error arises from the information processing limitations of human decision makers. These limitations can generate imprecise responses which can cause noisy MAUFs to be elicited (Barron 1983). To simulate imprecise respondent inputs, the synthetic subject's "true" preferences were perturbed by a randomly generated error component. The structural ability of each elicitation procedure to convert these noisy inputs into reliable MAUFs was then investigated.

Two types of model misspecification error exist. One results when the composite form of the assessed model is not a correct representation of the decision maker's internal preference structure

(Barron 1983). The structural inability of an encoding procedure to capture the underlying functional form, analyst misjudgment, and noisy subject inputs can all produce incorrectly specified models. The second occurs when relevant attributes are excluded from the elicitation exercise. An attribute set is incompletely specified when the decision maker cannot articulate all attributes describing his internal evaluation space, or when suitable attributes cannot be found to depict a decision objective. Throughout the experiment, misspecified models were elicited and their preference predictions compared to the synthetic subjects' "true" evaluations. By comparing preference predictions within and among techniques for different elicitation scenarios, this analysis extends the works of Fischer (1977), Barron and Person (1979), Barron (1980, 1987), and Farmer (1987).

1.4 Outline of the Dissertation

Although several composite forms for aggregating single attribute utility functions exist, in practice consideration is generally given to the multilinear, multiplicative, and additive representations (Fischer 1979). Chapter 2 discusses these three functional forms and the utility independence conditions necessary to establish their existence. Chapter 3 contains an overview of several elicitation procedures, as well as a detailed discussion of the specific assessment techniques examined. Chapter 4 summarizes the literature relevant to this research. A detailed discussion of the research methodology is given in Chapter 5. The results of the study are

presented in Chapter 6. Finally, Chapter 7 summarizes the research findings and presents guidelines for selecting an elicitation methodology.

CHAPTER 2 MULTIATTRIBUTE UTILITY FUNCTION FORMS AND INDEPENDENCE CONDITIONS

2.1 Overview.

The decomposition procedures systematically combine single attribute utility functions into a composite model to express a decision maker's preferences for multidimensional outcomes. For convenience, some assessment procedures (e.g., Edwards' simple multiattribute rating technique) apply the same rule to every decision task for aggregating conditional utility functions. Other procedures require the analyst to determine the MAUF functional form by directly verifying specific attribute independence properties. Fortunately, a substantial amount of theoretical research establishing the relationship between MAUF composite forms and independence properties has been conducted (Keeney and Raiffa 1976; Fishburn 1970; and Farquhar 1977).

While several MAUF aggregate expressions exist, the multilinear, multiplicative, and additive composite models have received the most attention in practical applications. These aggregate forms (presented in Table 2-1 for those cases comprised of three attributes) have proven to be quite robust when applied to a variety of multicriteria decision making scenarios (Keeney and Raiffa 1976). A mathematical account of the relationship among the underlying independence properties and these functional forms is presented in Keeney and Raiffa (1976), French (1986), and von Winterfeldt and

Edwards (1986). A nontechnical summary of these results is presented in Sections 2.2, 2.3, and 2.4.

Table 2-1

Model Forms for Aggregating Single Attribute Utility Functions (n=3)

Model Form	Expression
Additive	$U(\underline{x}) = k_1u_1(x_1) + k_2u_2(x_2) + k_3u_3(x_3)$
Multiplicative	$1+U(\underline{x}) = (1+Kk_1u_1(x_1))(1+Kk_2u_2(x_2))(1+Kk_3u_3(x_3))$
Multilinear	$U(\underline{x}) = k_1u_1(x_1) + k_2u_2(x_2) + k_3u_3(x_3) + k_{12}u_1(x_1)u_2(x_2) + k_{13}u_1(x_1)u_3(x_3) + k_{23}u_2(x_2)u_3(x_3) + k_{123}u_1(x_1)u_2(x_2)u_3(x_3)$

notation:

- $U(\cdot)$ overall utility function normalized by $U(\underline{x}^*)=1$ and $U(\underline{x}^o)=0$
- $\underline{x}$ multidimensional consequence, $\underline{x} = (x_1,x_2,x_3)$
- x_i attribute i in $\underline{x}$
- x_i^* best level of attribute i in $\underline{x}$
- x_i^o worst level of attribute i in $\underline{x}$
- $u_i(\cdot)$ single attribute utility function normalized by $u_i(x_i^*)=1$ and $u_i(x_i^o)=0$
- K, k_i, k_{ij}, k_{ijl} are scaling parameters

2.2 The Multilinear Model and Utility Independence

The multilinear MAUF model is the most general of the three composite expressions. Its mathematical representation is as follows:

$$\begin{aligned} U(\underline{x}) &= k_1u_1(x_1) + \ldots + k_nu_n(x_n) \\ &\quad + \Sigma\Sigma k_{ij}u_i(x_i)u_j(x_j) + \Sigma\Sigma\Sigma k_{ijl}u_i(x_i)u_j(x_j)u_l(x_l) \\ &\quad + \ldots + k_{123\ldots n}u_1(x_1)u_2(x_2)\ldots u_n(x_n) \qquad (1) \end{aligned}$$

where the sum of all scaling parameters equals one.

The multilinear functional form requires verification of the least restrictive of the independence properties, namely utility independence. A decision attribute x_i is defined to be utility independent of its complement $(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$ if the conditional preference order for lotteries involving x_i does not depend on the levels at which the complementary attributes are fixed. In other words, preferences for lotteries on a single attribute must be independent of fixed identical levels in the other attributes. The multilinear form is appropriate if and only if each attribute is utility independent of its complement.

As an example of this property, consider the task of evaluating new automobiles on three attributes: price, horsepower, and miles per gallon. Further, let the decision maker indicate a preference for the lottery on the left to the lottery on the right

($7500,90hp,30mpg)   ($9000,90hp,30mpg)
p   p
             $>_p$
1-p   1-p
($8000,90hp,30mpg)   ($7000,90hp,30mpg)

where $>_p$ is read "is preferred to". The attribute price is utility independent of the attributes horsepower and miles per gallon if and only if the following preference exists for all changes Δ_2 and Δ_3:

$$\left\langle \begin{array}{l} (\$7500, (90+\Delta_2)\text{hp}, (30+\Delta_3)\text{mpg}) \\ p \\ 1-p \\ (\$8000, (90+\Delta_2)\text{hp}, (30+\Delta_3)\text{mpg}) \end{array} \right. \quad >_p \quad \left\langle \begin{array}{l} (\$9000, (90+\Delta_2)\text{hp}, (30+\Delta_3)\text{mpg}) \\ p \\ 1-p \\ (\$7000, (90+\Delta_2)\text{hp}, (30+\Delta_3)\text{mpg}) \end{array} \right.$$

Thus, preferences over lotteries in one attribute must be independent of fixed common values in other attributes.

The multilinear model requires that $2^n - 1$ scaling parameters be assessed. Even when the number of decision criteria is small, this produces a large number of parameter elicitations. Therefore, while the multilinear model is the most general of the functional representations, its practical application is restricted to decisions having fewer than four attributes (Keeney and Raiffa 1976).

2.3 The Multiplicative Model and Multiplicative Utility Independence

A special case of the multilinear model is the multiplicative functional form:

$$1+KU(\underline{x}) = (1+Kk_1u_1(x_1))(1+Kk_2u_2(x_2))\ldots(1+Kk_nu_n(x_n)) \qquad (2)$$

where K is a parameter that satisfies

$$(1+K) = (1+Kk_1)(1+Kk_2)\ldots(1+Kk_n) \qquad (3)$$

This form requires that decision attributes be mutually, or multiplicatively, utility independent. The attributes $x_1, x_2, \ldots, x_n$ are said to be mutually utility independent if every subset of $\{x_1,$

$x_2,\ldots,x_n\}$ is utility independent of its complement. In other words, preferences among lotteries with uncertain attribute levels should be independent of identical fixed levels in the other attributes.

For example, suppose that the decision maker in the automobile decision task prefers the lottery on the left to the lottery on the right:

$$\left\langle \begin{array}{l} (\$7500,90hp,30mpg) \\ p \\ 1-p \\ (\$8000,100hp,30mpg) \end{array} \right. >_p \left\langle \begin{array}{l} (\$9000,110hp,30mpg) \\ p \\ 1-p \\ (\$7000,75hp,30mpg) \end{array} \right.$$

The attributes selling price and horsepower are mutually utility independent of the attribute miles per gallon if and only if the following preference holds for all values of Δ_3.

$$\left\langle \begin{array}{l} (\$7500,90hp, \\ \quad (30+\Delta_3)mpg) \\ p \\ 1-p \\ (\$8000,100hp, \\ \quad (30+\Delta_3)mpg) \end{array} \right. >_p \left\langle \begin{array}{l} (\$9000,110hp, \\ \quad (30+\Delta_3)mpg) \\ p \\ 1-p \\ (\$7000,75hp, \\ \quad (30+\Delta_3)mpg) \end{array} \right.$$

Unlike the multilinear model, the multiplicative form has only one interactive parameter, K, which determines the scaling of all interaction terms. Thus, when mutual utility independence is verified only $n + 1$ scaling constants must be assessed. The multiplicative and multilinear model forms are identical when $n = 2$.

2.4 The Additive Model and Additive Independence

The additive aggregate model is a special case of both the multilinear and multiplicative functional forms and is expressed as:

$$U(\underline{x}) = k_1 u_1(x_1) + \ldots + k_n u_n(x_n) \tag{4}$$

where the sum of the scaling constants (k_i's) equals one.

A prerequisite to the additive utility model is additive utility independence. Attributes $x_1, x_2, \ldots, x_n$ are defined as additive utility independent if preferences for lotteries on $x_1, x_2, \ldots, x_n$ depend on their marginal probability distributions only and not on their joint probability distributions. This requires that the decision maker be indifferent between two lotteries having equal marginal (single attribute) probability distributions.

Additive utility independence can be illustrated using the automobile example and the following lotteries:

$$\left\langle \begin{array}{l} (\$7500, 90hp, 30mpg) \\ 0.5 \\ 0.5 \\ (\$9000, 100hp, 30mpg) \end{array} \right. \sim \left\langle \begin{array}{l} (\$7500, 100hp, 30mpg) \\ 0.5 \\ 0.5 \\ (\$9000, 90hp, 30mpg) \end{array} \right.$$

where $\sim$ indicates "is indifferent to". Since the additive model requires that n single attribute utility functions and only n-1 independent scaling constants be assessed, its primary advantage relative to the other model forms is its simplicity.

2.5 Summary

This chapter described the three most common MAUF functional expressions. The independence properties underlying these composite models were also presented. Chapter 3 provides an overview of various MAUF assessment techniques and a detailed description of the procedures examined in this study.

CHAPTER 3 PROCEDURES FOR ASSESSING MULTIATTRIBUTE UTILITY FUNCTIONS

3.1 Overview

Several MAUF elicitation procedures are available for analyzing real-world complex decision problems. Among these are the assessment technologies of Keeney and Raiffa (1976); Klein, Moskowitz, Mahesh, and Ravindran (1985); Barron and Person (1979); Briskin (1973); Edwards (1977); Einhorn and McCoach (1977); and Amoli and Ciampi (1983). While these procedures employ either a holistic or decomposed assessment framework, each differs in its analytical approach to modeling the MAUF. The salient characteristics of these procedures are summarized in Table 3-1.

The Keeney-Raiffa procedure decomposes the overall decision task into its individual attributes. Standard decision analytic techniques are used to encode a conditional utility function for each decision criterion. The MAUF scaling parameters are derived from decision maker indifference evaluations of hypothetical multiattributed consequences. The single attribute utility functions and the scaling parameters are combined to form the composite model whose functional form is determined by directly verifying specific utility independence properties. Consistency checks are conducted throughout the assessment exercise to ensure a reliable MAUF is obtained. In spite of the large investment of time and effort required to implement the

TABLE 3-1

SALIENT CHARACTERISTICS OF ALTERNATIVE MAUF ELICITATION PROCEDURES

	Keeney-Raiffa	Mathematical Programming	Differential Equations	HOPE	SEE	SMART	SMAUP
General Approach	Decomposed framework using standard decision analytic techniques.	Decomposed framework using mathematical programming techniques.	Uses Differential Equations to model tradeoffs among attributes.	Holistic evaluations of hypothetical orthogonal profiles.	Nonlinear least squares procedure using holistic evaluations. Procedure builds on HOPE.	Decomposed framework based on simplifying assumptions.	Decomposed framework based on simplifying assumptions similar to SMART.
Assessment of Unidimensional Utility Functions	Assessed for each attribute using decision maker lottery-based evaluations.	Nonlinear programming model incorporating qualitative and quantitative responses.	Computes the utility of each alternative.	Not directly assessed. Utility values for attribute levels inferred.	Unidimensional utility functions are not directly assessed.	Assumes a linear function is a reasonable approximation.	Fits a linear function by standardizing attribute values.
Estimation of Scaling Parameters	Derived from DM indifference evaluations.	Solves goal programming problem.	Solves a set of differential equations.	Inferred from subject's holistic evaluations.	Estimated using nonlinear least squares.	DM ranks and rates attributes. Normalized ratings serve as attribute weights.	Standardized ratings or standardized rankings.
Determination of MAUF Functional Form	Determined through direct verification of independence assumptions.	Determined through direct verification of independence assumptions.	Specified by established differential equations.	Complimentary profile responses indicates either multiplicative or additive form.	Statistical test determines either multiplicative or additive form.	Assumes additive model is a good approximation to underlying form.	Weighted linear function used.
Procedure's Advantages	General methodology with close ties to underlying theory.	Reduces decision maker inputs and employs a general univariate utility equation.	General procedure which can model a variety of decision problems.	Ease of administration and computation.	Statistical test to determine if K = 0.	Simple to implement.	Simple to implement.
Procedure's Disadvantages	Arduous and time consuming to implement.	Nonlinear programming problems may be hard to solve.	Conceptually difficult. Inseparable equations can be hard to solve.	Excludes the general multilinear model and considers only orthogonal profiles.	Increased computational burden over HOPE since orthogonality of design is destroyed.	Method's performance depends on extent to which simplifying assumptions hold.	Method's performance depends on extent to which simplifying assumptions hold.

Keeney-Raiffa methodology, its application has dominated the applied complex decision making literature.

Klein, Moskowitz, Mahesh, and Ravindran (1985) solve a series of mathematical programming problems to estimate both the conditional utility functions and the MAUF scaling parameters. Decision maker certainty equivalent responses are used to formulate a nonlinear programming problem whose solution provides a univariate utility expression. Scaling parameters for the composite function are determined from a goal programming problem formulated from decision maker expressed preferences for multiattributed consequences. By reducing the number of decision maker inputs, and by fitting a general univariate utility equation, the procedure overcomes specific difficulties associated with the Keeney-Raiffa technology.

Briskin (1973) solves a set of partial differential equations to estimate a general multidimensional utility function. The tradeoffs among conflicting objectives for each decision alternative are expressed in terms of attribute x_1. This produces a set of differential equations whose solutions yield a general multivariate utility model. While no utility independence assumptions are made, the decision analyst must explain the meaning of the various model forms to the decision maker. The procedure can be computationally overwhelming, especially when the formulated differential equations are inseparable.

The holistic orthogonal parameter estimation (HOPE) technique of Barron and Person (1979), and the simultaneous estimation of everything (SEE) approach of Amoli and Ciampi (1983) represent holis-

tic MAUF assessment procedures. HOPE models are derived from overall decision maker judgments of consequence profiles defined by an orthogonal experimental design. SEE uses a nonlinear least squares procedure to convert a nonorthogonal set of profile evaluations into an MAUF. Unlike HOPE, SEE employs a statistical test to determine the aggregate expression's functional representation.

Both Edward's (1977) simple multiattribute rating technique (SMART), and Einhorn and McCoach's (1977) simple multiattribute utility procedure (SMAUP) impose assumptions to facilitate the elicitation exercise. The two methods are based on linear univariate utility expressions and additive aggregate models. The reliability of the assessed MAUFs depend on the extent to which the two simplifying assumptions can be justified.

This dissertation is concerned with the structural abilities of competing MAUF elicitation procedures to provide meaningful preference predictions. The methodologies of Briskin, Amoli and Ciampi, and Einhorn and McCoach were (for specific reasons) not considered. Because of its complexity, decision makers are reluctant to employ the differential equations approach. Furthermore, this technique has received only limited mention in the applied decision analysis literature. In terms of general structure, underlying assumptions, and implementation, SMAUP and SEE are closely related to SMART and HOPE, respectively. Consequently, an examination of all four procedures would not likely provide additional insight into the structural capability of any pair to elicit valid preference predictions. For these reasons, the optimal strategy called for an investigation of

the Keeney-Raiffa, HOPE, SMART, and mathematical programming MAUF elicitation technologies. These four techniques are discussed in greater detail in Sections 3.2, 3.3, 3.4, and 3.5.

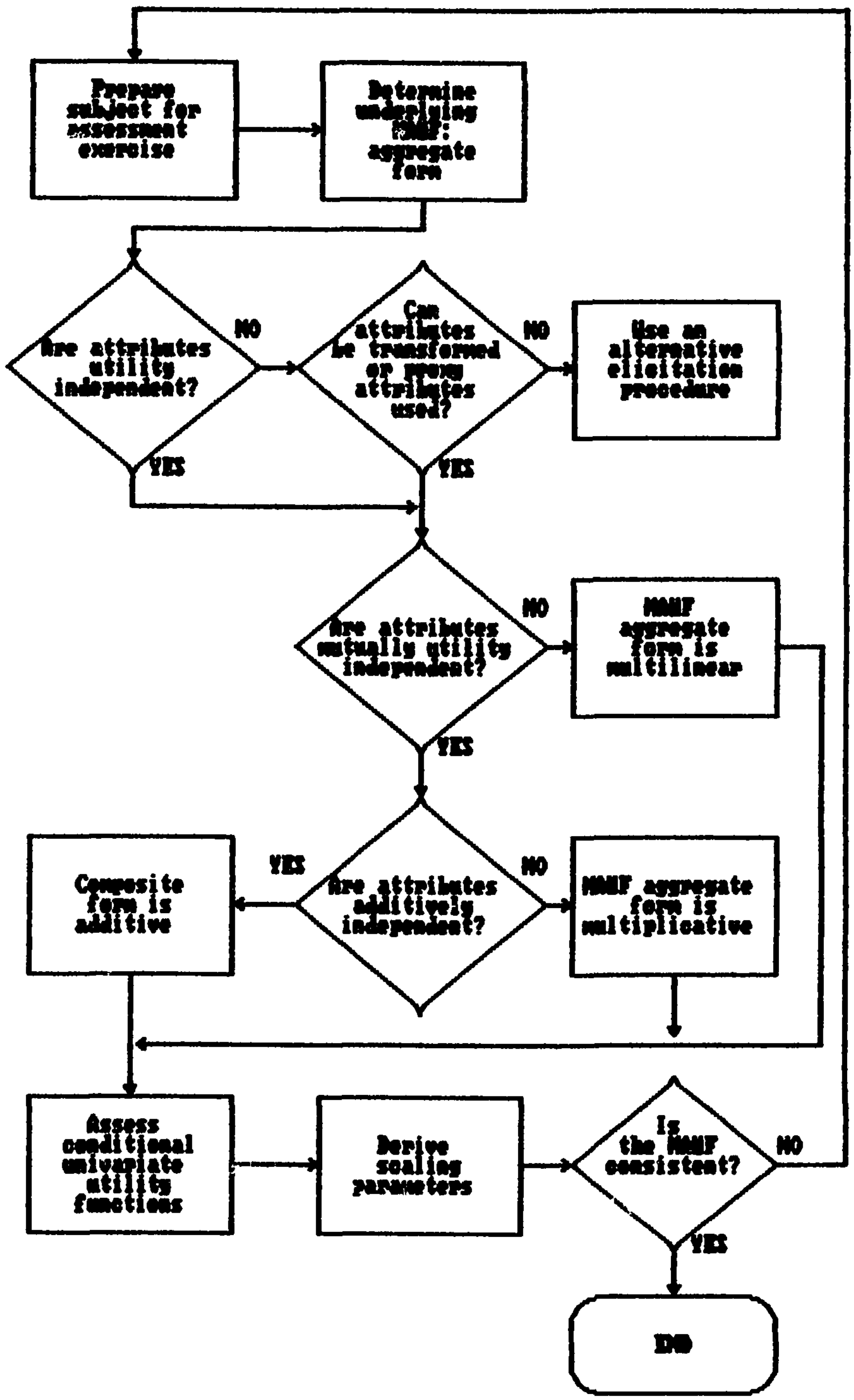
3.2 The Keeney-Raiffa Elicitation Procedure

Keeney (1972, 1977, 1980) and Keeney and Raiffa (1976) have developed a general five step approach for encoding a decision maker's MAUF. The Keeney-Raiffa (KR) methodology decomposes multi-valued decision consequences and uses standard decision analytic techniques to model a multilinear, a multiplicative, or an additive composite preference expression. The basic components of the KR procedure are depicted in Figure 3-1.

First, the decision analyst prepares the subject for the assessment exercise. The framework of the methodology is introduced. A general discussion of the decision context is conducted to define the decision problem and the objectives to be measured. Relevant decision attributes, their ranges, and the direction of increasing attribute preferences are identified. Finally, pairwise preference evaluations are performed to verify the decision maker's understanding of the decision consequence space.

The second step determines the functional form of the MAUF by directly verifying relevant independence properties. Multiattributed lotteries, designed to determine the decision maker's underlying conditional preferences for attributes comprising the decision task, are presented to the decision maker for evaluation. The decision

Figure 3-1 Keeney-Raiffa Elicitation Procedure Schematic



maker's responses reflect the presence of specific independence properties which determine the composite MAUF model form.

The third stage of the KR procedure assesses a single attribute utility function for each decision objective. First, the subject's relevant qualitative preference characteristics, such as monotonicity and risk attitude, are identified. Second, a few quantifiable points on the utility curve are established from the subject's certainty equivalent responses to a set of 50-50 uniattributed lotteries. A functional expression which captures the decision maker's qualitative preference characteristics is then fit to these elicited points. Finally, the consistency of the assessed conditional utility models must be verified. This can be accomplished by presenting the decision maker with an additional set of uniattributed lotteries and comparing the elicited responses to those revealed by the assessed function. The correction of any observed inconsistencies may require previous steps of the overall procedure to be repeated.

Because the conditional single attribute utility functions are assessed on an arbitrary scale, the fourth step of the KR procedure rescales each function to a common origin and unit of measurement. Although in principle several rescaling plans exist, the process is facilitated if n-2 attributes are fixed while two are varied to obtain pairs of indifference consequences for every attribute. Since indifference implies equal levels of utility, a set of independent equations is developed to compute the relative values of the scaling parameters. Once a consistent scaling is established, the decision maker provides the probability such that he is indifferent between:

1. the lottery with probability p_i for consequence $(x_1^*, x_2^*, \ldots x_n^*)$ and $1-p_i$ for consequence $(x_1^o, x_2^o, \ldots, x_n^o)$ and
2. the certain consequence $(x_1^o, x_2^o, \ldots, x_{i-1}^o, x_i^*, x_{i+1}^o, \ldots, x_n^o)$.

Because $U(x_1^*, x_2^*, \ldots, x_n^*)$ and $U(x_1^o, x_2^o \ldots, x_n^o)$ are scaled to be 1 and 0, respectively, $p_i = k_i$. This value for k_i, together with the previously derived set of independent indifference equations, is solved for the MAUF scaling parameters.

The final stage of the procedure checks the elicited MAUF for errors. An error is defined to occur whenever the assessed model fails to reproduce the subject's preferences when tested by an hypothetical example. Errors are typically detected through paired comparisons of multivalued decision outcomes. For instance, assume that the three attribute utility model, $U(x_1, x_2, x_3)$, has been encoded. If the decision maker states a preference for consequence (x_1', x_2', x_3') to consequence (x_1'', x_2'', x_3''), and $U(x_1', x_2', x_3')$ is greater than $U(x_1'', x_2'', x_3'')$, then no error is said to occur. This process is repeated for different consequences until the analyst is satisfied that the choices implied by the encoded MAUF are consistent with the decision maker's expressed preferences. To rectify any observed inconsistencies, some steps in the overall procedure will have to be retraced.

The KR procedure is a theoretically sound methodology which has been applied to several real-world decision tasks (e.g., Keeney 1972, 1973, 1975, 1977, 1979; Keeney and Wood 1977; Keeney and Sicherman

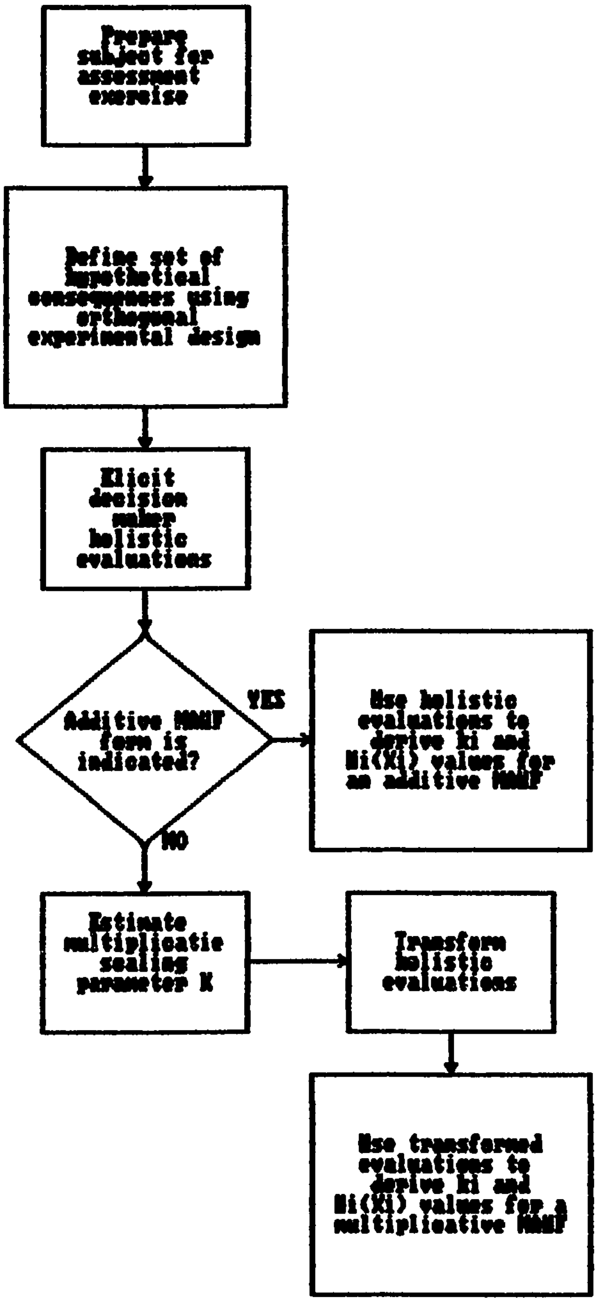
1983; and Keeney, Lathrop, and Sicherman 1986). The method's successful application, however, demands a substantial investment of time and effort from both the decision analyst and the decision maker. As a result, the number of practitioners of this procedure has remained relatively small. In an attempt to reduce the time and effort required to assess an individual decision maker's MAUF, alternative procedures have been developed. Three such techniques are discussed in Sections 3.3, 3.4, and 3.5.

3.3 The HOPE Elicitation Procedure

The holistic orthogonal parameter estimation (HOPE) procedure of Barron and Person (1979) converts a decision maker's holistic utility assessments for hypothetical vector valued decision consequences into an MAUF. The hypothetical consequences are specified by an orthogonal experimental design, and HOPE assumes the underlying utility functional form is either multiplicative or additive. The essential elements of the technique are depicted in Figure 3-2.

HOPE can be divided into three phases. A preparatory phase (similar to the initial stage of the KR methodology) provides the subject with an overview of the elicitation exercise, defines relevant attributes, establishes their bounds, and verifies necessary independence assumptions. After this phase, the two procedures follow separate paths. Unlike the KR procedure, HOPE does not directly estimate single attribute utility functions and scaling parameters. Rather, these values are inferred from the decision maker's global utility evaluations.

Figure 3-2 HOPE Elicitation Procedure Schematic



HOPE's second phase elicits the inputs required to operationalize the technique. To reduce the number of overall utility judgments needed, holistic responses are obtained for profiles defined by an orthogonal experimental design. Such designs have been shown to minimize the number of necessary evaluations (Addelman 1962). Once the profiles are specified, the subject assigns one value from the interval (0,1) to each. For risky decision scenarios, holistic responses represent indifference probabilities for a gamble where each profile is viewed as a sure thing consequence to be compared to a lottery offering the least and most preferred reference cases as probabilistic outcomes.

The third phase converts the subject's holistic responses into either an additive or multiplicative MAUF. Scaling parameter values and conditional utility values for each attribute level specified by the experimental design are inferred from the overall assessments. The necessary computations are simplified by HOPE's reliance on an orthogonal set of consequence profiles.

The MAUF functional form is determined from the subject's evaluations of complimentary decision profiles. When the holistic utilities of these two profiles sum to one, K equals zero, and the additive composite form is established. Since the evaluated consequences are orthogonal, each attribute's main effect at any specific level is the average of all holistic responses containing the attribute at that level, less the average of all evaluations containing the attribute at its least preferred level. These calculations provide an estimate of $k_i u_i(x_i^j)$ for every level j of attribute x_i included in

the design. Because the utility of an attribute at its most desirable level is scaled to equal one, a value for k_i when x_i equals x_i^* can be inferred. When necessary, the derived scaling constants are normalized to satisfy the restrictions of the additive expression.

The multiplicative functional form is encoded whenever the sum of the k_i's substantially differs from one. An estimate of K, K', is first calculated using two complimentary profiles. Next, the holistic utility responses for each consequence i, U_i, are transformed by the expression $\ln(1+K'U_i)$. This converts the right hand side of the multiplicative model to an additive form. Calculations on the transformed inputs then proceed as in the additive case, except the scaling parameters are not normalized. The expressions $\ln(1+K'k_iu_i(x_i^j))$ are available for each level j of attribute x_i specified by the experimental design. When x_i equals x_i^*, an estimate of $\ln(1+K'k_i)$ is generated. Since the estimate for K is determined separately from these expressions, values for all k_i and $u_i(x_i^j)$ terms defined by the analysis can be inferred.

The most commonly cited shortcomings of HOPE concern the exclusion of the multilinear functional form, the single use of an orthogonal experimental design, and the cognitive burden that global utility evaluations place on the decision maker. Barron and Person (1979) note that the exclusion of the multilinear model is not unduly restrictive given the general robustness of the additive and multiplicative expressions in practical decision contexts (see Keeney and Raiffa 1976, p.298). Furthermore, the use of an orthogonal set of consequence profiles offers the advantage of having a complete set

of easily computed and uncorrelated parameter values. Finally, because only a small number (e.g. a four attribute utility function on four levels requires sixteen evaluations) of holistic evaluations are required, Barron and Person consider the cognitive burden argument to be generally irrelevant.

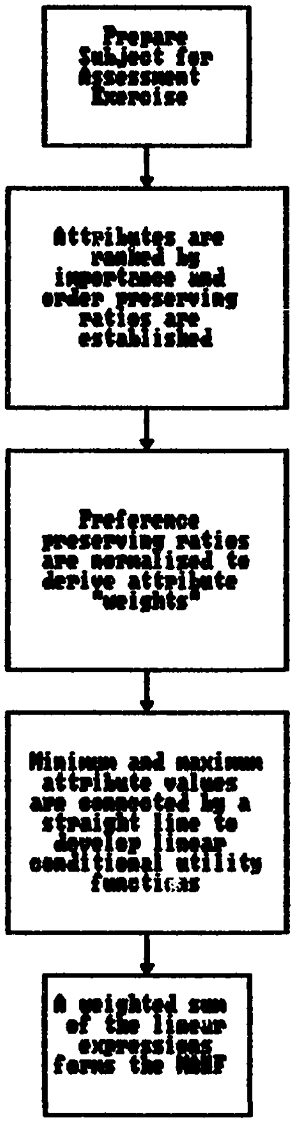
3.4 The SMART Elicitation Procedure

In an attempt to simplify the MAUF assessment process, Edwards (1977) developed an elicitation procedure based on simple weighting rules. While his simple multiattribute rating technique (SMART) lacks the theoretical elegance of the KR methodology, SMART's avoidance of lottery based evaluations makes it easy to implement. The technique's essential steps are described in Figure 3-3.

As is common to all elicitation procedures, SMART's initial step sets the stage for the assessment session. The decision problem, the purpose for the assessment, and the alternatives to be evaluated are discussed with the subject in general terms. Relevant objectives (attributes) are identified, and bounds are established on their values. A set of importance weights for each attribute is then determined.

While various weighting plans are available (e.g., rank sum, rank reciprocal, or rank exponent), the earliest versions of SMART assessed weights in the following manner. First, the relevant attributes are ranked in order of importance. Next, ratio estimates of the importance of each attribute relative to the least important attribute are determined. These ratio estimates are computed as

Figure 3-3 SMART Elicitation Procedure Schematic



follows. The decision maker assigns a value of 10 to the least significant decision criterion. The next least important criterion is given a value reflecting its relative importance to the one ranked lowest in significance. A similar comparison is made for each attribute with a ratio preserving value indicating relative importance assigned to each. For instance, an attribute which is assigned a value of 90 is considered to be three times as important to the subject as an attribute rated 30. Throughout the rating, the subject is allowed to modify assigned values to ensure consistency. Finally, importance ratios are normalized by dividing each by the sum of all the preference preserving values.

At this stage of the exercise, Edwards makes two simplifying assumptions. First, linear expressions are assumed to offer reasonable approximations for all underlying univariate utility functions. Every attribute's minimum and maximum values are scaled to have utilities of 0 and 1, respectively. A conditional utility function is then estimated by a straight line connecting these scaled endpoints.

The second assumption defines the rule for aggregating the linear single attribute utility expressions into a composite model. Edwards believes that the decision maker's internal preference space can be reasonably approximated by an additive functional form. Thus, the SMART assessed MAUF is a weighted sum of the linear utility expressions with the normalized importance ratios serving as weights.

Since the decision maker provides only a ranking of attributes and a set of importance preserving ratios, SMART can be easily implemented. The predictive validity of the encoded MAUF depends on

whether SMART's two critical assumptions (linear univariate conditional functions and additive aggregation rule) are tenable. Edwards (1977) held that if only small amounts of measurement error are present, then substantial deviations from the independence assumptions will not adversely affect the multiattributed evaluations, and will have even less impact on the rank orderings of the decision consequences. He further argued that the additive approximation will perform well as long as the subject's preferences are, or can be transformed to be, conditionally monotonic.

3.5 The Mathematical Programming Elicitation Procedure

Klein, Moskowitz, Mahesh, and Ravindran (1985) recommend the use of mathematical programming techniques to estimate both the single attribute utility functions and the MAUF scaling parameters. The univariate utility functions are assessed by solving a nonlinear mathematical programming problem formulated from decision maker preference responses. Composite scaling parameters are determined by a goal programming model formulated from the subject's expressed preferences for pairs of consequence profiles. The procedure is designed to simplify the KR methodology by reducing the number of lottery based evaluations, and by eliminating the need for choosing a suitable form for modeling the conditional utility functions.

While their technique can be used to estimate an MAUF directly, the authors argue that more precise MAUF scaling constants are obtained by converting a multiattribute measurable value function (MAMVF) into an MAUF. As a result, they recommend that single attri-

bute measurable value functions and MAMVF scaling constants first be estimated and then transformed mathematically to produce an MAUF. The basic steps of the procedure are depicted in Figure 3-4.

The univariate value functions are estimated by solving a non-linear programming problem (NLP) of the form:

$$\text{Minimize} \sum_{i=1}^{n} (v(x_i) - o(x_i)) \tag{5}$$

Subject To:

$$m(x_i) < (\text{or} >)\ m(x_{i+1}),\ i=1,2,\ldots,n-1, \tag{6}$$

$$m(x_i) > (< \text{or} =)\ 0, \quad i=1,2,\ldots,n, \tag{7}$$

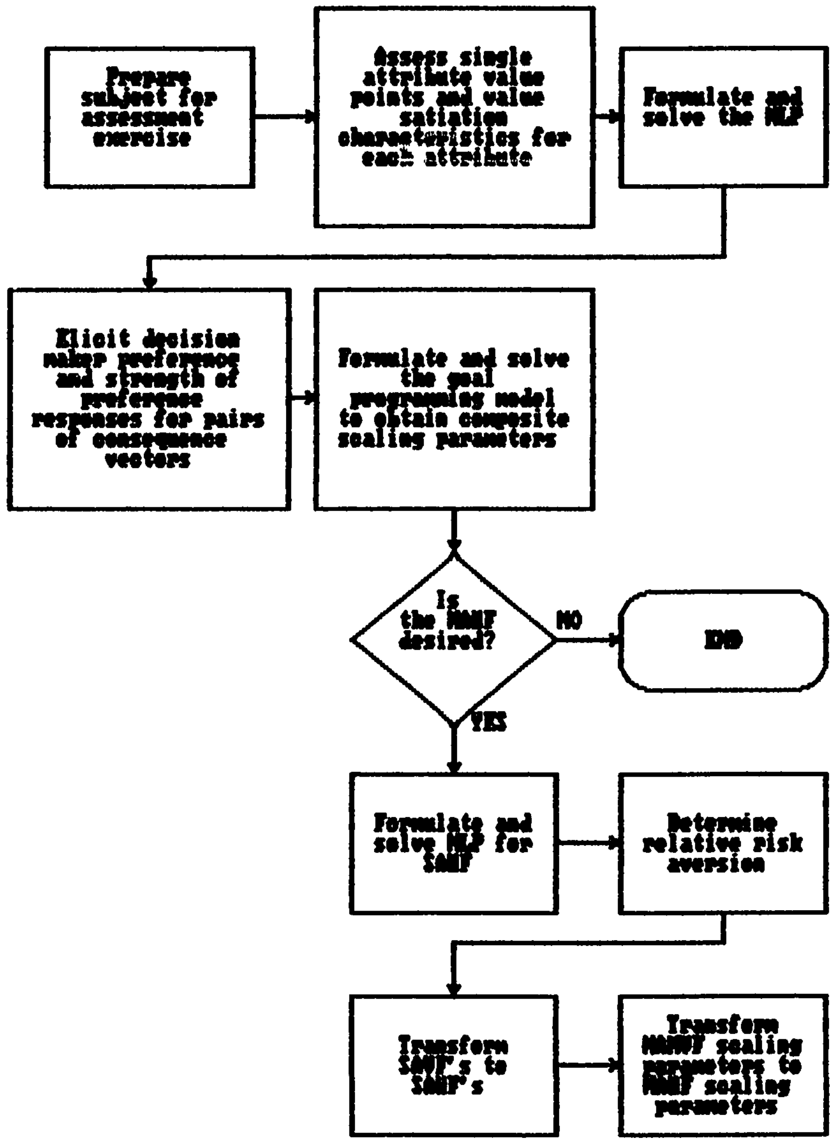
$$\text{Max } v'(x) > (\text{or} <)\ 0, \tag{8}$$

where $o(x_i)$ is the observed value of the single attribute value function at x_i, n is the number of observations, $v(\cdot)$ is the equation selected to model the conditional value functions, and $m(x_i)$ is a measure of the decision maker's preference attitudes.

The objective function (5) serves as the curve fitting criterion. Its formulation requires the analyst to assess a minimum of three measurable value points, $o(x_i)$. The best and worst attribute levels are assigned values of 1 and 0, respectively. The level having a value of 0.5 is determined to be the amount of the attribute for which the subject is indifferent between exchanges of the worst level for this amount and the best level for this amount. Additional points, if desired, can be elicited by repeated application of the midvalue splitting technique.

The observed points are then fitted to a single attribute functional equation, $v(\cdot)$, subject to specific decision maker preference attitudes. The authors essentially eliminate the selection of an

Figure 3-4 Mathematical Programming Elicitation Procedure Schematic



appropriate functional equation by employing a summed exponential function to express the value for an attribute x as

$$v(x) = a - b(\exp(cx)) - d(\exp(ex)). \tag{9}$$

While the summed exponential expression does not uniquely capture a decision maker's preference attitudes, it is recommended because its parameters are easily manipulated analytically to model a variety of decision maker preference characteristics. The parameters a, b, c, d, and e of (9) become the NLP decision variables.

The constraint set (6) enforces the increasing or decreasing nature of the value satiation coefficient m(x). For monotonically increasing value functions, Dyer and Sarin (1982) defined this measure as

$$m(x) = - v''(x)/v'(x), \tag{10}$$

where $v''(x)$ and $v'(x)$ are the second and first derivatives of the univariate value function, respectively, with respect to a specific level of the attribute x. The sign and relative magnitudes of m(x) are computed for each of the assessed measurable value points using formulas provided by Dyer and Sarin (1982).

Constraint set (7) models whether the subject's value satiation attitudes are increasing ($m(x)>0$), decreasing ($m(x)<0$), or constant ($m(x)=0$). The monotonically increasing or decreasing nature of the value function is assured by constraint (8), which is itself an optimization problem. Constraints of this type are discussed by Bracken and McGill (1973) and require special consideration. The maximum of (8) must occur at either an extreme point or at an interior point with $v''(x) = 0$. Since the summed exponential function is used to

model the univariate value functions, all possible maxima and minima are easily derived. For a monotonically decreasing value function, the following constraints, derived from the extreme points and the first order conditions, replace constraint (8):

$$v'(x^*) \leq 0, \tag{11}$$

$$v'(x^0) \leq 0, \tag{12}$$

$$v'(x = \ln(-bc^2/de^2)/(e-c)) \leq 0, \tag{13}$$

$$v'(x = \ln(-de^2/bc^2)/(c-e)) \leq 0, \tag{14}$$

For cases where values are monotonically increasing in x, the inequality signs are reversed.

The NLP provides substantial flexibility in estimating the single attribute measurable value functions. The curve fitting criterion (5) can be modified to impose different weights on the deviations and constraints can be added or deleted to redefine value satiation attitudes. The NLP also provides for an automatic consistency check of the subject's inputs. As long as a feasible region exists, the decision maker's responses are considered to be "practically consistent", and small assessment inconsistencies will be "smoothed" when the NLP is solved. The presence of significantly inconsistent and incoherent responses will, however, produce an infeasible solution space which will require the NLP to be reformulated.

Because both the objective function and feasible region of the NLP are nonconvex, the possibility of converging to a local, rather than a global, minimum exists. To reduce the risk of obtaining an inferior solution, the authors suggest that alternative starting points be used to solve the NLP.

After all single attribute value functions are obtained, the composite scaling parameters are determined by solving a goal programming (GP) model designed to minimize violations in the decision maker's revealed preferences. The approach differs from other mathematical programming parameter estimation techniques (Srinvasan and Shocker 1973, Horsky and Rao 1984) because it directly incorporates both decision maker preference and strength of preference responses in an attempt to improve the reliability of the estimated scaling constants.

The general GP model for estimating the scaling constants for the additive or multiplicative multiattribute measurable value function is:

$$\text{Min } W_1\left(\sum_{i=1}^{q} S_i E_i^-\right) + W_2\left(\sum_{i=1}^{q} S_i (y_i^- + z_i^+)\right) \tag{15}$$

Subject To:

$$V(\underline{x}_{1i}) - V(\underline{x}_{2i}) + E_i^- - F_i^+ = 0, \quad i=1,2,\ldots q, \tag{16}$$

$$F_i^+ - S_i r + y_i^- - z_i^+ = 0, \quad i=1,2,\ldots q, \tag{17}$$

$$0 \le k_j \le 1, \quad j=1,2,\ldots n, \tag{18}$$

and, either

$$k_1 + k_2 + \ldots + k_n = 1 \text{ for the additive model,} \tag{19}$$

or

$$-1 < K \tag{20}$$

and

$$(1+K) = (1+Kk_1)(1+Kk_2)\ldots(1+Kk_n) \tag{21}$$

for the multiplicative model,

where q represents the number of preference and strength of preference responses and n is the number of attributes.

The weighted objective function (15) minimizes decision maker preference inconsistencies (E_i^-) and negative and positive deviations (y_i^-, z_i^+) from the strength of preference responses. The objective weights, W_1 and W_2, are analogous to the weights of a weighted goal programming problem. Their values are determined at the analyst's discretion.

Constraint set (16) is formulated from decision maker pairwise preference comparisons. Every comparison produces an inequality constraint. For example, if consequence $\underline{x}_1$ is preferred to consequence $\underline{x}_2$, then $V(\underline{x}_1) > V(\underline{x}_2)$ when measured by the overall value function, $V(\underline{x})$. These inequalities are modeled as

$$V(\underline{x}_{1i}) - V(\underline{x}_{2i}) + E_i^- - F_i^+ = 0 \qquad (22)$$

where F_i^+ is > 0 if profile $\underline{x}_{1i}$ is preferred to profile $\underline{x}_{2i}$, and E_i^- denotes a preference inconsistency.

Constraint set (17) enforces the decision maker's intensity of preference responses. Without these constraints, the possibility exists that some preferences will be more closely satisfied than others. To insure the subject's indicated preferences and underlying preference representation are accurately preserved, a strength of preference response is formulated for each paired comparison as follows:

$$F_i^+ - S(r) + y_i^- - z_i^+ = 0, \qquad (23)$$

where F_i^+ represents the positive deviation in the paired comparisons, y_i^- and z_i^+ are the negative and positive deviations from the strength of preference response S_i, and r denotes the relative magnitude of the value difference. The consistency of the composite form

being modeled, is guaranteed by the inclusion of constraint (19) for additive MAMVF's and the inclusion of constraints (20) and (21) for multiplicative MAMVF's.

The MAUF is obtained by exploiting the mathematical relationship between value and utility functions as discussed by Dyer and Sarin (1979, 1982). A single attribute utility function for any one of the attributes is estimated using the following NLP:

$$\text{Minimize} \quad \sum_{i=1}^{n} (u_i(x_i) - o(x_i))^2 \qquad (24)$$

Subject To:

$$r(x_i) >,=,< r(x_{i+1}) \quad \text{for } i=1,2,\ldots n-1 \qquad (25)$$

$$u_i''(x_i) >,=,< 0 \quad \text{for } i=1,2,\ldots n \qquad (26)$$

$$\text{Max } u_i'(x) < (\text{or} >)\ 0 \qquad (27)$$

where,

n = the number of certainty equivalent responses provided by the DM.

$r(x_i)$ = the risk function evaluated at the i^{th} value of the attribute being considered.
$r(x) = -u''(x)/u'(x)$ for utilities increasing in x

$u_i'(x)$ = the first derivative of the utility function, u_i, evaluated at x.

$u_i''(x_i)$ = the second derivative of the utility function, u_i, evaluated at x_i.

$o(x_i)$ = the observed utility value at x_i.

The process is similar to the procedure for estimating single attribute value functions with certainty equivalent responses replacing exchange responses, and decision maker risk attitudes replacing value satiation characteristics.

First, the best and worst attribute levels are determined. Then, the attribute level having a utility value of 0.5 is estab-

lished from a certainty equivalent response for a 50-50 lottery offering the best and worst levels as payoffs. Once this level is determined, attribute values having utilities of 0.25 and 0.75 can be similarly measured, if desired. The objective function defines the criterion by which a summed exponential equation is fitted to the observed utility values.

Constraint set (25) captures the increasing, decreasing, or constant risk attitude of the decision maker (as defined by Pratt 1964), while constraints (26) indicate whether the subject is risk averse, risk prone, or risk neutral. Both types of constraints are determined from the decision maker's risk premiums for 50-50 lotteries having outcomes over comparable ranges. Finally, the monotonically increasing or decreasing nature of the univariate functions are ensured by (27).

The decision maker's coefficient of relative risk aversion is computed and used to estimate the other utility functions from their corresponding single attribute measurable value functions using the appropriate transformations (Dyer and Sarin 1982). A set of MAUF scaling parameters can be derived from an additive or multiplicative MAMVF using the relationships developed by Dyer and Sarin (1979). Thus, by exploiting the theoretical relationship between value and utility, the mathematical programming procedure attempts to reduce the number of decision maker inputs required by the KR methodology to encode an MAUF.

3.6 Summary

This chapter presented an overview of several MAUF elicitation techniques. In addition, a detailed description was provided of the four assessment technologies examined in this study. A discussion of the research relevant to the comparison of alternative MAUF assessment procedures is presented in Chapter 4.

CHAPTER 4 A REVIEW OF THE LITERATURE COMPARING ALTERNATIVE MAUF ASSESSMENT PROCEDURES

4.1 Overview

Given the importance of selecting an elicitation procedure capable of encoding a reliable MAUF, it is surprising that few studies have systematically compared the validity of alternatively elicited MAUFs. Most of the research available has focused on either the sensitivity of choices to different weighting rules, the effects of model form on preference predictions, or the potential sources and effects of assessment error on the modeling process. Sections 4.2, 4.3, and 4.4 provide surveys of this research.

4.2 Studies Of Weighting Methodologies

Many studies have investigated the sensitivity of multiattribute evaluations to choice of weighting methodology for the additive composite model. Of these, several have compared the predictions of models based on differential and equal attribute weighting plans. Beckwith and Lehmann (1973) modeled consumer attitudes for various television programs. They determined that differentially weighted attributes provided only limited benefits when compared to attitudes predicted by equally weighted criteria. Einhorn and Hogarth (1975) found that equal or unit weights did a good job of predicting utilities constructed from differential weights whenever attributes were both small in number and positively correlated. The case for equal weights received further support from Wainer (1976) who established

their effectiveness whenever attributes were conditionally monotonic and positively correlated.

Other studies have shown how equal weights can lead to unreliable results. Newman (1977) determined that differential and equal weighting plans could produce alternate preference orderings if attributes are negatively correlated. While Barron (1980, 1987) found that equal weights can perform well when attributes are positively correlated, he demonstrated that they are comparatively poorer performers when consideration is given only to selecting alternatives on the efficient frontier.

In an attempt to establish an empirical justification for simple weighting procedures, Stillwell, Seaver, and Edwards (1981) analyzed the sensitivity of multiattribute evaluations to different weighting schemes. Models based on rank sum weights, rank reciprocal weights, rank exponent weights, decision rule weights, and unit weights were correlated with a "true" ratio weighting model. In three applications they determined that when equal weights performed well, so did the rank weight models. More importantly, they discovered that in those cases where the equal weights model correlated poorly (e.g., negatively correlated attribute pairs), the models based on rank weights did much better.

Encouraged by the findings of Stillwell, Seaver, and Edwards, Beach and Barnes (1983) investigated the ability of three alternative simple weighting schemes to produce orderings consistent with ratio measurement weights. Their findings indicated that proper orderings could be achieved using point allocation, voting, and rating plan

weights, with voting rule weights the poorest performer. They concluded that while simple weighting rules are useful, additional work is needed to more fully identify the circumstances where simple weighting plans do and do not perform well.

Horsky and Rao (1984) evaluated alternative weight estimation techniques for a four attribute additive composite model. When the number of alternatives considered was between five and nine, weights estimated by linear programming showed greater stability and predictive accuracy than equal weights.

Nutt (1980) conducted a field study to compare weighting plans for additively combined attributes. Nine subjects employed one indirect and four direct weighting procedures to evaluate ten alternatives described by five attributes. Indirect weights were derived from hypothetical decision tasks using analysis of variance techniques. Direct weights were determined using anchored rating scales, logarithmic scales, rank weights, and point assignments.

Weighting methodologies were compared by measuring the consistency of each method's assigned weights and the similarity of decision priorities established by each set of weights. The distinct set of weights provided by the indirect plan was consistent with earlier studies which indicated that such plans can produce a wide range of attribute weights. Anchored rating scale weights were the most consistent and reliable set of weights assigned. However, no definitive set of recommendations were offered. Rather, Nutt suggested that the performances of the different weighting rules be examined under different decision contexts and circumstances.

Hobbs (1980) argued that ranking and rating procedures were unlikely to derive theoretically valid weights because they failed to incorporate the decision maker's willingness to make tradeoffs. To illustrate his point, a hypothetical nuclear power plant siting decision was used to empirically compare two weighting methodologies. One method derived attribute weights from decision maker tradeoffs. The other method had decision makers select weights on a scale from 0 to 10. Subjects were divided in their opinions on the worthiness of the extra effort needed to derive tradeoff based weights. Using an additive model, the study found that the two weighting procedures produced significantly different site selections. The choice of weights was important in this decision context, argued Hobbs, because the decision maker was concerned with only the best few alternatives and not a ranking of all candidate sites. Thus, in order to accurately model the decision maker's preferences, he believed that the theoretically more rigorous weighting procedures should be used.

To date, the most extensive study comparing weight assessments for additive utility functions was conducted by Schoemaker and Waid (1982). Five weighting procedures were examined: (1) multiple linear and nonlinear regression analyses of holistic judgments; (2) direct decomposed tradeoffs based on the Keeney-Raiffa procedure; (3) Saaty's (1980) analytic hierarchy method; (4) a 100 point allocation process; and (5) unit weighting.

The decision task required the experimental subjects to evaluate a hypothetical set of college applicants described by four attributes. Linear and nonlinear single attribute utility functions were

used to construct the additive composite models. For twenty pairwise comparisons, the predictive ability of each model was judged against each subject's strength and direction of preference. The quality of predictions was measured using the paired comparisons of the proportion of correct predictions and correlations. While each method assigned different weights to the attributes, all methods, except unit weighting (which was clearly inferior) predicted equally well on the average. Furthermore, an inverse correlation was exhibited between the subjects' perception of a procedure's complexity and level of reliability.

Schoemaker's and Waid's study differed from previous research regarding weighting techniques because it:

(1) focused on the models' predictive ability and not on its explanatory power as did Einhorn and Hogarth (1975);

(2) used binary choices, not rankings as did Newman (1977);

(3) did not restrict its analysis to a few methods as in Fischer (1977), and did not examine only variations within a basic method;

(4) summarized subjects' impressions regarding a method's difficulty and perceived reliability; and

(5) tested many subjects using a familiar task with a monetary incentive.

In spite of their extensive investigation, the authors concluded that the question of the appropriateness of a weighting technique remained unresolved, since performances could vary across subjects and decision tasks.

Stillwell, Barron, and Edwards (1983) conducted an empirical study of alternative multiattribute weight elicitation techniques in

a situation offering a "meaningful external criterion" of comparison. Various value models (riskless utility models) were tested to determine whether or not complex weight elicitation procedures afforded the user an advantage. By definition, an advantage existed if, despite additional computational burden, a weighting scheme altered the preference relationship among decision choices in a manner producing more nearly correct conclusions.

Three rank weighting procedures (rank sum, rank reciprocal, and rank exponent) along with three ratio weighting techniques (SMART, additive HOPE, and relative importance of value) were examined. In terms of the quality of the decisions rendered, little difference was detected among the weight elicitation techniques. A simple equal weighting of the attributes performed well. However, results using the additive HOPE procedure were generally poorer. This finding conflicted with earlier studies supporting HOPE derived weights (see Fischer 1977 and Barron and Person 1979). The authors recommended additional research, since differences among weight elicitation techniques could vary with the decision context.

Adelman, Sticha, and Donnell (1984) believed that differences among multiattribute weighting techniques had not been detected in earlier studies because most previous research had failed to systematically vary either the number of attributes or the distribution of the correct attribute weights. They hypothesized that fewer attributes would reduce the information processing requirements of the decision maker, thereby improving any weighting plan's effectiveness. They also believed that the peakedness of the distribution of correct

weights could influence the performance of the various weighting rules.

Two experiments were analyzed, one designed to test the accuracy of weights assigned by individuals, the other designed to test the accuracy of group assigned weights. Five weighting techniques were considered: (1) SMART; (2) paired comparisons; (3) policy capturing; (4) Keeney-Raiffa lotteries for risky choices; and (5) 100 point allocation among attributes. A significant main effect for both weighting technique and the number of attributes was detected. All weighting methods were more effective whenever a smaller number of attributes were considered. The authors suggested that a simulation analysis be conducted to assess the accuracy of different weighting techniques in different decision environments. Presumably, subsequent results could be used to generate a general statement regarding the effect of attribute properties on the accuracy of weighted linear composites.

These studies provide useful information regarding specific advantages to various weighting methodologies. However, they fail to address the more pressing issue of whether or not the MAUF assessment technique itself is structurally capable of generating a utility model which provides preference orderings in concert with the decision maker's internal preference structure.

4.3 Research Comparing the Sensitivity of Preference Predictions to MAUF Model Forms and Encoding Technique

A limited, but growing body of research has examined the

sensitivity of preference predictions to the encoded multivariate utility functional form and the assessment procedure used. Huber, Daneshgar, and Ford (1971) investigated the validity of different utility models as predictors of job ratings and job selections. Their findings confirmed, to a limited extent, their hypothesis that validity is a function of the predictive model's form. They also suggested that a model's validity could be dependent on the subject whose preferences were being assessed. Their results, however, were unable to support a firm conclusion since their work suffered from a small sample size and poor field conditions.

Huber (1974) provided a review of field and field like studies, as opposed to laboratory studies, of multiattribute utility models under certainty. He concluded that research in behavioral science has validated the use of subjective values as parameters in such models. Furthermore, simple additive forms often structured outcomes as well as, or better than, more complex models.

Fischer (1976) investigated whether or not simple additive and multiplicative utility models could adequately capture preferences for multidimensional consequences for riskless and risky decision making environments. His study can be considered an improvement on earlier works because it considered preferences for both riskless and risky decision situations using statistical and conjoint measurement techniques.

Ten subjects evaluated 27 hypothetical job offers which were defined by three attributes (type of work, annual salary, and city of employment) varying on three levels. An analysis of the evaluations

revealed that additive statistical models produced good approximations to both the riskless and risky case responses for every subject. However, fifty percent of the subjects revealed a significant departure from additivity; while the simple linear models offered good preference predictions, they did not always correctly characterize the structure of the decision maker's internal decision space. Several subjects violated certain ordinal properties shared by both additive and multiplicative models. All subjects displayed some minor violations of the conjoint measurement axioms (i.e., independence properties for additive models of riskless choice). An analysis of the results suggested that many of the minor violations were due to random error.

As an extension of his 1976 study, Fischer (1977) compared holistic and decomposed multiattribute utility assessment techniques. The 1977 work was specifically designed to determine the validity of the Keeney-Raiffa risky utility decomposition procedure. As before, subjects were asked to evaluate 27 hypothetical job descriptions defined by three levels of three attributes. Each subject evaluated the alternatives using five procedures: (1) holistic riskless rating scales; (2) holistic risky utility assessments; (3) riskless additive decomposed rating scales; (4) risky Keeney-Raiffa decomposed additive models; and (5) risky Keeney-Raiffa decomposed multiplicative models. Fischer reasoned that if both decomposed and holistic procedures offered valid preference measures, the two methods should provide linearly related utility values. The median within-subject rank order correlations between the various procedures were about 0.85,

whereas correlations between the additive and multiplicative Keeney-Raiffa decompositions approached 1.0. Fischer offered two possible explanations for the high correlations between the two Keeney-Raiffa models. First, the correlations could have been inflated since identical sets of judgments were used to elicit the two models. Second, additive models can often provide good approximations to multiplicative ones (Yntema and Torgerson 1961).

Barron and Person (1979) simulated the HOPE procedure using a known multiplicative MAUF to examine the effects of model misspecification on preference predictions. They discovered that larger prediction errors resulted from incorrectly specified additive models, with or without random error components, than from correctly specified models with additive random error terms.

The experimental job choice data of Fischer (1976, 1977) was reexamined by Barron (1980) to assess the usefulness of HOPE derived utility values for decision making. Three additive and six multiplicative HOPE models were estimated per subject for either riskless values or risky utilities. The coefficient of determination was computed between the set of holistic judgments not used to fit the HOPE models and the utility predictions derived using the HOPE technique. Higher levels of convergence than those found by Fischer for riskless and risky decomposition procedures were observed between holistic utilities and HOPE based utility predictions. Barron concluded that HOPE elicited utilities are valid and useful for decision making, and perform as well as decomposition procedures when a small number of attributes are involved.

To validate additive utility assessments under uncertainty, Eils and John (1980) used an external criterion to compare SMART evaluations to those of a group communication strategy. The decision task required group utility assessments for ten hypothetical bank credit card applicants defined over ten dimensions. When decisions from the alternative procedures were compared against the judgments generated by a bank model for potential credit applicants, it was discovered that SMART assessments, not the holistic evaluations, improved the group's decision making ability. No comparison of alternative weight elicitation plans was conducted. The study's findings supported Edwards' contention that decomposed methods are better suited for resolving complex decision tasks than are holistic methods. However, findings that a single decomposition procedure outperformed a holistic judgment approach in one specific instance offer no general conclusions concerning the relative performances of the two methodologies.

Currim and Sarin (1984) modeled consumer decisions under riskless and risky situations using statistical estimation and algebraic solution procedures. Their study asked 100 students to evaluate postgraduation job choices over three attributes. The preference predictions considered additive conjoint, additive and multiplicative measurable value models, and additive and multiplicative utility models. The statistical estimation procedures exhibited greater predictive accuracy than did the algebraic models. For risky decision tasks, utility models outperformed the conjoint procedures. Because multiplicative models did not improve decisions in the cer-

tainty case, they were excluded from the analysis of risky decision making.

Farmer (1987) conducted a study to measure the robustness of the Keeney-Raiffa procedure for eliciting multiattribute utility functions. His research was designed specifically to reveal the insensitivity of simplified MAUFs to violations of attribute independence assumptions and scaling constant simplifications in an accounting decision scenario. However, Farmer made no general conclusions concerning multiattribute utility theory and its application to complex decision making.

Fifteen auditors participated in the study. A Keeney-Raiffa MAUF for each auditor's preferences for internal auditing control systems was assessed by combining four discrete unidimensional utility functions multiplicatively. Two additive models were compared to the Keeney-Raiffa preference predictions. One additive model was a scale-weighted model employing both the Keeney-Raiffa conditional utility functions and the scaling constants. The second model of comparison was a simple aggregation of the unidimensional utility functions (i.e., an equal weights model). Subjects ranked 42 client cases according to their perception of the reliability of a hypothetical firm's internal accounting control system. System reliability ratings were predicted for each model and correlated with the subjects' unaided judgments. Wilcoxon's two-sided signed rank test for differences in paired data indicated no statistical differences in the predictive ability attributable to the more complex Keeney-Raiffa encoded MAUF. For the context examined, Farmer concluded that util-

ity functions modified from appropriately encoded MAUFs do a good job of predicting auditor judgments. However, limitations of the study prohibited Farmer from extending this conclusion to more general decision contexts.

The studies reviewed here address an important issue in applied decision analysis, namely the sensitivity of preference predictions to model form and elicitation methodology. However, this research has typically examined only limited decision contexts using a few assessment procedures. Furthermore, the conclusions offered by these studies often conflict. As a result, the available research is unable to provide either definitive guidelines or general recommendations for deciding when the form of the assessed model and the choice of an elicitation technique matters.

4.4 Studies of Potential Sources of MAUF Elicitation Error

Recent interest in potential sources of MAUF assessment error and their effects on preference modeling has produced several studies. Pitz, Heerboth, and Sachs (1980) compared the sensitivity of decomposed evaluations and holistic judgments to variations of relevant and irrelevant decision variables to examine the usefulness of multiattribute utility analysis. Subjects evaluated four hypothetical apartments described by six attributes. Holistic evaluations were compared to decomposed values derived from an additive multiattribute utility model. The holistic judgments displayed little sensitivity to differences among apartment profiles, but led subjects to order apartments in a manner suggesting an oversimplification of the

information provided. This linear ordering of preferability (labeled a linearity effect) could not be supported solely by the information available to the study's subjects. While the decomposed evaluations exhibited no linearity effect, they were influenced by systematic differences in apartment profiles. The authors concluded that if decision makers had a tendency to simplify information when making overall judgments, then the decomposition techniques might be sensitive to small changes in attributes that would be overlooked by unaided holistic evaluations.

Eliashberg (1980) gauged the preferences of 85 undergraduate students for housing location measured on two criteria. All the elicited models were shown to be better predictors than chance models. The data also implied that the elicited cardinal utility functions were predictively robust given deviations from the necessary utility independence assumptions. Eliashberg noted that further research would need to consider more than two attributes to establish general conclusions.

A formal approach to error analysis in assessing MAUFs was discussed by Barron (1983). Four potential sources of error were identified and related to four separate utility elicitation methods within the framework of the general multiplicative multiattribute utility model. Two decomposition procedures (Keeney-Raiffa and SMART) and two holistic procedures (Social Judgment Theory and HOPE) were discussed.

Barron noted that errors could easily arise in the preparation phase of any assessment exercise. These errors, which are generally

attributable to the elicitor, include the misidentification of relevant attributes, the incorrect specification of attribute ranges, and the improper framing of the decision context. Throughout his discussion, Barron assumed that these effects were neutral across assessment procedures. Four possible sources of preference prediction error: (1) systematic error; (2) model specification error; (3) random error; and (4) substantial error were then identified and discussed.

Potential sources of systematic error are described in the works of Yates and Jagacinski (1979), Kahneman and Tversky (1979), Krzysztowicz and Duckstein (1980), and Hershey, Kunreuther and Schoemaker (1982). The effects of model specification error within elicitation methods are discussed by Fischer (1977), Barron and Person (1979), and Barron (1980).

The possible effects of random error (noise) on the four assessment procedures was discussed in a general manner. Since the Keeney-Raiffa and SMART procedures estimate conditional utility functions and scaling constants separately, any noise contained in the estimate of one need not perturb estimates of the other. SMART estimated weights are normalized by the sum of the estimated weight values. As a result, they are only sensitive to errors in the relative values of the estimates. Keeney-Raiffa scaling constants, which are estimated using standard gambles, are subject to Kahneman and Tversky's certainty effect. If direct tradeoffs are used to estimate scaling parameters, then noise in the single attribute Keeney-Raiffa utility functions will produce noisy scaling parameter values. Social Judg-

ment Theory infers a set of scaling constants by regressing standardized attribute levels against holistic evaluations. As a result, noise can be introduced through the reliance on linear conditional utility functions as well as through the holistic evaluations. Because HOPE infers a set of scaling constants and univariate utility values from a set of global evaluations, noise is introduced through errors inherent in holistic judgments.

Finally, substantial errors in judgment on the part of the decision maker are a possibility with any assessment technique. Generally, each elicitation procedure detects and corrects for such errors through a series of consistency checks conducted at various points in the multiattribute utility encoding process. Barron described these verification tests for the Keeney-Raiffa, SMART, and HOPE procedures. Practical considerations (e.g., ease of use and flexibility) were highlighted for each method. However, no applied analysis of the methods was conducted. Barron's work was a general discussion, not an empirical study, and he noted that a procedure's specific advantages or disadvantages would have to be determined through applied analyses and empirical research.

Recently, Barron (1987a) investigated the effects of incomplete attribute sets on the selection of the preferred multivariate alternative. His study employed 18 data sets each consisting of 15 alternatives described by nine attributes. Attributes were combined additively using equal weights. Each case determined the best alternative based on six attributes. The value of this best alternative was then calculated using all nine attributes and compared to the

best alternative selected using all nine attributes. Measure of value loss and correlation were used to compare rankings for the incomplete and complete attribute set evaluations. Value losses ranging from 1.6% to 69.1% of the true value were found in 36% of the cases considered. All other cases had no loss in value. While the study's results cannot be extended to all decision contexts, Barron concluded that frequent and/or sizeable value losses may accompany high correlation values, and that missing attributes may lead to incorrect choices even in the presence of high correlations (e.g., greater than 0.90).

Laskey and Fischer (1987) analyzed the effects of response error on the assessment of preferences for multiattributed consequences. Decision maker preference responses were assumed to be composed of two components: a systematic or "true" component, and an additive random component reflecting response error. The study analyzed the ability of statistical models to filter out response error and capture the decision maker's systematic preferences.

Specifically, the authors explored whether:

(1) preferences for multiattributed decision alternatives are stable over a two week time horizon;

(2) statistically estimated multiattribute utility models can accurately predict preferences two weeks into the future;

(3) ranking procedures designed to generate more consistent utility assessments produce more accurate statistical preference models; and

(4) ranking procedures induce serially correlated errors in judgment, and if so, whether better preference models can be estimated by statistically adjusting for serially correlated response errors.

Fifteen experimental subjects assumed the role of an air pollution control regulator, and evaluated alternative pollution control policies for coal burning electricity generating facilities in a hypothetical city. All decisions were evaluated on three attributes:

(1) the annual cost per household for electricity, including pollution control costs;

(2) the annual number of chronic respiratory illnesses, including those caused by coal fired power plants; and

(3) the annual respiratory mortality rate, including those caused by coal fired power plants.

Each decision attribute was restricted to three values: an average, maximum, and minimum value for cities similar to the experimental city.

Every subject evaluated 25 policy alternatives in two experimental sessions separated by two weeks. In each session, consequences were ranked and utilities assigned to each policy alternative. The order in which the rankings and utility assignments were performed was varied across the two sessions. To measure the predictive validity of the additive representation, additive multiattribute utility models were statistically fitted to each subject's utility evaluations.

The authors reported five major findings. First, the correlation of between session utility responses suggested that preferences remained stable over the two week period. Second, the statistically estimated additive models accounted for a high percentage of the variance of the utility assessments. Furthermore, the parameters of the additive model were precise (i.e., had small standard errors) and stable over the experimental time horizon. Third, high correlations

were observed between the additive models' predicted choices and direct judgments obtained earlier. Fourth, high levels of serially correlated errors were found when subjects ranked alternatives prior to assessing utility values. Fifth, the use of a first order autoregressive estimation procedure to account for the serial correlation had a negligible influence on the model's preference predictions in a different time period.

Since the experimental subjects were relatively unfamiliar with the experimental decision scenario, Laskey and Fischer felt that improved preference stability and predictability could be realized for more familiar decision tasks. In addition, the authors concluded that the effects of response error would be minimal if statistical models were fitted to a relatively large number of responses. Thus, they argue that statistical modeling procedures provide an attractive alternative to conventional decision analytic techniques when eliciting individual preferences. However, they recommend that comparative studies of statistical modeling and standard assessment technologies be conducted to determine whether statistical procedures are better than traditional assessment methodologies in filtering out response errors.

The currently available research represents an important, but as yet incomplete, examination of the effects of potential assessment errors on the modeling of multiattributed preferences. To a large extent the existing research has not systematically compared the sensitivity of alternative assessment technologies to various sources of elicitation error. Thus, there is a need for describing the pre-

ference prediction performances of alternatively encoded MAUFs in the presence of noisy responses, incomplete descriptions of the attribute set, and misspecified model forms and parameters.

4.5 Summary

This chapter reviewed the available research describing the effects of different attribute weighting schemes, alternative assessment procedures, and potential sources of elicitation error on preference predictions for multidimensional decision outcomes. While the available research is important, a need also exists for studies comparing the preference prediction performances of alternatively encoded MAUFs under various situations. This dissertation both complements and extends the existing literature by comparing the structural ability of four assessment methodologies to replicate a known preference order. Chapter 5 describes this research in detail.

CHAPTER 5 COMPARATIVE STUDY OF MAUF ELICITATION PROCEDURES

5.1 Introduction

The purpose of this research was to examine the structural abilities of the KR, HOPE, SMART, and mathematical programming procedures to model an experimental decision maker's internal preferences for disparate choice alternatives. The experimental elicitation exercises were designed to determine the relative sensitivity of each technique's performance to specific assessment errors.

This chapter describes the study in detail. Section 5.2 presents the experimental MAUF elicitation paradigm. The types of assessment error included in the analysis are described in Section 5.3. Details of each elicitation technique's implementation and the assessment of the experimental subject's MAUFs are provided in Section 5.4. Finally, Section 5.5 presents the criteria used to evaluate the encoded models.

5.2 The Experimental MAUF Elicitation Paradigm

5.2.1 Overview

Because actual complex decision problems do not have a "correct" solution in a real-world setting, it is difficult to compare the performances of alternative elicitation procedures, or to evaluate a procedure's sensitivity to specific elicitation errors. To overcome these difficulties, this study employed a hypothetical decision making environment using a computer to simulate the decision maker. Keeney-Raiffa encoded MAUFs were selected from the applied

decision analysis literature and used to configure the artificial subject's internal preference space for different decision contexts. The researcher assumed the role of the decision analyst. The four elicitation procedures were each used to model the synthetic decision maker's known underlying preferences. To neutralize elicitor induced biases, the analyst performed only the basic tasks necessary to implement each assessment technique. Different assessment errors of known type and magnitude were, however, systematically incorporated into the experimental elicitations to determine their effects on a procedure's performance.

The advantages of a simulated elicitation environment for investigating an assessment procedure's structural robustness are discussed in Subsection 5.2.2. Subsection 5.2.3 provides a detailed description of the hypothetical decision maker's underlying preference constructs.

5.2.2 Advantages of a Simulated Elicitation Environment

In addition to facilitating the comparison of several elicitation techniques over different decision contexts, a simulated decision making environment eliminates those effects which can compromise an elicitation technique's ability to encode valid MAUFs. Behavioral influences, such as intersubject value differences, personal involvement in the decision task, the order in which the assessment techniques are applied, and the analyst's facility in administering a procedure, can easily prejudice a technique's performance (Hobbs 1986). Further, changes in the framing of the decision task have been demonstrated to introduce systematic error into the utility

modeling process (Kahneman and Tversky 1979; Hershey, Kunreuther, and Schoemaker 1982). By eliminating behavioral effects from the experimental elicitation exercises, the structural differences among techniques are separated from other sources of variation that can influence a procedure's ability to assess valid MAUFs. For this reason, a simulation analysis was judged to be superior to a field or laboratory study for investigating the research objective.

5.2.3 Establishing the Experimental Decision Maker's Internal Preference Structures

As noted by von Winterfeldt, Griffin, and Edwards (1984, p.27), "A major problem in the normative study of utility and value is the lack of an external validation criterion against which models of choice can be compared." To overcome this difficulty, the Keeney-Raiffa assessed MAUFs specified in Tables 5-1 and 5-2 were used to formulate the experimental decision maker's "true" preferences for different decision tasks. KR assessed MAUFs were chosen to represent the experimental subject's "true" preferences because of the procedures close ties to the theory underlying MAUA (Barron 1983) and because the technique has been applied to several real-world decision tasks (Keeney 1973, 1977; Keeney and Wood 1977; Keeney and Sicherman 1983; Keeney, Lathrop, and Sicherman 1986).

The four attribute multiplicative model, designated KR4M and listed in Table 5-1, was elicited by Keeney (1979) to rank ten potential pumped storage sites for electricity generation in the southwestern United States. The six attribute multiplicative model, designated KR6M and specified in Table 5-2, was assessed by Keeney

TABLE 5-1

KR4 Multiplicative Utility Function[a]

Attribute	Measure	Range Best	Range Worst
x_1	First Year Cost (1976 Millions)	50	75
x_2	Transmission Line Distance Miles Equivalent	0	800
x_3	Pinyon Juniper Forest (Acres)	0	800
x_4	Riparian Community (Yards)	0	2000

Single Attribute Utility Functions

$$u_1(x_1) = 1.096[1-\exp(0.0975(x_1-75))]$$

$$u_2(x_2) = 4.521[1-\exp(0.000313(x_2-800))]$$

$$u_3(x_3) = 2.519[1-\exp(0.000632(x_3-800))]$$

$$u_4(x_4) = 2.019[\exp(-0.000201(x_4-2000))-1]$$

Scaling Factors:

$k_1 = 0.716$ $k_2 = 0.382$

$k_3 = 0.014$ $k_4 = 0.077$

$K = -0.534$

[a]Assesssed by Keeney (1979) to evaluate pumped storage sites for electricity generation.

TABLE 5-2

KR6 Multiplicative Utility Function[a]

Attribute	Measure	Range Best	Range Worst
x_1	Total Cost (Millions of Pesos)	500	4000
x_2	Capacity (Number of Aircraft Operations/hour)	130	50
x_3	Access Time (Minutes)	12	90
x_4	# people killed or seriously injured/aircraft accident	1	1000
x_5	# people displaced by airport development	2500	250,000
x_6	# people (in thousands) subjected to a high noise level	2	1500

$$u_1(x_1) = 1.2399[1-\exp(0.000469(x_1-4000))]$$

$$u_2(x_2) = 1.0635 - 7.96\exp(-0.040071(x_2))$$

$$u_3(x_3) = 1.369282[1-\exp(0.016799(x_3-90))]$$

$$u_4(x_4) = 1.001001 - 0.001001(x_4)$$

$$u_5(x_5) = 1.01 - 0.000004(x_5)$$

$$u_6(x_6) = 2.307011[1-\exp(0.000379(x_6-1500))]$$

Scaling Constants:

$k_1 = 0.48$ $\quad$ $k_2 = 0.6$

$k_3 = 0.1$ $\quad$ $k_4 = 0.35$

$k_5 = 0.18$ $\quad$ $k_6 = 0.18$

$K = -0.877$

[a]Assessed by Keeney (1973) to analyze airline service to the metropolitan area of Mexico City.

(1973) to evaluate competing air service strategies for the Mexico City metropolitan area. The six univariate utility functions listed here are numerical approximations to the graphical representations provided by Keeney. While the individual expression for capacity, attribute x_2, was initially vector valued, it was modeled as a scalar valued function for this research.

Additive composite models were created by normalizing the original KR4M and KR6M scaling parameters. These additive underlying preference constructs, denoted as KR4A and KR6A, are displayed in Tables 5-3 and 5-4.

Together, the KR4M, KR4A, KR6M, and KR6A multivariate models represented the experimental decision maker's basic preferences for four separate decision contexts. Since the hypothetical subject's "true" underlying evaluations were known, a criterion was available against which the preference predictions of the alternatively encoded MAUFs could be compared.

5.3 Sources of MAUF Elicitation Error

5.3.1 Overview

Of the four types of elicitation error discussed by Barron (1983), two types, random error and model specification error, were incorporated into the hypothetical MAUF assessment exercises. The effects of systematic and substantial error were assumed to be neutral throughout the experimental elicitations. Important sources of systematic error include elicitor effects (Fischoff, Slovic, and Lichenstein 1980), range effects (Krzysztofowicz and Duckstein 1980),

TABLE 5-3

KR4 Additive Utility Function

Attribute	Measure	Range Best	Range Worst
x_1	First Year Cost (1976 Millions)	50	75
x_2	Transmission Line Distance Miles Equivalent	0	800
x_3	Pinyon Juniper Forest (Acres)	0	800
x_4	Riparian Community (Yards)	0	2000

Single Attribute Utility Functions

$$u_1(x_1) = 1.096[1-\exp(0.0975(x_1-75))]$$

$$u_2(x_2) = 4.521[1-\exp(0.000313(x_2-800))]$$

$$u_3(x_3) = 2.519[1-\exp(0.000632(x_3-800))]$$

$$u_4(x_4) = 2.019[\exp(-0.000201(x_4-2000))-1]$$

Scaling Factors:

$k_1 = 0.602$ $k_2 = 0.321$

$k_3 = 0.012$ $k_4 = 0.065$

TABLE 5-4

KR6 Additive Utility Function

Attribute	Measure	Range Best	Range Worst
x_1	Total Cost (Millions of Pesos)	500	4000
x_2	Capacity (Number of Aircraft Operations/hour)	130	50
x_3	Access Time (Minutes)	12	90
x_4	# people killed or seriously injured/aircraft accident	1	1000
x_5	# people displaced by airport development	2500	250,000
x_6	# people (in thousands) subjected to a high noise level	2	1500

$$u_1(x_1) = 1.2399[1-\exp(0.000469(x_1-4000))]$$

$$u_2(x_2) = 1.0635 - 7.96\exp(-0.040071(x_2))$$

$$u_3(x_3) = 1.369282[1-\exp(0.016799(x_3-90))]$$

$$u_4(x_4) = 1.001001 - 0.001001(x_4)$$

$$u_5(x_5) = 1.01 - 0.000004(x_5)$$

$$u_6(x_6) = 2.307011[1-\exp(0.000379(x_6-1500))]$$

Scaling Constants:

$k_1 = 0.254$ $\quad$ $k_2 = 0.317$
$k_3 = 0.053$ $\quad$ $k_4 = 0.185$
$k_5 = 0.095$ $\quad$ $k_6 = 0.095$

and certainty effects (Kahneman and Tversky 1979). Substantial errors represent significant decision maker misjudgments. Regardless of the elicitation technique, errors of this type are potentially present in any application. An analyst can, however, reduce the effects of these errors through careful preparation and administration of the elicitation process, and by conducting consistency checks throughout the modeling exercise.

A more basic issue is whether, in the absence of systematic and substantial error, an MAUF assessment technology is structurally capable of encoding reliable preference models. It is important to determine a procedure's performance in the presence of these errors because decision maker responses are inherently noisy, and underlying utility functional forms and attribute sets cannot always be correctly and completely identified.

A few studies have examined the effects of specification error within elicitation methods. Fischer (1977) observed high correlations between additive and multiplicative KR model evaluations for hypothetical job offers described by city, salary, and type of work. Barron and Person (1979) found that incorrectly specified additive HOPE models (assessed with or without random error) produced larger prediction errors than did noisy, but correctly specified HOPE models. Farmer (1987) determined that for one task the simple additive (but theoretically inappropriate) models predicted as well as, or better, than the "true" multiplicative MAUFs for a single hypothetical decision problem. Using an equally weighted additive model, Barron (1987) discovered that the exclusion of relevant at-

tributes from an assessed model produced incorrect choices. The present study extended the existing research by examining the sensitivity of an elicitation technique's preference predictions to noisy respondent inputs, misspecified composite forms, and missing relevant attributes.

5.3.2 Simulating Noisy Decision Maker Responses

Decision maker preference responses represent the inputs required to implement each MAUF elicitation methodology. However, because an individual's preferences can be imprecise and ephemeral, preference responses can be inconsistent or ambiguous (Fischoff, Slovic, and Lichenstein 1980). Also, because humans are imperfect information processors, internally coherent preferences can be misarticulated (Lindley, Tversky, and Brown 1979). Consequently, a decision maker's expressed judgments can be viewed as consisting of a "true" preference component and an error component (Laskey and Fischer 1987).

For purposes of this study, noisy decision maker inputs were simulated by disturbing the experimental subject's "true" internal preference evaluations. The GGNML subroutine of the International Mathematical and Statistical Library was used to generate random error terms from a normal distribution having a mean of 0, and standard deviations of 0.025 and 0.05. Justification for the standard deviation values is provided in Appendix A. The randomly generated error components were combined with the artificial decision maker's "true" evaluations to yield noisy preference responses.

These noisy responses served as the inputs necessary for each assessment technique to encode the experimental MAUFs.

5.3.3 Model Specification Error

By definition, the assessed MAUF is incorrectly specified when it misrepresents the decision maker's underlying utility form (e.g., encoding an additive expression to model preferences which are actually multiplicative). This error can result from a technique's structural limitations, such as SMART's sole reliance on an additive model, assessor misjudgment, or from noisy inputs which imply an aggregate expression different from the subject's basic preference configuration. To determine the effects of model misspecification on a procedure's preference predictions, incorrect functional forms (i.e.,additive models) were systematically elicited for the KR4M and KR6M underlying preference structures.

The encoded MAUF is incompletely specified when relevant decision attributes are missing. This error results when a decision maker is unable or unwilling to fully articulate every objective comprising his or her internal decision space, or when a direct or proxy attribute cannot be found to measure an identified decision objective.

The effects of missing attributes on preference predictions was investigated by deliberately excluding "minor" attributes from certain MAUF assessments. "Minor" attributes were defined as those having the smallest scaling parameters in the underlying MAUF. Specifically, for the decision contexts involving four attributes, incomplete MAUFs were assessed with attribute x_3 and attributes x_3

and x_4 missing. For the six attribute problems, incomplete models were encoded by dropping attribute x_3 and attributes x_3, x_5, and x_6 from the elicitation process.

5.4 Implementation of the Elicitation Procedures

5.4.1 Overview

In actual decision applications, specific decision maker preference responses provide the necessary inputs for the assessment procedure to encode an MAUF. To assess the experimental MAUFs, the inputs needed to implement each elicitation technology were simulated using the synthetic subject's basic preference models described in Subsection 5.2.3.

This section describes in detail how the experimental KR, HOPE, mathematical programming, and SMART MAUFs were assessed. Each technique was used to model the subject's underlying preferences for the KR4M, KR4A, KR6M, and KR6A decision contexts. Both noisy decision maker responses and model specification error were incorporated into the assessment exercises to determine their effects on a procedure's performance to encode valid preference models. The elicitation scenarios examined are illustrated in Table 5-5.

5.4.2 Assessing the Experimental HOPE MAUFs

The assessment of the experimental HOPE MAUFs followed the study of Barron and Person (1979). For each elicitation scenario, a set of hypothetical consequence profiles defined by an appropriate orthogonal experimental design (see Appendix B) were presented to the subject for evaluation. Using an underlying MAUF, the synthetic

TABLE 5-5

Experimental MAUF Elicitation Scenarios

HOPE, Mathematical Programming and
Keeney-Raiffa Assessed MAUFs

Underlying MAUF	Decision Maker Inputs	Model Specification	
	Noisy	Incorrect	Incomplete
KR4A and KR6A	X		
	X		X
KR4M and KR6M	X		
	X	X	
	X		X
	X	X	X

SMART Assessed MAUFs

Underlying MAUF	Noisy	Incorrect	Incomplete
KR4A and KR6A	X		
	X		X
KR4M and KR6M	X	X	
	X	X	X

subject computed a "true" utility for each profile. Noisy holistic utility responses were simulated by perturbing the error-free judgments with normally distributed additive error components having mean 0 and standard deviations of 0.025 and 0.05. In other words,

Holistic Utility Response = "True" Evaluation + Error.

Since utility values are scaled to fall between 0 and 1, any noisy evaluation outside these limits was set at its appropriate bound. The simulated holistic responses provided the inputs necessary for HOPE to estimate an MAUF.

Additive aggregate models were forced on responses simulated from a multiplicatively configured decision space to examine the sensitivity of HOPE preference predictions to incorrectly specified composite forms. The experimental subject's noisy holistic responses for profiles defined by an orthogonal design appropriate for the reduced attribute sets were used to encode incompletely specified MAUFs.

5.4.3 Assessing the Experimental Mathematical Programming MAUFs

Rather than convert a multiattribute measurable value function to an MAUF, the experimental MAUFs were estimated directly. Justification for this approach is provided by the mathematical relationship between single attribute value and utility functions when the coefficient of relative risk attitude, rv(v), equals 0. In such cases $u(x_i)$ is linearly related to $v(x_i)$. One instance where rv(v) = 0 occurs when r(x) and m(x) both equal 0, implying risk neutrality and constant marginal value at x, respectively. Another instance when rv(v) = 0 is when r(x) and m(x) are constant and equal to one

another. Dyer and Sarin (1982) conjectured that constant but different values for r(x) and m(x) are empirically unlikely. Since the experimental subject's underlying univariate utility functions reflect either a neutral or constant risk attitude, it was assumed that the subject's value and utility functions were simple linear transformations of one another. If the functional form of the two expressions is assumed to be the same, then for all practical purposes, the two expressions are identical. Thus, single attribute utility functions, rather than single attribute measurable value functions, were estimated directly using the NLP model described in (24)-(27).

NLP problems were formulated from certainty equivalent responses for 50-50 scalar valued lotteries established for each decision attribute. The certainty equivalent for a lottery is that amount of the attribute which makes the decision maker indifferent between accepting the gamble implied by the lottery and receiving that amount of the attribute for sure (Moskowitz and Wright 1979). Noisy certainty equivalent responses were simulated as follows:

1) the expected utility of each lottery and the absolute difference in utilities between the lottery's outcomes were computed using the appropriate underlying single attribute utility equation;

2) an error component, equal to a randomly generated percentage of the absolute difference in utilities, was computed and added to each lottery's expected utility, to create a noisy expected utility value; and

3) using the appropriate underlying univariate utility function, the experimental subject calculated an attribute value whose utility was equal to the lottery's 'noisy' expected utility. This attribute value became the decision maker's 'noisy' certainty equivalent response.

This approach is illustrated in Figure 5-1.

The NLP objective function was formulated using certainty equivalent responses with observed utility values of 1, 0.75, 0.5, 0.25, and 0. The best and worst attribute values were arbitrarily assigned utilities of 1 and 0, respectively. The three intermediate values were obtained using the variable certainty equivalent method (see von Winterfeldt and Edwards, 1986, pp. 253-254, or Keeney and Raiffa, 1976, p.252). The NLP constraint set was formulated by computing the decision maker's risk premiums for lotteries with comparable outcome ranges. The NLP was solved using the General Interactive Optimizer (GINO) software of Liebman, Lasdon, Schrage, and Waren (1984).

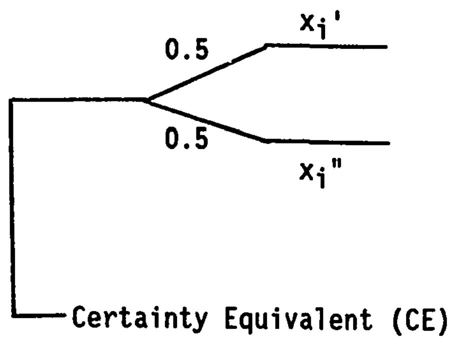
To determine the composite scaling parameters, the experimental subject provided preference and strength of preference responses for pairs of randomly generated consequence profiles as described below. These responses became the inputs of the goal programming problem described in (15)-(21). The resulting linear and nonlinear GP problems were solved using the Linear INteractive Discrete Optimizer (LINDO) software (Schrage 1986) and GINO, respectively.

To simulate the necessary preference responses, the decision maker compared each pair of profiles using an underlying MAUF. To reflect the imprecision inherent in such comparisons, each profile's

FIGURE 5.1

Simulation of Noisy Certainty Equivalent Responses

The decision maker provided a certainty equivalent response to the 50-50 lottery having outcomes x_i' and x_i''.



The "true" Expected Utility of the Lottery, $E[u_i(L)]$,

$$= 0.5u_i(x_i') + 0.5u_i(x_i'')$$

where u_i is the univariate utility function for attribute i.

By definition,

$$u_i(CE) = E[u_i(L)]$$

which implies that

$$CE = u_i^{-1}E[u_i(L)].$$

The noisy expected utility of the lottery was calculated as follows:

$$\text{Noisy } E[u_i(L)] = E[u_i(L)] + (N * |u_i(x_i') - u_i(x_i'')|)$$

where N is a normally distributed random variable having mean 0 and standard deviation of 0.025 or 0.05.

Therefore, the noisy CE reponse

$$= u_i^{-1}[0.5u_i(x_i') + 0.5u_i(x_i'') + (N * |u_i(x_i') - u_i(x_i'')|)]$$

"true" utility was perturbed by a normally distributed additive error term having mean 0 and standard deviation of 0.025 and 0.05. The subject's revealed preference became the profile in each pair having the largest noisy utility value.

Strength of preference responses were simulated from the differences in the noisy utilities for each pair of profiles evaluated. Differences between 0 and 0.09 simulated an intensity of preference response of 1. Differences between 0.1 and 0.19 simulated an intensity of preference of 2, etc. While there is no a priori reason to believe that differences between profiles computed from an underlying MAUF would be identical to those computed from an underlying MAMVF, studies do exist which suggest that practical distinctions between utility and value are small (Barron, von Winterfeldt, and Fischer 1984; von Winterfeldt, Griffin, and Edwards 1984; von Winterfeldt and Edwards 1986).

Barron, von Winterfeldt, and Fischer (1984) found that in most instances a linear relation between utility and value offered a good fit that could not be significantly improved upon when subjected to an exponential transformation derived from theoretical relationships. An experiment by von Winterfeldt, Griffin, and Edwards (1984) revealed that for most subjects, utility and value were equal in multiattributed conditions. Von Winterfeldt and Edwards (1986) take the position that there is little experimental evidence to suggest that drastic differences between utility and value exist. For these reasons, it is believed that strength of preference responses simulated from an underlying MAUF do not adversely affect the procedure's

operationalization nor its structural ability to derive a set of scaling parameters.

Incorrectly specified mathematical programming MAUFs were encoded by modeling additive aggregate forms whenever the underlying representation was multiplicative. Incompletely specified models were assessed by omitting the attributes described in Subsection 5.3.3 from the pairwise preference comparisons and hence from the GP problems.

5.4.4 Assessing The Experimental Keeney-Raiffa MAUFs

The KR conditional utility functions were encoded by fitting an equation to a set of observed utility values. The observed utilities were revealed by the subject's noisy certainty equivalent responses to a set of 50-50 lotteries as described in Subsection 5.4.3.

A set of independent equations, established from decision maker indifference responses and the estimated conditional utility functions, were solved to determine each MAUF's scaling parameters. All "true" responses were distorted by a percentage error term (having mean 0 and standard deviation of 0.025 and 0.05) to simulate noisy inputs. A generic illustration of the approach used to formulate the equations is given below for the three attribute case.

The largest scaling parameter is first identified from the subject's simulated preferences for the following profiles:

$$(x_1^*,x_2^o,x_3^o),\ (x_1^o,x_2^*,x_3^o),(x_1^o,x_2^o,x_3^*).$$

Assume the decision maker prefers the profile (x_1^*,x_2^o,x_3^o). This implies that k_1 is the largest scaling parameter.

Next, the decision maker is asked to provide a value for attribute x_1 (e.g., x_1') so that the alternatives (x_1', x_2^o, x_3^o) and (x_1^o, x_2^*, x_3^o) are equally preferable. Using an underlying MAUF, a value which establishes this indifference is simulated. Since utility values have been scaled between 0 and 1, this implies

$$k_1 u_1(x_1') = k_2. \tag{28}$$

Likewise, the synthetic subject simulates another value for attribute x_1 (e.g., x_1'') so that he or she is indifferent between the profiles (x_1'', x_2^o, x_3^o) and (x_1^o, x_2^o, x_3^*). Equating utilities yields

$$k_1 u_1(x_1'') = k_3. \tag{29}$$

If the parameters are being elicited for an additive composite model, then consistency requires that

$$k_1 + k_2 + k_3 = 1, \tag{30}$$

and equations (28)-(30) are solved for the values of the k_i's.

If the composite utility form is multiplicative, then consistency requires that (30) be replaced by

$$1 + K = (1 + Kk_1)(1 + Kk_2)(1 + Kk_3). \tag{31}$$

Since equations (28), (29), and (31) contain four variables, an additional equation must be formulated to determine a solution for k_1, k_2, k_3 and K. This equation is established by having the subject simulate a probability, p_1, so that he or she is indifferent between the certain outcome (x_1^*, x_2^o, x_3^o) and the lottery offering outcome $\underline{x}^*$ with a probability p_1 and outcome $\underline{x}^o$ with probability $1 - p_1$. By equating expected utilities, it can be seen that

$$k_1 = p_1. \tag{32}$$

The system of equations (28), (29), (31), and (32) are then solved to yield the scaling parameters for a multiplicative MAUF.

Incorrectly specified KR models were elicited by fitting additive composite forms to underlying multiplicative preference configurations. Incompletely structured models were assessed by modeling reduced attribute sets as described in Subsection 5.3.2.

5.4.5 Assessing the Experimental SMART MAUFs

The SMART univariate utility expressions were estimated by fitting an equation to a straight line connecting each attribute's least and most preferred values. The linear expressions were additively combined with decision maker simulated attribute weights to form the experimental SMART MAUFs.

Various weighting techniques have been used by SMART. Unlike the KR methodology which views the k_i's as scaling constants necessary to match the units of one attribute with the units of another, SMART parameters explicitly involve the concept of attribute importance (von Winterfeldt and Edwards 1986). To reflect this difference, SMART assessed weights (parameters) are denoted w_i.

The rank exponent weighting technique, a variation of the preference preserving ratio estimation technique, was used to derive the synthetic decision maker's attribute importance weights. The rank exponent weighting plan was chosen because it considers the decision maker's judgments about the dispersion of weights. Rank exponent weights are defined by the following (see von Winterfeldt and Edwards, 1986, p.284):

$$w_i = (n + 1 - R_i)^z / \sum_{i=1}^{n} R_i^z, \tag{33}$$

where w_i is the weight for the ith attribute, n is the number of attributes considered, R_i is the importance order rank for attribute i with the most important attribute assigned a rank of 1 and the least important attribute assigned a rank of n, and z is estimated from

$$w_i/w_j = (n + 1 - R_i)^z/(n + 1 - R_j)^z \qquad (34)$$

for some pair of attributes (e.g., the most and least important).

To implement the weighting procedure the underlying MAUF scaling parameters (i.e., the k_i's) were degraded by an error component reflecting a percentage of the "true" scaling constant. That is,

Noisy Parameter Value = "True" Parameter * (1 + N)

where N is a normally distributed random variable with mean 0 and standard deviation of either 0.025 or 0.05. The largest noisy parameter was assigned a rank of 1, the next largest a rank of 2, etc., until all n attributes were ordered. The ratio of the most to least important attribute was used to calculate an estimate of z using (34). Noisy weights for every attribute were simulated using (33). These noisy weights were combined with the linear univariate utility equations to form the MAUF.

Because SMART assumes an additive representation, the procedure incorrectly specifies any MAUF whose underlying form is nonadditive. Thus, SMART assessed MAUFs were incorrectly specified for those decision contexts where the synthetic subject's internal decision space was ordered by the KR4M and KR6M expressions. Incompletely specified SMART models were encoded by excluding attributes from the elicitation exercise as discussed in Subsection 5.3.2.

5.4.6 Replicating the Experimental Elicitations

Findings derived from the inputs of a single decision maker could be biased by atypical responses. To reduce the effects of anomalous inputs, a field or laboratory study would assign each elicitation procedure to several subjects. To lessen the influence of anomalous simulated decision maker responses on this study's findings, the experimental elicitations were replicated. Specifically, ten MAUFs were encoded at both levels of noisy decision maker response using each assessment procedure for every elicitation scenario examined.

5.5 Validating the Assessed MAUFs

5.5.1 Overview

While decision scientists recognize the importance of validating alternative elicitation procedures, no universal validation criterion has yet been developed. A primary problem in validating MAUFs is the general lack of external criteria by which inherently subjective inputs and models can be judged. Because objective validation standards are not readily available, applied decision researchers have devised substitute validation plans: convergent validation and predictive validation. This section describes these two validation strategies, and the criteria by which the preference predictions of the experimental MAUFs were analyzed.

5.5.2 Convergent Validation

Convergent validation is based on the principle that different procedures for measuring an underlying construct (i.e., utility)

should be highly correlated. Thus, high correlations are said to validate both models, while low correlations invalidate at least one of the procedures. Researchers have employed the convergent validation strategy to determine the reliability of an elicited model, and to investigate whether or not predictions are sensitive to the assessment procedure used. Fischer (1977) computed the coefficient of determination to measure the degree of linear convergence between holistic and decomposed utilities. Barron (1980) employed the same measure to demonstrate the degree of convergence between holistic responses and HOPE predictions. A major shortcoming of the convergent validation criterion for comparing alternative assessment technologies is that the approach does not guarantee validity since competing techniques can possess the same systematic bias.

5.5.3 Predictive Validation

Predictive or estimative validation measures how well an elicited function replicates a "true" set of preference evaluations. The "true" evaluations can be represented by an external model or criterion used by expert decision makers (Stillwell, Seaver, and Edwards 1983; Adelman, Sticha, and Donnell 1984), a function taught to experimental subjects (von Winterfeldt, Griffin, and Edwards 1984), or an assumed "true" function used in a simulation analysis (Barron and Person 1979).

The researcher's definition of what constitutes a significant difference in preference predictions is a key component of any predictive validation study. Hobbs (1986) noted that significant differences in preference predictions can be defined to occur when the

assessed MAUF incorrectly ranks the most preferred alternative, the best few alternatives, or all decision alternatives. In part, the researcher's choice of a definition will determine the criterion by which predictive validity is measured. For example, some studies have used correlation analysis to gauge an elicited model's predictive reliability (Stillwell, Barron, and Edwards 1983; Farmer 1987). While high correlations reflect a model's ability to correctly rank the entire set of decision alternatives, they can be a misleading indicator of a model's ability to predict a decision maker's most preferred choice (Hobbs, 1986; and Barron, 1987a, 1987b).

From a practical point of view, the problem of selecting a best alternative may be how close the elicited model's predicted choice is in value to the foregone "true" preference. In other words, predicting a first preference which is the decision maker's underlying second choice may not produce a serious loss in value. What is important, however, is avoiding large losses in utility which can occur when the assessed model selects a "truly" inferior alternative as a first preference. Measures which have computed the proportion of utility or value loss to quantify the degree of conformity between predicted and "true" evaluations have been used by Fischer, Damodaran, Laskey, and Lincoln (1987), Johnson and Payne (1985), and Barron (1980, 1987a).

5.5.4 Analyzing The Preference Predictions Of The Elicited MAUFs

The assessed MAUFs evaluated five separate sets of decision consequences consisting of eight randomly generated disparate alternatives. No alternative within a set was dominated by (i.e., in-

ferior to) its competitors. The sets of consequences evaluated are contained in Appendix C. The performance of the elicited models was determined by comparing their predicted preferences to the synthetic subject's known "true" choices.

The degree of conformity between the assessed and "true" evaluations was quantified using several measures. To determine an assessed model's overall predictive capabilities, nonparametric measures of correlation, Spearman's rho and Kendall's tau, were computed. To quantify differences between an MAUF's predicted first preferences and the decision maker's "true" first choice, the proportion of utility loss (PUL) was determined. This measure is defined as follows:

$$PUL = \frac{\text{True Utility } (C_i^*) - \text{True Utility } (C_i')}{\text{True Utility } (C_i^*) - \text{True Utility } (C_i^0)} \tag{35}$$

where, C_i^* and C_i^0 denote the most and least preferred consequences selected by the "true" MAUF, and C_i' the alternative selected by the encoded model. The numerator represents the loss in utility resulting from selecting decision alternative C_i' rather than C_i^*. The denominator normalizes this loss relative to the difference between the expected utilities of the most preferred and least preferred alternatives according to the "true" MAUF. This measure was used to assess the first preference prediction capabilities of the alternatively elicited MAUFs. Table 5-6 illustrates the calculation of Spearman's rho, Kendall's tau, and the proportion of utility loss for a hypothetical case.

TABLE 5-6

Measuring the Assessed MAUF's Predictive Validity

Consequence	Utility Value: True	Utility Value: Predicted	Ranking[a]: True	Ranking[a]: Predicted	Pairs[b]: Concordant	Pairs[b]: Discordant
1	0.894	0.880	4	5	6	1
2	0.861	0.855	6	6	6	0
3	0.900	0.920**	2	1	4	1
4	0.898	0.890	3	4	3	1
5	0.915*	0.902	1	2	3	0
6	0.880	0.892	5	3	2	0
7	0.812°	0.829	8	7	0	1
8	0.820	0.805	7	8	0	0

* Best Consequence According to True Utility
° Worst Consequence According to True Utility
** Best Consequence According to Predicted Utility

[a] Spearman's Rho = 0.881

[b] Kendall's Tau = 0.714

Proportion of Utility Loss for the Most Preferred Consequence:

$$= \frac{\text{TRUE } U(C_i^*) - \text{TRUE } U(C_i')}{\text{TRUE } U(C_i^*) - \text{TRUE } U(C_i^o)}$$

$$= (0.915 - 0.900)/(0.915 - 0.812) = 0.1456$$

where C_i^* and C_i^o denote the most and least preferred consequences selected by the "true" MAUF, and C_i' is the consequence selected by the predicted model.

5.6 Summary

This chapter presented the plan for examining the structural abilities of four MAUF elicitation procedures to recover a synthetic decision maker's underlying preferences for four decision contexts. The experimental elicitation paradigm, the advantages of a simulation study, and the sources of elicitation error to be included in the assessment exercises were discussed. The implementation of each assessment technology was described in detail. Also presented were the measures by which the performances of the assessed MAUFs were judged. The results of this study are presented in Chapter 6.

CHAPTER 6 RESULTS AND DISCUSSION

6.1 Introduction

This chapter presents the results of the study described in Chapter 5. Section 6.2 discusses the predictive precision of the alternatively encoded MAUFs. Section 6.3 describes the ability of the elicited scaling constants to characterize the underlying willingness of the experimental decision maker to enact tradeoffs between pairs of competing objectives. A summary of the study's findings is provided in Section 6.4.

6.2 Assessed MAUF Preference Predictions

6.2.1 Overview

Differences in the predictive performances of the competing assessment technologies were judged by (1) how well the assessed models revealed the subject's "true" ordinal preferences, and (2) how well the assessed models correctly identified the subject's most preferred decision alternative. Performance criterion (1) was determined using both Kendall's Tau and Spearman's Rho. However, since both measures led to the same conclusions, findings are reported in terms of Spearman's Rho only. Each procedure's performance on criterion (2) was evaluated by the percentage of first preferences correctly identified, while the "regret" associated with an incorrectly predicted first choice was measured by the proportion of utility loss (PUL).

Results are presented separately for the four and six attribute decision tasks by elicitation scenario. Findings are separated by the level of decision maker response error only for those cases where mean performances were judged to be sensitive to the magnitude of noise involved.

6.2.2 Ordinal Predictive Performances

Results are reported in both graphical and tabular form. Figures 6-1 and 6-2 graphically display the overall ability of each elicitation technique to capture the subject's "true" ordinal preferences for the four and six attribute tasks, respectively. The mean measure of performance is Spearman's Rho. The horizontal axis of each graph lists the elicitation scenario examined. Thus, "complete-correct" denotes completely and correctly specified models, "complete-incorrect" denotes completely but incorrectly specified models, etc. Tables 6-1 and 6-2 list, also by elicitation scenario, the mean correlation, standard deviation, and range for each assessment methodology for the four and six attribute decision problems, respectively.

A Kruskal-Wallis one-way analysis of variance revealed significant differences ($p < 0.0001$) in predictive performances among assessment procedures within each elicitation scenario. To determine which mean performances were different, pairwise comparisons between procedures within an elicitation scenario were conducted using the Wilcoxon rank sum test. Results of the paired comparison tests are given in Tables 6-3 and 6-4.

Figure 6.1

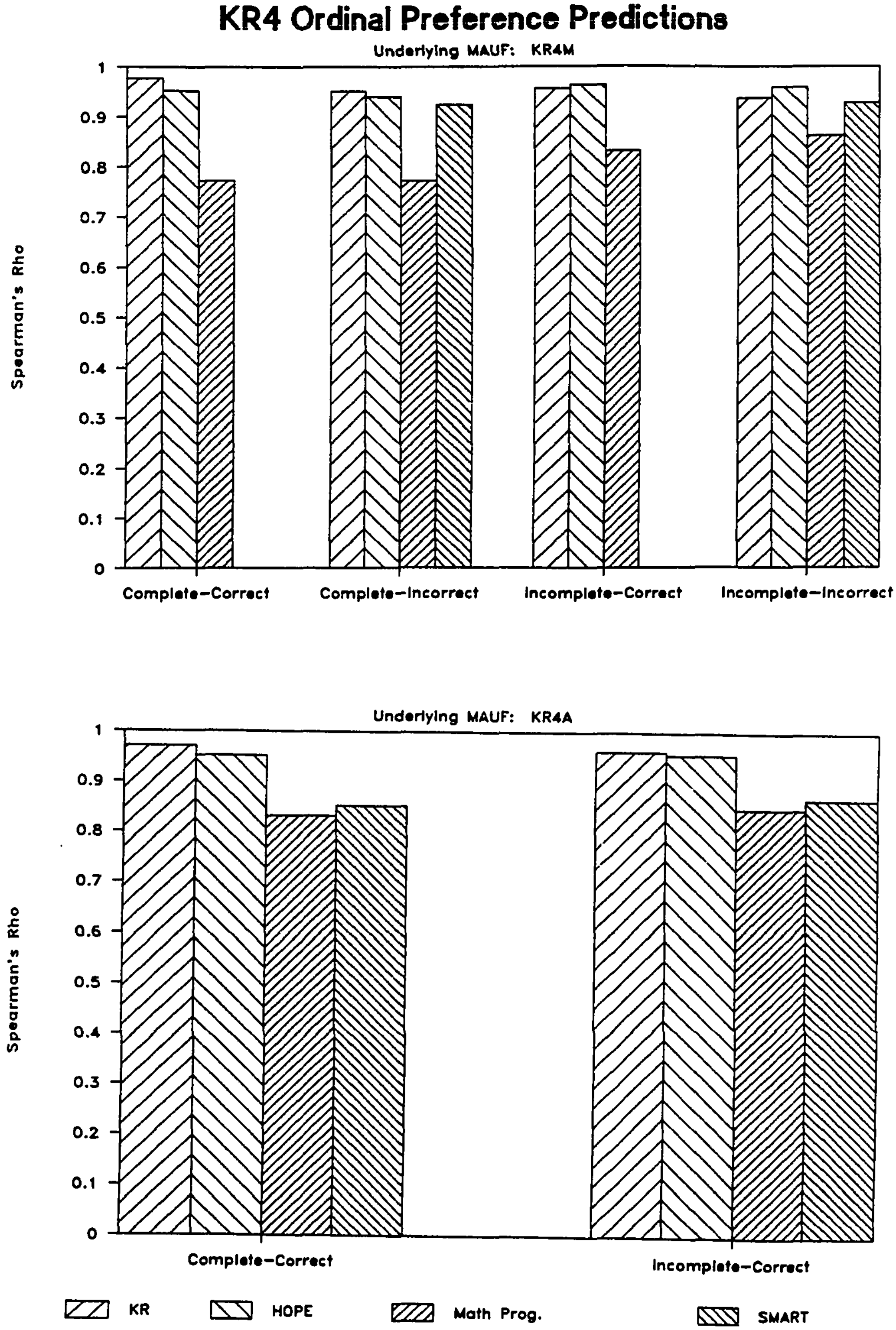

KR4 Ordinal Preference Predictions
Underlying MAUF: KR4M
Spearman's Rho
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Complete–Correct
Complete–Incorrect
Incomplete–Correct
Incomplete–Incorrect
Underlying MAUF: KR4A
Spearman's Rho
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Complete–Correct
Incomplete–Correct
KR
HOPE
Math Prog.
SMART

Figure 6.2

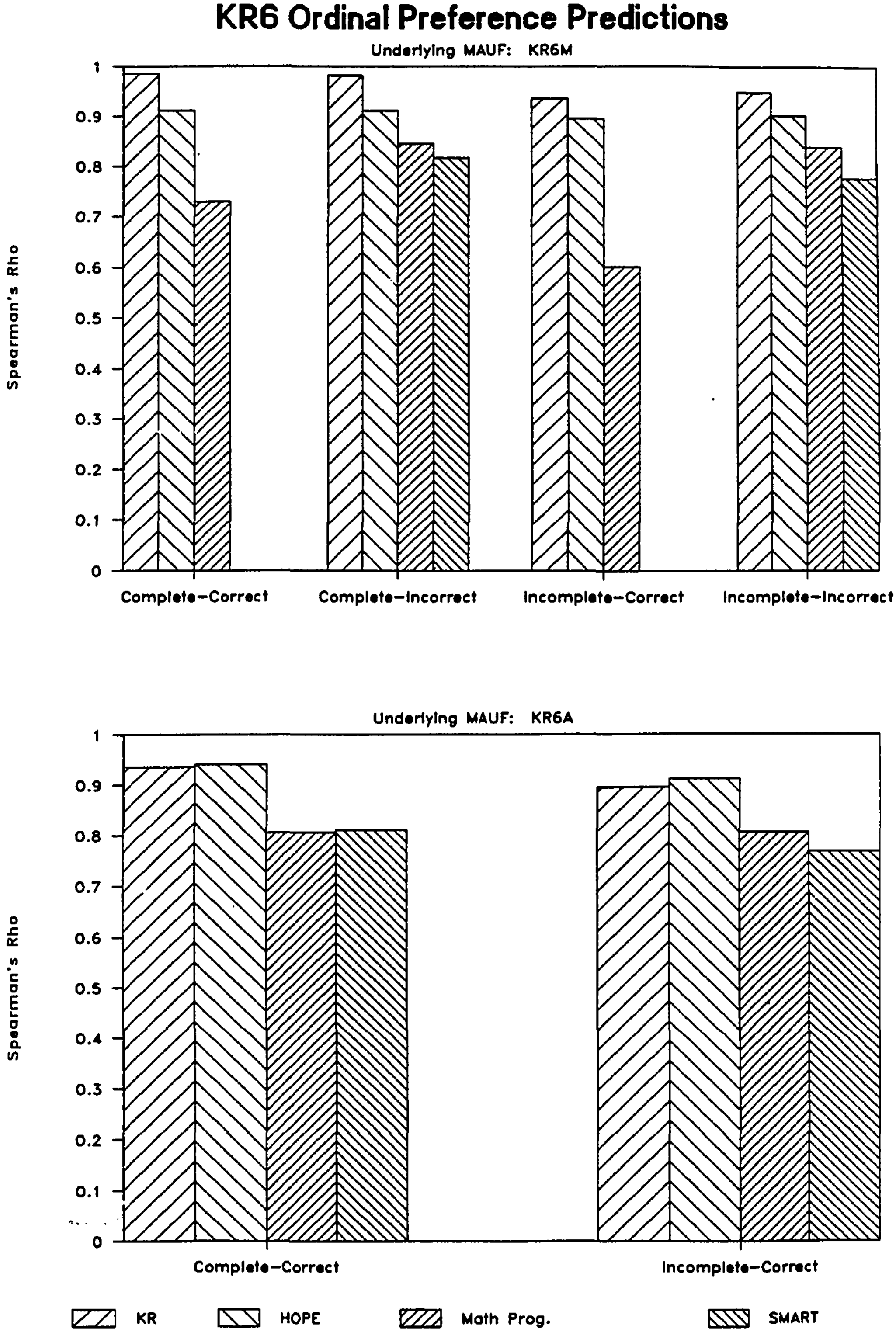

KR6 Ordinal Preference Predictions
Underlying MAUF: KR6M
Spearman's Rho
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Complete–Correct
Complete–Incorrect
Incomplete–Correct
Incomplete–Incorrect
Underlying MAUF: KR6A
Spearman's Rho
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Complete–Correct
Incomplete–Correct
KR
HOPE
Math Prog.
SMART

TABLE 6-1

KR4 Ordinal Preference Predictions

Elicitation Scenario Technique	Mean	Standard Deviation	Range
	Underlying MAUF: KR4M		
Complete-Correct			
Keeney-Raiffa	0.977	0.038	0.190
HOPE	0.953	0.044	0.238
Math Prog.	0.774	0.072	1.424
Complete-Incorrect			
Keeney-Raiffa	0.951	0.058	0.286
HOPE	0.940	0.054	0.333
Math Prog.	0.773	0.251	1.714
SMART	0.924	0.046	0.143
Incomplete-Correct			
Keeney-Raiffa	0.955	0.060	0.212
HOPE	0.963	0.043	0.236
Math Prog.	0.831	0.184	1.048
Incomplete-Incorrect			
Keeney-Raiffa	0.937	0.070	0.414
HOPE	0.959	0.045	0.236
Math Prog.	0.861	0.123	0.619
SMART	0.927	0.054	0.178
	Underlying MAUF: KR4A		
Complete-Correct			
Keeney-Raiffa	0.971	0.043	0.214
HOPE	0.951	0.056	0.238
Math Prog.	0.832	0.038	1.643
SMART	0.852	0.115	0.333
Incomplete-Correct			
Keeney-Raiffa	0.961	0.053	0.398
HOPE	0.955	0.051	0.333
Math Prog.	0.849	0.142	0.810
SMART	0.868	0.091	0.333

TABLE 6-2

KR6 Ordinal Preference Predictions

Elicitation Scenario Technique	Mean	Standard Deviation	Range
	Underlying MAUF: KR6M		
Complete-Correct			
Keeney-Raiffa	0.986	0.015	0.071
HOPE	0.911	0.080	0.381
Math Prog.	0.729	0.095	1.047
Complete-Incorrect			
Keeney-Raiffa	0.982	0.016	0.048
HOPE	0.912	0.079	0.381
Math Prog.	0.845	0.093	0.452
SMART	0.817	0.132	0.367
Incomplete-Correct			
Keeney-Raiffa	0.936	0.055	0.167
HOPE	0.895	0.091	0.476
Math Prog.	0.600	0.284	1.495
Incomplete-Incorrect			
Keeney-Raiffa	0.949	0.052	0.310
HOPE	0.904	0.081	0.476
Math Prog.	0.840	0.110	0.548
SMART	0.777	0.179	0.762
	Underlying MAUF: KR6A		
Complete-Correct			
Keeney-Raiffa	0.936	0.038	0.119
HOPE	0.941	0.063	0.262
Math Prog.	0.806	0.117	0.428
SMART	0.812	0.117	0.357
Incomplete-Correct			
Keeney-Raiffa	0.895	0.090	0.381
HOPE	0.912	0.077	0.357
Math Prog.	0.807	0.140	0.857
SMART	0.769	0.168	0.595

TABLE 6-3

KR4 Statistical Analysis - Wilcoxon Rank Sum Test

Elicitation Scenario Paired Analysis	T-Test Approximate KR4M	Significance[a] KR4A	Sample Size
Complete-Correct			
KR - HOPE	0.0001	0.0009	100
KR - Math Prog.	0.0001	0.0001	100
KR - SMART	*	0.0001	100
HOPE - Math Prog.	0.0001	0.0001	100
HOPE - SMART	*	0.0001	100
Math Prog. - SMART	*	0.0551	100
Complete-Incorrect			
KR - HOPE	0.0024		100
KR - Math Prog.	0.0001		100
KR - SMART	0.0001		100
HOPE - Math Prog.	0.0001		100
HOPE - SMART	0.0003		100
Math Prog. - SMART	0.0256		100
Incomplete-Correct			
KR - HOPE	0.4197	0.1008	200
KR - Math Prog.	0.0001	0.0001	200
KR - SMART	*	0.0001	200
HOPE - Math Prog.	0.0001	0.0001	200
HOPE - SMART	*	0.0001	200
Math Prog. - SMART	*	0.8983	200
Incomplete-Incorrect			
KR - HOPE	0.0042		200
KR - Math Prog.	0.0001		200
KR - SMART	0.0004		200
HOPE - Math Prog.	0.0001		200
HOPE - SMART	0.0001		200
Math Prog. - SMART	0.0001		200

[a]test from NPAR1WAY SAS program

* No data for these scenarios because the SMART technique is structured for solving only additive models.

TABLE 6-4

KR6 Statistical Analysis - Wilcoxon Rank Sum Test

Elicitation Scenario Paired Analysis	T-Test Approximate Significance[a] KR6M	KR6A	Sample Size
Complete-Correct			
KR - HOPE	0.0001	0.0121	100
KR - Math Prog.	0.0001	0.0001	100
KR - SMART	*	0.0001	100
HOPE - Math Prog.	0.0001	0.0001	100
HOPE - SMART	*	0.0001	100
Math Prog. - SMART	*	0.7090	100
Complete-Incorrect			
KR - HOPE	0.0001		100
KR - Math Prog.	0.0001		100
KR - SMART	0.0001		100
HOPE - Math Prog.	0.0001		100
HOPE - SMART	0.0001		100
Math Prog. - SMART	0.3346		100
Incomplete-Correct			
KR - HOPE	0.0001	0.0937	200
KR - Math Prog.	0.0001	0.0001	200
KR - SMART	*	0.0001	200
HOPE - Math Prog.	0.0001	0.0001	200
HOPE - SMART	*	0.0001	200
Math Prog. - SMART	*	0.0088	200
Incomplete-Incorrect			
KR - HOPE	0.0001		200
KR - Math Prog.	0.0001		200
KR - SMART	0.0001		200
HOPE - Math Prog.	0.0001		200
HOPE - SMART	0.0001		200
Math Prog. - SMART	0.0041		200

[a]test calculated from NPAR1WAY SAS program

* No data for these scenarios because the SMART technique is structured for solving only additive models.

KR and HOPE assessment techniques were significantly ($p < 0.0001$) better performers than SMART and mathematical programming regardless of the elicitation scenario considered. Although some differences were detected in the mean performances of the KR and HOPE procedures for the four attribute decision tasks, both techniques produced mean correlations in the mid to high 0.90's. SMART was also a good performer; producing mean correlations above 0.92 when underlying preferences were defined multiplicatively, and above 0.85 when underlying preferences were additive. All three procedures generated relatively consistent predictions as indicated by the range and standard deviations (Table 6-1) of the correlations.

For the six attribute alternatives, KR and HOPE continued to provide accurate recoveries of the subject's underlying preference structures. KR MAUFs significantly ($p < 0.0001$) outperformed all other techniques in each elicitation scenario when the subject's internal preference judgments were configured multiplicatively. No significant differences were detected between KR and HOPE models when preferences were described additively.

SMART also provided good ordinal rankings (i.e., mean correlations between 0.77 and 0.82). However, a decrease in SMART's overall performance from the four attribute evaluations was reflected in lower mean correlations and higher standard deviations. This reduced performance may be due to the "reduced additivity" of the KR6M as compared to the KR4M preferences, or because the six attribute decision profiles were closer in terms of "true" utility and therefore less distinguishable than the four attribute consequences.

The lower mean correlations and wider ranges typically produced by the mathematical programming procedure suggest that its performance was less reliable than those of the other methodologies. In addition, the technique was generally more sensitive to the magnitude of response noise, missing attributes, and the form of the assessed aggregate model than the other procedures. Tables 6-5 and 6-6 list the mean performances and standard deviations, separated by both error level and elicitation scenario, for the mathematical programming predictions for the four and six attribute decision alternatives, respectively.

When choices involved four attributes, the procedure provided accurate preference recoveries (mean correlations above 0.8) at the 0.025 level of error regardless of the elicitation scenario or underlying composite form. At the 0.05 magnitude of response noise, the quality of the ordinal preference predictions deteriorated noticeably. At this noise level, both lower mean correlations (between 0.62 and 0.855) and increased instability in predictions were noted over the 0.025 error rate. Most surprising, were the observed improvements in preference predictions at the 0.05 error level when attributes were omitted from the assessed models.

A possible explanation for this unexpected finding rests in the structural composition of the mathematical programming technique itself. Weber (1987) argued that error based elicitation methods which permit inconsistent decision maker inputs can create difficulties when extended to the case of incomplete information. The mathematical programming procedure is essentially an error based method. The

Table 6-5

Mathematical Programming Technique
KR4 - Ordinal Preference Predictions

Underlying Function Elicitation Scenario	Complete Attributes	Attribute x_3 Missing	Attributes x_3 and x_4 Missing
	0.025 Response Error		
KR4M			
Correctly Specified			
Mean	0.928	0.915	0.839
(SD)	(0.073)	(0.082)	(0.594)
Misspecified			
Mean	0.847	0.903	0.850
(SD)	(0.252)	(0.045)	(0.137)
KR4A			
Correctly Specified			
Mean	0.946	0.896	0.870
(SD)	(0.039)	(0.077)	(0.116)
	0.05 Response Error		
KR4M			
Correctly Specified			
Mean	0.619	0.721	0.743
(SD)	(0.371)	(0.254)	(0.149)
Misspecified			
Mean	0.672	0.837	0.855
(SD)	(0.409)	(0.152)	(0.124)
KR4A			
Correctly Specified			
Mean	0.718	0.801	0.830
(SD)	(0.321)	(0.190)	(0.144)

TABLE 6-6

Mathematical Programming Technique
KR6M - Ordinal Preference Predictions

Elicitation Scenario	Complete Attributes	Attribute x_3 Missing	Attributes x_3 x_5 and x_6 Missing
	0.025 Response Error		
Correctly Specified			
Mean	0.830	0.704	0.488
(SD)	(0.096)	(0.195)	(0.318)
	0.05 Response Error		
Correctly Specified			
Mean	0.627	0.711	0.497
(SD)	(0.254)	(0.261)	(0.269)

technique derives a model's scaling constants by solving a goal programming problem to minimize decision maker preference inconsistencies. In certain instances, it is possible that no utility function will conform exactly to the inconsistent information provided. When this occurs, a function is estimated which best fits the inconsistent inputs according to some error measure. When less inconsistent information is given, the errors are smaller, and the analyst can better fit a function to the available (inconsistent) information.

This effect was most evident in the four attribute decision tasks at the 0.05 level of response error. Here, the omission of "minor" attributes increased the number of "correct" (i.e. reduced the number of inconsistent) inputs to the goal programming model.

This resulted in an increase in the number of replications where the orderings of the scaling parameters of the elicited and the "true" models agreed. Table 6-7 displays the rank orderings of the scaling constants elicited by the mathematical programming technique for the cases where no attributes, one attribute, and two attributes were omitted from the assessed functions. For the KR4M underlying preference structure, when completely specified, the procedure estimated scaling constant values which were correctly ordered, when compared to the "true" rankings, in only 3 of 10 cases. When one attribute was omitted from the assessed functions, the incidence of correct orderings increased to 6 of 10 replications, and to 8 of 10 replications when two attributes were omitted. Similar results were also found for the KR4A underlying decision space. Because the omitted attributes represented relatively "minor" objectives, any deleterious effect their absence had on preference predictions was more than offset by the improved ordering of the assessed parameter values.

For the six attribute decision scenarios (Table 6-6), the omission of one attribute at the 0.05 error level improved the preference predictions of the assessed models from mean correlations of 0.627 to 0.711. This effect was not repeated, however, when three objectives were excluded from the assessed functions. In this case the scaling parameter values of the omitted attributes were sufficiently large that their absence had an adverse influence on the decision maker's inputs. This created two problems. First, an increase in inconsistent information which led to reduced accuracy in assessed parameter estimates. Second, a sufficiently large loss of relevant information

TABLE 6-7

SCALING PARAMETER ORDERING
KR4 - Mathematical Programming Procedure
0.05 Error Level

Replication	k_i's Complete Attributes	k_i's Attribute x_3 Missing	k_i's Attribute x_3 and x_4 Missing
Correct Ordering	1, 2, 4, 3	1, 2, 4	1, 2
	Underlying MAUF: KR4M		
Correctly Specified			
Rep. 1	2, 1, 3 and 4*	1, 4, 2	2, 1
Rep. 2	2, 1, 3 and 4*	1, 2, 4	1, 2
Rep. 3	2, 1, 4, 3	1, 2, 4	1, 2
Rep. 4	1, 2, 3 and 4*	1, 2, 4	1, 2
Rep. 5	1, 2, 3, 4	1, 2, 4	1, 2
Rep. 6	1, 3, 4, 2	1, 4, 2	1, 2
Rep. 7	1, 4, 3, 2	1, 4, 2	1, 2
Rep. 8	1, 2, 4, 3	1, 2, 4	1, 2
Rep. 9	1, 2, 4, 3	1, 2, 4	1, 2
Rep. 10	1, 2, 4, 3	2, 4, 1	2, 1
Incorrectly Specified			
Rep. 1	1, 2, 4, 3	1, 2, 4	2, 1
Rep. 2	1, 2, 3 and 4*	1, 2, 4	1, 2
Rep. 3	1, 2, 4, 3	1, 2, 4	1, 2
Rep. 4	1, 2, 3, 4	1, 2, 4	1, 2
Rep. 5	1, 3, 2, 4	1, 2, 4	1, 2
Rep. 6	1, 3, 2, 4	1, 4, 2	1, 2
Rep. 7	1, 3, 2 and 4*	1, 4, 2	1, 2
Rep. 8	1, 2, 4, 3	1, 2, 4	1, 2
Rep. 9	1, 2, 3 and 4*	1, 2, 4	1, 2
Rep. 10	1, 2, 4, 3	1, 2, 4	1, 2
	Underlying MAUF: KR4A		
Rep. 1	1, 2, 4, 3	2, 4, 1	2, 1
Rep. 2	1, 2, 3, 4	2, 1, 4	2, 1
Rep. 3	1, 2, 4, 3	1, 2, 4	1, 2
Rep. 4	1, 3, 2, 4	1, 2, 4	1, 2
Rep. 5	1, 2, 3 and 4*	1, 2, 4	1, 2
Rep. 6	1, 3, 2, 4	1, 4, 2	1, 2
Rep. 7	1, 2, 4, 3	1, 2, 4	1, 2
Rep. 8	1, 2, 4, 3	1, 2, 4	1, 2
Rep. 9	2, 1, 4, 3	2, 1, 4	2, 1
Rep. 10	1, 2, 4, 3	1, 2, 4	2, 1

*denotes tie between scaling parameters

necessary for the encoded models to accurately discriminate among decision consequences. The result was a decrease in ordinal predictive performance from 0.711 with one attribute missing to 0.497 with three attributes excluded.

In summary, the KR procedure typically outperformed the other techniques in terms of ordinal preference predictions. HOPE was a close second, and in some instances provided slightly more accurate predictions than KR. SMART and mathematical programming were the least reliable predictors of ordinal preferences. For the four attribute decision tasks, SMART provided substantially better replications of the underlying multiplicative decision space and marginally better replications of the additive space than did mathematical programming. When decision alternatives involved six attributes, however, mathematical programming was slightly better than SMART. No instances existed, however, where the preferences revealed by SMART or mathematical programming were superior to those revealed by KR or HOPE.

6.2.3 First Preference Predictive Performance

Again, results are presented in both graphical and tabular form. Because the performance of each procedure was sensitive to differences in error levels, findings are displayed separately for the 0.025 and 0.05 magnitudes of response noise.

Figures 6-3 and 6-4 illustrate each procedure's ability to correctly identify the simulated subject's most preferred four attribute alternative at the 0.025 and 0.05 error levels, respectively. The measure of performance is the percentage of first preferences

Figure 6.3

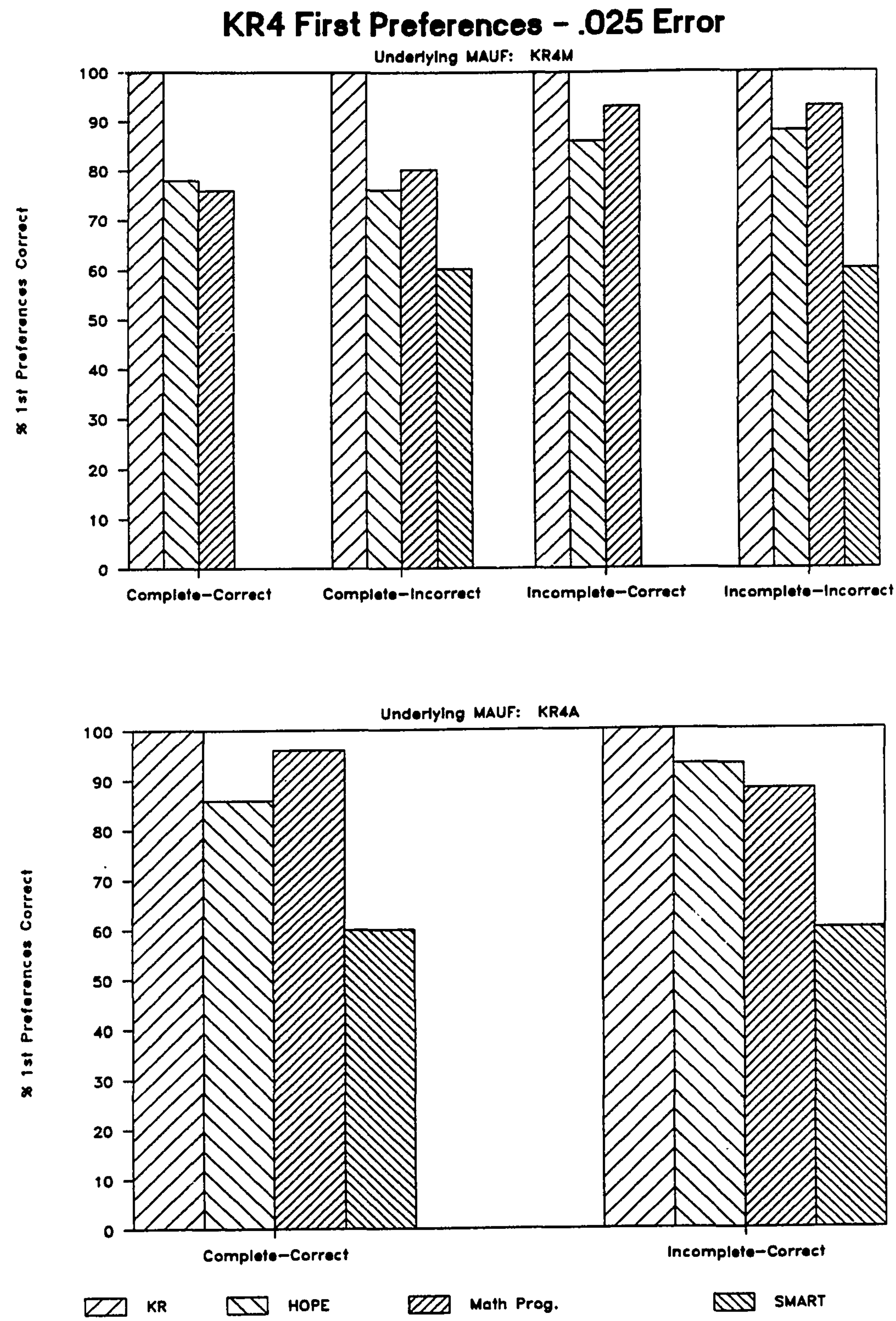
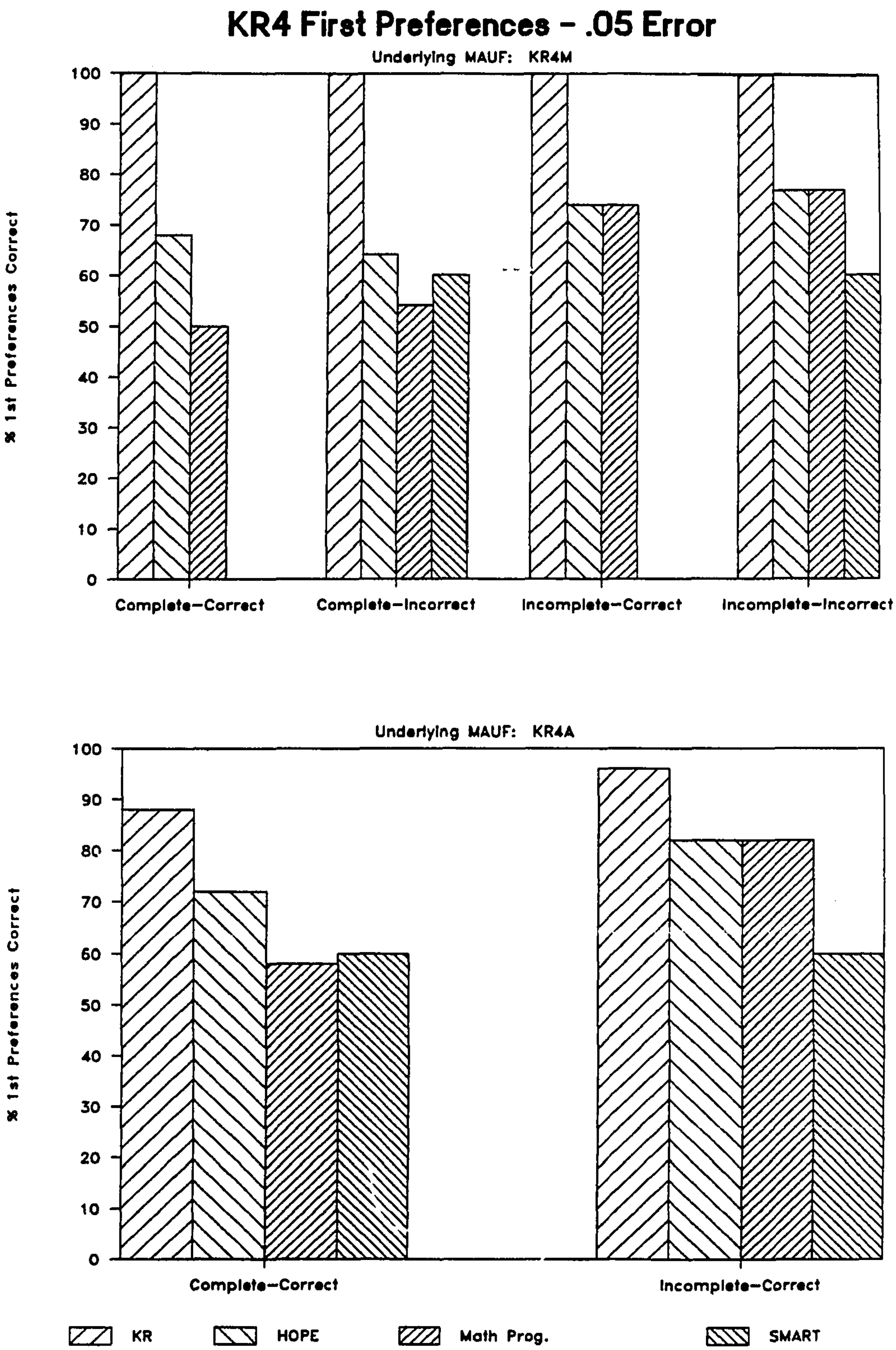

KR4 First Preferences - .025 Error
Underlying MAUF: KR4M
% 1st Preferences Correct
0
10
20
30
40
50
60
70
80
90
100
Complete-Correct
Complete-Incorrect
Incomplete-Correct
Incomplete-Incorrect
Underlying MAUF: KR4A
% 1st Preferences Correct
Complete-Correct
Incomplete-Correct
KR
HOPE
Math Prog.
SMART

Figure 6.4


KR4 First Preferences - .05 Error
Underlying MAUF: KR4M
% 1st Preferences Correct
100
90
80
70
60
50
40
30
20
10
0
Complete—Correct
Complete—Incorrect
Incomplete—Correct
Incomplete—Incorrect
Underlying MAUF: KR4A
% 1st Preferences Correct
100
90
80
70
60
50
40
30
20
10
0
Complete—Correct
Incomplete—Correct
KR
HOPE
Math Prog.
SMART

correctly predicted. The horizontal axis of each graph lists the elicitation scenario examined. In addition to these results, the mean PUL, its largest value, and its frequency, by error level, are shown in Tables 6-8 and 6-9.

As shown, KR assessed models strongly dominated the other techniques. At the 0.025 level of noisy decision maker inputs, KR encoded models always correctly identified the subject's most preferred choice. While some incorrect first preferences were revealed by KR assessed models at the 0.05 level of error, the procedure continued to produce more correct predictions and lower mean PULs than the other three methodologies.

Selecting a second best predictor of first choices was more difficult. At the 0.025 level of error, the percentage of correct first preferences was usually higher for mathematical programming than for HOPE models. Both methods outperformed SMART, whose 60% accuracy was insensitive to the elicitation scenario examined. Differences between the percentage of correct predictions for HOPE and mathematical programming encoded models were, however, small, and mean PULs were relatively insignificant for both procedures. Furthermore, "substantial regret" (i.e. a high PUL) from incorrect first choices occurred only when a completely but incorrectly encoded mathematical programming model led to a PUL of 0.860.

The picture is somewhat clearer at the 0.05 error level where, with respect to the number of correctly revealed first choices, HOPE assessed models weakly dominated those of mathematical programming, and strongly dominated those of SMART. In addition, HOPE's incorrect

TABLE 6-8

KR4 First Preference Predictions - 0.025 Error Response

Elicitation Scenario Technique	% of 1st Preferences Correct	Mean PUL	High PUL	No. of Occurrences
	Underlying MAUF: KR4M			
Complete-Correct				
Keeney-Raiffa	100	0.000	*	
HOPE	78	0.009	0.094	1
Math Prog.	76	0.025	0.224	3
Complete-Incorrect				
Keeney-Raiffa	100	0.000	*	
HOPE	76	0.019	0.315	2
Math Prog.	80	0.056	0.860	1
SMART	60	0.032	0.094	10
Incomplete-Correct				
Keeney-Raiffa	100	0.000	*	
HOPE	86	0.014	0.176	1
Math Prog.	93	0.006	0.175	1
Incomplete-Incorrect				
Keeney-Raiffa	100	0.000	*	
HOPE	88	0.011	0.176	2
Math Prog.	93	0.005	0.094	15
SMART	60	0.032	0.094	20
	Underlying MAUF: KR4A			
Complete-Correct				
Keeney-Raiffa	100	0.000	*	
HOPE	86	0.026	0.200	6
Math Prog.	96	0.001	0.020	2
SMART	60	0.063	0.177	10
Incomplete-Correct				
Keeney-Raiffa	100	0.000	*	
HOPE	93	0.005	0.207	1
Math Prog.	88	0.024	0.298	5
SMART	60	0.063	0.177	20

* Indicates accurate first preference predictions in all cases

TABLE 6-9

KR4 First Preference Predictions - 0.05 Error Response

Elicitation Scenario Technique	% of 1st Preferences Correct	Mean PUL	High PUL	No. of Occurrences
	Underlying MAUF: KR4M			
Complete-Correct				
Keeney-Raiffa	100	0.000	*	
HOPE	68	0.028	0.315	1
Math Prog.	50	0.213	1.000	2
Complete-Incorrect				
Keeney-Raiffa	100	0.000	*	
HOPE	64	0.036	0.315	3
Math Prog.	54	0.147	0.860	3
SMART	60	0.032	0.094	10
Incomplete-Correct				
Keeney-Raiffa	100	0.000	*	
HOPE	74	0.019	0.176	1
Math Prog.	74	0.089	1.000	1
Incomplete-Incorrect				
Keeney-Raiffa	100	0.000	*	
HOPE	77	0.024	0.315	1
Math Prog.	77	0.066	0.701	2
SMART	60	0.032	0.094	20
	Underlying MAUF: KR4A			
Complete-Correct				
Keeney-Raiffa	88	0.012	0.136	4
HOPE	72	0.039	0.298	1
Math Prog.	58	0.110	0.941	1
SMART	60	0.063	0.177	10
Incomplete-Correct				
Keeney-Raiffa	96	0.006	0.136	4
HOPE	82	0.021	0.207	2
Math Prog.	82	0.062	0.849	1
SMART	60	0.063	0.177	20

* Indicates accurate first preference predictions in all cases

recommendations produced, on average, less decision maker "regret" than incorrectly revealed mathematical programming or SMART choices. Finally, unlike HOPE or SMART encoded models, mathematical programming MAUFs revealed, for each elicitation scenario, at least one incorrect first preference which produced a large PUL.

Results for the six attribute cases are presented in Figures 6-5 and 6-6 as well as in Tables 6-10 and 6-11 for the 0.025 and 0.05 error rates, respectively. In general, all techniques were less accurate predictors of the experimental subject's most preferred alternative when the underlying problem space was structured by six rather than four attributes. A possible explanation is that the alternatives described by six attributes were closer in their "true" attractiveness than the consequences described by four attributes.

The relative performances of the techniques when MAUFs were completely and correctly assessed was unaffected by the amount of noise introduced into the elicitation process. Generally, the KR procedure produced more correct first preferences and lower mean PULs than its competitors. HOPE was the next best performer. Noticeably poorer predictions were offered by mathematical programming and SMART. Neither of these methods could achieve a 50% success rate; and substantial "regret" (e.g., PUL of 1.00) was produced by multiplicatively assessed mathematical programming models at the 0.05 error level.

The findings for the remaining elicitation scenarios do not readily reveal a clear winner. For instance, SMART models accurately identified an equal or greater number of KR6M first choices, at both

Figure 6.5

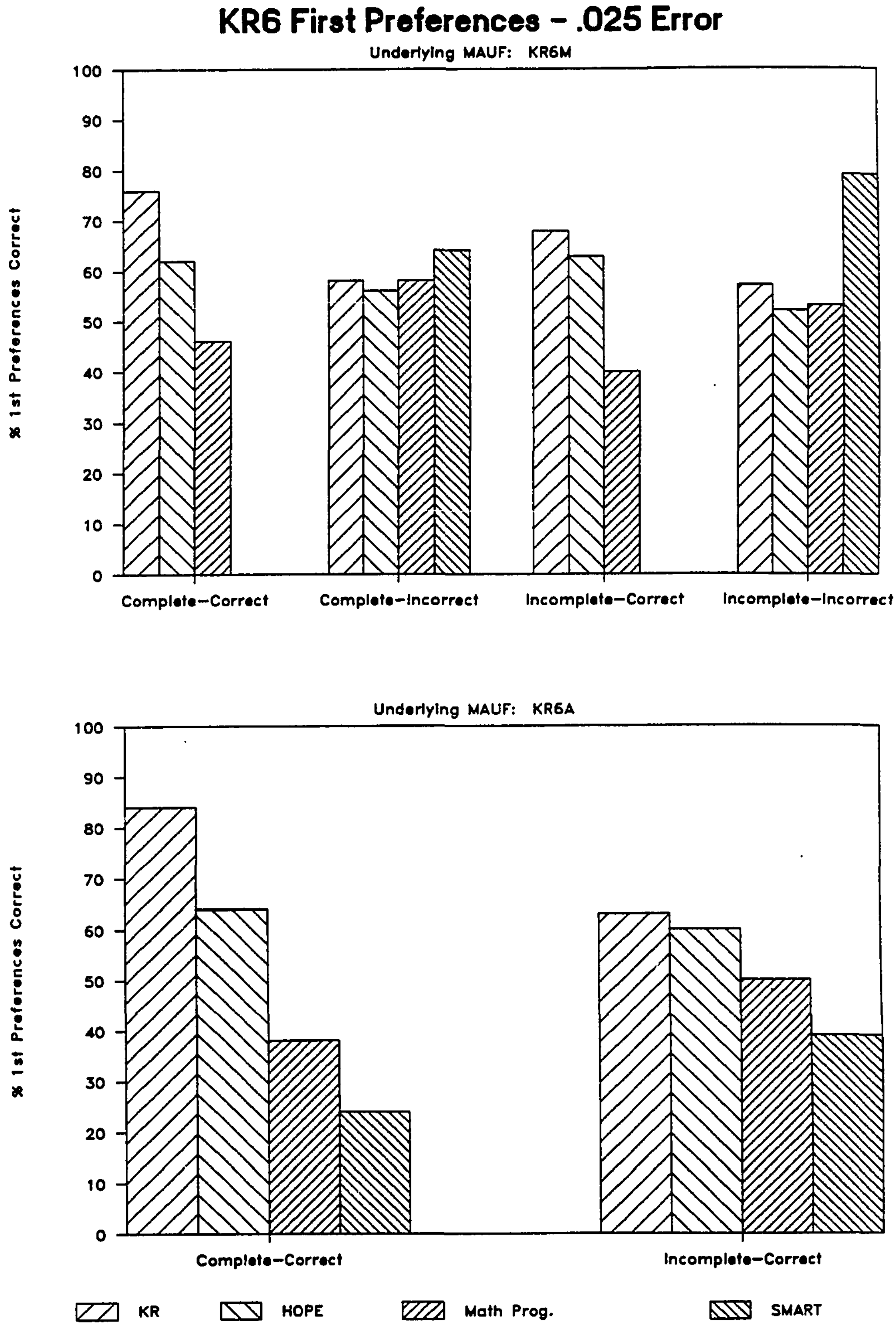

KR6 First Preferences - .025 Error
Underlying MAUF: KR6M
100
90
80
70
60
50
40
30
20
10
0
% 1st Preferences Correct
Complete-Correct
Complete-Incorrect
Incomplete-Correct
Incomplete-Incorrect
Underlying MAUF: KR6A
100
90
80
70
60
50
40
30
20
10
0
% 1st Preferences Correct
Complete-Correct
Incomplete-Correct
KR
HOPE
Math Prog.
SMART

Figure 6.6

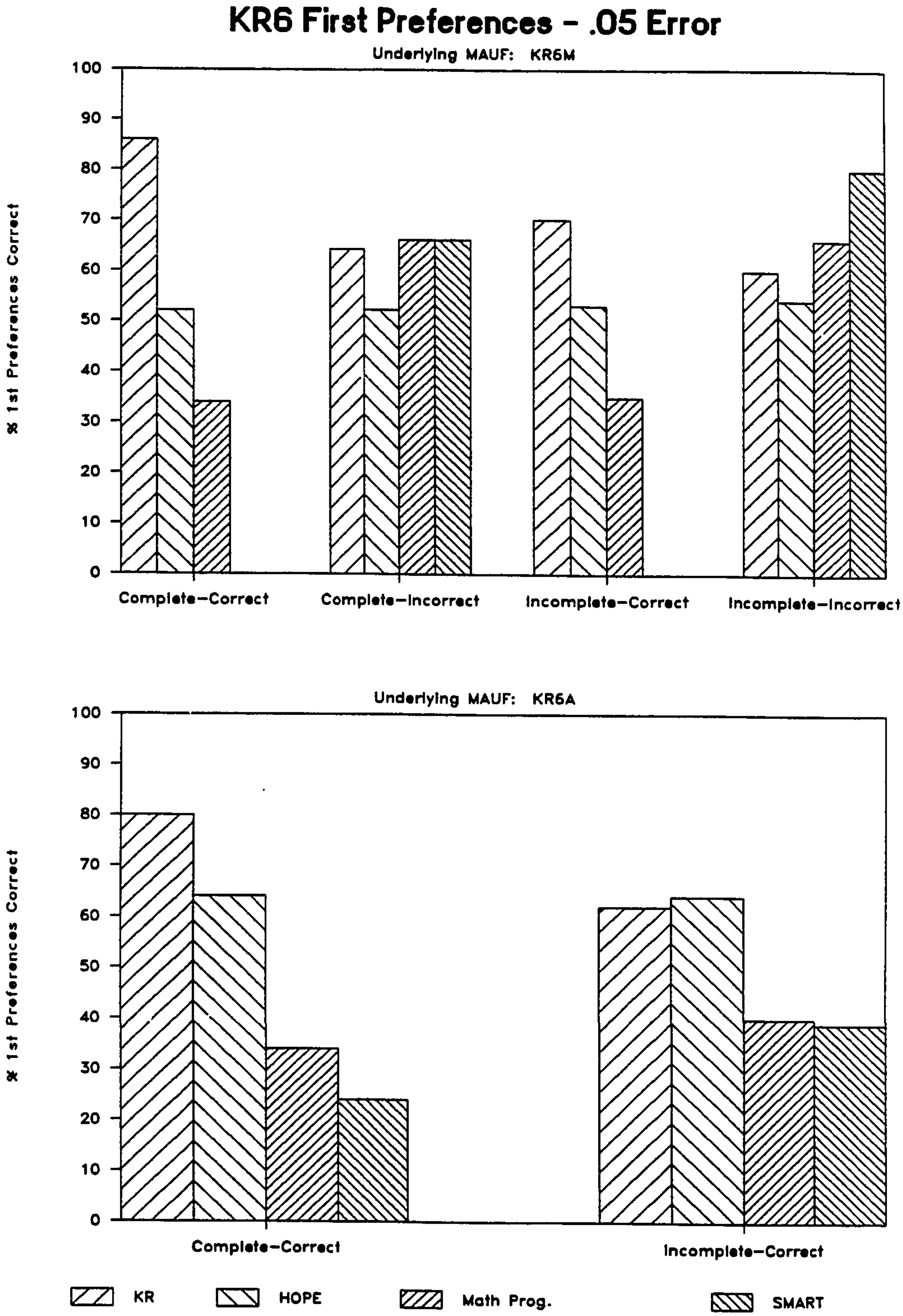

KR6 First Preferences - .05 Error
Underlying MAUF: KR6M
% 1st Preferences Correct
100
90
80
70
60
50
40
30
20
10
0
Complete-Correct
Complete-Incorrect
Incomplete-Correct
Incomplete-Incorrect
Underlying MAUF: KR6A
Complete-Correct
Incomplete-Correct
KR
HOPE
Math Prog.
SMART

TABLE 6-10

KR6 First Preference Predictions - 0.025 Error Response

Elicitation Scenario Technique	% of 1st Preferences Correct	Mean PUL	High PUL	No. of Occurrences
	Underlying MAUF:	KR6M		
Complete-Correct				
Keeney-Raiffa	76	0.004	0.047	3
HOPE	62	0.012	0.174	1
Math Prog.	46	0.032	0.236	2
Complete-Incorrect				
Keeney-Raiffa	58	0.011	0.047	10
HOPE	56	0.006	0.097	1
Math Prog.	58	0.027	0.287	2
SMART	64	0.083	0.504	8
Incomplete-Correct				
Keeney-Raiffa	68	0.015	0.114	10
HOPE	63	0.019	0.114	10
Math Prog.	40	0.097	1.000	9
Incomplete-Incorrect				
Keeney-Raiffa	57	0.018	0.114	10
HOPE	52	0.026	0.114	10
Math Prog.	53	0.030	0.326	3
SMART	79	0.019	0.504	2
	Underlying MAUF:	KR6A		
Complete-Correct				
Keeney-Raiffa	84	0.008	0.055	4
HOPE	64	0.018	0.094	3
Math Prog.	38	0.079	0.451	1
SMART	24	0.103	0.486	8
Incomplete-Correct				
Keeney-Raiffa	63	0.030	0.163	10
HOPE	60	0.033	0.163	10
Math Prog.	50	0.053	0.444	1
SMART	39	0.036	0.094	10

TABLE 6-11

KR6 First Preference Predictions - 0.05 Error Response

Elicitation Scenario Technique	% of 1st Preferences Correct	Mean PUL	High PUL	No. of Occurrences
	Underlying MAUF:		KR6M	
Complete-Correct				
Keeney-Raiffa	82	0.003	0.047	2
HOPE	52	0.026	0.326	1
Math Prog.	34	0.236	1.000	4
Complete-Incorrect				
Keeney-Raiffa	64	0.008	0.047	7
HOPE	52	0.026	0.326	1
Math Prog.	66	0.033	0.504	1
SMART	66	0.073	0.504	7
Incomplete-Correct				
Keeney-Raiffa	70	0.014	0.114	10
HOPE	53	0.023	0.420	1
Math Prog.	35	0.284	1.000	11
Incomplete-Incorrect				
Keeney-Raiffa	60	0.017	0.114	2
HOPE	54	0.029	0.420	2
Math Prog.	66	0.033	0.420	1
SMART	80	0.019	0.504	2
	Underlying MAUF:		KR6A	
Complete-Correct				
Keeney-Raiffa	80	0.009	0.055	4
HOPE	64	0.027	0.270	2
Math Prog.	34	0.061	0.348	2
SMART	24	0.103	0.486	8
Incomplete-Correct				
Keeney-Raiffa	62	0.029	0.163	9
HOPE	64	0.040	0.592	1
Math Prog.	40	0.067	0.444	2
SMART	39	0.036	0.094	10

error levels, than other completely but incorrectly specified models. At the same time, however, when the consequences of incorrect predictions were evaluated, SMART recorded higher mean PULs.

When the elicited functions were incompletely but correctly modeled, KR and HOPE registered the best performances. Both procedures recorded a larger number of correct first preference predictions than mathematical programming or mathematical programming and SMART for the KR6M and KR6A scenarios, respectively. For the KR6M case, mathematical programming's inability to accurately identify a most preferred alternative was exacerbated by the large decision maker "regret" which sometimes accompanied the technique's incorrect recommendations. On the other hand, when encoded MAUFs were incorrectly and incompletely specified, SMART provided the highest percentage of correct first preference predictions, albeit the largest PUL for an incorrect choice.

To summarize, KR encoded models were generally better predictors of first preferences than the models of the other elicitation methodologies. Even when mathematical programming and/or SMART elicited functions provided more correct predictions, the KR procedure minimized the consequences of an incorrect choice (e.g., lower mean and individual PUL values). Determining a second best performer was more difficult. No clear choice was evident solely on the basis of the number of correct first preference predictions offered. When the consequences of an incorrect recommendation were considered, HOPE typically provided lower mean PULs than either mathematical programming or SMART. Furthermore, inaccurate HOPE first preference predic-

tions avoided the substantial "regret" that accompanied some incorrectly revealed SMART and mathematical programming first choices. Thus, on the basis of these measures it appears that HOPE was a better performer than both SMART and mathematical programming.

6.3 Relative Tradeoff Evaluations

6.3.1 Overview

Multiattribute utility theory cautions against interpreting the MAUF scaling parameters as indicators of an attribute's (objective's) relative importance. Rather, the appropriate theoretical interpretation is that the parameters reflect the willingness of the decision maker to effect tradeoffs between conflicting objectives. Such information could prove valuable in analyzing complex decision tasks.

Of the elicitation procedures examined, only KR openly claims to recover this tradeoff information. However, the other procedures do not always formally disassociate themselves from this property. As a consequence, the user of an assessment technology might believe that the encoded MAUF parameters reveal tradeoff information, when in fact, they do not.

This section discusses the extent to which the alternatively elicited scaling constants reliably recover tradeoff information. Relative tradeoff ratios were calculated for each pair of parameter estimates as follows:

$$\text{Relative Tradeoff Ratio} = \frac{\text{Assessed } k_i/\text{Assessed } k_j}{\text{True } k_i/\text{True } k_j} \qquad (36)$$

When unbiased, this ratio will equal 1.0. Ratios greater (less) than

1.0 indicate that the assessed parameters have overweighted (underweighted) the decision maker's willingness to make tradeoffs between two objectives. Unfortunately, this ratio is asymmetric about the value 1.0. This can result in misleading mean values of the ratio scores. To overcome this problem, the natural logarithm of the ratios was computed, averaged, and then the averages were retransformed to yield a measure of the mean relative tradeoff values.

A technique's ability to capture relative tradeoff information should be greatest for those cases where assessed models are correctly and completely specified. Thus, with the exception of SMART, ratios were computed only for the scaling constants of completely and correctly modeled functions. An exception was granted for SMART since the procedure is structurally incapable of encoding a multiplicative composite form.

6.3.2 Relative Tradeoff Ratio Performance

The mean relative tradeoff ratios produced by each technique when the assessed MAUF was fully specified are presented in Tables 6-12 to 6-17. Standard deviations (SD) are also listed. The results are presented by underlying MAUF and rate of decision maker response error.

With respect to the underlying KR4M and KR4A preference structures (Tables 6-12 and 6-13), the derived KR scaling parameters provided consistently accurate relative tradeoffs as evidenced by mean values near 1.0 with small standard deviations. The HOPE procedure did not perform as well as KR when tradeoffs involved the attribute x_3. One explanation could be HOPE's inability to recover a reason-

TABLE 6-12

Relative Tradeoff Ratios - KR4M

Relative Tradeoff Ratios		Keeney-Raiffa	HOPE	Mathematical Programming	SMART[a]
		0.025 Response Error			
k_1-k_2	Mean	1.005	1.038	0.949	1.194
	(SD)	(0.023)	(0.066)	(0.445)	(0.005)
k_1-k_3	Mean	0.967	0.548[b]	239.537[b]	0.941
	(SD)	(0.041)	(0.606)	(1.032)	(0.027)
k_1-k_4	Mean	0.989	1.169	3.807	0.749
	(SD)	(0.039)	(0.291)	(1.588)	(0.016)
k_2-k_3	Mean	0.962	0.530[b]	134.711[b]	0.788
	(SD)	(0.034)	(0.630)	(1.362)	(0.025)
k_2-k_4	Mean	0.984	1.127	2.141	0.627
	(SD)	(0.041)	(0.280)	(1.327)	(0.012)
k_3-k_4	Mean	1.023	2.138[b]	0.008[b]	0.796
	(SD)	(0.029)	(0.773)	(1.989)	(0.025)
		0.05 Response Error			
k_1-k_2	Mean	1.001	1.058	2.021[b]	1.191
	(SD)	(0.094)	(0.092)	(2.755)	(0.011)
k_1-k_3	Mean	1.044	0.291[b]	66.469[b]	0.933
	(SD)	(0.145)	(0.532)	(2.588)	(0.057)
k_1-k_4	Mean	1.011	1.075	9.150[b]	0.746
	(SD)	(0.157)	(0.285)	(2.771)	(0.026)
k_2-k_3	Mean	1.043	0.274[b]	17.546[b]	0.784
	(SD)	(0.070)	(0.561)	(5.052)	(0.047)
k_2-k_4	Mean	1.010	1.016	2.415[b]	0.626
	(SD)	(0.091)	(0.317)	(4.430)	(0.015)
k_3-k_4	Mean	0.969	3.506[b]	0.073[b]	0.799
	(SD)	(0.045)	(0.464)	(3.033)	(0.034)

[a]SMART is misspecified since it must always be additive

[b]Some replications have negative or near-zero value estimated scaling parameters

TABLE 6-13

Relative Tradeoff Ratios - KR4A

Relative Tradeoff Ratios		Keeney-Raiffa	HOPE	Mathematical Programming	SMART
		0.025 Response Error			
k_1-k_2	Mean	1.193	1.015	0.791	1.189
	(SD)	(0.025)	(0.066)	(0.034)	(0.006)
k_1-k_3	Mean	1.153	0.792[a]	1.980[a]	0.936
	(SD)	(0.034)	(0.719)	(1.847)	(0.034)
k_1-k_4	Mean	1.177	1.125	0.560	0.745
	(SD)	(0.044)	(0.236)	(0.107)	(0.012)
k_2-k_3	Mean	0.967	0.780[a]	2.504[a]	0.787
	(SD)	(0.032)	(0.714)	(1.863)	(0.029)
k_2-k_4	Mean	0.986	1.108	0.708	0.627
	(SD)	(0.042)	(0.220)	(0.093)	(0.007)
k_3-k_4	Mean	1.020	1.448[a]	0.283[a]	0.797
	(SD)	(0.030)	(0.862)	(1.928)	(0.024)
		0.05 Response Error			
k_1-k_2	Mean	1.191	1.072	0.828	1.185
	(SD)	(0.104)	(0.125)	(0.460)	(0.010)
k_1-k_3	Mean	1.230	0.348[a]	1.672[a]	0.942
	(SD)	(0.164)	(0.489)	(2.446)	(0.063)
k_1-k_4	Mean	1.203	1.061	1.838[a]	0.740
	(SD)	(0.159)	(0.154)	(2.446)	(0.027)
k_2-k_3	Mean	1.032	0.322[a]	2.020[a]	0.795
	(SD)	(0.085)	(0.523)	(2.779)	(0.053)
k_2-k_4	Mean	1.010	0.990	2.220[a]	0.624
	(SD)	(0.085)	(0.207)	(2.469)	(0.016)
k_3-k_4	Mean	0.978	2.918[a]	1.099[a]	0.785
	(SD)	(0.042)	(0.456)	(4.003)	(0.037)

[a]Some replications have negative or near-zero value estimated scaling parameters

able parameter value for the "minor" objective (attribute) x_3 when holistic responses were noisy. For some replications, HOPE produced a negative value for k_3. While this shortcoming of HOPE has been acknowledged by Weber (1985), it has not been widely reported in the literature, and is not mentioned in the original work developing the procedure. Although highly consistent results were obtained from rank exponent weights elicited by SMART, the procedure tended to underweight tradeoffs for most attribute pairs.

As can be seen, the mathematical programming scores indicate the presence of a substantial bias. The ratios are also highly variable. In many replications the technique determined a near zero value for one or more of the parameters modeled which resulted in large tradeoff ratios. It should be noted, however, that the procedure makes no claim to capture tradeoff information since the pairwise comparisons required to implement the procedure do not involve such judgments.

Results for the KR6M underlying construct are exhibited in Tables 6-14 and 6-15. In this instance both KR and HOPE produced reasonably good results. The HOPE generated ratios did, however, display greater instability than the KR scores. As before, HOPE did not always generate theoretically correct (i.e., >0) scaling parameters for each attribute. The mathematical programming procedure was again ineffective in recovering useful information regarding the subject's willingness to make tradeoffs. SMART produced good ratios for some pairs of attributes, but had a tendency to significantly underweight others.

TABLE 6-14
Relative Tradeoff Ratios - KR6M (0.025 Response Error)

Relative Tradeoff Ratios		Keeney-Raiffa	HOPE	Mathematical Programming	SMART[a]
k_1-k_2	Mean	0.961	0.962	1.046	1.043
	(SD)	(0.025)	(0.046)	(0.172)	(0.004)
k_1-k_3	Mean	0.966	0.922[b]	8.558[b]	1.032
	(SD)	(0.061)	(0.453)	(2.699)	(0.032)
k_1-k_4	Mean	0.950	1.027	15.132[b]	0.912
	(SD)	(0.030)	(0.078)	(2.880)	(0.005)
k_1-k_5	Mean	0.951	1.091	3.122[b]	0.733
	(SD)	(0.038)	(0.172)	(2.105)	(0.209)
k_1-k_6	Mean	0.946	0.971	2.037[b]	0.793
	(SD)	(0.042)	(0.220)	(1.936)	(0.208)
k_2-k_3	Mean	1.005	0.966[b]	8.178[b]	0.990
	(SD)	(0.082)	(0.459)	(2.720)	(0.036)
k_2-k_4	Mean	0.988	1.068	14.459[b]	0.875
	(SD)	(0.051)	(0.074)	(2.901)	(0.009)
k_2-k_5	Mean	0.990	1.134	2.984[b]	0.703
	(SD)	(0.061)	(0.189)	(2.136)	(0.208)
k_2-k_6	Mean	0.984	1.009	1.946[b]	0.761
	(SD)	(0.058)	(0.221)	(1.855)	(0.209)
k_3-k_4	Mean	0.984	1.001[b]	1.768[b]	0.884
	(SD)	(0.043)	(0.502)	(2.662)	(0.028)
k_3-k_5	Mean	0.985	1.186[b]	0.365[b]	0.710
	(SD)	(0.033)	(0.437)	(4.447)	(0.209)
k_3-k_6	Mean	0.980	1.033[b]	0.238[b]	0.769
	(SD)	(0.043)	(0.382)	(4.105)	(0.208)
k_4-k_5	Mean	1.002	1.062	0.206[b]	0.804
	(SD)	(0.025)	(0.196)	(4.324)	(0.208)
k_4-k_6	Mean	0.996	0.945	0.135[b]	0.870
	(SD)	(0.028)	(0.248)	(4.360)	(0.208)
k_5-k_6	Mean	0.995	0.890	0.652[b]	1.082
	(SD)	(0.037)	(0.274)	(2.616)	(0.416)

[a]SMART is misspecified since it must always be additive

[b]Some replications have negative or near-zero value estimated scaling parameters

TABLE 6-15
Relative Tradeoff Ratios - KR6M (0.05 Response Error)

Relative Tradeoff Ratios		Keeney-Raiffa	HOPE	Mathematical Programming	SMART[a]
k_1-k_2	Mean	0.903	0.961	0.702	1.042
	(SD)	(0.046)	(0.129)	(1.330)	(0.005)
k_1-k_3	Mean	0.927	0.679[b]	14.491[b]	1.041
	(SD)	(0.082)	(0.412)	(2.862)	(0.046)
k_1-k_4	Mean	0.896	1.031	66.844[b]	0.914
	(SD)	(0.054)	(0.206)	(2.976)	(0.007)
k_1-k_5	Mean	0.939	1.033	5.316[b]	0.828
	(SD)	(0.082)	(0.412)	(2.865)	(0.193)
k_1-k_6	Mean	0.893	1.174	2.538[b]	0.706
	(SD)	(0.068)	(0.597)	(3.481)	(0.203)
k_2-k_3	Mean	1.027	0.698[b]	20.637[b]	0.999
	(SD)	(0.115)	(0.424)	(2.603)	(0.051)
k_2-k_4	Mean	0.992	1.072	95.190[b]	0.877
	(SD)	(0.075)	(0.137)	(2.184)	(0.012)
k_2-k_5	Mean	1.040	1.075	7.572[b]	0.795
	(SD)	(0.105)	(0.422)	(2.945)	(0.193)
k_2-k_6	Mean	0.990	1.222	3.615[b]	0.678
	(SD)	(0.086)	(0.500)	(2.965)	(0.204)
k_3-k_4	Mean	0.967	1.583[b]	4.613[b]	0.878
	(SD)	(0.064)	(0.534)	(3.427)	(0.040)
k_3-k_5	Mean	1.013	1.634[b]	0.367[b]	0.796
	(SD)	(0.078)	(0.489)	(4.237)	(0.206)
k_3-k_6	Mean	0.963	1.796[b]	0.175[b]	0.679
	(SD)	(0.060)	(0.826)	(4.872)	(0.191)
k_4-k_5	Mean	1.048	1.002	0.080[b]	0.907
	(SD)	(0.077)	(0.421)	(3.861)	(0.194)
k_4-k_6	Mean	0.997	1.139	0.038[b]	0.773
	(SD)	(0.056)	(0.493)	(3.487)	(0.201)
k_5-k_6	Mean	0.951	1.137	0.477[b]	0.852
	(SD)	(0.100)	(0.693)	(4.230)	(0.394)

[a]SMART is misspecified since it must always be additive

[b]Some replications have negative or near-zero value estimated scaling parameters

Tables 6-16 and 6-17 list the relative tradeoff ratios from the KR6A underlying structure. The KR tradeoff information was not as reliable for this scenario. This appears to be due to the procedure's underweighting of the "true" parameter x_1. HOPE produced accurate, but highly variable results. The mathematical programming procedure remained a poor performer. No difference was observed between the SMART ratios for the KR6M and KR6A cases.

6.4 Summary

This chapter presented the results of the study. The quality of the elicited functions' predictive performances were judged in two ways. First, the ordinal predictions of the alternatively assessed MAUFs were compared to the synthetic decision maker's "true" preference rankings. The most consistently reliable performances were demonstrated by the KR and HOPE methodologies. SMART's performance, while consistent, was less accurate than the KR and HOPE elicited rankings. Mathematical programming ordinal predictions exhibited less precision and greater variability than the orderings revealed by the other assessment technologies.

The ability of the assessed models to correctly identify the subject's most preferred decision alternative was then investigated. KR encoded models dominated the other procedures when choices involved four attributes. When the internal problem space was ordered by six attributes, it became more difficult to determine a best procedure. KR predictions did not always excel in terms of the number of first choices correctly identified. However, incorrect

TABLE 6-16
Relative Tradeoff Ratios - KR6A (0.025 Response Error)

Relative Tradeoff Ratios		Keeney-Raiffa	HOPE	Mathematical Programming	SMART
k_1-k_2	Mean	0.514	0.965	0.884	1.045
	(SD)	(0.049)	(0.054)	(0.349)	(0.004)
k_1-k_3	Mean	0.972	1.063	3.435[a]	1.027
	(SD)	(0.042)	(0.528)	(2.396)	(0.036)
k_1-k_4	Mean	0.953	1.021	15.891[a]	0.911
	(SD)	(0.025)	(0.073)	(2.585)	(0.005)
k_1-k_5	Mean	0.955	1.071	8.333[a]	0.730
	(SD)	(0.021)	(0.119)	(2.303)	(0.208)
k_1-k_6	Mean	0.950	0.956	0.637	0.791
	(SD)	(0.034)	(0.176)	(0.681)	(0.209)
k_2-k_3	Mean	1.893	1.101	3.886[a]	0.983
	(SD)	(0.085)	(0.511)	(2.565)	(0.039)
k_2-k_4	Mean	1.855	1.057	17.977[a]	0.872
	(SD)	(0.066)	(0.065)	(2.727)	(0.008)
k_2-k_5	Mean	1.860	1.109	9.427[a]	0.699
	(SD)	(0.066)	(0.139)	(2.136)	(0.208)
k_2-k_6	Mean	1.850	0.990	0.721	0.757
	(SD)	(0.067)	(0.175)	(0.408)	(0.210)
k_3-k_4	Mean	0.980	0.960	4.626[a]	0.887
	(SD)	(0.034)	(0.520)	(0.981)	(0.031)
k_3-k_5	Mean	0.982	1.008	2.426[a]	0.711
	(SD)	(0.031)	(0.544)	(4.532)	(0.208)
k_3-k_6	Mean	0.977	0.900	0.185[a]	0.770
	(SD)	(0.039)	(0.498)	(2.909)	(0.210)
k_4-k_5	Mean	1.003	1.049	0.524[a]	0.801
	(SD)	(0.025)	(0.149)	(4.778)	(0.207)
k_4-k_6	Mean	0.997	0.937	0.040[a]	0.868
	(SD)	(0.028)	(0.186)	(3.088)	(0.209)
k_5-k_6	Mean	0.995	0.893	0.076[a]	1.083
	(SD)	(0.039)	(0.191)	(1.823)	(0.416)

[a]Some replications have negative or near-zero value estimated scaling parameters

TABLE 6-17
Relative Tradeoff Ratios - KR6A (0.05 Response Error)

Relative Tradeoff Ratios		Keeney-Raiffa	HOPE	Mathematical Programming	SMART
k_1-k_2	Mean	0.476	0.970	0.778	1.044
	(SD)	(0.077)	(0.131)	(0.466)	(0.005)
k_1-k_3	Mean	0.922	0.816[a]	4.467[a]	1.039
	(SD)	(0.060)	(0.438)	(2.280)	(0.049)
k_1-k_4	Mean	0.873	0.984	24.438[a]	0.912
	(SD)	(0.070)	(0.176)	(2.838)	(0.007)
k_1-k_5	Mean	0.936	0.959	8.559[a]	0.825
	(SD)	(0.071)	(0.275)	(2.362)	(0.192)
k_1-k_6	Mean	0.918	1.078	0.514	0.703
	(SD)	(0.080)	(0.468)	(0.506)	(0.204)
k_2-k_3	Mean	1.936	0.838[a]	5.746[a]	0.996
	(SD)	(0.120)	(0.487)	(2.227)	(0.054)
k_2-k_4	Mean	1.833	1.015	31.431[a]	0.874
	(SD)	(0.088)	(0.109)	(2.574)	(0.011)
k_2-k_5	Mean	1.965	0.989	11.008[a]	0.790
	(SD)	(0.111)	(0.268)	(2.531)	(0.191)
k_2-k_6	Mean	1.927	1.112	0.661	0.674
	(SD)	(0.098)	(0.371)	(0.316)	(0.205)
k_3-k_4	Mean	0.947	1.223[a]	5.470[a]	0.877
	(SD)	(0.098)	(0.538)	(2.491)	(0.043)
k_3-k_5	Mean	1.015	1.241[a]	1.916[a]	0.794
	(SD)	(0.075)	(0.457)	(3.473)	(0.202)
k_3-k_6	Mean	0.995	1.346[a]	0.115[a]	0.677
	(SD)	(0.108)	(0.772)	(2.293)	(0.195)
k_4-k_5	Mean	1.072	0.974	0.350[a]	0.904
	(SD)	(0.097)	(0.247)	(4.801)	(0.193)
k_4-k_6	Mean	1.051	1.096	0.021[a]	0.771
	(SD)	(0.088)	(0.373)	(2.640)	(0.201)
k_5-k_6	Mean	0.981	1.125	0.060[a]	0.853
	(SD)	(0.106)	(0.457)	(2.609)	(0.393)

[a]Some replications have negative or near-zero value estimated scaling parameters

selections were not accompanied by large proportions of utility loss. Alternatively, in those cases where SMART or mathematical programming led to a greater number of correctly identified first choices than either KR or HOPE, their incorrect recommendations often produced greater decision maker regret.

Finally, the ability of the elicited scaling parameters to represent the decision maker's willingness to make tradeoffs between pairs of objectives was measured. It was determined that KR assessed scaling parameters (with some exceptions) generally recovered this information. Parameters assessed by HOPE were also relatively good at providing useful results. Exceptions were noted for HOPE, however, when the underlying MAUF contained "minor" attributes and decision maker holistic evaluations were noisy. In such cases, derived scaling parameters were sometimes negative. Mathematical programming does not profess to offer this information and was unable to do so for the models estimated. Scaling parameters elicited by SMART consistently underweighted the decision maker's relative tradeoffs.

CHAPTER 7 SUMMARY AND CONCLUSION

7.1 Overview

This research investigated the structural abilities of four MAUF assessment procedures to capture an experimental decision maker's underlying preferences for disparate choice alternatives. The assessment procedures examined were: the Keeney-Raiffa (KR) procedure (Keeney and Raiffa 1976); the holistic orthogonal parameter estimation, or HOPE, procedure (Barron and Person 1979); the simple multiattribute rating technique, or SMART (Edwards 1977); and the mathematical programming procedure (Klein et al. 1985).

A hypothetical decision making environment was created to compare the sensitivity of each procedure's performance to elicitation errors of known type and magnitude. While behavioral effects were neutralized, the experimental elicitation exercises contained one or more of the following assessment errors: noisy respondent inputs; incorrectly specified model forms; and incompletely defined attribute sets. A procedure's robustness was determined, in a general manner, by comparing the revealed preferences of the assessed functions to the artificial subject's known internal preference constructs.

The general findings of the research are contained in Section 7.2. Section 7.3 presents guidelines to assist an analyst in selecting an assessment technology. The limitations of the study and suggestions for future research are given in Section 7.4. Finally, conclusions are drawn in Section 7.5

7.2 Summary of General Findings

Each assessment procedure modeled four distinct underlying MAUFs. The relative performance of each technique was evaluated in terms of its ability to capture the subject's ordinal preferences, most preferred preference, and willingness to effect tradeoffs between attributes.

On the basis of its overall performance across elicitation scenarios, the KR procedure was the most robust technique examined. KR models consistently provided accurate ordinal preference recoveries. In addition, the technique revealed a higher percentage of correct first preferences than did its competitors for every assessment context except one. Perhaps, just as importantly, was KR's ability to encode models which avoided substantial "regret" when the subject's most desired alternative was incorrectly revealed. Finally, KR derived scaling parameters reliably portrayed the internal tradeoffs made by the decision maker in evaluating choice alternatives.

While HOPE encoded useful MAUFs, its preference predictions were typically more inconsistent and less accurate than those of KR. Both KR and HOPE ordinal preference predictions were relatively insensitive to the type and magnitude of errors introduced into the assessment process. However, unlike KR, the accuracy of HOPE's identified first choices often exhibited marked deterioration with increases in the level of response noise from the decision maker. Also, the relative tradeoff information gleaned from HOPE's derived scaling parameters was more variable and less reliable than the

information provided by the KR elicited parameters. This was most evident when tradeoffs involved "minor" attributes and/or when holistic inputs were subjected to moderate levels of noise.

The structure of the mathematical programming procedure makes implementation relatively expensive in terms of analyst time and effort. Because the nonlinear programming problems used to estimate the single attribute utility functions are nonconvex, care must be taken to avoid inferior solutions. This will generally require the analyst to solve each problem several times using different starting points. Furthermore, the procedure does not provide a specific approach for conducting consistency checks. Rather, inputs are considered "practically consistent" as long as a feasible region for the nonlinear programming problems exists. Finally, the analyst has no prior knowledge as to the number of pairwise preference comparisons needed to estimate a set of scaling parameters; nor does the analyst know when a sufficient number of such comparisons have been elicited.

The increased effort required of the elicitor to implement the mathematical programming technique is not justified considering the procedure's less accurate and substantially more variable preference predictions. The procedure demonstrated a greater sensitivity to increased respondent noise than either the KR or HOPE methodologies. In several instances, the accuracy of the technique's ordinal predictions decreased by 25%-30% when the subject's inputs became noisier. Similar rates of decrease in the percentage of correctly identified first preferences were also found to accompany increases in noisy

response rates. The mathematical programming procedure was also the only technique to select as the most preferred choice the subject's least desired alternative. The procedure's elicited scaling constants could not accurately reveal the subject's internal relative tradeoffs between objectives. However, this technique is not structured to recover this tradeoff information since it is based solely on decision maker pairwise preference comparisons.

Similarly, SMART ordinal preferences were less accurate than those generated by the KR and HOPE methodologies. In addition, a decrease in ordinal preference predictions was observed when the subject's internal decision space deviated more severely from an additive construct. SMART also appeared to be more successful at representing an overall preference structure than it was in revealing the subject's most desired alternative. While this may be an artifact of the underlying preferences and choices evaluated, it does suggest the technique is structurally better at providing gross preference orderings than at making fine distinctions among alternatives. Finally, scaling parameters elicited by SMART repeatedly underweighted the decision maker's actual relative tradeoffs among attributes.

7.3 Guidelines for Selecting an Elicitation Procedure

The suitability of an assessment technique depends on the participants, the purpose, and the consequences of the decision problem being analyzed. At a minimum, the technique selected must be structurally capable of recovering the information needed to fulfill

the requirements of the analysis. The following guidelines can assist an analyst in deciding if a technique is structurally appropriate for a specific decision task.

1. While the implementation of the KR procedure can require a sizeable investment of analyst and decision maker time, its close structural link to the basic theory of multiattribute utility analysis yields robust preference models. In the presence of various types of assessment errors, the technique consistently and accurately captures the respondent's underlying preference construct. Therefore, the KR technique is recommended for those decisions where selecting a most preferred alternative is critical. Since KR accurately modeled the subject's willingness to effect tradeoffs between objectives, it is also recommended whenever such information is needed.

2. HOPE appears to be structurally capable of providing good to excellent recoveries of a decision maker's ordinal preferences. Its ability to accurately predict first preferences, however, appears to be sensitive to the magnitude of noise contained in the holistic inputs. The structural composition of the technique is such that noise in one input can affect the estimated value of at least one scaling constant and several individual utility values. Thus, even moderate amounts of noise can reduce HOPE's accuracy in predicting a most preferred alternative. This makes HOPE slightly less attractive structurally than KR when the most desirable decision alternative must be correctly identified. In addition, noisy inputs can yield theoretically invalid scaling parameters and meaningless single

objective utility values when "minor" objectives are present. While this does not appear to significantly alter HOPE's ordinal preference predictions, it can reduce the method's effectiveness when fine distinctions among competing strategies must be made to identify the best course of action.

3. Unlike HOPE, SMART estimates scaling constants and single attribute utility functions separately. Thus, errors in the estimate of conditional utility functions and scaling parameters are independent of one another. However, SMART's sole reliance on linear conditional utility functions and an additive construct inherently introduces error into the assessment process when the underlying single attribute functions are highly nonlinear and/or the composite form is nonadditive. The potential therefore exists for a less robust model to be assessed whenever the underlying preference construct deviates substantially from an additive representation. SMART's structural simplicity may also explain why it appears to be more successful in providing overall preference rankings than in correctly identifying a first preference. Consequently, when a subject's most desired alternative must be identified, the analyst should select either the KR or HOPE procedure. However, one advantage of SMART's structural simplicity is the ease with which the technique can be implemented. This advantage increases the suitability of SMART for those decision tasks where contemplated choices must be quickly separated into categories of acceptable and unacceptable alternatives.

4. Not only is the mathematical programming procedure difficult to implement (from the analyst's standpoint) it appears to be

the least robust of the methods tested. Its structural weaknesses are especially evident when noisy responses create inconsistent inputs. In such instances, estimated scaling parameters can be so distorted that preference predictions can reverse the decision maker's least and most preferred choices. Even overall preference rankings can be highly unstable and/or inaccurate. Since noisy and inconsistent responses are potentially present in every assessment exercise, the procedure is not recommended.

7.4 Limitations and Suggestions for Future Research

The generalization of this study's findings to other decision contexts is limited by several factors. First, this study was conducted using a hypothetical decision maker whose preferences were elicited in an environment where various assessment errors could be either neutralized or systematically introduced and controlled. Because this study was concerned with the structural ability of different elicitation procedures to encode valid preferences this approach seemed prudent. However, in reality the assessment process does not take place in a behavioral vacuum, and the presence of various analyst and elicitor influences can also affect a procedure's performance.

Second, to study as wide a range of assessment errors as possible the study examined only four underlying preference structures. A larger and more varied study represents an important step in extending this study's recommendations to a broader set of underlying preferences and decision tasks. Closely related to this limi-

tation is the constraint imposed by using a single hypothetical experimental subject. The suitability of any procedure may vary across decision makers and decision problems. Information regarding individual perceptions of a technique's effectiveness in real decision making situations would therefore be beneficial in establishing a procedure's appropriateness for a specific decision task.

This study developed a set of guidelines to assist an analyst's choice of a structurally appropriate MAUF elicitation technique. However, because a procedure's effectiveness can depend on the decision problems and its participants, the recommendations presented here are preliminary ones. To establish a more definitive set of guidelines, future research must address the limitations discussed above. For example, many different underlying preference constructs could be developed by sampling from a distribution of simulated scaling parameters and univariate utility functions. This would provide a more extensive examination of a technique's structural ability to accurately model a wider and more varied set of preferences than those considered here.

In addition, research extending the recently completed works of Laskey and Fischer (1987) and Fischer et al. (1987) to compare the robustness of statistical models to traditionally encoded functions, and to evaluate the effects of proxy attributes on a procedure's performance, respectively, should be conducted. Finally, to more fully delineate the influences of behavioral effects on a technique's performance, studies employing live subjects in both laboratory and real decision making settings should be conducted. Such research

would provide greater insight into a subject's perception of a methodology's worthiness and effectiveness under different assessment environments.

7.5 Conclusion

The findings of this research revealed that alternatively encoded MAUFs can yield significantly different decision rules for evaluating experimental choice alternatives. This suggests that an analyst deliberating over the choice of an assessment methodology should relate a technique's structural abilities to the requirements of the decision being contemplated. While the recommendations provided by this study may be applicable only to decision contexts similar to those examined, the dissertation provides an initial set of guidelines to assist an analyst in choosing an MAUF elicitation procedure.

REFERENCES

Addelman, Sidney (1962), "Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments," Technometrics, Vol. 4, 1, 21-46.

Adelman, Leonard; Sticha, Paul J.; and Donnell, Michael L. (1984), "The Role of Task Properties in Determining the Relative Effectiveness of Multiattribute Weighting Techniques," Organizational Behavior and Human Performance, 33, 243-262.

Amoli, J. Etezadi; and Ciampai, A. , "Simultaneous Parameter Estimation for the Multiplicative Multiattribute Utility Model," Organizational Behavior and Human Performance, 32, 232-248.

Aschenbrenner, Michael K.; and Kasubek, Wolfgang (1978), "Challenging the Cushing Syndrome: Multiattribute Evaluation of Cortisone Drugs," Organizational Behavior and Human Performance, 22, 216-234.

Barron, F. Hutton (1980), "Empirical Validation of Elicited Utilities for Prediction and Decision Making," Decision Sciences, 11, 411-424.

_______ (1983), "Validation and Error in Multiplicative Utility Functions," Acta Psychologica, 53, 99-117.

_______ (1987a), "Influence of Missing Attributes on Selecting a Best Multiattributed Alternative," Decision Sciences, Vol. 18, 2, 194-205.

_______ (1987b), "Limits of Equal Weights in Additive Multi-Attribute Models," presented at the ORSA-TIMS National Conference, New Orleans, Louisiana.

_______; and Henry B. Person (1979), "Assessment of Multiplicative Utility Functions via Holistic Judgments," Organizational Behavior and Human Performance, 24, 147-166.

_______; and John, Richard (1980), "Reference Effects: A Sheep in Wolf's Clothing," Organizational Behavior and Human Performance, 25, 365-374.

_______; von Winterfeldt, Detlof; and Fischer, Gregory W. (1984), "Empirical and Theoretical Relationships Between Value and Utility Functions," Acta Psychologica, 56, 233-244.

_______; and Kleinmuntz, Don N. (1986), "Sensitivity in Value Loss of Linear Multiattribute Value Models to Attribute Completeness," in New Directions in Research on Decision Making, ed. B. Brehmer, H. Jungermann, P. Lourens, and G. Sevan, Elsevier Science Publishers B. V. (North-Holland), 291-301.

Beach, Lee Roy; and Barnes, Valerie (1983), "Approximate Measurement in a Multiattribute Utility Context," Organizational Behavior and Human Performance, 32, 417-424.

Beckwith, Neil E.; and Lehmann, Donald R. (1973), "The Importance of Differential Weights in Multiple Attribute Models of Consumer Attitude," Journal of Marketing Research, X, 141-145.

Bell, David E. (1979), "Multiattribute Utility Functions: Decompositions Using Interpolation," Management Science, Vol. 25, 8, 744-753.

_______ (1983), "Potential Contributions to Decision Analysis," Decision Sciences, 13, 534-553.

Bowman, E. H. (1963), "Consistency and Optimality in Managerial Decision-Making," Management Science, 310-321.

Bracken, Jerome; and McGill, James T. (1973), "Mathematical Programs with Optimization Problems in the Constraints," Operations Research, 37-44.

Brehmer, Berndt; and Qvarnstrom, Goran (1976), "Information Integration and Subjective Weights in Multiple-Cue Judgments," Organizational Behavior and Human Performance, 17, 118-126.

Briskin, Lawrence E. (1966), "A Method of Unifying Multiple Objective Functions," Management Science, Vol. 12, 10, B-406 - B-416.

_______ (1973), "Establishing a Generalized Multi-Attribute Utility Function," in Multiple Criteria Decision Making, J. L. Cochrane and M. Zeleny (eds.), University of South Carolina Press, Columbia.

Cook, Richard L.; and Stewart, Thomas R. (1975), "A Comparison of Seven Methods for Obtaining Subjective Descriptions of Judgmental Policy," Organizational Behavior and Human Performance, 13, 31-45.

Crosby, Michael A.; Moskowitz, Herbert; and Mahesh, Sathiadev (1986), "Modeling Auditor Utility Functions," Omega, Vol. 14, 1, 57-68.

Currim, Imran S.; and Sarin, Rakesh K. (1984), "A Comparative Evaluation of Multiattribute Consumer Preference Models," Management Science, Vol. 30, 5, 543-561.

Dyer, James S.; and Sarin, Rakesh K. (1979), "Measurable Multiattribute Value Functions," Operations Research, 22, 810-822.

_______; and _______ (1982) "Relative Risk Aversion," Management Science, Vol. 28, 8, 875-886.

Eckenrode, Robert T. (1965), "Weighting Multiple Criteria," Management Science, Vol. 12, 3, 180-192.

Edwards, Ward (1977), "How to Use Multiattribute Utility Measurement for Social Decisionmaking," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 7, 5, 326-340.

Eils, Lee C. III; and John, Richard S. (1980), "A Criterion Validation of Multiattribute Utility Analysis and of Group Communication Strategy," Organizational Behavior and Human Performance, 25, 268-288.

Einhorn, Hillel J.; and Hogarth, Robin M. (1975), "Unit Weighting Schemes for Decision Making," Organizational Behavior and Human Performance, 13, 171-192.

_______; and McCoach, William (1977), "A Simple Multiattribute Utility Procedure for Evaluation," Behavioral Science, 22, 270-282.

Eliashberg, Jehoshua (1980), "Consumer Preference Judgments: An Exposition With Empirical Applications," Management Science, Vol. 31, 1, 1-25.

Farmer, Timothy A. (1987), "Testing the Robustness of Multiattribute Utility Theory in an Applied Setting," Decision Sciences, 18, 178-193.

Farquhar, Peter H. (1977), "A Survey of Multiattribute Utility Theory and Applications," Management Science, 6, 59-89 (Errata, Vol. 24, 7, 785).

_______ (1984), "Utility Assessment Methods," Management Science, Vol. 30, 11, 1283-1300.

Fischer, Gregory, W. (1976), "Multidimensional Utility Models for Risky and Riskless Choice, Organizational Behavior and Human Performance, 17, 127-146.

_______ (1977), "Convergent Validation of Decomposed Multi-Attribute Utility Assessment Procedures for Risky and Riskless Decision," Organizational Behavior and Human Performance, 18, 295-315.

_______ (1979), "Utility Models for Multiple Objective Decisions: Do They Accurately Represent Human Preferences?" Decision Sciences, 10, 451-479.

_______; Kamlet, Mark S.; Fienberg, Stephen E.; and Schkade, David (1986), "Risk Preferences for Gains and Losses in Multiple Objective Decision Making," Management Science, Vol. 32, 9, 1065-1086.

_______; Damodoran, N.; Laksey, K.; and Lincoln, D. (1987), "Preferences for Proxy Attributes," Management Science, Vol. 33, 2, 198-214.

Fishburn, Peter C. (1967), "Methods of Estimating Additive Utilities," Management Science, Vol. 13, 7, 435-453.

_______ (1968), "Utility Theory," Management Science, Vol. 14, 5, 335-378.

_______ (1970), "Intransitive Indifference in Preference Theory: A Survey," Operations Research, 18, 207-228.

_______ (1977), "Multiattribute Utilities in Expected Utility Theory," in Conflicting Objectives in Decisions, ed. Bell, Keeney, and Raiffa, Wiley International Series on Applied Systems Analysis.

Fischoff, B.; Slovic, P.; and Lichtenstein, S. (1980), "Knowing What You Want: Measuring Labile Values," in Cognitive Processes in Choice and Decision Behavior, T. Wallsten (Ed.), Erlbaum, Hillsdale, NJ 64-85.

French, Simon (1986), Decision Theory: An Introduction to the Mathematics of Rationality, John Wiley & Sons, New York.

Green, Paul E.; and Srinivasan, V. (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," Journal of Consumer Research, 5, 103-123.

Hannon, Edward L.; Smith, Joseph A.; and Gilbert, Ronald A. (1983), "A Multiattribute Decision-Making Approach to the Selection of an Auxiliary Device for Icebreakers," Decision Sciences, 14, 240-252.

Hauser, John R.; and Urban, Glen L. (1979) "Assessment of Attribute Importance and Consumer Utility Functions: von Neumann-Morgenstern Theory Applied to Consumer Behavior," Journal of Consumer Research, 5, 251-262.

Hershey, John C.; Kunreuther, Howard C.; and Schoemaker, Paul J. H. (1982), "Sources of Bias in Assessment Procedures for Utility Functions," Management Science, Vol. 28, 8, 926-954.

_______; and Schoemaker, Paul J. H. (1985), "Probability Versus Certainty Equivalence Methods in Utility Measurement: Are They Equivalent?" Management Science, Vol. 31, 10, 1213-1231.

Hobbs, Benjamin F. (1980), "A Comparison of Weighting Methods in Power Plant Siting," Decision Sciences, Vol. 11, 725-737.

_______ (1986), "What Can We Learn From Experiments in Multiobjective Decision Analysis," IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-16, 3, 384-394.

Horsky, Dan; and Rao, M. R. (1984), "Estimation of Attribute Weights From Preference Comparisons," Management Science, Vol. 30, 7, 801-822.

Hovsepian, Beverly A.; Bond, Nicholas L.; and Ludlow, Larry (1982), "Multiattribute Evaluation in Baseball All-Star Predictions," Behavioral Science, 27, 273-280.

Huber, George P. (1974), "Methods for Quantifying Subjective Probabilities and Multi-Attribute Utilities," Decision Sciences, 5, 430-458.

_______ (1974), "Multi-Attribute Utility Models: A Review of Field and Field-like Studies," Management Science, Vol. 20, 10, 1393-1402.

_______; Daneshgar, Rahmna; and Ford, David L. (1971), "An Empirical Comparison of Five Utility Models for Predicting Job Preferences," Organizational Behavior and Human Performance, 6, 267-282.

International Mathematical and Statistical Library (1984), User's Manual, Edition 9.2, IMSL, Inc., Houston, TX.

Johnson, Edgar M.; and Huber, George P. (1977), "The Technology of Utility Assessment," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-7, 5, 311-325.

Johnson, Eric J.; and Payne, John W. (1985), "Effort and Accuracy in Choice," Management Science, Vol. 31, 4, 395-414.

Kahneman, Daniel; and Tversky, Amos (1979), "Prospect Theory: An Analysis of Decision Under Risk," Econometrica, Vol. 47, 2, 263-291.

Karmarkar, Uday S. (1978), "Subjectively Weighted Utility: A Descriptive Extension of the Expected Utility Model," Organizational Behavior and Human Performance, 21, 61-72.

Keelin, Thomas W. (1981), "A Parametric Representation of Additive Value Functions," Management Science, Vol. 27, 10, 1200-1208.

Keeney, Ralph L. (1972), "An Illustrated Procedure for Assessing Multiattributed Utility Functions," Sloan Management Review, Vol. 14, 1, 37-50.

_______ (1972), "Utility Functions for Multiattributed Consequences," Management Science, Vol. 18, 5, 276-287.

_______ (1973), "A Decision Analysis With Multiple Objectives: The Mexico City Airport," Bell Journal of Economics and Management Science, 4, 101-117.

_______ (1975), "Examining Corporate Policy Using Multiattribute Utility Functions," Sloan Management Review, 17, 63-76.

_______ (1977), "The Art of Assessing Multiattribute Utility Functions," Organizational Behavior and Human Performance, 19, 267-310.

_______ (1979), "Evaluation of Proposed Storage Sites," Operations Research, Vol. 17, 48-64.

_______ (1980), Siting Energy Facilities, Academic Press, New York.

_______ (1982), "Decision Analysis: An Overview," Operations Research, Vol. 30, 5, 803-838.

_______;Kirkwood, C. W.; Ford, C. K.; Robinson, J. A.; and Gottleib, P. (1979), "An Evaluation and Comparison of Nuclear Power Plant Siting Methodologies," tech. rep. NUREG/CR-0407, SAND 78-1284, Sandia Laboratories, Albuquerque, NM

_______; Lathrop, John F.; and Sicherman, Alan (1986), "An Analysis of Baltimore Gas and Electric Company's Technology Choice," Operations Research, Vol. 34, 1, 18-39.

_______; and Raiffa, Howard (1976), Decision with Multiple Objectives: Preferences and Value Tradeoffs, John Wiley & Sons, New York.

_______; and Sicherman, Alan (1976), "Assessing and Analyzing Preferences Concerning Multiple Objectives: An Interactive Computer Program," Behavioral Science, 21, 173-182.

_______; and _______ (1983), "Illustrative Comparison of One Utility's Coal and Nuclear Choices," Operations Research, 31, 50-83.

_______; and Wood, E. F. (1977), "An Illustrative Example of the Use of Multiattribute Utility Theory for Water Resource Planning," Water Resources Research, Vol. 13, 4, 705-712.

Klein, Gary; Moskowitz, Herbert; Mahesh, Sathiadev; and Ravindran, A. (1985), "Assessment of Multiattributed Measurable Value and Utility Functions via Mathematical Programming, Decision Sciences, Vol. 16, 309-324.

Krzysztofowicz, Roman (1983), "Strength of Preference and Risk Attitude in Utility Measurement," Organizational Behavior and Human Performance, 31, 88-113.

_______; and Duckstein, Lucien (1980), "Assessment Errors in Multiattribute Utility Functions," Organizational Behavior and Human Performance, 26, 326-348.

Laskey, Kathryn Blackmond; and Fischer, Gregory W. (1987), "Estimating Utility Functions in the Presence of Response Error," Management Science, Vol. 33, 8, 965-980.

Laughhunn, Dan J.; Payne, John W.; and Crum, Roy (1980), "Managerial Risk Preferences for Below-Target Returns," Management Science, Vol. 26, 12, 1238-1249.

Libby, Robert; and Fishburn, Peter C. (1977), "Behavioral Models of Risk Taking in Business Decisions: A Survey and Evaluations," Journal of Accounting Research, 15, 272-292.

Liebman, Judith; Lasdon, Leon; Schrage, Linus; and Waren, Allen (1986), Modeling and Optimization with GINO, The Scientific Press, Palo Alto, CA.

Lindley, D. V. (1986), "The Reconciliation of Decision Analysis," Operations Research, 34, 289-295.

_______; Tversky, A.; and Brown, R. V. (1979), "On the Reconciliation of Probability Assessments," Journal of the Royal Statistical Society, 142, Part 2, 146-180.

Moskowitz, Herbert; and Wright, Gordon P. (1979), Operations Research Techniques, Prentice-Hall, Inc., Englewood Sliffs, NJ.

de Neufville, Richard; and McCord, Mark (1984), "Unreliable Measurement of Utility: Significant Problems for Decision Analysis," Operational Research, 354-366.

Newman, J. Robert (1977), "Differential Weighting in Multiattribute Utility Measurement: When It Should Not and When It Does Make a Difference," Organizational Behavior and Human Performance, 20, 312-325.

Nutt, Paul C. (1980), "Comparing Methods for Weighting Decision Criteria," Omega, Vol. 8, 2, 163-172.

Payne, John W.; Laughhunn, Dan J.; and Crum, Roy (1981), "Further Tests of Aspiration Level Effects in Risky Choice Behavior," Management Science, Vol. 27, 8, 953-958.

_______; _______; and _______ (1984), "Multiattribute Risky Choice Behavior: The Editing of Complex Prospects," Management Science, Vol. 30, 11, 1350-1361.

Pekelman, Dov; and Sen, Subrata K. (1974), "Mathematical Programming Models for the Determination of Attribute Weights," Management Science, Vol. 20, 8, 1217-1229.

Pitz, Gordon F.; Heerboth, Joel; and Sachs, Natalie J. (1980), "Assessing the Utility of Multiattribute Utility Assessments," Organizational Behavior and Human Performance, 26, 65-80.

_______; _______; and _______ (1980), "Procedures for Eliciting Choices in the Analysis of Individual Decisions," Organizational Behavior and Human Performance, 26, 396-408.

_______; and Sachs, Natalie J. (1984), "Judgment and Decision: Theory and Application," Annual Review of Psychology, 35, 139-163.

Pratt, J. W. (1964), "Risk Aversion in the Small and the Large," Econometrica, Vol. 32, 122-136.

Richard, Scott F. (1975), "Multivariate Risk Aversion, Utility Independence and Separable Utility Functions," Management Science, Vol. 22, 1, 12-21.

Rowe, Michael D.; and Pierce, Barbara L. (1982), "Some Tests of Analytical Multiobjective Decision-Making Methods," Socio-Economic Planning Science, Vol. 16, 3, 133-140.

Russo, J. Edward; and Dosher, Barbara Anne (1983), "Strategies for Multiattribute Binary Choice," Journal of Experimental Psychology: Learning, Memory and Cognition, Vol. 9, 4, 676-696.

Saaty, T. L., (1980), The Analytic Hierarchy Process, McGraw-Hill, New York.

Sarin, Rakesh K. (1977), "Experimental Results of an Approach for Evaluating Multiattribute Alternatives," Decision Sciences, 8, 722-733.

SAS User's Guide: Basics, Version 5 Edition, SAS Institute, Inc., Cary, NC.

SAS User's Guide: Statistics, Version 5 Edition, SAS Institute, Inc., Cary, NC.

Schmitt, Neal (1978), "Comparison of Subjective and Objective Weighting Strategies in Changing Task Situations," Organizational Behavior and Human Performance, 21, 171-188.

Schoemaker, Paul J. H.; and Waid, C. Carter (1982), "An Experimental Comparison of Different Approaches to Determining Weights in Additive Utility Models," Management Science, Vol. 28, 2, 182-196.

Schrage, Linus (1986), Linear, Integer and Quadratric Programming with LINDO, The Scientific Press, Palo Alto, CA.

Simon, et al. (1987), "Decision Making and Problem Solving," Interfaces, Vol. 17, 5, 11-31.

Slovic, Paul (1972), "From Shakespeare to Simon: Speculations--and Some Evidence--About Man's Ability to Process Information," ORI Research, 12(2).

_______; and Lichtenstein, Sarah (1971), "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment," Organizational Behavior and Human Performance, 6, 649-744.

_______; Fischhoff, Baruch; and Lichtenstein, Sarah (1977), "Behavioral Decision Theory," Annual Review of Psychology, 28, 1-39.

Srinvasan, V.; and Shocker, Allen D. (1973), "Linear Programming Techniques for Multidimensional Analysis of Preferences," Psychometrika, Vol. 38, 3, 337-369.

Steuer, Ralph E. (1986), Multiple Criteria Optimization: Theory, Computation, and Application, Wiley & Sons, New York.

Stillwell, William G.; Seaver, David A.; and Edwards, Ward (1981), "A Comparison of Weight Approximation Techniques in Multiattribute Utility Decision Making," Organizational Behavior and Human Performance, 28, 62-77.

_______; Barron, F. Hutton; and Edwards, Ward (1983), "Evaluating Credit Applications: A Validation of Multiattribute Utility Weight Elicitation Techniques," Organizational Behavior and Human Performance, 32, 87-108.

Stillwell, William G.; von Winterfeldt, Detlof; and John, Richard S. (1987), "Comparing Hierarchical and Nonhierarchical Weighting Methods for Eliciting Multiattribute Value Models," Management Science, Vol. 33, 4, 442-450.

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

von Winterfeldt, Detlof (1980), "Additivity and Expected Utility in Risky Multiattribute Preferences," Journal of Mathematical Psychology, 21, 66-82.

_______ (1980), "Structuring Decision Problems for Decision Analysis," Acta Psychologica, 45, 71-93.

_______; and Edwards, Ward (1986), Decision Analysis and Behavioral Research, Cambridge University Press, New York.

_______; and Fisher, Gregory W. (1975), "Multi-attribute Utility Theory: Models and Assessment Procedures," in Utility, Probability, and Human Decision Making, Wendt/Vlek (eds.), D. Reidel Publishing Company, Dordrecht, Holland, 47-85.

_______; Griffin, Gregory M.; and Edwards, Ward (1984), "On the Relevance of the Utility Value Distinction for Decision Analysis," SSRI Research Report 84-1, University of Southern California.

Wainer, H. (1976), "Estimating Coefficients in Linear Models: It Don't Make No Nevermind," Psychological Bulletin, 83, 713-717.

Weber, Martin (1987), "Decision Making with Incomplete Information", European Journal of Operational Research, 28, 44-57.

_______ (1985), "A Method of Multiattribute Decision Making with Incomplete Information," Management Science, Vol. 31, 11, 1365-1371.

Yates, J. Frank; and Jagacinski, Carolyn M. (1979), "Reference Effects in Multiattribute Evaluations," Organizational Behavior and Human Performance, 24, 400-410.

_______; and _______ (1980), "Multiattribute Evaluation Reference Effects: A Reply to Barron and John," Organizational Behavior and Human Performance, 25, 375-383.

Yntema, D. B.; and Torgerson, W. S. (1961), "Man-Computer Cooperation in Decision Requiring Common Sense," IRE Transactions on Human Factors in Electronics, HFE-2, 20-26.

Zeleny, Milan (1982), Multiple Criteria Decision Making, McGraw-Hill Book Company, New York.

APPENDIX A

DETERMINATION OF THE MAGNITUDE OF THE NOISY RESPONSES

The magnitude of the response error to be introduced into the experimental elicitation exercises was determined using the approach of Barron and Person (1979). Four attribute levels (1,2,3,4), with levels 1 and 4 representing the least and most preferred attribute values, respectively, were defined for every decision criterion. The numerical values corresponding to each attribute level are presented in Table A-1. For the underlying KR4 (KR6) MAUF, all attribute values were fixed at level 3, except the two (three) attributes with the smallest scaling parameters. The "true" MAUFs were used to compute the utility for each of the 16 (64) profiles consisting of the four levels of the two (three) free attributes. The process was repeated with the two smallest KR4 (three smallest KR6) attributes held constant at level 2. Table A-2 contains the standard deviation values calculated from the "true" utility responses over the 16 (64) utility values. Standard deviations ranging from 0.017 to 0.054 were derived. The values of 0.025 and 0.05 were selected as representative standard deviation levels for purposes of this study.

Table A-1 Attribute Values at Four Levels

KR4

	Attribute x_1	x_2	x_3	x_4
Level 1	75	800	800	2000
Level 2	68	550	550	1400
Level 3	57	250	250	700
Level 4	50	0	0	0

KR6

	Attribute x_1	x_2	x_3	x_4	x_5	x_6
Level 1	4000	50	90	1000	250,000	1500
Level 2	2834	77	64	667	167,500	1000
Level 3	1667	104	38	334	85,000	501
Level 4	500	130	12	1	2,500	2

Table A-2. Standard Deviation of "TRUE" Holistic Responses

Underlying MAUF	Attributes Fixed at Level 3	Attributes Fixed at Level 2
KR4A	0.0256	0.0256
KR4M	0.0169	0.0223
KR6A	0.0541	0.0541
KR6M	0.0234	0.0395

APPENDIX B

ORTHOGONAL EXPERIMENTAL DESIGN

Table B-1. Orthogonal Experimental Design for 4 Attributes On 4 Levels[a]

Attribute x_1	x_2	x_3	x_4
1	1	1	1
1	2	2	3
1	3	3	4
1	4	4	2
2	1	2	2
2	2	1	4
2	3	4	3
2	4	3	1
3	1	3	3
3	2	4	1
3	3	1	2
3	4	2	4
4	1	4	4
4	2	3	2
4	3	2	1
4	4	1	3

[a] Source Addelman (1962)

Table B-2. Orthogonal Experimental Design for 6 Attributes On 4 Levels[a]

Attribute					
x_1	x_2	x_3	x_4	x_5	x_6
1	1	1	1	1	1
1	2	2	3	4	1
1	3	3	1	2	4
1	4	4	2	1	3
1	1	1	4	3	2
2	1	2	2	2	2
2	2	3	4	1	1
2	3	4	1	3	1
2	4	1	3	1	4
2	1	1	1	4	3
3	1	3	3	3	3
3	2	4	1	1	2
3	3	1	2	4	1
3	4	1	4	2	1
3	1	2	1	1	4
4	1	4	4	4	4
4	2	1	1	2	3
4	3	1	3	1	2
4	4	2	1	3	1
4	1	3	2	1	1
1	1	1	1	1	1
1	2	1	2	3	4
1	3	2	4	1	3
1	4	3	1	4	2
1	1	4	3	2	1

[a] Source Addelman (1962)

APPENDIX C

FOUR ATTRIBUTE AND SIX ATTRIBUTE PROFILES

TABLE C-1. FOUR ATTRIBUTE PROFILES

Value x_1	x_2	x_3	x_4	Level x_1	x_2	x_3	x_4
			SET 1				
50	550	550	2000	4	2	2	1
57	250	800	1400	3	3	1	2
57	550	250	2000	3	2	3	1
50	800	800	700	4	1	1	3
68	250	250	2000	2	3	3	1
68	250	800	0	2	3	1	4
68	800	800	700	2	1	1	3
75	0	800	700	1	4	1	3
			SET 2				
50	250	550	1400	4	3	2	2
57	250	0	2000	3	3	4	1
50	800	250	1400	4	1	3	2
50	800	800	0	4	1	1	4
68	0	800	700	2	4	1	3
57	0	550	1400	3	4	2	2
68	0	0	2000	2	4	4	1
57	250	800	700	3	3	1	3
			SET 3				
50	0	550	1400	4	4	2	2
50	550	0	1400	4	2	4	2
50	800	250	700	4	1	3	3
68	550	0	700	2	2	4	3
50	550	550	700	4	2	2	3
68	550	250	0	2	2	3	4
68	0	250	700	2	4	3	3
50	0	0	2000	4	4	4	1
			SET 4				
50	800	0	2000	4	1	4	1
57	800	250	700	3	1	3	3
50	550	550	1400	4	2	2	2
50	550	800	0	4	2	1	4
57	0	800	700	3	4	1	3
68	0	250	2000	2	4	3	1
57	550	0	2000	3	2	4	1
			SET 5				
68	0	550	0	2	4	2	4
57	250	0	0	3	3	4	4
50	0	550	2000	4	4	2	1
50	550	550	0	4	2	2	4
50	550	0	2000	4	2	4	1
68	0	250	1400	2	4	3	2
50	550	250	700	4	2	3	3
50	250	550	700	4	3	2	3

TABLE C-2. SIX ATTRIBUTE PROFILES

Value						Level					
x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6
				SET 1							
4000	104	12	1000	85,000	501	1	3	4	1	3	3
2834	104	64	1	167,500	501	2	3	2	4	2	3
1667	77	90	667	2,500	501	3	2	1	2	4	3
4000	130	90	667	250,000	1000	1	4	1	2	1	2
2834	104	90	334	85,000	2	2	3	1	3	3	4
500	77	64	1000	2,500	1000	4	2	2	1	4	2
500	50	64	1	167,500	2	4	1	2	4	2	4
500	50	12	334	250,000	1000	4	1	4	3	1	2
				SET 2							
1667	130	90	1	250,000	1000	3	4	1	4	1	2
4000	130	38	334	167,500	1000	1	4	3	3	2	2
2834	77	64	1000	85,000	1500	2	2	2	1	3	1
500	130	38	667	167,500	1000	4	4	3	2	2	2
4000	104	90	334	167,500	2	1	3	1	3	2	4
4000	130	38	667	250,000	2	1	4	3	2	1	4
1667	50	90	334	85,000	2	3	1	1	3	3	4
2834	104	12	667	250,000	501	2	3	4	2	1	3
				SET 3							
2834	130	38	667	85,000	2	2	4	3	2	3	4
500	130	38	334	85,000	1000	4	4	3	3	3	2
2834	77	12	334	2,500	1000	2	2	4	3	4	2
4000	77	64	1	167,500	501	1	2	2	4	2	3
1667	104	38	1	2,500	1000	3	3	3	4	4	2
4000	130	12	667	250,000	1500	1	4	4	2	1	1
500	50	90	334	85,000	501	4	1	1	3	3	3
2834	77	12	667	2,500	501	2	2	4	2	4	3
				SET 4							
4000	130	90	334	250,000	1500	1	4	1	3	1	1
500	130	64	667	250,000	501	4	4	2	2	1	3
2834	104	90	1	250,000	1500	2	3	1	4	1	1
1667	130	64	667	85,000	1500	3	4	2	2	3	1
500	104	64	334	250,000	1500	4	3	2	3	1	1
1667	104	64	334	2,500	1500	3	3	2	3	4	1
1667	50	64	1	250,000	2	3	1	2	4	1	4
500	104	90	1000	85,000	1500	4	3	1	1	3	1
				SET 5							
500	130	38	1	85,000	1500	4	4	3	4	3	1
500	77	64	334	250,000	1000	4	2	2	3	1	2
2834	130	64	1000	167,500	1000	2	4	2	1	2	2
1667	77	38	667	250,000	501	3	2	3	2	1	3
500	130	90	334	2,500	2	4	4	1	3	4	4
1667	104	12	1000	250,000	501	3	3	4	1	1	3
2834	50	64	667	250,000	2	2	1	2	2	1	4
2834	104	38	334	2,500	1500	2	3	3	3	4	1

VITA

Phillip C. Fry was born in Fort Smith, Arkansas on December 22, 1953. After spending several years in Santa Fe, New Mexico, he returned to Arkansas and graduated from Fayetteville High School, Fayetteville, Arkansas, in May, 1971. In 1971, he entered the University of Arkansas at Fayetteville and received a Bachelor of Arts degree, Phi Beta Kappa, in 1975 and a Master of Business Administration degree in 1977. After working for the Arkansas Public Service Commission and the Kansas Corporation Commission, he resumed his studies in 1981 at Louisiana State University (LSU) in Baton Rouge, Louisiana. In 1983, he earned a Master of Science degree in Economics. While a student, he served as a research and teaching assistant in the Department of Economics. From 1984 to 1987, he served as a teaching assistant in the Quantitative Business Analysis Department at Lousiana State University while working toward a degree of Doctor of Philosophy in Business Administration. During the course of his studies, he majored in quantitative business analysis and minored in economics.

Mr. Fry is married to the former Susan K. Wallace of Minneapolis, Kansas. They currently reside in Boise, Idaho, where Mr. Fry is an Assistant Professor of Decision Sciences at Boise State University.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Phillip C. Fry

Major Field: Business Administration

Title of Dissertation: "A Comparison of Four Multiattribute Utility Function Elicitation Procedures for Preference Predictions"

Approved:

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Stephen W. Looney

Date of Examination:

July 1, 1988