Using expected values to simplify decision making under uncertainty

Ian N. Durbach and Theodor J. Stewart
Department of Statistical Sciences, University of Cape Town

Full contact details of authors:

Ian N. Durbach	Theodor J. Stewart
Department of Statistical Sciences	Department of Statistical Sciences
PD Hahn (North) building	PD Hahn (North) building
University of Cape Town	University of Cape Town
Rondebosch	Rondebosch
7701	7701
Cape Town	Cape Town
South Africa	South Africa
ian.durbach@uct.ac.za	tjstew@stats.uct.ac.za
TEL: +27-21-6505058	TEL: +27-21-6503224
FAX: +27-21-6504773	FAX: +27-21-6504773

The corresponding author is Ian Durbach.

Abstract

A simulation study examines the impact of a simplification strategy that replaces distributional attribute evaluations with their expected values and uses those expectations in an additive value model. Several alternate simplified forms and approximation approaches are investigated, with results showing that in general the simplified models are able to provide acceptable performance that is fairly robust to a variety of internal and external environmental changes, including changes to the distributional forms of the attribute evaluations, errors in the assessment of the expected values, and problem size. Certain of the simplified models are shown to be highly sensitive to the form of the underlying preference functions, and in particular to extreme non-linearity in these preferences.

keywords: multicriteria, risk, decision process, simulation

1 Introduction

Although it is by now well-established that expected utility theory in its traditional form does not accurately describe human decision making (e.g. [1, 9, 25]), the shadow cast by the theory and its multiattribute counterpart in the normative or prescriptive decision sciences is large and it remains a substantial influence on theoretical and practical developments today (see [6] for a recent example). Although there are at present several alternate methodologies for dealing with uncertainties in the context of multiple criteria decision analysis (MCDA) e.g. [18], arguably none have incorporated axiomatic bases that have as broad a normative appeal as that of the expected utility axioms, despite their descriptive failures. When a 1989 gathering of top researchers in expected utility and non-expected utility theories was asked whether the maximisation of expected utility was an appropriate normative rule for decision making under uncertainty, affirmation was unanimous [8], and although agreement more than 15 years later might not be as overwhelming, a scan of the MCDA literature in the past five years will indicate the enduring popularity of multiattribute utility theory as a normative tool. Moreover, there is evidence that the majority of decision makers are at least in tentative agreement with the axiom of independence [11], and that some apparently non-expected utility behaviour can be classified as decision making with error [7] rather than axiomatic violation of expected utility theory.

The multiattribute forms of the expected utility model become increasingly complex in the face of violations of certain rationality axioms, most notably various forms of independence (see [16] for details). Leaving aside the acknowledged descriptive violations of the expected utility axioms, continued debate about the normative appeal of certain independence axioms (e.g. [21]) has led to a search for more and more complex models, which naturally increase the demand on both the decision maker and analyst. Perhaps the best-known example of this is the prospect theory of Kahneman and Tversky [26], some aspects of which we make use of in later simulations. Yet Stewart [23] has shown that an additive approximation to the multiplicative form of the utility function i.e. a model ignoring interactions between attributes, only marginally affects preference orderings, and has written of a 'fundamental contradiction' facing MCDA in which apparently rational violations lead to a search for increasingly complex models which, though always more difficult to implement practically, only seldom make recommendations that are different from those of simpler models. Von Winterfeldt and Edwards discuss similar 'flat maximum' examples in which simplifications of certain decision models do not alter their basic recommendations [28], and Kirkwood has proposed a model testing whether uncertainty would change a set of evaluation results, finding that it often does not [17]. If the aim of the decision analysis is to help the decision maker to construct and learn about their own preferences in the search for a solution, then there are distinct advantages to using a simple and easy-to-understand decision model.

The aim of this paper is to test the impact of applying a simplification to the multiattribute expected utility model in which the full distribution of attribute evaluations for each alternative is replaced by its expected value. This rather drastically simplified model, in which an alternative is not evaluated by the utility of all possible attribute values that may be assumed, but only by a function of the expected values for each attribute, has been explicitly proposed in [17] and applied in [3] to the problem of selecting venture capital projects. However, inasmuch as most decision problems will contain at least some uncertainty, it might be argued that there are many more implicit applications of the simplification of attribute evaluations to an expectation form, including many applications of multiattribute value theory. This widespread use, mostly implicitly, of an expected value simplification increases the need for a formal evaluation of its possible effects. It is this need which the current paper attempts to address by extending [17] to include a more detailed testing of the simplified model in a controlled set of simulation experiments. The current paper is thus in the tradition of several pieces of research that use Monte Carlo simulation in order to investigate the effects of various practical variables in a controlled environment, for example the effect of using piecewise utility functions [22], missing attributes [2], weight assessment difficulties [13, 5], and violations of preferential independence [24].

Formally, suppose a decision problem consists of n alternatives ($i = 1, 2, \ldots, n$) evaluated on m attributes ($j = 1, 2, \ldots, m$), with Z_{ij} a random variable denoting the (stochastic) attribute evaluations of alternative i on attribute j, with mean $\mathrm{E}[Z_{ij}]$ and variance σ^2_{ij}. The $u_j(Z_{ij})$ are single-attribute utility functions, and w_j denotes the importance weight for attribute j. Since Stewart has already shown that an additive form can be an excellent approximation of a multiplicative utility function, the focus in this paper is on further simplifications of the additive form. This additive form of the global utility of alternative i, denoted U_i, can be simplified by writing it as a function of the expected attribute values and the variability of attribute values, according to the following general form

$$\mathrm{E}[U_i] \approx \sum_{j=1}^{m} w_j u_j\big(\mathrm{E}[Z_{ij}]\big) + w^R_{ij}\sigma^2_{ij} \quad (1)$$

where w^R_{ij} is a weight to be attached to the variability of the attribute evaluations. The real question of interest then becomes what form to use for this weight w^R_{ij}. Kirkwood has shown in [17] that using $w^R_{ij} = (1/2)w_j u''_j(\mathrm{E}[Z_{ij}])$ can lead to close approximations of $\mathrm{E}[U_i]$, under the important conditions that the Z_{ij} be normally distributed or numerous enough for the central limit theorem to apply, and the underlying utility functions u_j 'not deviate too much from linear'. An even simpler approximation might employ $w^R_{ij} = 0$ i.e. might ignore the attribute variability altogether, which could conceivably lead to superior performance in those circumstances in which the restrictions on the Z_{ij} and u_j are strongly violated.

In fact, both the above simplifications reduce the problem to a deterministic framework in the sense that the expected values, once obtained, may be thought of in the same way as deterministic attribute evaluations when no uncertainty is present: that is, simply as evaluations of alternative i on attribute j. The practical implication of the simplified structure is that not all possible outcomes need be elicited if the decision maker feels comfortable directly estimating the expected attribute value. The decision maker may want to explicitly specify a number of possible outcomes to the extent that it helps him or her learn about the problem at hand and form an estimate of the expected attribute value, but the number of outcomes would typically be small. For example, Keefer and Bodily [15] present a range of three-point approximations to the mean of the generalised form $w_a z_a + w_b z_b + w_c z_c$, where a, b and c denote the three selected quantiles. In the simulations to follow, we approximate the mean using the best-performing of the Keefer and Bodily approximations, and contrast the results obtained with those using other quantiles and other weights. It is also worth noting here that later research in [14] found that the best of the Keefer and Bodily approximations was also able to accurately approximate the certainty equivalents of various bell-shaped beta PDF's, in addition to the mean.

In this paper we use Monte Carlo simulation to investigate the quality of the preference order generated by different forms of the simplified model under a variety of external environmental influences, including different problem sizes, underlying distributional forms for the uncertain attribute evaluations, and the presence of dominated alternatives, and internal environmental influences, including different forms of underlying preferences and varying degrees of accuracy in the estimation of the expected attribute values by the decision maker. Our specific aims are three-fold:

1. To assess whether the simplified model gives a sufficiently good approximation to the decision maker's preference order to be used in practice,
2. To assess the degree of accuracy that is necessary in the estimation of expected attribute values in order to give acceptable results
3. To assess the robustness of the simplified model to various external and internal environmental conditions.

The remainder of the paper is organised as follows. In section 2 the details of the simulation experiment investigating the simplified model are provided, following which sections 3 and 4 discuss the results of the experiment. A final section draws together conclusions and discusses some implications for practical decision aiding and future research.

2 Implementation of the simplified approach

The basic outline of the structure used in the simulations is shown in Figure 1. The first two levels of the diagram correspond roughly to a first stage in which basic data for the decision problem are generated, essentially consisting of attribute evaluations, attribute weights, and scenario probabilities. The second stage involves using the available information to construct a hypothetical 'true' set of preferences and hence a 'true' rank order to be used as a basis for evaluation, while the third stage involves limiting the amount of information considered in order to simulate the utilities and preference orderings that might be obtained from an application of the simplified models. These two stages are made up of various parts of the third, four, and fifth levels of Figure 1. In the final stage, the results of the simplified models are evaluated by comparing them to the results that were obtained from the hypothetical 'true' preferences. Each of these stages is discussed in more detail below.

2.1 Generating the problem context

The problem context is represented by a set of n alternatives evaluated over m attributes for each of p scenarios, and a set of probabilities $\Pr(k)$ on the set of scenarios $k = \{1, 2, \ldots, p\}$.

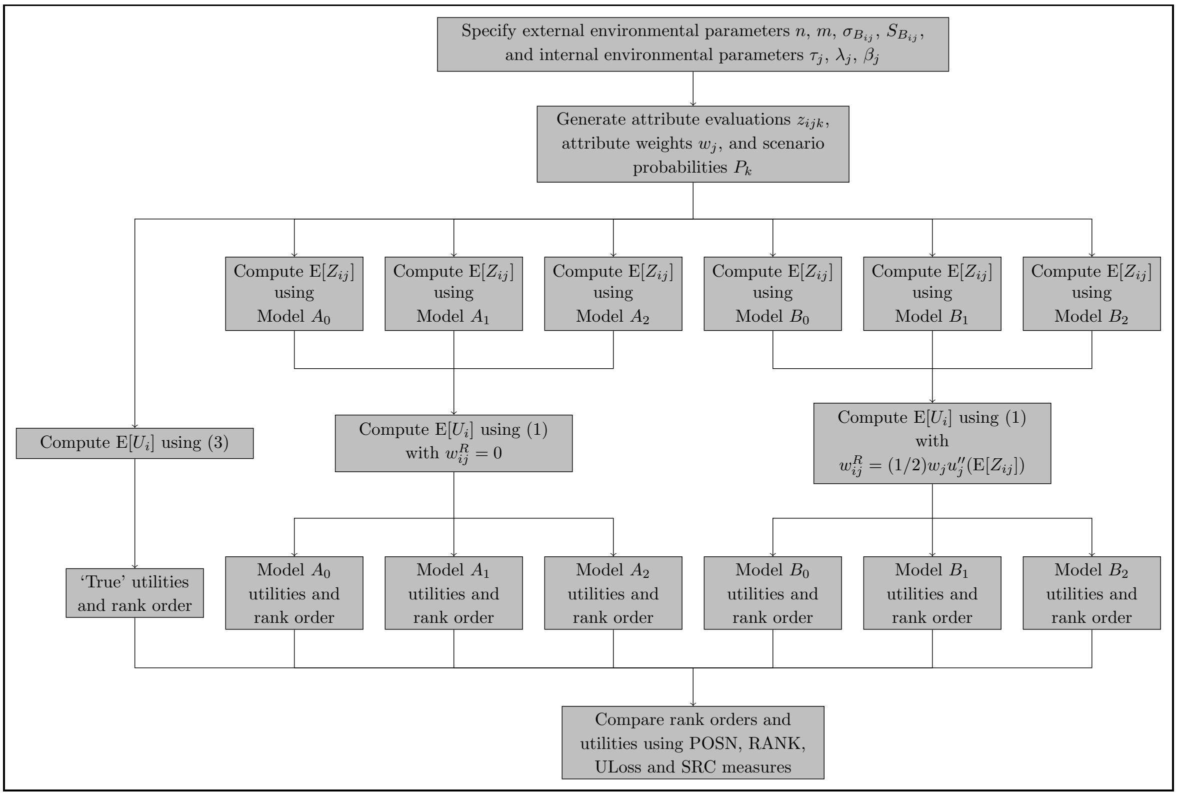


Figure 1: Outline of simulation structure used

The assumption is thus that the true underlying distribution of outcomes for each alternative is discrete. The assumption is made for convenience, but is not particularly restrictive in that for sufficiently large p, continuous distributions can be approximated to any desired degree of accuracy.

Attribute Evaluations

The evaluations z_{ijk}, representing the performance of alternative i on attribute j when scenario k occurs, will in practice occur almost exclusively in a highly complex manner. For the purposes of a simulation experiment it is necessary to limit the complexity of the governing process to the point where a meaningful interpretation of the simplified structure is possible, without compromising the non-trivial aspects of the process. Many possible simplified structures might be used and our intention is purely to gain a rich enough variety in the decision space for a non-trivial evaluation of the simplified methodology; at no stage is the simulated structure employed here intended as an even tentative description of reality.

The evaluations are generated according to the process $z_{ijk} = a_{ij} + b_{ijk}$. For alternative i and attribute j, a_{ij} is an initial value of a random variable Z_{ij} giving rise to the attribute evaluations z_{ijk}, taken across all p scenarios. The a_{ij} for each alternative i are standardised to lie on the unit hypersphere i.e. $a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2 = 1$, $\forall i$, which ensures that the resulting alternatives are non-dominated in the sense that no alternative will have a smaller attribute value than another alternative on every attribute. The form of the random variable a_{ij} was selected after a number of trial-and-error experiments to be normally distributed with mean 0 and standard deviation 1. Several other forms were considered, with the form finally selected providing a balanced but reasonably diverse set of alternatives. The effect of assuming a non-dominated alternative set was investigated by running a separate set of simulations in which 20% of the generated alternatives were allowed to be dominated and then comparing the results to those obtained with non-dominated alternatives.

Uncertainty is introduced via the b_{ijk}, which cause the z_{ijk} to differ over scenarios. For a particular scenario k^*, the resulting b_{ijk^*} is modelled as a realisation of a random variable B_{ij} following a gamma distribution with shape parameter ξ_{ij} and scale parameter ω_{ij}, so that the expectation, variance, and skewness of B_{ij} are given by the quantities $\xi_{ij}\omega_{ij}$, $\xi_{ij}\omega_{ij}^2$, and $2/\sqrt{\xi_{ij}}$ respectively. One of the stated aims of the simulation experiment is to test the robustness of the simplified strategy under environmental conditions in which (a) certain alternatives have more variable evaluations on some attributes than other alternatives, and (b) the expected attribute value for a particular alternative is closer to the extreme attribute evaluations i.e. attribute evaluations are either positively or negatively skewed.

The gamma distribution has several properties that make it a desirable choice for investigating a simplified decision making approach. The normal distribution is included as a special case of the gamma distribution with $\xi_{ij} = \infty$, allowing one to use a perfectly symmetric distribution as a basis for evaluating the extent to which increasing skewness affects the simplification strategy. With the desired variance and coefficient of skewness specified, one can easily solve for the two unknown parameters ξ_{ij} and ω_{ij}. The empirical mean across all scenarios is then subtracted from the random variate in each of the scenarios to obtain the desired expected value of zero without affecting the other moments of interest. If random variates from a negatively-skewed distribution are desired, a gamma distribution with the same magnitude of skewness (but positive) is used, following which the centered random variates are reflected about the origin – allowing the scenario-based deviations to be either generally positive or generally negative. For each attribute, the resulting attribute evaluations are then standardised over all alternatives and scenarios to lie between 0 and 1.

Scenario Probabilities
The relative scenario probabilities were generated in such a way that they are uniformly distributed and have a minimum normalised value of 0.01, using the approach outlined in [5]. For $p = 50$ scenarios, this requires the generation of 49 $U(0, 1)$ variates which are then sorted in descending order to obtain the sequence $r_1, r_2, \ldots, r_{49}$. The scenario probabilities are then obtained using $Pr(k) = 0.01 + (1 - 0.01p)(r_{k-1} - r_k)$ for $k = 1, \ldots, 50$, where $r_0 = 1$ and $r_{50} = 0$. We appreciate the input of a reviewer here, who pointed out that an earlier version was not giving uniform probabilities and suggested this approach.

2.2 Constructing the idealised preference order

The construction of an idealised rank order follows the structure used by Stewart [23, 24], and is based on the assumption that the decision maker possesses an idealised underlying preference structure which exists as a goal which is aimed toward even if its actual form is not consciously known, and that this idealised structure may be represented by an additive utility function. This conjecture should not be associated with the (behavioural) idea that underlying every decision maker is such a function, but rather considered as a modelling mechanism for creating an idealised preference structure resulting in a 'true' rank ordering, which can then be compared to the rank orderings produced by the simplified models making use of expected attribute values.

Utility Functions

The utility functions used in the construction of the idealised preference structure are based upon the characteristics of diminishing sensitivity i.e. risk proneness for losses and risk aversion for gains relative to a reference level, and loss aversion [26]. This implies a utility function which is convex below a reference level and concave above it, and which is steeper below the reference point. Each marginal utility function is fully described by four parameters, following the ideas of Stewart [24]: the reference level, τ_j, the value of the utility function at the reference level, λ_j, the curvature of the utility function below the reference level, α_j, and the curvature of the utility function above the reference level, β_j, and is of the standardised exponential form

$$u_j(x) = \begin{cases} \lambda_j(\mathrm{e}^{\alpha_j x} - 1)/(\mathrm{e}^{\alpha_j \tau_j} - 1) & \text{for } 0 \le x \le \tau_j \\ \lambda_j + \left[(1-\lambda_j)(1-\mathrm{e}^{-\beta_j(x-\tau_j)})/(1-\mathrm{e}^{-\beta_j(1-\tau_j)})\right] & \text{for } \tau_j \le x \le 1 \end{cases} \tag{2}$$

Quite a diverse set of preference types may be simulated by adjusting values for τ_j, λ_j, and β_j. The parameter λ_j is an indication of the strength of preference for avoiding performances below the reference level τ_j for attribute j, so that the severity of the preference threshold separating losses and gains increases in λ_j, while β_j determines the nonlinearity of preferences below and above the reference level. Also note here that utility functions that are concave or convex in the entire domain of attribute values can be obtained as special cases of $\tau_j = \lambda_j = 0$ and $\tau_j = \lambda_j = 1$ respectively. Further details are outlined in section 2.5.

Construction of the Idealised Preference Order

The global utilities are determined in an additive model

$$U_i = \sum_{j=1}^{m} w_j u_{ij} \tag{3}$$

where the attribute weights w_j are generated in a similar way to the scenario probabilities to be uniformly distributed and sum to one, with a sub-minimum of $1/2m$, and are held constant over scenarios. The final calculation of the utilities is performed by taking expectations over the p scenarios, which produces a complete order, termed the true rank order.

2.3 Constructing the preference order in each of the simplified models

In practical applications of a simplified strategy making use of expected values, the expected values used would in all likelihood merely be approximations to the true expected values, since the only way to arrive at the true expected values would be through complete enumeration of all possibilities, which would negate the need for a simplified model. For the purposes of simulating the approximation of expected value, we have used the following approach which aims at capturing some of the complexities of eliciting expected values in real-world applications: the expected values themselves are approximated using the Pearson-Tukey approximations advocated by Keefer and Bodily in [15], given for a particular attribute by $\mathrm{E}[Z] = 0.185z_{0.05} + 0.63z_{0.5} + 0.185z_{0.95}$ and $\mathrm{Var}[Z] = 0.185z_{0.05}^2 + 0.63z_{0.5}^2 + 0.185z_{0.95}^2 - (\mathrm{E}[Z])^2$, where $z_{0.05}, z_{0.5}$, and $z_{0.95}$ are those attribute values which exceed the attribute values in 5%, 50%, and 95% of the simulated scenarios respectively. This scenario-based quantile construction is based upon ranking the attribute values over scenarios, without making use of relative likelihood information, in line with the approach advocated in scenario planning (e.g. [27]). In cases where scenario probabilities are uniformly generated, such as in our simulations, the two approaches should lead to similar quantiles – a preliminary set of 5000 simulations showed that the scenario-based quantiles exceeded the true quantiles by 0.018, 0.005, and 0.006 for $z_{0.05}, z_{0.5}$, and $z_{0.95}$ respectively. Two broad types of models have been simulated: a type-B model using the simplification $w_{ij}^R = (1/2)w_j u_j''(\mathrm{E}[Z_{ij}])$ used by Kirkwood, and a type-A model using the more simplified version $w_{ij}^R = 0$, which ignores the variance component altogether.

As it seems reasonable to suggest that the results obtained from the simplified strategy will depend partly on the accuracy of the approximations, we have made use of three different degrees of accuracy in computing the Pearson-Tukey approximations, basing the modelling of accuracy of approximation on two well-documented results from cognitive psychology. First, there is substantial evidence that when asked to give confidence intervals around their estimates of a mean value, people generally err by making their estimates of these confidence intervals too narrow [25, 4]. This cognitive approximation error is modelled by replacing $z_{0.05}$ and $z_{0.95}$ with quantiles closer to the median. Secondly, when asked to estimate the median attribute value, people may anchor their estimate on what they consider to be most likely, in an application of the well-known anchoring heuristic [25, 10]. This heuristic is modelled by replacing $z_{0.5}$ with a convex combination of $z_{0.5}$ and z_{ijP^*}, where P^* is the scenario with the highest simulated probability of occurring. The six simulated models are shown in Table 1.

Model	Lower quantile	Median quantile	Upper quantile	Variance component
A_0	$z_{0.05}$	$z_{0.5}$	$z_{0.95}$	absent
A_1	$z_{0.10}$	$0.8z_{0.5} + 0.2z_{ijP^*}$	$z_{0.90}$	absent
A_2	$z_{0.15}$	$0.5z_{0.5} + 0.5z_{ijP^*}$	$z_{0.85}$	absent
B_0	$z_{0.05}$	$z_{0.5}$	$z_{0.95}$	present
B_1	$z_{0.10}$	$0.8z_{0.5} + 0.2z_{ijP^*}$	$z_{0.90}$	present
B_2	$z_{0.15}$	$0.5z_{0.5} + 0.5z_{ijP^*}$	$z_{0.85}$	present

Table 1: Simplified models used in the simulations

Models A_0 and B_0 use the exact approximations, and can be considered ‘best case’ approximations. Models A_1 and B_1 incorporate moderate errors in both the estimation of extreme quantiles, and in the estimation of the median value, while the same errors are present in Models A_2 and B_2 but to a greater degree. It should be emphasised that no claim is made that decision makers think strictly in terms of convex combinations or that they consider only the various quantities outlined above; the three models merely provide a mechanism for the examination of what is in all likelihood a highly complex process. Note also that the effects of quantile approximation and median approximation are confounded in our models; the relative contribution of each of the approximations is discussed in Appendix B.

2.4 Comparing the results obtained from the idealised and simplified models

For each set of environmental parameters considered, six rank orders are produced corresponding to each of the type-A and type-B models. The results of the simulations are presented in the form of the following four comparative measures.

1. POSN. The average position of the true best alternative (according to the idealised rank ordering) in the model rank order.

2. RANK. The average rank of the best alternative selected by the model in the true rank order.

3. ULoss. Utility loss has been proposed in [2] as

$$\text{ULoss} = \frac{U_{i^*} - U_{i^{sel}}}{U_{i^*} - U_{i_*}}$$

   where U_{i^*} is the utility of the true best alternative, U_{i_*} is the utility of the true worst alternative, and $U_{i^{sel}}$ is the utility of the alternative selected by the model. The measure is an indication of the relative drop in quality resulting from the use of a simplified model, and is particularly suited to multicriteria choice problems i.e. where the goal is to select the best alternative.

4. Spearman's Rank Correlation coefficient calculated between the model rank order and the true rank order. This measure is better suited to situations in which a full rank ordering of alternatives is desired.

2.5 Parameters of the simulation experiment

The effect of the size of a decision problem is investigated by making use of six problem sizes ($n = 5, m = 4$), ($n = 9, m = 8$), ($n = 9, m = 16$), ($n = 19, m = 8$), ($n = 19, m = 16$), and ($n = 39, m = 24$). The first and last of these are intended to be extreme decision problems (small and large respectively), with the intermediate four allowing for an independent investigation of alternatives and attributes in decision environments that are located – as far as this can be said at all of MCDA problems – somewhere between 'typically small' and 'typically big'. All simulations make use of $p = 50$ scenarios. Changes to the attribute values over scenarios were simulated using the coefficient of skewness of the gamma distribution, denoted by $S_{B_{ij}}$, and the standard deviation of the B_{ij}, denoted by $\sigma_{B_{ij}}$. As discussed above, the gamma distribution includes the symmetric normal distribution as a special case, which forms a base case for evaluation. Cases 2 and 3 in Table 2 model skewness in a single direction only (taken as positive), differing in the order of magnitude of the skewness, while in cases 4 and 5 the deviations may be either positively or negatively skewed.

Case	$S_{B_{ij}}$	Description
1	0	No skewness
2	U(0, 1)	Small positive skewness
3	U(0, 2)	Large positive skewness
4	U(-1, 1)	Small mixed (positive and negative) skewness
5	U(-2, 2)	Large mixed (positive and negative) skewness

Table 2: Simulated levels of attribute skewness

Appropriate values were chosen for the standard deviations $\sigma_{B_{ij}}$ by considering the correlation between rank orderings of the utilities in the different scenarios, on a particular attribute. For example, a high correlation between rank orders over scenarios implies that the rank order of alternatives (on a particular attribute) is similar in all scenarios. The moderate variability cases (cases 1 and 2 in Table 3) lead to rank correlations of the order of 0.3, and the high variability cases (cases 3 and 4 in Table 3) to correlations of the order of 0.1. Further, the effect of not only the magnitude of the variability but also its distribution over alternatives is investigated. That is, in cases 2 and 4 there is a greater range of possible variabilities relative to cases 1 and 3. Finally, the generated random variates are centred so that $\mathrm{E}[B_{ij}] = 0$.

Case	$\sigma_{B_{ij}}$	Description
1	U(0.3, 0.4)	Consistent moderate variability
2	U(0.2, 0.5)	Moderate variability, but differs between alternatives
3	U(0.5, 0.6)	Consistent high variability
4	U(0.4, 0.7)	High variability, but differs between alternatives

Table 3: Simulated levels of attribute variability

The main simulation study uses eight different basic preference structures, four of which are taken directly from Stewart [24]. Each of the utility function parameters τ_j, λ_j, and β_j takes on a 'low' and 'high' condition, with values for the low condition of both τ_j and λ_j generated uniformly on $[0.15, 0.4]$, and values for the high condition uniformly on $[0.6, 0.85]$. Values for the curvature of the utility function above the reference level, β_j, were generated uniformly on the interval $[1, 4]$ in the low condition and $[2, 8]$ in the high condition, and the curvature below the reference level, α_j, was generated uniformly on the interval $\beta_j + [1, 4]$ in the low condition and $\beta_j + [2, 8]$ in the high condition. For the sake of completeness, additional simulations were run using fully convex ($\tau_j = 1, \lambda_j = 1$), fully concave ($\tau_j = \lambda_j = 0$) and linear utility functions, but these are reported separately from the eight main structures discussed above, which employ the more general S-shaped utility function, and are only discussed under section 3.2. The six possible problem sizes (varying parameters n, m), twenty degrees of change to the attribute values (varying parameters $S_{B_{ij}}$ and $\sigma_{B_{ij}}$), together with the eight basic preference structures define the 960 basic decision problems for the simulated implementation of the simplified approach. For each combination of simulation parameters, 100 simulations were run and average statistics gathered. Although no standard errors are given in the results below, they were usually of the order of 0.05 for the average POSN and RANK scores. Since in general any differences in POSN or RANK scores less than 0.2 positions are not considered substantial enough to be considered of practical importance, and are thus not discussed in the results, all differences that are discussed would, by virtue of being greater than 0.2, be statistically significant at the 5% level, and usually considerably beyond.

3 Main results: approximation method, approximation accuracy, and preference structure

The analysis of results is contained in three sections: this first section contains what we consider to be the most important of our findings, regarding the effects of expected value approximation, incorporating the method and the accuracy of approximation; and the effects of the form of the underlying utility functions, incorporating the location of reference levels, utility at the reference level, and curvature of the utility function above and below the reference level. In section 4, we discuss the effects of the distributional form of the attribute evaluations, incorporating the variability and skewness of those evaluations; the effects of problem size, incorporating the number of alternatives and attributes. Finally, in appendix B we briefly decompose the earlier approximation results to analyse the relative influence of median and quantile approximations. Significant and substantial interaction effects, where they exist, are discussed in the main effect section to which they apply. The results of a multivariate analysis of variance is provided in the appendix (as Table 7) as a quick and concise way to summarise the statistical significance of the main effects and all second- and third-order interactions on POSN and RANK scores. The table shows F-statistics for all main and interaction effects that are significant in more than one model at the 0.5% level, where our convention is to identify those effects that are significant at the 5% level by a single asterisk and those effects that are significant at the 0.5% level by a double asterisk. In order to simplify the interpretation of results, the problem size in the discussion of all other effects has been fixed to $n = 19$ alternatives, so that model results can be compared to a completely random selection strategy, which would return average POSN and RANK scores of 10. Where any of the main or interaction effects differ as a function of the number of alternatives, this is highlighted under the discussion of the effect.

3.1 Effect of approximation method and accuracy

The models used to approximate the global utilities differ in two respects: firstly, whether the assumed form of the w_{ij}^R allows the variances of attribute evaluations to be included or not; and secondly, how close to the Pearson-Tukey values the approximations of expected value are. Table 4 shows the overall mean POSN and RANK scores, utility losses, and Spearman's rank correlations for the six models considered, under either the absence or presence of dominated alternatives.

	No dominated alternatives				With dominated alternatives			
	POSN	RANK	SRC	ULoss	POSN	RANK	SRC	ULoss
Model A_0	3.12	3.12	0.73	0.13	1.82	1.80	0.83	0.04
Model A_1	3.20	3.16	0.72	0.14	1.87	1.87	0.82	0.04
Model A_2	4.32	4.29	0.57	0.20	2.51	2.61	0.71	0.09
Model B_0	3.53	4.10	0.69	0.18	2.15	2.71	0.79	0.08
Model B_1	3.25	3.76	0.71	0.16	1.96	2.35	0.81	0.07
Model B_2	4.31	4.69	0.57	0.22	2.56	2.99	0.70	0.11

Table 4: Effect of model type and accuracy of approximation

We first consider the effect of model type and accuracy in the absence of dominated alternatives, before asking how dominance modifies these results. An initially surprising result is that the simpler model setting $w_{ij}^R = 0$ does not perform any worse than the more complex model which uses an approximation of the variance of the attribute evaluations in addition to their expected values, and in fact sometimes performs substantially better, most notably in the RANK scores when few or no errors are made: in this case, Model A_0 has a RANK score of 3.12, while Model B_0's score is 4.10. This and other important results relating to the method and accuracy of approximation are summarised below:

Result 1: *Type-A models offer equal or better overall performance than their type-B counterparts, regardless of the accuracy of approximation.*
Result 2: *Deteriorations in model accuracy only occur for large errors in approximation.*
Result 3: *Small errors in approximation, linked to the use of less extreme quantiles, can improve the accuracy of the type-B models.*

The relatively poor performance of the more complex type-B models can be traced to their sensitivity to different forms of preference functions, and in particular to their sensitivity to non-linearity pointed out by Kirkwood in [17]. As the utility functions are defined in (2), their second derivative is negative above the reference point and positive below it, with the result that, because the form $w_{ij}^R = (1/2)w_j u_j''(\mathrm{E}[Z_{ij}])$ used in [17] adds back the second derivative of the utility function evaluated at the expected attribute value, better-performing expected values incur relatively greater penalties than poorer-performing ones. Particularly with a view to the behavioural justifications used in the definition of the utility functions, this appears to be a genuine weakness of the type-B models. The robustness of the type-A models to the form of preference function used is perhaps the most important of our simulation results, and is elaborated on in section 3.2.

It is clear from Table 4 that the degree of accuracy exerts a substantial influence over the quality of results, more so than the type of model chosen. If sufficient care is taken to approximate the Pearson-Tukey values closely, then the results of the type-A models are reasonably good even if some anchoring or other contamination occurs. Model A_0 and A_1 locate the true best alternative near the third position of the model rank order (out of 19), with the alternative selected by the model also being ranked around third in the hypothesised true rank order. Even if a lack of rigour and effort result in a large degree of contamination occurring, average POSN and RANK scores remain in the top 20–25% of the rank order, which may be acceptable for some applications. The utility loss statistics also indicate that relatively little is given up in terms of the quality of the selected alternative until substantial approximation errors are made, with the utility of the selected alternative between 85% and 90% that of the true best alternative for the type-A models. The rank correlations reflect further that the accuracy of the approximation extends beyond the first one or two positions in the rank order.

A second significant and interesting result here is that small errors of the type incurred in Model B_1 in fact appear to *improve* the RANK scores of the type-B model, in this case from 4.10 to 3.76 (similar improvements are observed for the POSN scores). Bearing in mind the scenario-based approach used in the construction of quantiles, this suggests that the poor performance in model results observed under certain conditions may be improved by using less extreme quantiles to approximate the expected values, with the intriguing possibility that approximation methods might be adapted for various decision-making environments (in particular, for different preference structures). While a more detailed investigation of alternate treatments of risk, including alternate forms for w_{ij}^R and the use of alternate weights and quantiles in the approximation of expected values, lies beyond the scope of the current paper, it is clear that these constitute interesting and important areas for research. Appendix B, which examines the relative influence of the approximation of the median and the use of extreme quantiles in the Pearson-Tukey approximations, shows that the relative improvement in Model B_1 is largely due to the use of less extreme quantiles, and that the approximation of the median has relatively little impact. Other additional simulations run as preliminary investigations showed that the results obtained by Model A_0 can also be improved, but only by a practically insignificant margin (less than 0.1 position), by slightly increasing the weight attached to the two extreme quantiles.

The general results highlight the need for some care whenever risk is reduced to a single measure, but do not rule out or even discourage the use of such simplifying heuristics. On the contrary, it appears as if a reasonably diligent estimate of the expected value may under certain conditions produce results that are good enough in a large variety of contexts. As a basis for comparison, we simulated a model which directly uses the exact expected attribute values rather than the Pearson-Tukey approximation. Those results show a moderate improvement of 0.5 to 0.7 positions over the error-free approximation of Model A_0, providing further evidence that model performance is sensitive to the approximation of the expected values to be used in the simplified models, and suggesting that some improvements over the Pearson-Tukey approximations may be possible.

Turning briefly to the effect of having dominated alternatives in the choice set, Table 4 shows that the presence of dominated alternatives substantially improves the quality of the selected alternative but does not alter the direction or magnitude of the effects of model type or model accuracy. Thus the models are able to identify and exclude dominated alternatives, and this ability is independent of the type of model used. In fact, the effect of the presence of dominated alternatives turns out not to interact with any of the simulated external or internal environmental conditions, and we therefore restrict our attention in the remainder of the paper to results obtained from the non-dominated alternative sets.

3.2 Shape of utility functions

As mentioned in the previous subsection, the performance of the simplified approach is heavily influenced by the shape of the underlying preference functions. This is indicated by the ANOVA results in Table 7, which show that all the preference function parameters τ_j, λ_j, and β_j are highly significant, as are all of their two-way and three-way interactions. Because all interactions are significant, results are presented for all combinations of the three main effects in Table 5, rather than for each effect individually as is the case with most of the other main effects. Though the subsequent discussion focuses on non-linear preferences, results for linear utility functions are included for completeness; by definition these are unaffected by β or by the model type used, since the second derivative of linear u_j is zero. The main results are summarised as follows:

Result 4: *Model accuracy progressively deteriorates as preferences become more S-shaped, and is especially bad when τ and λ are very different so that a sharp preference threshold exists.*
Result 5: *Type-A models are more robust to changes in preference structure than type-B models, which perform better under convex or concave preferences but can be extremely poor when reference levels are low.*
Result 6: *The effects of changes in preference structures have the same direction in error-free and error-containing models, though sensitivity to those changes increases as approximation accuracy increases.*

We begin by examining the effect of various preference functions on the error-free Models A_0 and B_0, before looking at how approximation errors can modify these effects. The most basic results are (a) that model accuracy tends to be lowest when $\tau \neq \lambda$, (b) that model accuracy is higher when τ and λ are both low than when they are both high, (c) that model accuracy is higher (or at least equal) when preferences are fully concave or convex relative to when preferences are mostly concave (i.e. low τ and λ) or mostly convex (i.e. high τ and λ). Table 5 shows, for example, how under low β conditions Model B_0's POSN scores improve from 2.98 to 2.62 to 2.23 as the shape of the preference function changes from low τ, high λ to low τ, low λ to $\tau = \lambda = 0$, and improves from 4.23 to 2.54 to 2.05 as the shape of the preference function changes from high τ, lowλ to high τ, high λ to $\tau = \lambda = 1$. In most cases it is the first of these transitions (from different values of τ and λ to the same or similar values) that brings the greatest increase in accuracy, with a further transition to fully concave or convex preferences yielding marginal or no further improvements. Note also that while POSN and RANK scores are generally best under linear preference functions, the scores are generally not substantially better than POSN and RANK scores obtained when preferences are fully concave because of the various non-idealities that exist (e.g. use of the Pearson-Tukey approximation, use of scenario-based quantiles)

Before discussing possible reasons for these results, it is important to note that there are two exceptions to the general observation that accuracy tends to be deteriorate as preference functions become more sharply S-shaped. Firstly, Model A_0 shows a strong *increase* i.e. worsening, of POSN score from 3.64 to 4.54 when preferences become fully convex and the curvature parameter β is large. Note that no such change happens to Model A_0's RANK score, which is relatively constant around 3.5 whether preferences are convex or represented by high τ and λ. Secondly, RANK scores in Model B_0 *increase* (i.e. worsen) from 2.51 when τ is high and λ is low (under low β) to 2.90 when τ and λ are both high. Subsequent to this deterioration, RANK scores improve substantially to 2.03 (i.e. the expected result) when preferences become fully convex, but it is this earlier deterioration that requires explanation. Note that a similar exception occurs when β is high, but that none occur in the POSN scores.

Turning now to reasons behind the basic effects and their associated exceptions, we offer the following observations. Firstly, the finding that accuracy deteriorates as preference functions change from fully concave or convex to steeply S-shaped can be explained by the increased departure from linearity observed in the most S-shaped preference function (i.e. $\tau \neq \lambda$) in the region where most of the attribute values fall (i.e. in the interval $[0.25, 0.75]$). Since expected utility will equal the utility of the expected attribute value exactly in the case of a linear utility function, this result should not be particularly surprising.

Secondly, the exception observed in the increase in Model A_0's POSN scores when preferences become fully convex and curvature is high reflects the extreme noncompensatory nature of this particular utility function, which is effectively flat over a large portion of the attribute value domain and severely disadvantages any compromise alternatives offering moderately good performance across attributes. Thus while these compromise alternatives, which may figure prominently in the true rank order, fall far down the model rank order (and hence the observed increase in POSN score from 3.64 to 4.54), the selected alternative will tend to be one that performs excellently on one or two attributes. Since performing excellently on a small number of attributes does not preclude an alternative from occupying a good spot in the true rank order, there in no associated increase in RANK scores, which for Model A_0 moves inconsequentially from 3.51 to 3.57. A final question on this exception is: why do Model B_0's POSN scores not increase as Model A_0's do? The answer here may lie in the relatively greater penalties applied to alternatives with higher expected values by the type-B models (since the second derivative of the convex preference function is negative, becoming more negative as $\mathrm{E}[Z_{ij}]$ increases). These penalties appear to rein in the tendency of the highly nonlinear convex preference functions to select an alternative on the basis of a single outstanding expected value.

The second exception, in which Model B_0's RANK scores initially deteriorate when τ and λ are both high, is more difficult to explain. Here, the problem may be the actual discontinuity in the preference function at the reference level τ. Since Model B_0 adds back a weighted function of the $u_j''(\mathrm{E}[Z_{ij}])$, and this expression changes from positive to negative as one moves from below τ to above τ, an alternative with an expected value just below τ might receive an undeserved boost above a second alternative with an expected value just above τ. The extent of the discontinuity, i.e. the difference between the second derivative just above and below τ, and hence the propensity to cause such reversals, is greater when τ and λ are both high than when τ is high and λ is low. The extent of the discontinuity is also greater when τ is low and λ is high than when both are low, so that the low τ, high λ condition is in fact disadvantaged on both fronts: it is more non-linear in the domain of most activity, and it suffers from the effects of the discontinuity more. Evidence in support of these two speculative explanations are provided by the lack of reversals and the terrible performance of Model B_0 in this second case.

Having considered the impact of preference structures on the two error-free models, we might now ask whether those effects are any different when approximation errors creep in. In fact, all the observed findings for the error-free models hold true for the error-prone models as well, though sensitivity to changes in the shape of the utility functions decreases as approximation errors increase. Thus, model accuracy tends to deteriorate as preference functions become more S-shaped, and the same two exceptions exist i.e. an increase in the POSN scores of the type-A models when preferences become fully convex under the high β condition, and an initial increase in the type-B models' RANK scores when preferences change from high τ, low λ to high τ, high λ. The only additional comment is that the first exception, which was not observed in Model B_0, is clearly observed in the error-prone type-B models, and thus can be said to apply to both types of models in the presence of any approximation errors.

A final point is worth making regarding the relative performances of the type-A and type-B models. When preferences are either fully concave or convex, the more complex type-B models consistently outperform the type-A models, albeit by an amount that is only substantial when no assessment errors are made. For the other four preference structures, evidence supporting the selection of one or the other model is mixed, though a few themes emerge. Under highly non-linear preferences, the type-A models substantially outperform the type-B models when reference levels are low. When the curvature parameter β is low, type-A models are only notably superior under the joint condition that τ is low and λ is high. If these conditions are switched, so that τ is high and λ low, then it is the type-B models that are superior, with the other two conditions i.e. when $\tau = \lambda$, leading to roughly the same accuracy in the two types of models. As mentioned before, it also sometimes occurs that the use of less extreme quantiles results in better model performance for the type-B models. A final conclusion might therefore be that the more complex approximation used by the type-B models are particularly suited to fully convex or concave utility functions, but that in the case of S-shaped utility functions, these models can behave quite strangely. It also seems worthwhile to note that the sensitivity to non-linearity highlighted in [17] may in fact interact in a fairly subtle way with the forms of simplified model and expectation approximation that is used.

		POSN							RANK						
	τ	n/a	0	low	low	high	high	1	n/a	0	low	low	high	high	1
	λ	(linear)	0	low	high	low	high	1	(linear)	0	low	high	low	high	1
Low β	Model A_0	2.14	2.22	2.28	3.10	3.44	2.88	2.98	2.15	2.34	2.31	3.35	3.19	2.82	2.63
	Model A_1	2.35	2.43	2.40	3.39	3.51	2.85	3.19	2.38	2.58	2.50	3.66	3.10	2.70	2.75
	Model A_2	3.89	3.93	3.89	4.59	4.50	4.11	4.33	4.03	4.28	4.10	5.06	3.87	3.84	3.68
	Model B_0	2.14	2.05	2.54	4.23	2.89	2.62	2.23	2.15	2.05	2.58	6.30	2.51	2.90	2.03
	Model B_1	2.35	2.27	2.38	3.91	3.02	2.56	2.73	2.38	2.35	2.41	5.29	2.68	2.63	2.37
	Model B_2	3.89	3.81	3.81	4.77	4.27	3.95	4.12	4.03	4.04	3.93	5.58	3.78	4.01	3.60
High β	Model A_0	2.14	2.41	2.54	3.20	3.91	3.64	4.54	2.15	2.64	2.67	3.45	3.64	3.51	3.57
	Model A_1	2.35	2.64	2.60	3.39	3.89	3.55	4.72	2.38	2.94	2.79	3.68	3.50	3.34	3.67
	Model A_2	3.89	4.04	3.94	4.45	4.63	4.42	5.34	4.03	4.60	4.34	5.06	4.00	4.08	4.05
	Model B_0	2.14	1.96	3.34	5.41	3.76	3.43	3.46	2.15	2.00	3.80	7.83	3.19	3.72	2.65
	Model B_1	2.35	2.26	2.72	4.83	3.45	3.13	4.02	2.38	2.44	2.98	7.54	3.04	3.47	3.04
	Model B_2	3.89	3.79	3.81	5.27	4.39	4.22	4.95	4.03	4.19	4.02	7.69	3.96	4.57	3.81

Table 5: Effect of shape of preference functions $(\tau_j, \lambda_j, \beta_j)$

4 Secondary results: distribution of evaluations and problem size

4.1 Variability and skewness of attribute evaluations

The variability of attribute evaluations is determined by the four $\sigma_{B_{ij}}$ cases described previously, with results shown in Figure 2 for POSN scores (similar observations hold for the other outcome measures). The effect of attribute variability is summarised in the following results.

Result 7: *Model accuracy deteriorates substantially when attribute evaluations are highly variable within alternatives, though differences in variability between alternatives exercises little or no effect.*
Result 8: *Type-A models tend to be more sensitive to changes in attribute variability than type-B models. Sensitivity to changes in the variability of attribute evaluations also increases with approximation errors, particularly for type-B models.*

The performance of all models is generally worse when attribute evaluations are more variable, and potential deteriorations can be severe; Model A_0's POSN score, for example, deteriorates 1.24 positions from 2.50 to 3.74 as variability increases. Under moderate attribute variability, the performance of Models A_0 and A_1 is good enough that the strategy might be used with some confidence, with POSN and RANK scores below 3, while under highly variable attribute evaluations POSN scores remain below 4 in all models except those making the largest assessment errors. Two comments can be made about the variability results. Firstly, the less accurate models tend to exhibit greater sensitivity to attribute variability in the type-B models, with Model B_2 experiencing deteriorations in RANK scores (not shown here) from 4.19 to 5.19 – more than double that of Model B_0, which deteriorates from 3.91 to 4.30 (the effect is less pronounced in the POSN scores, and is not observed at all in the type-A models). Secondly, the relative variability among alternatives appears to exercise far less of an effect on performance than the amount of absolute variability, with the cases in which there is a greater variety of variabilities between alternatives (case 2 and case 4) performing much the same as those cases with equivalent average variability, but in which all alternatives were similarly variable (case 1 and case 3). This is again most clearly seen in the type-B models.

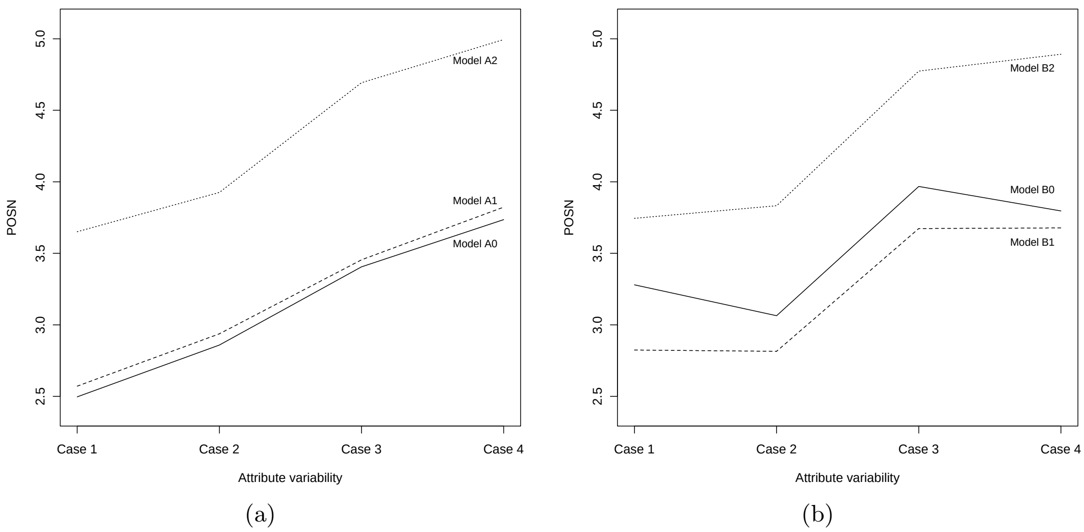
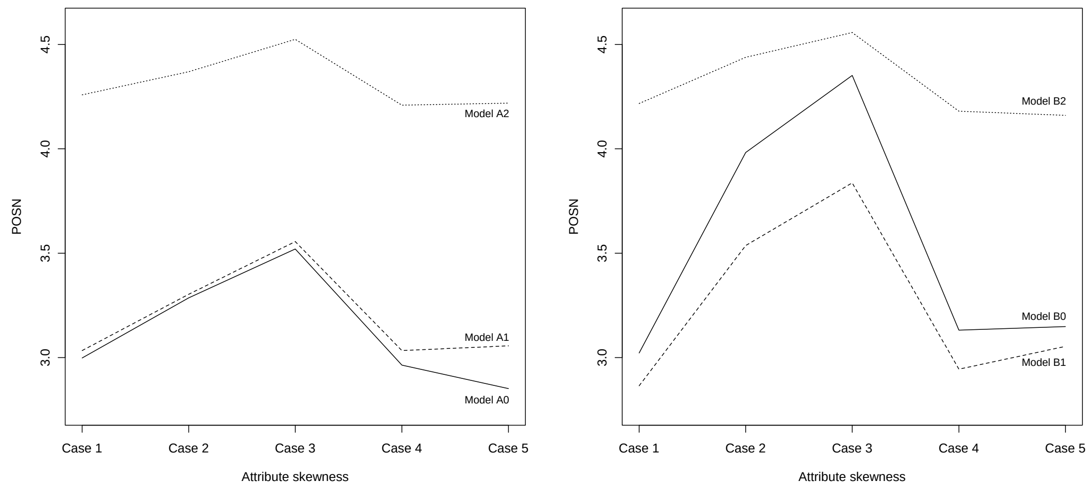


Figure 2: Effect of attribute variability ($\sigma_{B_{ij}}$) on (a) type-A models, (b) type-B models

An interesting feature of the results is the reasonable insensitivity of most models to the skewness of the attribute evaluations as determined by the $S_{B_{ij}}$ parameter. Results are shown in Figure 3 for POSN scores – observations are similar if the other outcome measures are used, and summarised as follows:

Result 9: *Type-A models are robust to changes in the skewness of attribute evaluations. Positively-skewed attribute evaluations can greatly harm the accuracy of type-B models if reference levels are low, since this increases the proportion of attribute evaluations that occur in the region of the discontinuity at the reference level, but under other preference structures the type-B models are also relatively insensitive to skewness.*
Result 10: *Sensitivity to attribute skewness in the type-B models increases as approximation accuracy increases. This is attributable to the deteriorations being due to the interaction of skewness with the effect of preference structure, which is more forcefully experienced by the error-free models.*

The results of Model A_0 and A_1 remain acceptable under all skewness conditions – all POSN scores are below 3.6 – providing further evidence that a strategy of using expected values may be appropriate in a reasonably wide variety of decision environments. There are four noteworthy features of the skewness results. Firstly, type-A models are more robust to changes in skewness than are the type-B models. Secondly, both types of models show increasing sensitivity to skewness as the accuracy of approximation improves, with deteriorations in POSN scores caused by skewness of $4.35 - 3.02 = 1.3$ positions for Model B_0, but only $4.56 - 4.16 = 0.4$ positions for Model B_2. Thirdly, the main determinant of model performance is not the magnitude of skewness, but whether all evaluations are skewed in the same direction or not. It is interesting to note that while increasing skewness in a single direction causes model performance to deteriorate, increasing skewness has little or no effect – and may even marginally improve model performance – if the skewness may be either positive or negative. The study of bias in the context of using decomposition approaches in subjective probability or utility assessment has returned similar findings, to the effect that certain positively- and negatively-biased inputs may interact to cancel one another out and provide less biased final estimates [20, 19]. Finally, differences in the directionality of skewness exert greater influence as the absolute magnitude of the skewness increases, becoming substantial when skewness is large.



(a) (b)

Figure 3: Effect of attribute skewness ($S_{B_{ij}}$) on (a) type-A models, (b) type-B models

This is not the whole story, however. There is an important interaction between attribute skewness and reference levels in which attribute skewness can have a disproportionately large impact when reference levels are low. This interaction is far more pronounced in type-B models; again, the type-A models show a certain robustness to changes in the problem context. Figure 4 shows the loss in utility resulting from using each of the type-B models under various skewness conditions, when reference levels are low and high respectively; the same effect is visible in other outcome measures. The net effect is that if reference levels are low, performance is severely downgraded in Models B_0 and B_1 if unidimensional attribute skewness is present (utility losses of 0.32 and 0.25 respectively) relative to when no skewness or both positive and negative skewness is present (utility losses of around 0.14 for both models). A similar interaction, with a similar magnitude of effect, is observed between λ_j and attribute skewness. This interaction, not shown here, means that attribute skewness can have a degrading effect on Model B_0 and B_1 results when λ_j is high, although it exercises little or no effect when λ_j is low. The reason for this sensitivity becomes clear once one considers the previous results indicating (a) that model accuracy in the type-B models is extremely poor when τ is low and λ is high, and (b) that one of the reasons for this is the strong discontinuity around the reference point τ. The effect of positive attribute skewness is to shift the mode of the attribute evaluations to the left, increasing the relative proportion of evaluations that are in the vicinity of the low reference point and hence exacerbating the detrimental effects of the discontinuity. Given that in most real-world problems both positive and negative skewness will exist, the problem may be less severe than suggested by the simulation results; nevertheless it is rewarding to find some supporting evidence for our speculations on the origin of the preference function effects in these interactions with attribute skewness. The results also suggest that external conditions (such as the distribution of attribute evaluations) may interact with internal conditions (such as decision maker preferences) in some quite subtle ways to influence the relative performances of the various simplified models.

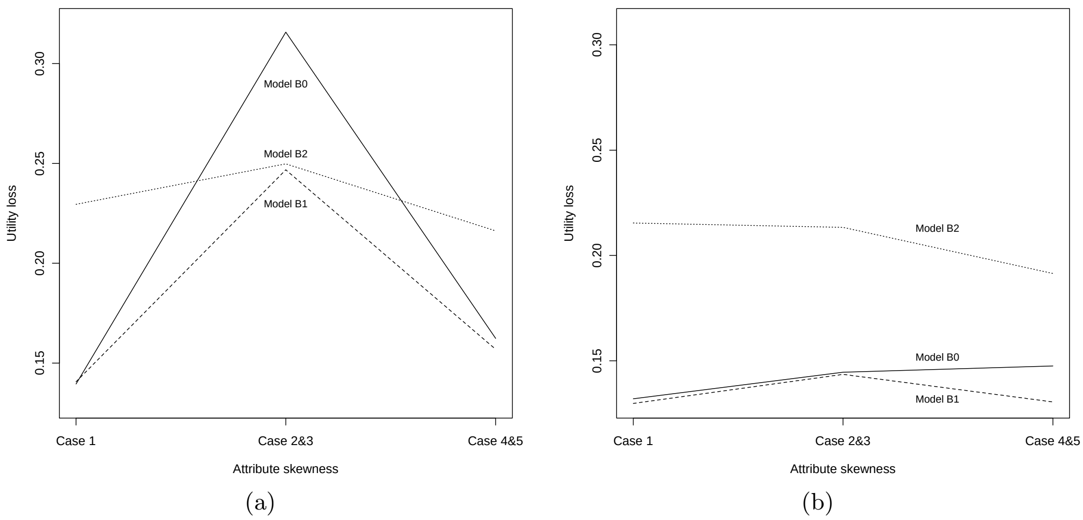


Figure 4: Effect of attribute skewness ($S_{B_{ij}}$) on type-B models when (a) reference levels (τ_j) are low, (b) reference levels are high

4.2 Problem size

The problem size is determined by the number of alternatives n and the number of attributes m. Table 6 shows the average utility loss scores for various combinations of alternatives and attributes. POSN and RANK scores, which will naturally change as the number of alternatives changes, are less useful here and are not shown. Interest is predominantly focused on columns 3 to 6 of the table, with column 2 and column 7 providing results for an extremely small and large decision problem respectively. The results can be summarised as follows:

Result 11: *Both types of models are robust (and equally so) to changes in the number of attributes and in particular alternatives, even when these values are extreme.*

The results in columns 3 to 6 of Table 6 show that all models are very robust to changes in the numbers of alternatives, with the largest change being an improvement in Model B_0's utility loss from 0.21 when $n = 9$ to 0.19 when $n = 19$ (for $m = 16$ attributes). Results are more sensitive to changes in the number of attributes, and deteriorate by up to 0.04 (from 0.12 to 0.16, a deterioration of about 33%, when $n = 9$ in Model A_0). Nevertheless, the performance of Models A_0, A_1, and to a certain extent B_1, are encouraging in that results appear good for even the quite large decision problem that is represented by $n = 19$ alternatives evaluated over $m = 16$ attributes. At the extremes of problem size, one continues to observe the same patterns, although the size of the improvement in moving to $m = 5$ attributes (or the deterioration in moving to $m = 24$ attributes) remains modest. Interestingly, and in contrast to many of the other results, the order of magnitude of the effect of problem size is not subject to substantial change over either model type or the accuracy of approximation.

	$n = 5$	$n = 9$		$n = 19$		$n = 39$
	$m = 5$	$m = 8$	$m = 16$	$m = 8$	$m = 16$	$m = 24$
Model A_0	0.11	0.12	0.16	0.12	0.15	0.17
Model A_1	0.12	0.13	0.16	0.12	0.15	0.17
Model A_2	0.17	0.19	0.23	0.18	0.22	0.24
Model B_0	0.17	0.18	0.21	0.17	0.19	0.20
Model B_1	0.14	0.16	0.19	0.15	0.17	0.19
Model B_2	0.19	0.21	0.24	0.20	0.24	0.25

Table 6: Effect of number of alternatives (n) and number of attributes (m) on utility loss

5 Conclusions

In this paper we have used a simulation experiment to evaluate a seemingly quite radical simplification of a decision problem under conditions of risk in which the distributions of attribute evaluations are replaced by their expectations. The use of this simplification may be more widespread than it would appear from a review of the decision aiding literature, in that most decision problems incorporate some degree of uncertainty which may be ignored or treated via sensitivity analysis instead of employing the theoretically more sound but far more demanding expected utility framework. In these cases, which would include many applications of multiattribute value theory, the reduction of distribution uncertainty to a deterministic value must come at some cost. This paper makes an attempt at evaluating the extent of the cost and the conditions that this cost is most sensitive to.

The most important of our simulation results were given in the six 'Results' highlighted in section 3. Those results indicate that the resulting simplified model can give results that would be acceptable in many, but not all, contexts. Provided that conditions are not too unfavourable, the alternative selected by the simplified model would on average appear between second and third position in the decision maker's hypothetical true preference order, which for the considered 19-alternative cases may be considered a good performance, particularly taking into account the inherent imprecisions and uncertainties in all stages of the decision process. The quality of results does deteriorate when highly inaccurate estimates of the Pearson-Tukey expected value approximations are used, but moderate inaccuracies are easily tolerated and in fact even lead to improvements in the type-B models under some conditions, a result due to the use of less extreme quantiles (10% rather than 5% in this case). This leaves the intriguing possibility open of tailoring approximation methods to the type of approach used, and various situational factors such as preference structure and the distribution of attribute evaluations. In this paper we are content to stop at the two main messages emerging on this theme from the simulations: that the use of expected values can provide acceptable results; and that model performance is aided by ensuring that the estimation itself is fairly accurate, but clearly the potential for context-dependent adjustment of the simplified models is worthy of future research.

Somewhat surprisingly, we identified the simpler type-A models as giving better overall accuracy than their more complex type-B counterparts, regardless of accuracy, which was later traced to the particularly bad performance of the type-B models under certain preference structures. The particularly severe sensitivity of the approximation using $w_{ij}^R = (1/2)w_j u_j''(\mathrm{E}[Z_{ij}])$ to non-linearity in the preference functions constitutes our other main conclusion. These results support the findings in [17], which we have in a sense extended by detailing the nature of the sensitivity and including the possibility of reducing it somewhat, either by choosing less extreme quantiles or by using the simpler approximation form $w_{ij}^R = 0$. Specifically, our three results highlighted (1) that model accuracy deteriorates as preferences become more S-shaped, and is especially bad when τ and λ are very different so that a sharp preference threshold exists, (2) that this is particularly true of the type-B models, such that they cannot be used with any confidence under these combinations of preference function parameters, especially if in addition both segments of the preference function are highly non-linear. The type-A models show the same patterns but are considerably more robust than the type-B models. We traced the sensitivity of the type-B models to two features: firstly, to the non-linearity of the preference functions in the region of the attribute evaluation domain where most evaluations fall (a sensitivity that the type-A models also suffer from); secondly, to the discontinuity of the preference functions at the reference level, where the second derivative changes abruptly from positive to negative (which only affects the type-B models). We do not have any firm proof for these speculations, but they are supported by evidence from other parts of the simulation. Finally, these preference structure effects are more strongly observed in those models making few or no approximation errors, although the direction of the effects remains the same.

The other effects investigated by the simulation study are the distribution of attribute evaluations and the number of alternatives and attributes constituting the decision problem. Our results there show that attribute variability can exert a substantial influence on the quality of the simplified model, although importantly, given sufficiently accurate estimation, the simplified model appears to perform acceptably even when attribute evaluations are quite variable: evaluations in our simulated 'moderate' variability condition in fact exhibited quite substantial variability, with correlations of rank orders between the scenarios averaging around 0.3. Under these conditions, average POSN and RANK scores for both Model A_0 and A_1 remained below 3, so that these models might be used with some confidence. The two other main results in this section indicate, firstly, that while model accuracy deteriorates substantially when attribute evaluations are highly variable within alternatives, differences in variability between alternatives exercises little or no effect; and secondly, that sensitivity to changes in variability is greatest in the type-A models. Sensitivity to changes in the variability of attribute evaluations also increases with approximation errors, most notably for the type-B models. The degree to which evaluations are skewed has less impact on results, and only really acts when the skewness is quite extreme. Here, only Models B_0 and B_1 exhibited any real sensitivity to the skewness of attribute evaluations, and this only when evaluations are skewed in a single direction (positively in this case), which appears to occur because the negative effect of a sharp threshold in the preference function (e.g.

low τ, high λ) is exacerbated. The variability and skewness results are encouraging for the use of expected values in cases where there is a fairly low degree of risk or uncertainty, which is likely to be the case in many of the decision problems to which multiattribute value theory is applied directly. Although one cannot in practice judge the quality of expectation estimation without fully enumerating all outcomes, the evidence suggests that one might be able to get away with modelling the expected values directly in a deterministic framework in quite a wide variety of decision contexts. Increasing the number of attributes in a decision problem has a small negative effect on model accuracy, but model performance remains fairly good even in the case of large problem sizes. Both types of models are very robust to changes in the number of alternatives present, even when this number is large, reinforcing the notion that a simplified model may be applicable across a wide variety of decision problems.

Further research might use a similar simulation approach to investigate other ways of incorporating the variability of attribute evaluations in a multicriteria context using only limited information, possibly via alternate forms for w_{ij}^R or use of other quantiles. It seems reasonable to infer from our simulation results that the success of a particular simplified model may depend quite heavily on the type of underlying preferences that are present. Other internal and external environmental variables peculiar to the decision aiding process might also be introduced: perhaps including the omission of attributes, errors in the estimation of attribute weights, violations of preferential independence, and the number of piecewise segments used to model the marginal utility function. Also, we have tested the simplification strategy in the context of value function models: similar tests can be envisaged for other MCDA methodologies, for example the outranking approaches, certain versions of goal programming, and the family of stochastic multicriteria acceptability analysis (SMAA) techniques. More broadly, simulation studies such as the one presented here provide ideal research opportunities for testing the effects of other simplified decision models, which can contribute to the development and refinement of models aiding decision making under conditions of risk or uncertainty: models based on a small number of state scenarios [12], in the spirit of scenario planning, come to mind as ideal candidates for such evaluation.

With reference to our aims stated in section 1, a summarised conclusion can be stated as follows: a simplified model replacing distributional attribute evaluations with their expected values provides a generally acceptable level of model performance, with the proviso that the estimation be performed with a reasonable degree of accuracy. Provided that the estimation is sufficiently accurate, the simplification afforded by the type-A models is robust to quite a wide variety of changes in both the external and internal decision environments, with results remaining good despite increasing problem sizes, more variable and skew attribute evaluations, and different preference structures. The performance of the more complex type-B models can also be good, but this is heavily dependent on the types of preference functions used, with the possibility of extremely poor performance under certain utility functions with sharp preference thresholds.

A Analysis of variance results for simulation experiment

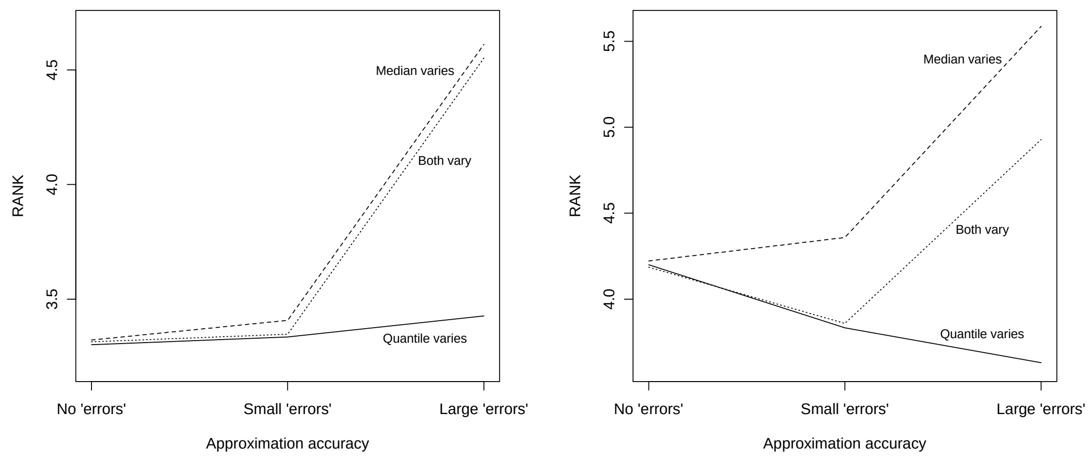
Effect	DoF	Model A_0	Model A_1	Model A_2	Model B_0	Model B_1	Model B_2
m	2	124.13**	117.92**	124.75**	58.42**	84.07**	112.20**
$S_{B_{ij}}$	8	35.25**	23.14**	4.60**	178.30**	95.45**	12.59**
$\sigma_{B_{ij}}$	6	223.17**	203.00**	159.12**	79.89**	132.76**	136.93**
τ_j	2	239.03**	146.66**	201.27**	1226.86**	877.04**	385.10**
λ_j	2	56.01**	68.53**	53.42**	1384.02**	1313.33**	597.03**
β_j	2	111.69**	73.51**	7.10**	507.99**	380.63**	135.20**
$m \times S_{B_{ij}}$	8	2.05*	0.92	1.81	3.10**	3.11**	1.69
$S_{B_{ij}} \times \sigma_{B_{ij}}$	24	2.27**	2.25**	2.28**	4.01**	4.12**	1.82*
$m \times \tau_j$	2	2.15	1.13	0.90	11.39**	7.60**	0.91
$S_{B_{ij}} \times \tau_j$	8	6.15**	16.55**	7.44**	326.05**	87.07**	3.16**
$\sigma_{B_{ij}} \times \tau_j$	6	3.60**	4.10**	2.36*	8.26**	2.80*	2.08
$S_{B_{ij}} \times \lambda_j$	8	3.37**	5.82**	4.22**	41.81**	23.58**	2.62*
$\sigma_{B_{ij}} \times \lambda_j$	6	0.70	1.26	0.45	13.31**	5.62**	0.62
$\tau_j \times \lambda_j$	2	237.68**	310.22**	78.27**	1013.49**	1247.97**	366.61**
$S_{B_{ij}} \times \beta_j$	8	6.55**	5.33**	1.80	9.18**	11.27**	3.35**
$\tau_j \times \beta_j$	2	27.78**	27.74**	4.91*	29.37**	57.84**	35.77**
$\lambda_j \times \beta_j$	2	0.85	0.55	0.36	4.79*	98.75**	89.73**
$S_{B_{ij}} \times \tau_j \times \lambda_j$	8	5.25**	3.72**	1.63	57.22**	52.95**	6.13**
$\sigma_{B_{ij}} \times \tau_j \times \lambda_j$	6	4.32**	8.34**	2.74*	12.73**	6.24**	2.13*
$S_{B_{ij}} \times \tau_j \times \beta_j$	8	3.01**	2.33*	0.52	31.51**	17.52**	3.51**
$\sigma_{B_{ij}} \times \lambda_j \times \beta_j$	6	1.17	1.21	1.07	3.85**	3.40**	2.00
$\tau_j \times \lambda_j \times \beta_j$	2	10.28**	10.76**	3.44*	4.97*	31.49**	42.23**

Table 7: MANOVA results for all simulated models ($n = 19$)

B Relative effects of median and quantile approximations

Earlier in section 3.1, we identified that the performance of the type-B models in fact improved in moving from Model B_0, which uses the exact Pearson-Tukey approximation, to Model B_1, which anchors the median estimate slightly to the most likely scenario and uses less extreme quantiles. Since Model B_1 (and Model B_2 for that matter) simultaneously vary both the median approximation and the quantile approximation, an interesting question is which of these approximations has the greater effect on results, and in particular which one is leading to the observed improvement. Figure 5 shows the changes in RANK scores that are observed when the median and quantile approximations are varied individually, and together. For each quantity, the approximation can be carried out with no 'errors' (using the true median or the 5% and 95% quantiles as in Model A_0 and B_0), small 'errors' (using the Model A_1 and B_1 median approximation and quantiles), or large 'errors' (using the Model A_2 and B_2 median approximation and quantiles), while holding the other quantity constant at its 'error'-free value.

The results clearly show that anchoring on the most likely scenario and adjusting insufficiently in the approximation of median attribute values consistently results in deteriorations in model performance, in both type-A and type-B models. These deteriorations are inconsequential for modest amounts of anchoring, but severe anchoring is severely punished. The use of less extreme quantiles, on the other hand, leads to a moderate but consistent improvement in the quality of the selected alternative in the type-B models, while having little or no effect on the type-A models. The net result is that using less extreme quantiles to some extent protects the type-B models from the harmful effects of anchoring and adjustment. In the main body of the paper we referred to changes to both median and quantile approximations as 'errors' on the basis that they could be justified using the well-known heuristics of anchoring and insufficient adjustment, and overconfidence respectively. The fact that they appear to sometimes lead to *improved* performance, and that one can potentially compensate for the other, implies that perhaps this label was unfair.



(a) (b)

Figure 5: Effect of median and extreme quantile approximation on RANK scores of (a) type-A models, (b) type-B models ($n = 19, m = 16$)

References

[1] M. Allais. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Americaine. *Econometrica*, 21:503–546, 1953.

[2] F. Barron. Influence of missing attributes on selecting a best multiattributed alternative. *Decision Sciences*, 18:194–205, 1987.

[3] G. Beim and M. Levesque. Selecting projects for venture capital funding : A multiple criteria decision approach. Technical Report 791, Weatherhead School of Management, Case Western University, 2004.

[4] L. Brenner, D. Koehler, V. Liberman, and A. Tversky. Overconfidence in probability and frequency judgments: A critical examination. *Organisation behaviour and human decision processes*, 65(3):212–219, 1996.

[5] J. Butler, J. Jia, and J. Dyer. Simulation techniques for the sensitivity of multi-criteria decision models. *European Journal of Operational Research*, 103:531–546, 1997.

[6] Y. Canbolat, K. Chelst, and N. Garg. Combining decision tree and MAUT for selecting a country for a global manufacturing facility. *Omega*, 35:312–325, 2007.

[7] E. Carbone and J. Hey. Estimation of expected utility and non-expected utility preference functionals using complete ranking data. In B. Munier and M. Machina, editors, *Models and experiments in risk and rationality*, pages 119–140. Kluwer Academic Publishers, Dordrecht, 1994.

[8] W. Edwards, editor. *Utility theories: measurements and applications*. Kluwer Academic Publishers, Boston, 1992.

[9] D. Ellsberg. Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics*, 75:643–669, 1961.

[10] N. Epley and T. Gilovich. The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, 17(4):311–318, 2006.

[11] S. French. *Decision Theory: An Introduction to the Mathematics of Rationality.* Ellis Horwood, Chichester, 1988.

[12] G. Islei, G. Lockett, and P. Naudé. Judgemental modelling as an aid to scenario planning and analysis. *Omega*, 27:61–73, 1999.

[13] J. Jia, G. Fisher, and J. Dyer. Attribute weighting methods and decision quality in the presence of response error: a simulation study. *Journal of Behavioural Decision Making*, 11:85–105, 1998.

[14] D. Keefer. Certainty equivalents for three-point discrete-distribution approximations. *Management Science*, 40(6):760–773, 1994.

[15] D. Keefer and S. Bodily. Three-point approximations for continuous random variables. *Management Science*, 29(5):595–609, 1983.

[16] R. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley & Sons, New York, 1976.

[17] C. Kirkwood. Estimating the impact of uncertainty on deterministic multiattribute evaluation. *Management Science*, 38(6):819–826, 1992.

[18] R. Lahdelma, S. Makkonen, and P. Salminen. Two ways to handle dependent uncertainties in multi-criteria decision problems. *Omega*, doi:10.1016/j.omega.2006.08.005, 2006.

[19] H. Ravinder. Bias in aggregation of subjective probability and utility. *Journal of the Operational Research Society*, 43(6):621–627, 1992.

[20] H. Ravinder, D. Kleinmuntz, and J. Dyer. The reliability of subjective probability assessments obtained through decomposition. *Management Science*, 34(2):186–199, 1988.

[21] G. Shafer. Savage revisited. *Statistical Science*, 1(4):463–485, 1986.

[22] T. Stewart. Use of piecewise linear value functions in interactive multicriteria decision support: a Monte Carlo study. *Management Science*, 39(11):1369–1381, 1993.

[23] T. Stewart. Simplified approaches for multicriteria decision making under uncertainty. *Journal of Multi-criteria Decision Analysis*, 4:246–258, 1995.

[24] T. Stewart. Robustness of additive value function methods in MCDM. *Journal of Multi-criteria Decision Analysis*, 5:301–309, 1996.

[25] A. Tversky and D. Kahneman. Judgment under uncertainty: heuristics and biases. *Science*, 185:1124–1131, 1974.

[26] A. Tversky and D. Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

[27] K. Van der Heijden. *Scenarios: The Art of Strategic Conversation*. John Wiley & Sons, New York, 1996.

[28] D. Von Winterfeldt and W. Edwards. *Decision Analysis and Behavioural Research*. Cambridge University Press, London, 1986.