Omega 40 (2012) 456–464

Contents lists available at SciVerse ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

A comparison of simplified value function approaches for treating uncertainty in multi-criteria decision analysis

Ian N. Durbach*, Theodor J. Stewart

Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, Cape Town, South Africa

ARTICLE INFO

Article history:
Received 28 December 2010
Accepted 23 August 2011
Processed by Associate Editor Triantaphyllou
Available online 28 September 2011

Keywords:
Decision making/process
Multicriteria
Risk

ABSTRACT

Uncertainty is present in many decisions where an action's consequences are unknown because they depend on future events. Multi-attribute utility theory (MAUT) offers an axiomatic basis for choice, but practitioners may prefer to use simpler decision models for transparency, ease of use, or other practical reasons. We identify some 'simplified' models currently in use and use a simulation experiment to evaluate their ability to approximate results obtained using MAUT. Our basic message is that avoiding assessment errors in the application of a simplified model is more important than the choice of a particular type of model, but that the best performance over a range of decision problems is from a model using a small number of quantiles.

1. Introduction

In decision-making the consequences of an action are often unknown because they depend on future events. Many models exist for multi-criteria decision analysis (MCDA) under such conditions of uncertainty; perhaps the most well-known are those based on multi-attribute utility theory MAUT, e.g. [25]. The great strength of MAUT is its axiomatic foundation "justifying the prescriptive approach provided the problem owners accept the related rationality assumptions" [21], but even in its simplest form the practical implementation of MAUT is formidable, requiring the assessment of probability distributions over each attribute as well as trade-offs involving single- and multi-attribute lotteries. Although there are several practical applications of MAUT reported in the literature e.g. [8,48], this number is small relative to MAUT's theoretical standing. Yet practitioners may prefer to use simpler decision models for transparency, ease of use, or other practical reasons. In this paper we identify a number of such 'simplified' decision models and evaluate their ability to replicate results obtained using MAUT. In doing so we hope to provide some guidance to practitioners about the types of simplified models that are being used for uncertain decision making.

Our evaluation uses a simulation experiment, acknowledging that while simulation allows us to assess how the simplification of MAUT models might impact results, it cannot evaluate critical issues like ease-of-use or insight generated (see [13] for an assessment of these). Simulation results are unable to provide general conclusions on the viability of different methods, but provide inputs to such discussion by identifying the potential trade-offs in accuracy that are implied when using a simplified model. Ultimately accuracy must be weighed against other factors to determine which decision model may be most appropriate for a problem.

The five models that we test represent uncertain attribute evaluations using (a) expected values, (b) expected values and variances, (c) expected values and the probabilities of obtaining performance below a cut-off, (d) quantiles, or (e) a small number of 'scenarios'. Since these methods all summarize aspects of probability distributions, they are referred to collectively as 'simplified' approaches; models (b) and (c) are sometimes referred to collectively as models using 'explicit risk attributes'. All models are based upon the principles of value function methods e.g. [6]; other MCDA methods are left to future research.

The paper is organized as follows. Section 2 provides an overview of the literature on uncertainty modeling in MCDA. Section 3 lists research hypotheses to be tested by the simulation. Section 4 outlines the simulation experiment, and Section 5 presents the results. A final section concludes with implications for practice.

2. Uncertainty modeling in multi-criteria decision analysis

Consider a decision problem consisting of I alternatives denoted by a_i, $i \in \{1,\ldots,I\}$, evaluated on J attributes denoted by c_j, $j \in \{1,\ldots J\}$. Let Z_{ij} be a random variable denoting the evaluation of a_i on c_j, and $u_j(Z_{ij})$ be single-attribute utility functions. Then the additive MAUT model [25] evaluates alternatives by their expected utilities

$$U_i = \sum_{j=1}^{J} w_j E[u_j(Z_{ij})] \tag{1}$$

* Corresponding author. Tel.: +27 216505058; fax: +27 216504773.
E-mail address: ian.durbach@uct.ac.za (I.N. Durbach).

0305-0483/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.omega.2011.08.004

where U_i is the expected utility of alternative a_i and w_j is an attribute importance weight indicating the relative importance of a one-unit change in attribute c_j e.g. [6]. The additive MAUT model requires additively independent preferences [25], although it can often closely approximate the multiplicative model when additive independence does not hold [41]. We consider further simplifications of the MAUT model; the following is not exhaustive but provides a broad coverage of simplifications appearing in the literature.

2.1. Models using expected values and explicit risk attributes

This approach provides a single or small number of risk measures indicating how 'risky' performance is. The fundamental notion is that uncertain evaluations can be expressed in terms of 'value' and 'risk' components. Models take the form

$$U_i^{(\mathrm{risk})} = \sum_{j=1}^{J} [w_j u_j(E[Z_{ij}]) - w_{ij}^R R_{ij}] \tag{2}$$

where R_{ij} is a measure of the 'risk' of Z_{ij} and w_{ij}^R is a 'risk weight' for R_{ij}. In this general formulation the risk weights may depend on alternatives as well as attributes. Several models are possible. The use of variances is standard in (single-attribute) portfolio optimization e.g. [29]; multi-attribute applications are reported in [11,26] and [3]. Kirkwood [26] has shown that using variances with $w_{ij}^R = (-1/2)w_j u_j''(E[Z_{ij}])$ can closely approximate (1) if the Z_{ij} are normally distributed (or numerous enough for the central limit theorem to apply) and the u_j "do not deviate too much from linear". However, simulation results [12] have suggested that under highly non-linear preferences the approximations in [26] do worse on average than a model using only expected values i.e. setting $w_{ij}^R = 0$. Examples using only expected values are [20] and [22], but it seems reasonable to suggest that a fair proportion of applications of multi-attribute *value* theory MAVT, e.g. [25] would also fall into this category. Probabilities of obtaining performance below a cut-off have also been used to measure risk e.g. [43,2], including the multi-attribute preference models using uncertain targets that have been shown to be equivalent to MAUT [7].

2.2. Models using quantiles

In practical decision analysis it is common to represent probability distributions using three to five quantiles [16]; this is the basis for the well-known bisection and interval elicitation methods e.g. [40]. Triples consisting of the minimum, median/mode, and maximum are also popular in fuzzy decision analysis e.g. [28,15]. Quantiles may be used in several ways–full distributions may be fitted to them, or they can be used to approximate moments e.g. [24]. The general quantile model evaluates a_i by

$$U_i^{(\mathrm{quan})} = \sum_{j=1}^{J} \left[w_j \sum_{k=1}^{N_q} w_{q_k} u_j(z_{ij}^{(q_k)}) \right] \tag{3}$$

where q_k refers to a specific quantile, $z_{ij}^{(q_k)}$ is the q_kth quantile of Z_{ij}, w_{q_k} denotes the weight associated with quantile q_k, and N_q is the number of quantiles used. Applications employing quantiles can be found in [43,34,37]. Note that the order of summation in (3) can be interchanged; the one used highlights that quantiles are introduced to approximate $E[u_j(Z_{ij})]$.

2.3. Models using scenarios

Uncertain outcomes can also be represented using a set of scenarios – incomplete but internally – consistent narratives of how the future might unfold. The use of 'scenario planning' e.g. [47,45] emphasizes gaining insight into the problem and generating novel actions rather than approximating MAUT. Multi-attribute scenario models e.g. [18, Chapter 14] apply a deterministic multi-attribute model within each scenario, followed (possibly) by an aggregation over scenarios [44]. The evaluation of a_i is given by

$$U_i^{(\mathrm{scen})} = \sum_{k=1}^{N_s} \left[w_{s_k} \sum_{j=1}^{J} w_j u_j(z_{ij}^{(s_k)}) \right] \tag{4}$$

where s_k refers to a specific scenario, $z_{ij}^{(s_k)}$ is the evaluation of alternative a_i on attribute c_j in scenario s_k, w_{s_k} is the weight associated with scenario s_k, and N_s is the number of scenarios used. The practical interpretation and assessment of the scenario weights w_{s_k} has not been fully resolved. Stewart [44] argues that the w_{s_k} should not be equated with scenario "probabilities" because the set of scenarios does not constitute a complete probability space, but should also not be equated with scenario "likelihoods" because the scenarios are incomplete descriptions and cannot in general be expected to represent the same dimensions in probability space. Rather, they should be interpreted as relative "swing" weights on performance in different scenarios. This is theoretically permissible, since the $\sum_{j=1}^{J} w_j u_j(z_{ij}^{(s_k)})$ values constitute an interval preference scale, but "it may be difficult to elicit appropriate values for the scenario weights" [44]. Max–min aggregation over scenarios has also been used [31]. Recent applications of multi-attribute scenario models can be found in [46,35,19].

3. Research aims and hypotheses

Our main aim is exploratory: To evaluate how closely the simplified models in Section 2 approximate MAUT. Nevertheless we do have some expectations which can be formalized as hypotheses. In the following we assume without loss of generality that utility increases in Z_{ij} and that each marginal utility function u_j has been scaled to have a minimum of 0 and a maximum of 1. All hypotheses and reported results in this paper use "utility loss" [4] to measure accuracy. Utility loss is defined as $UL = (U_{i^*} - U_{i^{sel}})/(U_{i^*} - U_{i_*})$ where U_{i^*} and U_{i_*} are the utilities (according to MAUT) of the best and worst alternatives in the MAUT rank order respectively, and $U_{i^{sel}}$ is the utility (according to MAUT) of the alternative selected by a simplified model. Although this measure is not entirely uncontentious (for example, the introduction of a weaker worst alternative will *ceteris paribus* reduce utility loss), it is widely used in simulation-based studies of decision making e.g [9,14]. Other metrics were also gathered, including the average rank of the alternative selected by a simplified model in the MAUT rank order, the average rank of the MAUT best alternative in a simplified model's rank order, and the rank correlation between the simplified model and MAUT rank orders. These did not provide any additional insights and are not reported.

Previous work [12] found that a model using expected values was on average more accurate than a variance model using Kirkwood's weights [26]. We expect worse accuracy from variance models in which only a general appetite or aversion for risk i.e. $w_{ij}^R = Cw_j$ with C a constant, is expressed.

Hypothesis 1. Variance models in which risk weights are fixed multiples of attribute importance weights will be less accurate than an expected value model.

Keefer and Bodily [24] have shown that expected values can be closely approximated using the 5%, 50% and 95% quantiles, so three or more quantiles could be used to approximate expected values and apply an expected value model. We expect better results if quantiles are transformed into utilities before aggregation i.e. from the quantile model.

Hypothesis 2. Quantile models will be more accurate than an expected value model.

The accuracy of all models will suffer from assessment errors, but theories of "error cancelation" e.g. [27] suggest that models that use multiple inputs will be more robust to random errors than those that provide more concise summaries of uncertainty.

Hypothesis 3. The robustness of model accuracy to assessment errors will be positively correlated with the number of inputs used to summarize probability distributions.

In [12], we noted that one of the key determinants of the accuracy of the expected value model was the steepness of the marginal utility functions in the region of the expected value approximation, because any differences between MAUT utilities (i.e. $E[u_j(Z_{ij})]$) and their approximations (i.e. $u_j(E[Z_{ij}])$) are more heavily penalized there [12]. Since all our simplified models make some use of central location measures, we expect the same result to apply.

Hypothesis 4. The accuracy of all simplified models will worsen as utility functions become steeper in the region of the attribute domain in which approximations are made.

The accuracy of the expected value model is not materially affected by the number of alternatives or attributes present [12]. We hypothesize the same relative insensitivity to problem size in the other simplified models.

Hypothesis 5. Problem size (the number of alternatives and attributes) will not have a material effect on the accuracy of any of the simplified models.

Because there has been little systematic research into the use of scenarios in decision analysis, no hypotheses are made regarding these models.

4. Design of the simulation experiment

Fig. 1 shows an outline of a single simulation run. Dashed boxes have been used to indicate those parts of the simulation applied iteratively to each alternative and attribute. Corresponding section numbers are shown to indicate where in the text further details can be found.

4.1. Generating realizations from Z_{ij}

The main difference between our and others' simulations of realizations from Z_{ij} e.g [5,36] is that we consider each Z_{ij} to be composed of $L=10$ normal distributions, denoted by $N(\mu_{ij\ell},\sigma^2_{ij\ell})$, $\ell \in \{1,\dots,L\}$ where μ and σ^2 are mean and variance respectively. The index ℓ is referred to as indexing the 'future' f_ℓ. Most previous studies have used a single distribution for each Z_{ij}, but multiple distributions are used here because they allow a parsimonious simulated application of scenario models. The choice of distribution, as well as the number of distributions to use, is somewhat arbitrary, but the robustness of our conclusions has been tested against these choices. We also simulated evaluations from the uniform distribution, and experimented with different numbers of distributions ($L \in \{6,10,20\}$), with no qualitative differences in the results.

We begin by generating means for the realizations in each future randomly between 0 and 1; although the uniform distributional form will be altered by later standardization, this does not materially affect results. A simulation parameter then determines whether:

- the $\mu_{ij\ell}$ are used directly, which allows alternatives to be dominated within one or more futures i.e. have a smaller mean than another alternative on every attribute e.g. [5,36], or
- alternatives are forced to be Pareto optimal in each future by standardizing within each alternative a_i so that $\sum_j \mu_{ij\ell}=1$ e.g. [12].

These cases are referred to as "with dominated alternatives" and "without dominated alternatives" respectively.

Within each future f_ℓ, the simulation then generates unstandardized realizations by:

1. Generating a standard deviation $\sigma_{ij\ell}$ randomly on Uni$[0.01,\sigma^{(d)}]$, where $\sigma^{(d)}$ is a parameter of the simulation.
2. Setting the number of realizations M_ℓ to be generated for future f_ℓ. A total of $K=400$ realizations for each Z_{ij} is used i.e. over all futures, although conclusions are insensitive to this value. These realizations are distributed over futures either "uniformly" ($\mathbf{M}=[40,40,\dots,40]$) or "non-uniformly" ($\mathbf{M}=[60,60,60,60,40,40,20,20,20,20]$). $\mathbf{M}$ is a parameter of the simulation.
3. Generating M_ℓ independent realizations from $N(\mu_{ij\ell},\sigma^2_{ij\ell})$. The $1\times M_\ell$ vector of realizations generated in future f_ℓ is denoted $\mathbf{z}^{(\ell)}_{ij}$.

Once realizations have been generated for each future, these are concatenated into a single $1\times K$ vector containing all the realizations for Z_{ij} i.e. $\mathbf{z}_{ij}=[\mathbf{z}^{(1)}_{ij},\mathbf{z}^{(2)}_{ij},\dots,\mathbf{z}^{(\ell)}_{ij},\dots,\mathbf{z}^{(L)}_{ij}]$. Realizations are scaled so that the largest realization on each attribute over all alternatives is one and the smallest is zero.

4.2. Generating inputs to the simplified models

Each simplified model uses a different summary of the realizations in $\mathbf{z}_{ij}$.

4.2.1. Expected value model

Uses the empirical mean of $\mathbf{z}_{ij}$, $E[\mathbf{z}_{ij}]$.

4.2.2. Explicit risk models

Two explicit risk models are simulated, both of which also use the expected values $E[\mathbf{z}_{ij}]$:

1. *Variance model*: uses empirical variances var$[\mathbf{z}_{ij}]$ to measure risk.
2. *Probability of poor performance model*: uses the proportion of realizations in $\mathbf{z}_{ij}$ that fall below a cut-off $\mathcal{L}$, a parameter of the simulation, to measure risk.

4.2.3. Quantile models

Uses empirical quantiles of the $\mathbf{z}_{ij}$. The number of quantiles N_q is a parameter of the simulation. If $N_q=3$, the 5%, 50%, and 95% quantiles are used; if $N_q=5$, the lower and upper quartiles are added.

4.2.4. Scenario models

Inputs to the simulated scenario models are generated in a two-step process. First, a random sample is drawn from the set of futures $\{f_1,f_2,\dots,f_L\}$ (the same random sample is used for all alternatives and attributes); then, realizations in each of the sampled futures are summarized, for each alternative and attribute. Scenario models differ with respect to how the random sampling and summarization are performed. The following three models are used:

- *'Mean scenario' model*
  1. Randomly draws a sample of size N_s from $\{f_1,f_2,\dots,f_L\}$ without replacement. Each future f_ℓ has an equal probability of selection.
  2. Uses the mean $\mu_{ij\ell}$ in each of the N_s selected futures.

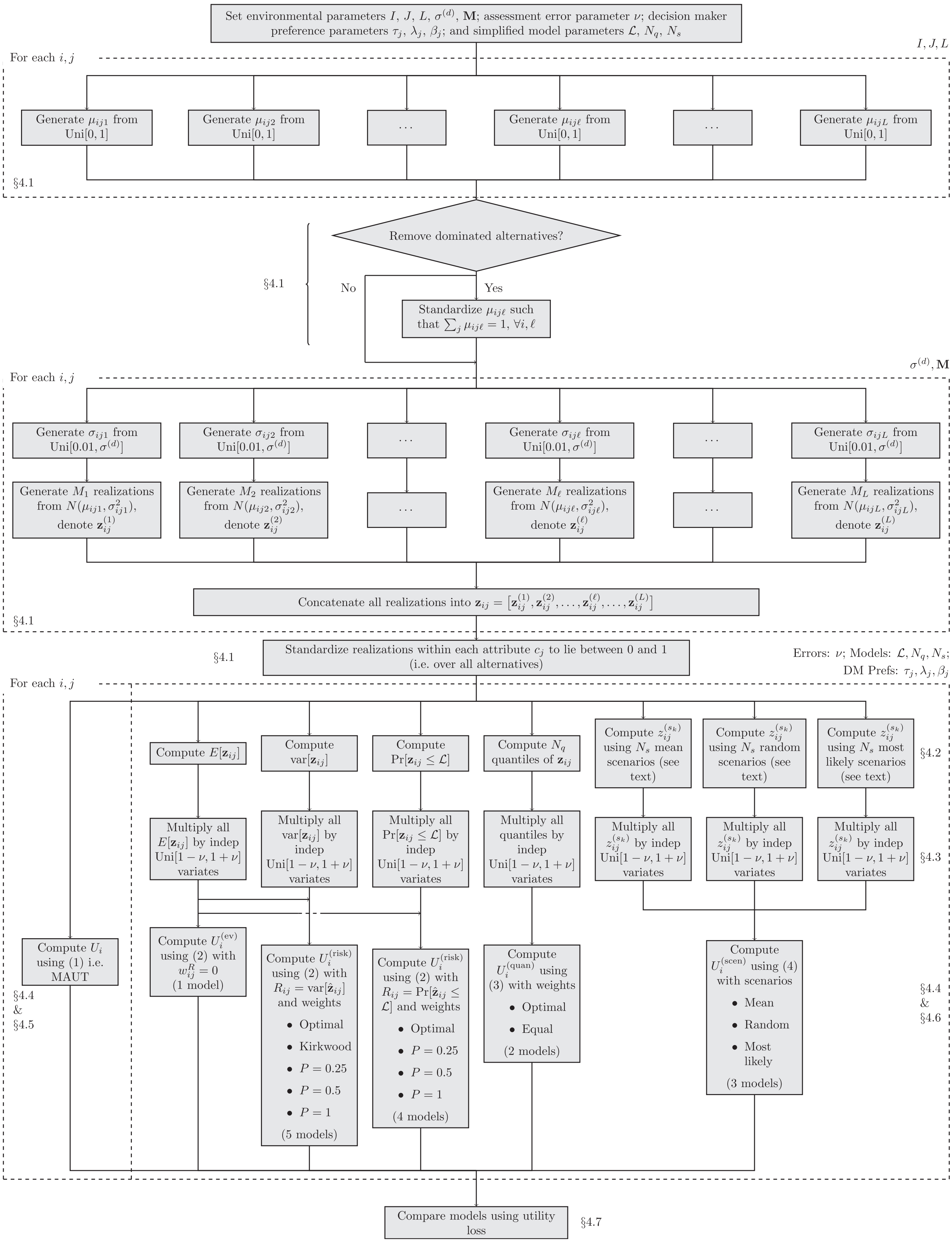

Fig. 1. Outline of a single simulation run.

- *'Random scenario' model*
  1. Randomly draws a sample of size N_s from $\{f_1,f_2,\ldots f_L\}$ without replacement. Each future f_ℓ has an equal probability of selection.
  2. Randomly selects one realization from $\mathbf{z}_{ij}^{(\ell)}$ in each of the N_s selected futures.
- *'Most likely scenario' model*
  1. Randomly draws a sample of size N_s from $\{f_1,f_2,\ldots f_L\}$ with replacement. Each future f_ℓ is selected with probability proportional to M_ℓ.
  2. Randomly selects one realization from $\mathbf{z}_{ij}^{(\ell)}$ in each of the N_s selected futures.

The proportion of futures selected i.e. N_s/L, is a simulation parameter termed the 'coverage' provided by a scenario model. Although attribute generation is somewhat biased in favor of a 'mean scenario' model using all L futures, the coverage parameter captures in an idealized way the scenario planning aim of constructing scenarios that "bound the future" e.g. [38]; a scenario model with 100% coverage is practically unrealistic but useful in giving an upper bound on accuracy. More realistic scenario models with less coverage (50% and 30%) are also simulated, and sensitivity to coverage and construction method are important results.

4.3. Generating errors in the assessment of uncertainty information

Assessment errors are simulated by multiplying all inputs to the simplified models (expected values, variances, probabilities of poor performance, quantiles, and realizations within selected futures) by independent and randomly generated realizations on Uni$[1-v,1+v]$, with v a simulation parameter. Final assessments are denoted using the 'hat' symbol e.g. expected values $\hat{E}[\mathbf{z}_{ij}]$.

4.4. Generating preference structures

The simulated marginal utility functions exhibit diminishing sensitivity and loss aversion [23] i.e. are convex below a reference level, concave above it, and steeper below the reference level. Each marginal utility function is described by four parameters: the reference level, τ_j, the value of the utility function at the reference level, λ_j, and the curvature of the utility function below and above the reference level, α_j and β_j respectively, using the standardized exponential form

$$u_j(x)=\begin{cases}\dfrac{\lambda_j(e^{\alpha_j x}-1)}{e^{\alpha_j\tau_j}-1} & \text{for } 0\le x\le\tau_j\\ \lambda_j+\dfrac{(1-\lambda_j)(1-e^{-\beta_j(x-\tau_j)})}{1-e^{-\beta_j(1-\tau_j)}} & \text{for } \tau_j<x\le 1\end{cases} \tag{5}$$

The same approach was used in [42] and [12]; a diverse set of preferences may be simulated by adjusting τ_j,λ_j and β_j (α_j is set to $\beta_j+U[0,2]$). Attribute importance weights are generated to be uniformly distributed with a minimum normalized value of $1/2J$, following [10].

4.5. Simulating the application of an additive MAUT model

The expected utility of a_i is given by (1), where as in other models the vector of realizations $\mathbf{z}_{ij}$ is used in place of the random variable Z_{ij}. Note that all probability information is taken into account in the generation of $\mathbf{z}_{ij}$.

4.6. Simulating the application of the simplified models

4.6.1. Expected value model

The evaluation of a_i is given by (2) with all $w_{ij}^R=0$.

4.6.2. Explicit risk models

The evaluation of a_i is given by (2). Risk weights are simulated using three approaches:

1. *Using a fixed risk multiplier*: Risk weights are set so that the average contribution made by the risk components over all alternatives is a proportion P (termed the 'fixed risk multiplier') of the average contribution made by the value components. 'Fixed risk weights' are given by $w_{ij}^R=Pw_ju_j(E_i[\hat{E}[\mathbf{z}_{ij}]])/E_i[\hat{R}_{ij}]\ \forall i$, where E_i denotes that expectations are taken over all alternatives and $P\in\{0.25,0.5,1\}$.
2. *Using an optimal risk multiplier*: as for the 'fixed' specification above, risk weights are a constant proportion P of the attribute weights w_j, but now with the constant P chosen to select as a winner an alternative that performs as well as possible in the rank order provided by MAUT. We implemented this using an integer program which minimizes the rank of the selected alternative in the MAUT rank order. This is equivalent to minimizing utility loss because, given generated problem data, the utility of the best and worst alternatives are fixed and so utility loss can only take on I possible values. Although the approach is not practically feasible because in practice one would not know the MAUT rank order, it provides an upper bound on the performance that can be achieved with a single multiplier P. The results are labeled 'optimal risk weights'.

When using variances we include a further approach:

3. *Using Kirkwood's weights:* Following [26] we use risk weights $w_{ij}^R=(-1/2)w_ju_j''(\hat{E}[\mathbf{z}_{ij}])$.

4.6.3. Quantile models

The evaluation of a_i is given by (3). Two approaches for generating quantile weights are used. The first uses equal weights for each quantile i.e. $w_{q_k}=1/N_q$. The second computes optimal quantile weights using a similar integer program to the one above, except that three or five weights can vary and quantile weights are constrained to sum to one.

4.6.4. Scenario models

The evaluation of a_i is given by (4). All scenarios are equally weighted i.e. $w_{s_k}=1/N_s$. We also tried weighting scenarios by their relative likelihoods i.e. M_ℓ/K, finding that this only improved accuracy substantively when one scenario was overwhelmingly more likely to occur e.g. 80% of all realizations generated from the same future. For brevity we restrict attention to the case of equal scenario weights.

4.7. Comparing results of the MAUT and simplified models

The accuracy of each simplified model is evaluated using utility loss as defined in Section 3.

4.8. Parameter values used in the simulations

Table 1 provides the parameter values used to simulate decision problems. The effect of problem size is investigated using $I=9$ or 29 alternatives, so that a random choice of alternative would appear on average 5th or 15th in the MAUT rank order respectively with $J=10$ or 20 attributes. The four combinations allow for an independent investigation of alternatives and attributes in decision environments located between 'fairly small' and 'fairly big' (in [5] and [36] smaller problems are also included but otherwise similar values are used).

Table 1
Parameter values used to simulate hypothetical decision problems.

Parameter	Description	Values
Problem context:		
I	Number of alternatives	9, 29
J	Number of attributes	10, 20
M	Distribution of realizations over futures	Uniform: [40, 40,40, . . . ,40] Non-uniform: [60,60,60,60,40,40,20,20,20,20]
Attribute evaluations:		
$\sigma^{(d)}$	Upper limit for $\sigma_{ij\ell}$	0.05 or 0.10
Errors in assessments of uncertainty information:		
v	Width of interval for random factor Uni$[1-v,1+v]$ generating errors	0, 0.1, 0.2, 0.3
Marginal utility functions:		
τ_j	Reference level for u_j	Uni[0.15,0.4] or Uni[0.6,0.85]
λ_j	Value of u_j at the reference level	Uni[0.15,0.4] or Uni[0.6,0.85]
β_j	Curvature of u_j above reference level	0, Uni[0,2] or Uni[0,5]

Table 2
Parameters used to simulate the application of simplified decision models.

Parameter	Description	Values
P	Fixed risk multiplier	0.25, 0.5, 1
$\mathcal{L}$	Poor performance cut-off	0.05, 0.1, 0.5
N_q	Number of quantiles used	3, 5
N_s/L	Coverage	30%, 50%, 100%

Parameter values for $\sigma^{(d)}$ were chosen by varying these until realizations in different futures could not be distinguished and appeared sufficiently uncertain. In the 'low' and 'high' variability conditions indicated in Table 1 the average difference between the 5th and 95th percentile of $\mathbf{z}_{ij}$ is 0.42 and 0.58 respectively. The assessment error parameter v is key and so is varied at four levels from 0% (error-free) to 30% (severely flawed assessment).

Parameter values for τ_j and λ_j can give preferences that are convex (e.g. $\tau_j=0.85, \lambda_j=0.15$) or concave (e.g. $\tau_j=0.15, \lambda_j=0.85$) in the majority of the domain, 'S-shaped' or linear (e.g. $\tau_j \approx \lambda_j$, with α_j and β_j both large or small respectively). A wide range of preferences can be simulated with relatively few parameters. The same parameter values have been used in [42] and [12].

Table 2 shows the values used for the parameters of the simplified models. The fixed risk multiplier P used in the explicit risk models is varied between 0 and 1 using intervals of 0.25. In assessing the probability of performing below a cut-off, we use two cut-offs ($\mathcal{L}=0.05, 0.10$) that represent very poor performance (between the 0.5% and 5% quantiles of performance, depending on attribute variability) and one cut-off ($\mathcal{L}=0.50$) representing mediocre performance (between 40% and 70% quantiles). Three- and five-quantile summaries are selected as standard summary statistics that are regularly used to approximate probability distributions [16] and moments [24]. Our main goal in selecting the number of scenarios is to investigate the effect of omitting futures—we simulate the selection of 10, 5 and 3 of the original $L=10$ futures, giving 'coverage' of 100%, 50% and 30% respectively. Note that this does not test the general effectiveness of using 3, 5, or 10 futures, even in the limited sense of accurate approximation of MAUT.

We use a resolution V fractional factorial design e.g. [30, Chapter 8], so all main effects are unconfounded with two- and three-factor interactions, and all two-factor interactions are unconfounded with other two-factor interactions. We perform 100 simulation runs for each combination of parameters, giving standard errors of at most 0.003 for mean utility losses in groups formed by combinations of two factors. This is small enough for any differences we discuss to be statistically significant at the 1% level.

5. Results

Fig. 2 shows the average utility loss of each simplified model under error-free assessments, indicated by unshaded circles, and erroneous assessments, indicated by shaded circles (10% error), squares (20% error) or triangles (30% error). Within each model type, utility losses are ordered from best to worst according to the error-free values. For comparison, a random selection policy gives a utility loss of approximately 0.50 with or without dominated alternatives. The mean utility of the MAUT best (worst) alternative is 0.43 (0.33) without dominated alternatives, and 0.52 (0.32) with them. The presence of dominated alternatives introduces weaker 'worst' alternatives while leaving the other components of utility loss relatively unaffected (because dominated alternatives are very rarely selected). This makes utility losses generally smaller when dominated alternatives are included, but our conclusions are unaffected and the discussion below focuses on results obtained without them.

In Fig. 2, Hypotheses 1 and 2 are strongly supported and Hypothesis 3 is conditionally supported. The average utility loss using expected values is at least as good as any explicit risk model using fixed risk multipliers (Hypothesis 1), provided assessment errors are not large. When assessment errors are large a model using probabilities of performing below a central quantile can be more accurate than one using expected values. This suggests that an explicit risk attribute may impart some robustness to assessment errors. In general though, it appears that a model using a fixed risk multiplier approximates MAUT relatively poorly. Sensitivity to assessment errors decreases as the risk multiplier is increased because 'risk' components (variances, probabilities of poor performance) are less sensitive to errors than (expected) 'value' components. If only variances are used, the utility loss varies from 0.40 to 0.45 depending on assessment errors. In contrast with the other simplified models, the utility losses of variance models with large risk multipliers do not improve when dominated alternatives are included because there are no substantial differences in the variances of dominated and non-dominated alternatives. A variance model using Kirkwood's weights [26] performs better than one using fixed risk multipliers, but it is only in the unrealistic case where the risk multiplier is optimally chosen that an explicit risk model gives consistently better results than expected values alone.

The average utility loss for the equal-weight quantile models over all assessment errors without dominated alternatives is 0.052 and 0.080 using five or three quantiles respectively, significantly lower than the 0.106 obtained using expected values (Hypothesis 2). The better accuracy is partly due to robustness to assessment errors, but even in the absence of errors average utility loss is 0.004 and 0.019 if five or three quantiles are used respectively, and 0.020 using expected values. If quantile weights are chosen optimally, the error-free mean utility loss is less than 0.001 using either five or three quantiles. The average weight allocated to the 5%, 50%, and 95% quantiles is 0.16, 0.65 and 0.19 respectively[1]; this is very close to the weights proposed in [24]: 0.185, 0.630 and 0.185. If five quantiles are used, the average quantile weights are 0.07, 0.12, 0.60, 0.12, and 0.09 for the 5%

[1] These are calculated based on error-free assessments only. When assessment errors are made the simulation chooses optimal quantile weights to compensate for the errors—something real-world decision makers cannot do.

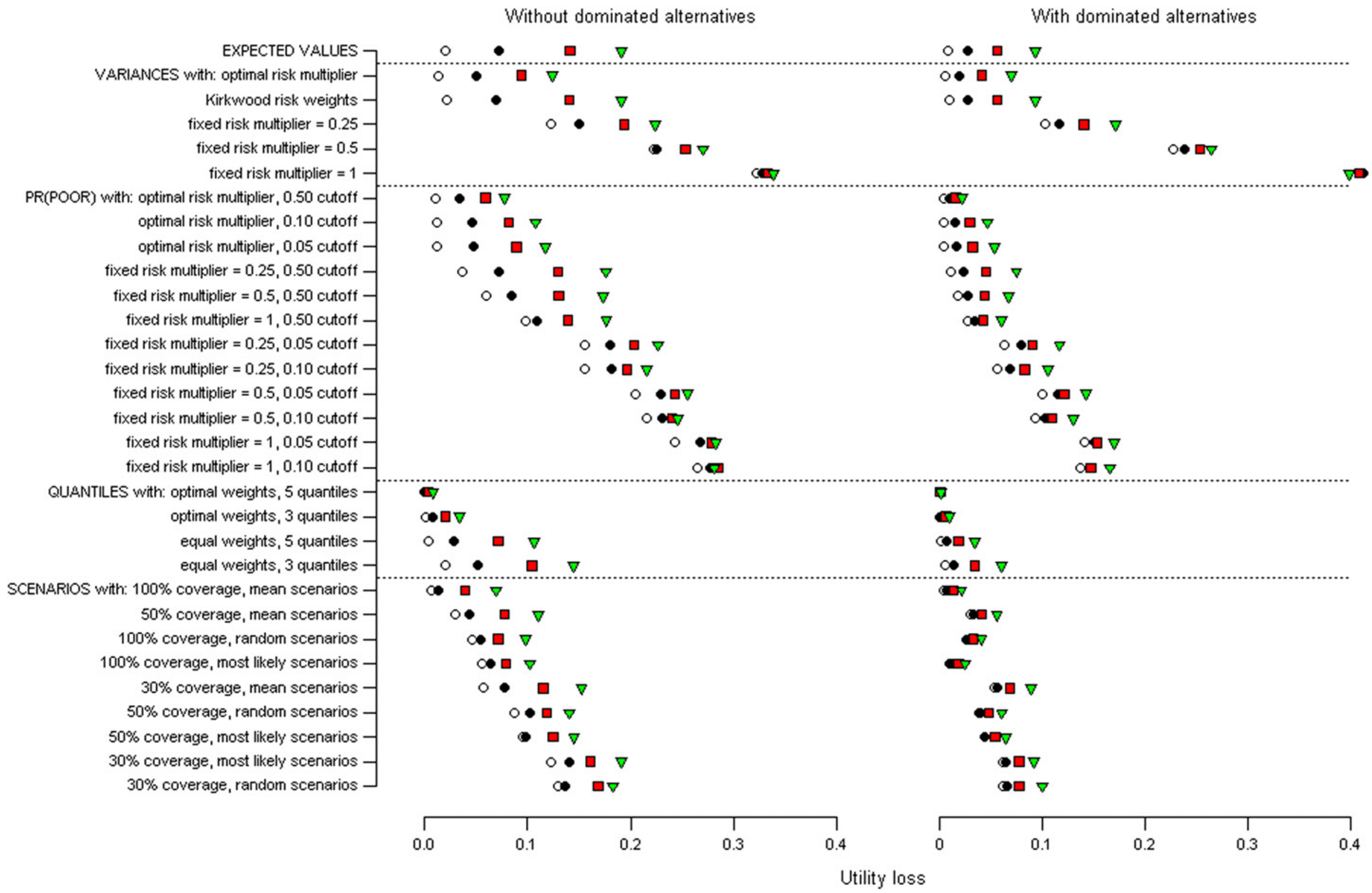

Fig. 2. Mean utility loss experienced by each simplified model. Unshaded and (three) shaded points show the average utility loss in error-free simulations and with assessment errors of 10%, 20%, and 30%.

through 95% quantiles respectively. Subsequent additional simulations confirmed the suggestion that good accuracy will be obtained if quantile weights are fixed at the Keefer–Bodily values.

Fig. 2 also shows that if one excludes explicit risk models which perform terribly there is a trend towards increased robustness in the quantile and scenario models, although robustness varies widely within model type. Table 3 shows the deteriorations in utility loss that occur due to assessment errors. The selected models are the best-performing versions of each simplified model (in terms of mean utility loss with no assessment error) that do not make use of optimal weights. These have been ranked from the smallest average increase in utility loss over all assessment errors to the largest. Although the three-scenario model is more robust than the five-quantile model, there is otherwise a clear association between the number of inputs used by a model and its robustness to error.

Scenario model accuracy is strongly influenced by both scenario construction and coverage. Substantially better accuracy is obtained if scenarios are constructed using mean values. The relatively poor results obtained when selecting realizations at random from each future highlights the importance of accurately assessing means. If no futures are omitted (100% coverage), then results can be excellent; but if coverage drops to 50% then accuracy when no assessment errors are made is worse than if expected values are used. Our view is that results using 50% and 30% coverage probably provide more appropriate indicators of the practical potential of scenario models. Coverage becomes more important relative to scenario construction when dominated alternatives are present; all models with 100% coverage outperform all those with 50% coverage, which in turn outperform all those with 30% coverage. This is because alternatives can perform terribly in omitted futures. An advantage of scenario models is increased robustness to assessment errors. When assessment errors are made, the mean scenario model with 50% coverage offers similar performance to a three-quantile model using equal quantile weights.

Table 4 shows how average utility loss differs over other simulation parameters. The utility function parameters τ_j and λ_j are shown jointly to evaluate Hypothesis 4. Accuracy is primarily affected by the shape of the utility functions, and is best when these are predominantly concave and worst when they are predominantly convex. Many of the approximations used in our simulations occur in the middle-to-upper part of the attribute domain, where convex utility functions are steeper than concave ones (Hypothesis 4). This occurs because any differences are more heavily penalized by a steeper utility function.[2] Further results show that both the approximations made by simplified models and assessment errors cause the deteriorations in accuracy. With no assessment errors, the greatest utility loss for all simplified models occurs with highly convex utility functions (high τ_j, low λ_j). The increase in average utility loss caused by the same size assessment error is also greatest when utility functions are highly

[2] Another explanation is that a sharp threshold at a low reference level implies a utility function that is relatively flat over a large portion of the attribute domain, which reduces the utility loss of selecting a 'good-but-not-the-best' alternative. However, the rank of the alternative selected by the simplified models in the MAUT rank order also deteriorates (from an average of 2.0 to 2.4 as utility functions shift from mostly concave to mostly convex), and thus this cannot completely explain the effect.

Table 3
Increases in mean utility loss caused by assessment errors. Results are only reported for simulations without dominated alternatives. Simplified models are represented by their best-performing versions, after excluding models using optimal risk or quantile weights.

Model	Conditions	Assessment error		
		10%	20%	30%
Scen	10 mean scenarios (100% coverage)	0.008	0.034	0.063
Scen	5 mean scenarios (50% coverage)	0.013	0.048	0.080
Scen	3 mean scenarios (30% coverage)	0.020	0.058	0.095
Quan	5 equally weighted quantiles	0.024	0.067	0.102
Quan	3 equally weighted quantiles	0.032	0.085	0.125
Pr(Poor)	0.50 cutoff, risk multiplier=0.25	0.036	0.093	0.140
Var	Kirkwood's weights	0.049	0.120	0.170
EV	–	0.052	0.121	0.172

Table 4
Average utility losses at different levels of the simulated decision problem parameters. Averages are calculated over all levels of the remaining parameters. Results are only reported for simulations without dominated alternatives. The same models are used as in Table 3: the variance model uses Kirkwood weights; the probability of poor performance model uses a 0.50 cutoff and a risk multiplier of 0.25, both quantile models use equal quantile weights, and all scenario models use mean scenarios.

Effect	Values	EV	Var	Pr	Quantiles		Scenarios: coverage=		
				(poor)	$N_q=3$	$N_q=5$	30%	50%	100%
I	9	0.10	0.10	0.10	0.07	0.05	0.10	0.06	0.03
	29	0.10	0.10	0.10	0.08	0.05	0.10	0.06	0.03
J	10	0.10	0.10	0.10	0.07	0.05	0.10	0.06	0.03
	20	0.10	0.10	0.10	0.08	0.05	0.09	0.06	0.03
M	Uniform	0.09	0.09	0.09	0.08	0.05	0.09	0.06	0.03
	Non-unif.	0.10	0.10	0.10	0.07	0.05	0.10	0.06	0.03
$\sigma^{(d)}$	Uni[0.01,0.05]	0.08	0.09	0.09	0.05	0.04	0.06	0.04	0.02
	Uni[0.01,0.10]	0.10	0.10	0.10	0.08	0.05	0.11	0.07	0.03
(τ,λ)	(Low,low)	0.08	0.08	0.07	0.05	0.03	0.09	0.05	0.02
	(Low,high)	0.07	0.08	0.09	0.04	0.02	0.07	0.04	0.01
	(High,low)	0.13	0.12	0.13	0.12	0.08	0.13	0.09	0.05
	(High,high)	0.10	0.10	0.09	0.08	0.05	0.10	0.06	0.03
β	0	0.09	0.09	0.09	0.07	0.04	0.09	0.06	0.03
	Uni[0,2]	0.10	0.09	0.10	0.07	0.05	0.09	0.06	0.03
	Uni[0,5]	0.11	0.11	0.10	0.08	0.05	0.11	0.07	0.03

convex. Table 4 also shows that the accuracy of all the simplified models is very nearly constant over the simulated problem sizes (Hypothesis 5), although this depends on utility loss being used to measure accuracy. Measures of accuracy based on ranks e.g. the average rank of the selected alternative in the MAUT rank order, differ significantly with the number of alternatives used, but an increasing rank does not necessarily imply a deterioration in decision quality because the size of the rank order i.e. number of alternatives, has also increased. The only other variable exerting a meaningful effect on accuracy is the variability of the attribute evaluations—as evaluations become more variable the accuracy of all the simplified models worsens.

6. Implications for decision analysis

Our simulation evaluates the ability of a number of simplified decision models to approximate MAUT. We do not intend to use these results to conclude a detailed apparatus prescribing rules for using particular models in particular situations. Rather the results suggest a general course of action for practitioners who for reasons of simplicity prefer not to use MAUT. We stress that all our findings are limited by the range of simulated cases, as all simulation experiments are. The complexity of the simulation apparatus is largely to ensure that a suitable range of problems have been covered, although doubtless there are counterexamples to our findings which could be constructed. Although the apparatus of the simulation experiment may be complex, our conclusions are fairly simple.

Our basic message is that – for a wide range of simulated problems – all of the simplified models are able to produce results that are close, on average, to MAUT. The best-performing of each of the simplified models have average utility losses less than 0.04 (where 0 is optimal, and randomly selecting an alternative returns an average utility loss of approximately 0.5). Given the time and effort required to implement MAUT, simplified models appear justifiable for many decision problems. Our results suggest that avoiding assessment errors in the application of a simplified model is more important than the type of model used. Analysts thus have considerable scope to choose the model that they or the decision makers they are facilitating are most comfortable with and least likely to apply poorly. One check that analysts should perform before using a simplified model is to test whether preferences are highly non-linear. Our results indicate that the accuracy of all simplified models deteriorates as preference thresholds become sharper in the region where approximations are made, because assessment errors are more heavily punished. Analysts wishing to use simplified models but finding strong preference thresholds should place extra effort in ensuring accurate assessments.

Although the performance of all the simplified models is good in the absence of assessment errors, a quantile model performs consistently best. The final choice of model will need to take into account other practical factors. Our results only suggest that if analysts wish to conform to MAUT but lack the resources to implement it, they should consider a quantile model first. These results echo those of the behavioral study in [13]. Accuracy is best served by aggregating evaluations over quantiles using the weights proposed in [24]: the weight on the median is 0.63 and the remainder is shared between the extreme quantiles. The quantile weights aim only to approximate a MAUT model; they do not provide preference information and so no assessment is required. Quantiles, and particularly non-extreme quantiles, can be assessed relatively accurately by decision makers e.g. [33,17] and are commonly used in the practical assessment of probability distributions using, for example, the bisection or interval methods e.g. [40]. Trade-off judgements are also likely to be relatively easy because quantiles use the same scale as attributes. This suggests that the three-quantile model may be a useful preliminary screening tool before a more detailed evaluation is carried out. This could select the most promising alternatives from a larger set, or assess whether a choice is possible without needing a full MAUT analysis. If probability distributions are to be assessed with the bisection or similar method, any quantiles used by a preliminary model could be used here too, so that a preliminary model might not add much time or effort to the analysis. If a quantile model is to be used on its own, our accuracy results as well as the relative inaccuracies reported when assessing extreme quantiles [1] suggest that upper and lower quartiles may be useful additions to the median and extremes. We suggest that it may be useful to structure the quantiles into scenario-like arrangements by collecting together all the attribute evaluations at a particular quantile. Although such an arrangement is not necessary, we conjecture that some insight may be gained from allowing decision makers to compare alternatives at their 'worst-case', 'best-case' and some 'intermediate' levels of performance.

Several authors have called for decision analysts to pay greater attention to scenario planning techniques [18,31], which are well-established in strategic decision making. Our results indicate that when a substantial number of futures are omitted a scenario model performs relatively poorly. If one only considers the best-performing versions of each simplified models (after excluding models using optimal risk or quantile weights, as in Tables 3 and 4), a scenario model with 30% coverage gives a worse average approximation to MAUT than any other simplified model. The time and cognitive effort involved in constructing scenarios means some omissions are probably inevitable. Scenario-based MCDA may possess other advantages – generating insights into uncertainty and novel actions are commonly cited benefits [39] – but analysts should be aware that a scenario model, even if correctly applied, will lead to a outcome that is more different to MAUT than other simplified models. Sometimes an analyst may for pragmatic reasons want to use scenarios but also wish to obtain results that are close to MAUT. Our results suggest that in such cases scenarios should attempt to capture mean performances in as many futures as is practically feasible. This is quite different to the philosophy of scenario planning, which often advocates taking a small number of extreme positions when constructing scenarios e.g. [39].

For the range of simulated problems, explicit risk models performed relatively poorly. When no assessment errors are made both explicit risk models can lead to poorer approximations of MAUT than a model using expected values only. The variance model does particularly poorly when risk weights are a fixed multiple of attribute importance weights. Behavioral research suggests that the elicitation and understanding of variance information is difficult e.g. [16], and the assessment of weights involving variances also seems a difficult prospect. In conjunction with our accuracy results, this suggests that the variance model must use Kirkwood's [26] risk weights if applied in practice.

Our results suggest some directions for future research. In behavioral decision research, potential influences on the 'accuracy' and 'effort' involved with decisions (operationalized in different ways; see [32] for an overview) are routinely tested in experimental settings. Future research might assess the effect of uncertainty format on accuracy and effort in a 'behavioral' setting, providing a bridge between simulation and real-world complexity. It also seems crucial to build an understanding of the practical issues surrounding the application of simplified models. Progress has been recently made in applications of scenario models [31]. Our results suggest that applications of other models would be equally welcome.

References

[1] Alpert M, Raiffa H. A progress report on the training of probability assessors. in: Judgment under uncertainty: heuristics and biases. Cambridge University Press; 1982. p. 306–34.
[2] Arcelus F, Kumar S, Srinivasan G. Risk tolerance and a retailer's pricing and ordering policies within a newsvendor framework. Omega 2012;40:188–98.
[3] Ballestero E. Stochastic goal programming: a mean-variance approach. European Journal of Operational Research 2001;131:476–81.
[4] Barron F. Influence of missing attributes on selecting a best multiattributed alternative. Decision Sciences 1987;18:194–205.
[5] Barron F, Barrett B. Decision quality using ranked attribute weights. Management Science 1996;42(11):1515–23.
[6] Belton V, Stewart T. Multiple criteria decision analysis: an integrated approach. Boston: Kluwer Academic Publishers; 2002.
[7] Bordley R, Kirkwood C. Multiattribute preference analysis with performance targets. Operations Research 2004;52(6):823.
[8] Bose Anne M, et al. Multi-attribute utility methods in group decision making: past applications and potential for inclusion in GDSS. Omega 1997;25(6):691–706.
[9] Butler J, Dyer J, Jia J. Using attributes to predict objectives in preference models. Decision Analysis 2006;3(2):100.
[10] Butler J, Jia J, Dyer J. Simulation techniques for the sensitivity of multi-criteria decision models. European Journal of Operational Research 1997;103: 531–46.
[11] Donckels R. Regional multiobjective planning under uncertainty: a stochastic goal programming formulation. Journal of Regional Science 1977;17(2):207–16.
[12] Durbach I, Stewart T. Using expected values to simplify decision making under uncertainty. Omega 2009;37(2):312–30.
[13] Durbach I, Stewart T. An experimental study of the effect of uncertainty representation on decision making. European Journal of Operational Research 2011;214:380–92.
[14] Fasolo B, McClelland G, Todd P. Escaping the tyranny of choice: when fewer attributes make choice easier. Marketing Theory 2007;7(1):13.
[15] Fenton N, Wang W. Risk and confidence analysis for fuzzy multicriteria decision making. Knowledge-Based Systems 2006;19(6):430–7.
[16] Garthwaite P, Kadane J, O'Hagan A. Statistical methods for eliciting probability distributions. Journal of the American Statistical Association 2005; 100(470):680–701.
[17] Garthwaite P, O'Hagan A. Quantifying expert opinion in the UK water industry: an experimental study. The Statistician 2000:455–77.
[18] Goodwin P, Wright G. Decision analysis for management judgement. 2nd ed. Chichester: John Wiley & Sons; 1998.
[19] Grigoroudis E, Orfanoudaki E, Zopounidis C. Strategic performance measurement in a healthcare organisation: a multiple criteria approach based on balanced scorecard. Omega 2012;40:104–19.
[20] Hajkowicz S, Higgins A. A comparison of multiple criteria analysis techniques for water resource management. European Journal of Operational Research 2008;184(1):255–65.
[21] Hamalainen R. Reversing the perspective on the applications of decision analysis (comment on Keefer et al. 2004). Decision Analysis 2004;1(1):26–31.
[22] Janssen R. On the use of multi-criteria analysis in environmental impact assessment in the Netherlands. Journal of Multi-criteria Decision Analysis 2001;10(2):101–9.
[23] Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979;47(2):263–91.
[24] Keefer D, Bodily S. Three-point approximations for continuous random variables. Management Science 1983;29(5):595–609.
[25] Keeney R, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. Cambridge: Cambridge University Press; 1993.
[26] Kirkwood C. Estimating the impact of uncertainty on deterministic multi-attribute evaluation. Management Science 1992;38(6):819–26.
[27] Kleinmuntz D. Decomposition and the control of error in decision analytic models. in: Einhorn H, Hogarth R, editors. Insights in decision making: a tribute to Hillel J. Einhorn. University of Chicago Press; 1990. p. 107–26.
[28] Lee J, Lee-Kwang H. Comparison of fuzzy values on a continuous domain. Fuzzy Sets and Systems 2001;118(3):419–28.
[29] Markowitz H. Portfolio selection. Journal of Finance 1952;7(1):77–91.
[30] Montgomery D. Design and analysis of experiments. Wiley; 1991.
[31] Montibeller G, Gummer H, Tumidei D. Combining scenario planning and multi-criteria decision analysis in practice. Journal of Multi-criteria Decision Analysis 2006;14:5–20.
[32] Payne J, Johnson E, Bettman J. The adaptive decision maker. Cambridge University Press; 1993.
[33] Peterson C, Snapper K, Murphy A. Credible interval temperature forecasts. Bulletin of the American Meteorological Society 1972;53(10):966–70.
[34] Rakes T, Deane J, Rees L. IT security planning under uncertainty for high-impact events. Omega 2012;40:79–88.
[35] Ram C, Montibeller G, Morton A. Extending the use of scenario planning and MCDA for the evaluation of strategic options. Journal of the Operational Research Society 2010;62(5):817–29.
[36] Salo A, Hamalainen R. Preference ratios in multiattribute evaluation (PRIME)—elicitation and decision procedures under incomplete information. IEEE Transactions on Systems, Man, and Cybernetics, Part A 2001;31(6):533–45.
[37] Sawik T. Selection of supply portfolio under disruption risks. Omega 2011;39(2):194–208.
[38] Schoemaker P. When and how to use scenario planning: a heuristic approach with illustration. Journal of Forecasting 1991;10(6):549–64.
[39] Schoemaker P. Scenario planning: a tool for strategic thinking. Sloan Management Review 1995;36:25–125.
[40] Spetzler C, Staël Von Holstein C-A. Probability encoding in decision analysis. Management Science 1975:340–58.
[41] Stewart T. Simplified approaches for multicriteria decision making under uncertainty. Journal of Multi-criteria Decision Analysis 1995;4:246–58.
[42] Stewart T. Robustness of additive value function methods in MCDM. Journal of Multi-criteria Decision Analysis 1996;5:301–9.
[43] Stewart T. Measurements of risk in fisheries management. ORiON 1998;14(1/2): 1–15.
[44] Stewart T. Dealing with uncertainties in MCDA. in: Figueira J, Greco S, Ehrgott M, editors. Multiple criteria decision analysis: state of the art surveys. Springer; 2005 [Chapter 11].
[45] Van der Heijden K. Scenarios: the art of strategic conversation. New York: John Wiley & Sons; 1996.
[46] van der Pas J, Walker W, Marchau V, Van Wee G, Agusdinata D. Exploratory MCDA for handling deep uncertainties: the case of intelligent speed adaptation implementation. Journal of Multi-Criteria Decision Analysis 2010;17(1/2):1–23.
[47] Wack P. Scenarios: uncharted water ahead. Harvard Business Review 1985(September/October):73–89.
[48] Wallenius J, Dyer J, Fishburn P, Steuer R, Zionts S, Deb K. Multiple criteria decision making, multiattribute utility theory: recent accomplishments and what lies ahead. Management Science 2008;54(7):1336–49.