DOI:
10.1002/ffo2.18

Publication date:
2019

Document Version
Peer reviewed version

Supporting multiattribute decisions in scenario planning using a simple method based on ranks

Short running title: Multiattribute decisions and scenarios

Paul Goodwin
School of Management
University of Bath
Bath
BA2 7AY
United Kingdom
Tel: +44 (0) 1373-461617
Email: p.goodwin@bath.ac.uk

Acknowledgement

The author would like to thank three anonymous referees for their helpful and insightful comments on an earlier version of this paper

Supporting multiattribute decisions in scenario planning using a simple method based on ranks

Abstract

A neglected area of scenario planning is the provision of support for strategic decisions that involve multiple attributes. When the number of scenarios and attributes is large, conventional multiattribute decision analysis methods require the elicitation of a large number of values and weights, which can be demanding and time consuming for decision makers. This paper examines the effectiveness of using a simple approximation to the simple multiattribute rating method (SMART) that is based purely on the ranking of options and attributes. The method was tested on 250,000 simulated decision problems and was found to perform well when assessed on the basis of its hit rate (the percentage of times it identified the same best option as SMART) and the utility loss resulting from the approximation. In large problems, where simplifications are likely to be most useful, it outperformed an alternative approximation method, SMARTER, which is more complex to apply, and it was almost as effective as SMARTER on smaller problems.

Keywords: Scenario planning, decision analysis, multiattribute decisions, SMART

Introduction

Proponents of scenario planning argue that it can provide effective guidance when strategic decisions have to be made under conditions of uncertainty (van der Heijden, K., 1996; Wright and Cairns, 2011). When the 'intuitive logics' approach to scenario planning is used it avoids the need to elicit subjective probabilities for future events and the psychological biases associated with such estimates (van der Heijden, K., 1996. pp28-29; Wright and Goodwin 2009). In a fast-changing world it does not rely on detailed 'single shot' forecasts, which are often based on extrapolations of past patterns (Eisenhardt and Sull, 2001). Moreover, the narrative nature of scenarios may be more naturally appealing to decision makers than quantitative assessments of future conditions. However, a neglected area of scenario planning is the provision of support for decision makers once the scenarios have been written. The range of scenarios have been compared to a wind tunnel in that they allow the performance of strategic options to be tested to assess their effectiveness under the different conditions that may prevail in the future (van der Heijden, 1996, p.57) But when this performance is dependent on multiple attributes, such as financial metrics, environmental impact, organisational reputation and stakeholder well-being, an accurate assessment of performance under a given scenario can be difficult without structured support.

This paper evaluates a simple structured support method for assessing the performance of alternative strategies when 'intuitive logics' scenario planning (Wright and Cairns 2011) is being applied. The support method is designed to take into account the time constraints faced by strategic planners and the cognitive difficulties they may encounter when asked to assess multiple strategies in multiple scenarios across multiple attributes. The paper is structured as follows. After a review of the relevant literature, the method is demonstrated and then tested on 250,000 simulated decisions that were designed to reflect the diversity of decisions that may need to be made in strategic planning. Finally, the paper reflects on the strengths and limitations of the method and indicates how any serious inaccuracies arising from its simplifications may be resolved.

Literature Review

Choosing between strategic options based on a scenario planning process can be challenging where there are multiple options than need to be evaluated across multiple criteria under the conditions described in multiple scenarios. There is evidence that under these circumstances poor choices can be made because of cognitive limitations which mean that unaided decision makers have difficulty in handling and comparing the large volumes of information involved (e.g. Tversky, 1972). This may motivate then to oversimplify the decision problem. For example, choices may be based on performances on a single attribute (such as cost) while the attractiveness of options on other attributes (e.g. company image or environmental impact) are ignored. Alternatively, even when performances on multiple dimensions are taken into account, non-compensatory strategies may be employed to make choices because of the cognitive difficulties of making trade-offs between attributes (Payne, 1976).

This has led a number of researchers to advocate the use of structured decision support methods in scenario planning (e.g., Schroeder and Lambert, 2011; see Stewart et al. (2013) for a review). These methods essentially decompose the decision problem into a series of simpler and less cognitively demanding tasks before recomposing the judgments made in these tasks, according to a set of explicit axioms, to indicate which strategic option should be chosen (e.g., see Goodwin and Wright, 2014, pp52-55). The benefits of decomposing judgments have been shown in a number of studies (e.g. MacGregor et al., 1988). These structured methods also have the advantage that they provide a documented and defensible rationale for any decision that is made.

Multiattribute utility theory (MAUT) (Keeney and Raiffa, 1976) is a classic decomposition approach to decision making under uncertainty where there are multiple attributes but its application can pose significant challenges to decision makers given the number and the demanding nature of the judgments it requires. Although suggestions for simplifying the process of applying MAUT have been put forward and tested (Durbach and Stewart, 2012) its role in intuitive logics scenario planning is problematical because it is based on the assumption that probabilities can be estimated for the states of nature. Scenarios cannot in most cases be regarded as a mutually exhaustive set of states of nature (Durbach, 2019) and, even if they could be, it is doubtful whether reliable probabilities could be estimated for them given their uniqueness and the high degree of uncertainty associated with scenario planning applications (Tversky and Kahneman, 1974; van der Heijden, K., 1996. pp28-29; but see Vilkkumaa et al., 2018 for a counter argument when scenarios can be assumed to be mutually exhaustive). Indeed, the probability of any given detailed scenario transpiring is infinitesimal –the scenarios are intended to bound the range of possible futures instead of representing outcomes that have a plausible chance of occurring.

This has led researchers to suggest much simpler methods of decision support. Some approaches have treated each scenario as a deterministic outcome, thereby avoiding the use of probabilities and allowing the use of deterministic multicriteria decision analysis (MCDA) methods. For example, Goodwin and Wright (2001) proposed the use of the Simple Multiattibute Rating Technique (SMART) (Edwards, 1977) to compare the performance of strategic options within each scenario in turn. Scores for the performance of options and weights reflecting the relative importance of the attributes are assessed across the scenarios. The result is a matrix displaying an aggregate score for each option-scenario combination. However, no formal process is used to make the final choice between options though Goodwin and Wright recommend that the decision should be based on considerations such as whether one option dominates the others or the robustness of the options' performances across the scenarios.

More recently, Durbach (2019) has recommended the use of the Analytic Hierarchy Process (AHP) in the evaluation of strategic options in scenario planning. He does so partly on the basis of the acceptability of the method to decision makers and its associated widespread use, though he acknowledges that the method has limitations (see also: Belton and Gear, 1983, Salo and Hämäläinen, 1997). Durbach's use of a structured method goes beyond that of Goodwin and Wright in that it provides an indication of the preferred final choice between options. It does this by treating combinations of attributes and scenarios as meta-attributes, allowing decision makers to express preferences, such as "I strongly prefer to reduce the product development time by 6 months in scenario A rather than scenario B" (see also Stewart et al, 2013).

Other proposals to integrate multicriteria decision making methods and scenario planning include the application of goal programming (Durbach and Stewart, 2003), the use of scenarios in the outranking method (Durbach, 2014) and the use of regret to choose between strategic options (Ram et al., 2011). Stewart et al. (2013) provide an overview and assessment of many of the approaches that have been suggested.

However, despite the relative simplicity of these methods compared to MAUT, in many scenario planning applications a large number of judgments will still be required from decision makers. For example, Montibeller et al. (2006) found this when they applied the method proposed by Goodwin and Wright to both an English insurance company and an Italian property development company. Comparing N strategies over S scenarios when there are A attributes requires N x S x A interval-scale values to be assessed - each reflecting the relative performance of a given strategy against a given objective if a given scenario prevails –together with A ratio-scale swing weights, that are designed to reflect the relative desirability of changes from the worst to the best performance on each attribute. For example, where there are 4 competing strategies being assessed on 4 attributes in each of 4 scenarios $4^3 + 4 = 68$ potentially demanding judgments will need to be elicited. Even then, the Goodwin and Wright method assumes that the same weights should apply in different scenarios. As Montibeller et al. found, decision makers may wish to apply different weights in different scenarios because priorities may change according to the environment an organisation is operating within. This could potentially lead to the requirement to elicit $4^3 + 16 = 80$ judgments in the example referred to above. In some cases, these judgments may be difficult and unfamiliar for many decision makers. In particular, in the assessment of swing weights in SMART, Borcherding et al. (1991) and others have demonstrated that a number of biases can emerge. These demands mean that the techniques may be unacceptable to decision makers, especially where large problems are involved. Alternatively, when they are used by inexperienced decision makers, or by decision makers who lack guidance on their use, the elicited judgments may be unreliable (Katsikopoulos and Fasalo, 2006).

This raises the question of whether it is possible to employ even simpler methods in scenario planning while still providing an accurate indication of the choice that should be preferred by decision makers. Edwards and Barron (1994) have argued for the strategy of 'heroic approximation'. This asserts that the inaccuracies associated with a simple approximate model of a decision problem are likely to be outweighed by the more reliable judgments that result because the technique poses simpler questions to the decision maker. In addition, the use of even a very simple technique may result in a better decision than unaided judgment. Moreover, simple techniques are more likely to be used because of the reduced effort and greater transparency associated with them. Ideally, a simple method will therefore not be overly time consuming, will be intuitively appealing and hence acceptable to decision makers, will provide a transparent rationale for its recommendations and will be able to support strategic decisions made by groups of planners. This suggests that it should meet the following four criteria: ì) it should allow for the simplification of the elicitation of values or utilities for given attributes, (ii) it should allow for the simplification of the estimation of weights for attributes (iii) it should allow for the simplification of the function used to aggregate the values or utilities, and (iv) it should permit a reduction of the number of judgments that are required from the decision maker.

The AHP involves pairwise comparisons of options and attributes at different levels of a hierarchy, including some redundancy to enable consistency checks to be made (Saaty, 1994). This inevitably involves a large number of judgments so we will not consider this further, given our quest for an approach that meets the above ideal conditions. The same is true for methods that involve fuzzy weights or other fuzzy judgments to allow for uncertainty in assessments (Durbach and Stewart, 2012). For example fuzzy weights can require three assessments of weights (low, average and upper) or at least an assessment of the two bounds of an interval (Arbel and Vargas, 1993; Salo and Punkka, 2005, Aguayo et al., 2014). The PRIME method of Salo and Jiminez et al. (2013) allows for imprecision in judgments and uses a dominance structure to reduce the number of options that need to be considered. However, it can require the assessment of ratios of value differences between adjacent alternatives, which would still be likely to be demanding in the context of scenario planning. Given these issues, the following discussion focuses on simplifications of the SMART method.

The most common way of simplifying value elicitation, where the performance of options on an attribute is measured on a physical scale, is to assume a linear value function across the range of options with values ranging from 0 for the worst performing option to 100 (or 1) for the best. For example, a value function for income that ranged from £20 million to $60 million would assign a value of 50 (or 0.5) to an income of $40 million. Edwards and Barron (1994) suggested that linear value functions provide an acceptable degree of approximation to nonlinear functions as long as the ratio of the increase (or decrease) in value at one of the scale to that at the other end does not exceed 2 to 1. For attributes where the performances of options are qualitative (e.g. the attractiveness of alternative corporate logos) Edwards and Barron recommend that decision makers indicate their preferences through direct rating on a 0 to 100 scale.

Most suggestions in the literature have focused on simplifying the elicitation of weights. In SMART these are designed to reflect decision makers' relative preference for swings between the worst and best performances of options on the different attributes (for example, an improvement from the worst market share to the best is only 40% as important as an improvement from the worst profit to the best profit). Stillwell et al. (1981) discussed approximating the weights by simply asking decision makers to rank their preferences for the swings. The weights can then be estimated using methods such as rank sum (RS) or rank reciprocal (RR) (see below). More recently, Edwards and Barron (1994) proposed converting ranks to approximate weights by using rank order centroid (ROC) weights. Roberts and Goodwin (2002) found that rank sum weights gave the most accurate approximations to decision maker's preference where more than two to three attributes were involved. However, this may reflect two different assumptions on how decision makers assign weights. The point allocation method assumes that a decision maker has 100 points and allocates these across the attributes to reflect their relative importance (in terms of swings). In this case, the weight that is assigned finally is predetermined by the earlier allocations. ROC and RR and weights perform better when this assumption is valid (Danielson and Ekenberg, 2017). In the direct rating method, decision makers assign a weight of 100 to the attribute offering the most desired swing and lower weights to less preferable swings, but there is no constraint on what the weights sum to. This assumption favours RS weights. Danielson and Ekenberg (2017) suggest the use of cardinal counterparts to RR,RS and ROC weights in order to reflect, not only the ranks of attribute swings, but the strengths of preference between them. While these weights may be appropriate to many multiattribute decision problems, by also requiring elicitations of strength of preference, they would add to the already large number of judgments required in scenario planning-based decisions.

The aggregation of values or utilities for each option can be simplified by assuming an additive function so that each option's aggregate score is a weighted sum of the performance scores on the different attributes. Such functions assume mutual preference independence between the attributes, but Stewart (1996) showed that they can provide robust approximations to decision makers' preferences when outcomes are deterministic.

There are number of ways of attempting to reduce the number of judgments that need to be elicited from decision makers. In the context of portfolio decision making by groups of decision makers, Keisler (2008) explored using equal weights and making the decision simply based on the performance of options on a randomly selected attribute or on the most important attribute. He found that the effectiveness of such simplifications depended on the nature of the decision and its context. In computer simulations involving ROC weights, Fasolo et al. (2007) found that reliable decisions could be made when only one or two attributes were weighted as long as the scores on attributes were positively related and the true weights of the attributes were unequal. Of course, to benefit from this simplification an initial assessment would need to be made across all attributes to see of the necessary conditions were fulfilled.

Given the challenges of applying multiattribute decision analysis in scenario planning, Wright and Cairns (2011) proposed an extremely simple method that meets the four criteria listed above. The method draws on many of the ideas presented above to simplify the SMART-based method for selecting strategic options in scenario planning proposed by Goodwin and Wright (2001). However, as shown below, when assessing the performance of options on attributes, ranks replace the interval-scaled values used in SMART. Given that the method places the minimum possible demands on decision makers in order to provide decision support, it is referred to here as MINIMOD (or Minimal Modelling). However, while the Wright and Cairns proposal self-evidently simplifies the application of multiattribute decision analysis in scenario planning, the extent to which its indications are likely to reflect the true preferences of decision makers has not been assessed. In this paper, following a discussion and illustration of MINIMOD, computer-based simulations are used to perform this assessment.

The MINIMOD method

We focus on the multiattribute value evaluation process that involves identifying individual attributes that are relevant to the decision problem. These attributes are non-redundant in the sense that they do not duplicate other attributes and are such that the performances of at least two options on a given attribute will differ. For a given scenario, the options are evaluated on each attribute to form an "attributes by options" matrix (Roberts and Goodwin, 2002) as shown in Table 1.

Please insert Table 1 about here

Here v_{ij} is the value (or score) assigned to the ith option and jth attribute, with minj(vij) = 0, and maxj (vij) = 100 for the worst and best performing option, respectively. The w_j re weights reflecting the relative importance of the ranges of the attribute values, with

$\sum_{j=1}^{n} w_j = 1$. In addition, $w_1 \geq w_2 \geq \ldots\ldots\ldots \geq w_n > 0$

and $v(O_i) = \sum_{j=1}^{n} w_j v_{ij}$ where $v(Oi)$ is the aggregate multi attribute value (or score) for

option i in the range $100 \geq v \geq 0$.

In this process the v_{ij} are measured on an interval scale while the w_j are measured on a ratio scale. An underlying assumption of the calculation of $v(O_i)$ is that the attributes are mutually preference independent. The option with the maximum value of $v(O_i)$ is taken to be the option that the decision maker should prefer given the values and weights that have been elicited.

The MINIMOD method involves two approximations to the above process. First, instead of obtaining the v_{ij} from the decision maker a set of ranks is elicited which reflect the decision maker's order of preferences for each option on a given attribute in descending order (i.e. a rank of 1 is given to the least preferred option on the attribute, 2 to the second least preferred and on). Of course, if the decision maker prefers to provide ranks in ascending order then these ranks can easily be converted to descending order ranks by simply subtracting each rank from the maximum rank +1. Essentially, the ranks are equivalent to a set of interval scaled scores when there are equal intervals between adjacent scores.

The second approximation is similar to that of the SMARTER method. Instead of eliciting the w_j the swings between the worst and best performers on each attribute are merely ranked in ascending order to reflect their relative importance or desirability. However, unlike SMARTER, the MINIMOD method then converts these ranks to weights using the rank sum procedure (Stillwell et al, 1981). In this procedure the weights, $w_i(RS)$, are simply the individual ranks normalized by dividing by the sum of the ranks. The formula producing the weights, in its simplest form can be written as:

$$W_i(RS) = 2R_i/(n^2+n)$$

where the i^{th} rank is denoted by R_i and the ranks are assigned in descending order so that the largest swing has the highest rank (i.e., the largest number). The use of rank sum weights has a number of advantages over ROC weights, for which the formula for the ith most important attribute is given below:

$$w_i(\mathrm{ROC}) = (1/n)\sum_{j=i}^{n} 1/j, \qquad i = 1,\ldots\ldots\ldots, n$$

First, they are more transparent in their derivation and are more easily calculated. Indeed, for any given decision problem even the normalization step is not strictly necessary as the ordering of the aggregate scores for the options will remain the same without it. More fundamentally, ROC weights are based on the assumption the decision maker's 'true' weights will naturally sum to a fixed total of 1 or 100 (Roberts and Goodwin, 2002). In this case the elicitation of the weights can be viewed as a points allocation procedure with (say) 100 points being distributed between the attributes to reflect the importance of the swings that are associated with them. As Doyle et al (1997) point out achieving this would impose greater cognitive demands on the decision maker who simultaneously has to assess the relative importance of swings between the worst and best performances on each attribute while also ensuring that the resulting weights sum to exactly 1 or 100.

In usual applications of SMART no such constraint is imposed and a direct rating procedure is used (Goodwin and Wright, 2014). Here decision makers first assess the most important swing and assign it a weight of 100 (or 1). Weights are then assigned to the other swings by directly comparing their importance with this 'benchmark' swing. The sum of these 'raw' weights will clearly exceed 100 (or 1) so they are normalized to ensure that they sum to this total. On the basis that people more naturally use direct rating, Roberts and Goodwin (2002) assumed that their true raw weights (for all but the most important attribute) will be uniformly distributed between 0 and 100, an assumption that is consistent with Bayes' criterion (e.g. Kwon, 1978 p142; Jia et al. 1998). They then derived the expected values for the normalized weights. These expected values, which they referred to as Rank

Order Distribution (ROD) weights, were found to be very similar to rank sum weights, with the similarity being greater as the number of attributes in a decision problem increased. Finally, as Belton and Stewart (2002, p142) point out, when ROC weights are used the ratio of the highest to the lowest weights is very large which means that that the lowest ranked attribute will only have a very marginal influence on the decision. Attributes with a relative importance as low as this would be eliminated from the decision model in many practical situations. The use of rank sum weights reduces this extreme ratio problem.

Illustration of the MINIMOD method

We illustrate the simplicity of the MINIMOD method compared to SMART with an example which relates to the choice of strategies under a given scenario. Five alternative strategies (A to E) have been identified and the their performance on each of five attributes is shown in table 2.

Please insert Table 2 about here

In the SMART method, values on a 0 (worst) to 100 (best) scale are assigned to the strategies to assess their performances on each attribute in turn. For the numeric attributes (NPV and market share) these may indicate a non-linear value function and this is the case in the values assigned in table 3.

Please insert Table 3 about here

The decision maker ranks the swings on each attribute from the worst to best performances in order of importance. The most desirable swing (an improvement in NPV from $586m to $790m) is assigned a weight of 1. Improving Risk from Very High to Very Low is considered to 90 percent as preferable as the improvement in NPV so it is assigned a weight of 90. These, and the remaining weights, are shown in Table 4. The weights are then normalised to sum to 1 and, assuming that mutual preference independence applies, the aggregate value for each strategy is calculated by taking a weighted sum of the individual values across the attributes. For example, for strategy A the aggregate value is (100 x 0.31) + (40 x 0.06) + (0 x 0.09) + (30 x 0.28) +(0 x 0.26) = 41.5. When this calculation is applied to all the strategies it reveals that strategy E achieves the highest aggregate value of 79.1, under the given scenario, with D a close second at 74.1.

Please insert Table 4 about here

In MINIMOD the swings are ranked, as before, with the most desirable swing receiving the highest rank, in this case 5 and rank sum weights are then calculated as shown in Table 5

Please insert Table 5 about here

Rather than being assigned values, as in SMART, the performances of the strategies on each attribute are simply ranked, as shown in Table 6 (where 5 = best performance).

Please insert Table 6 about here

Again, assuming that mutual preference independence applies, the overall performance of each strategy across the attributes is simply obtained by multiplying the performance ranks by the rank sum weights. For example, for strategy A the score for overall performance is: (5 x 0.33) + (2 x 0.07) + (1 x 0.13) + (2 x 0.27) + (1 x 0.20) = 2.66. Similar calculations indicate that strategy E, with an aggregate score of 3.94, achieves the highest performance under the given scenario as was the case with SMART. Strategy D is again a close second with an aggregate score of 3.54. It can be seen that MINIMOD avoids the need to assess ratios for the swing weights and values for the performance of the strategies. In both cases only ranks are required. Scenario planning exercises usually involve groups of planners. As Ahn and Park (2008) argue, "In the situation where there is a group of decision makers, it may be realistic to expect agreement only on a ranking of weights." The same may be true of the performances of strategies on attributes.

Sensitivity analysis can be a crucial element in applications of decision support methods, but for many existing methods it can be problematical. Ideally, sensitivity analysis should indicate to decision makers the effect of making changes to the judgments (e.g. the weights assigned to the attributes) that *they* have put forward. However, in SMART the relationship between the score achieved by an option and a given pre-normalized weight is non-linear. To avoid this complexity, software products normally provide sensitivity analysis on the normalized weights, but this makes the relationship between the analysis and the decision maker's original judgments indirect. In alternative multiattribute decision methods such as Even-Swaps (Hammond et al., 1998; Belton et al. 2008) sensitivity analysis is not possible at all, while, for the Analytic Hierarchy Process (Saaty, 1994) , specialist software is required.

Unlike these methods, MINIMOD easily lends itself to sensitivity analysis. Figure1 shows the effect on the score of the options as the rank of net present value (NPV) changes from 1 (the least important attribute) to 5 (the most important). The other attributes retain their relative ranks in the analysis. In this case it can be seen that strategy E is preferred in the given scenario, irrespective of the rank that is assigned to NPV.

** Please insert Figure 1 about here**

Testing the MINIMOD method

Given that the MINIMOD method is an approximation of a SMART model it is important to assess the extent to which this approximation is likely to give misleading indications to a decision maker on which option should be chosen. We therefore followed the procedure of other researchers (e.g., Barron and Barrett, 1996; Salo and Hämäläinen, 2001; Jimenez, et al., 2013) and used simulation to test the method -in our case on 250,000 simulated decisions. For each of these decisions we compared the 'true' results obtained from a full SMART model with the following four methods

1. MINIMOD using rank sum weights, as illustrated above (MINIMOD(RS)).

2. MINIMOD but using rank-reciprocal weights. (MINIMOD(RR)). Rank reciprocal weights were also suggested by Stillwell et al. (1981) and are obtained using the following formula

$$w_i(RR) = (1/i)/(\sum_{j=1}^{n} 1/j) \qquad i = 1,........,n$$

3. MINIMOD using rank order centroid (ROC) weights (MINIMOD(ROC)).

4. SMARTER -which, as we indicated earlier, uses ROC weights but also uses actual scores rather than ranks to evaluate the options on the attributes. A comparison of (3) and(4) will therefore show the effect of this simplification.

Following Barron and Barrett's procedure, the simulated decisions had 3, 6, 9, 12 or 15 attributes and 5, 10, 15, 20 or 25 options. This yielded 25 decision 'sizes'. For each of these decision sizes 10,000 simulated decisions were generated. As Danielson and Ekenberg (2017) point out, the results of such simulations are heavily dependent on the assumptions that underpin the mechanism that generates the simulations. For example, the assumptions underlying the simulations conducted by Edwards and Barron (1994) favoured the use of ROC weights. A common mechanism is to generate the 'true' weights in each decision by selecting them randomly from a uniform distribution on the basis that these true weights are generally unknown (Jia et al. 1998) . However, the efficacy of an approximation to SMART may depend on the skewness of the true weight distribution. Table 7 shows examples of five possible distributions of non-normalized weights for a five attribute decision. For example, in the case of positive skew the majority of the weights are at the lower end of the 0 to 1 scale. Equal weights apply when the decision maker is indifferent between the swings from worst to best performance on the different attributes.

Please insert Table 7 about here

The same distributions could also apply to the true values (or scores) associated with the performance of options on a given attribute, with the exception of the equal weights distribution which would imply that an attribute was redundant as it did not discriminate between the options. Generating simulated weights or values solely from a uniform distribution would bias the results in favour of the MINIMOD method as the use of ranks assumes equal intervals between the true individual values or weights. To counter this, and to make the results representative of a wide range of possible decisions, unless equal weights applied, the true weights and values were randomly sampled from beta distributions which allow a range of different shaped distributions to be modelled (Vose, 2000). The distributions used were beta (0.8, 02) for positive skew, beta(1,1) for a uniform distribution, beta(4,4) for a bell-shaped distribution and beta(0.2,0.8) for negative skew. The weights and values were generated using the inverse transform method based on uniformly distributed random numbers sampled from the range 0 to 1.

For a given sized decision the simulations proceeded as follows. For each attribute one of the four beta distributions was randomly selected and this was then used to randomly generate the true scores of the options on that attribute. These were then rescaled so that that the lowest score was 0 and the highest 100. A similar method was then used to generate the weights - though here the selection was randomly made from the four beta distributions plus equal weights. The weights were subsequently normalized to sum to 1. For each decision size the following results were recorded for each of the four methods after the 10,000 simulations:

1. The Hit Rate (if there was no tie between two 'best options) i.e. the percentage of decision where the approximation method selected the best option indicated by the full SMART model

2. The mean percentage utility loss (if no tie) i.e.

$$\frac{v^* - v_a}{v^* - v_w} \times 100$$

Where: v^* = the aggregate score of the best option in SMART

v_a = the aggregate SMART score of the option selected by the approximation method

v_w = the aggregate score of the worst option in SMART

For example, if in the SMART model the best option scores 80 and the worst 20, but the approximation selects an option with a SMART score of 60 the utility loss will be:

$$\frac{80 - 60}{} \times 100 = 33.3\%$$

80 – 20

The advantage of utility loss as a measure of performance is that, when an approximation method identifies an option which is a close second best in the SMART model, this will only result in a small utility loss.

Results of testing

Figure 2 shows the hit rates for decisions of different sizes while figure 3 shows the mean percentage utility losses for these decisions.

** Please insert Figure 2 about here**

** Please insert Figure 3 about here**

In terms of hit rates SMARTER tends to outperform the MINIMOD (RS) method for small problems, but the gap closes as the number of attributes increases and also as the number of options gets smaller. Indeed, for very large problems (which is precisely where its simplifications are likely to be most useful) the performance of MINIMOD(RS) is indistinguishable from SMARTER. Its hit rates are typically around 75%, but more importantly its mean % utility loss is very low –typically around 2%-4%. Other weighting systems are not so successful.

Clearly, where there are more options available there is a greater chance that an approximation method will miss the best options and figure 2, shows that the hit rate of all methods declines as the number of options increases. In contrast, the percentage utility loss declines for all methods as the number of options get larger (for a given number of attributes).

When more options are available their aggregate scores tend to fill the space between 0 and 100 more intensively. Hence a second or third best option is likely to have a score closer to the best score. For example for three options the scores might be 90, 40 and 10 so the second option has a utility loss of 62.5%. For 10 options we might have scores of 90, 85, 72, 64, 54, 45, 42, 32, 21 and 10 so the second option has a utility loss of only 6.25%.

Interestingly, the performance of MINIMOD(RS) on both hit rate and utility loss tends to be less affected by the number of attributes than that of SMARTER. The ROC method tends to assign a high percentage of the total of the weights to the first and second most important attributes. For example, when there are 15 attributes the top 2 attributes receive 38% of the total of the weights, compared to only 24% in the rank sum method. Thus the ROC method will tend to be most reliable when the top few attributes dominate the decision problem and the other attributes are relatively insignificant. This is less likely to be the case when there a large number of attributes. It is also interesting to note that the ROC weights do not combine well with the ranked scores. The performance of MINIMOD(ROC) was generally worse than both SMARTER and MINIMOD(RS).

Discussion

Overall, the MINIMOD(RS) approach appears to provide a reliable approximation to SMART, despite its extreme simplicity. On occasions where the decision maker's intuition is at variance with the recommendation of the model this simplicity and the transparency of MINIMOD(RS) lends itself to easy investigation so that the conflict can be explored, and hopefully resolved, and a requisite decision model obtained (Phillips, 1984). According to Phillips such investigations can lead to deeper understanding about the decision problem so that the final decision can be made with greater confidence and insight.

When discrepancies do arise between the decision maker's intuition and the recommendation of MINIMOD(RS) then this investigation may involve eliciting scores, rather than ranks for an attribute which is suspected to be the source of the conflict. For example, when the underlying performance variable is quantifiable (e.g. where performance is measured in monetary returns) the method can make implicit assumptions that may be in sharp contrast with a decision maker's preferences. For example, suppose that six options offer monetary returns of \$0, \$50m, \$75m, \$90m, \$99m and \$100m, respectively. MINIMOD would assume that an increase in returns from \$99 to \$100m is just as desirable as an increase from \$0 to \$50m, suggesting an increasing marginal utility for money, when diminishing marginal utility is the reality. In this case scores of 0, 70, 90, 98,99,100 might be elicited. These can be easily converted to a scale compatible with the ranks being used for the other attributes using the following formula (we will call the results 'rank scores').

$$\text{'Rank score'} = 1 + x(m-1)/100$$

where x is the elicited score and m is the number of options. In this case a score of 70, for example, would be converted to a 'rank score' of $1 + 70(6-1)/100 = 4.5$ and the new aggregate scores could then be calculated to see if the discrepancy with the decision maker's intuition has been resolved.

Another potential problem can emerge when the performances of strategies tie on particular attributes. For example, if there are five strategies and the three worst performers tie, their ranks (5= best) will be 5,4,3,3 and 3. This has two effects. It narrows the range of scores to one which will only be 50% of that for attributes with no ties. This will mean that the worst performers are being wrongly treated as if they are halfway between the worst and best performers in calculations. It also opens the possibility of rank reversal in the aggregate scores if some strategies are removed (Wang and Luo, 2009). In this case, as above, direct scores could be elicited and then converted to ‘rank scores’ using the above formula.

There are inevitably a number of limitations associated with the results of this study. First, the underlying assumption is that decision makers employ a direct rating method when assigning weights to attributes as opposed to a point allocation system where, say, 100 points are distributed between the attributes to reflect the relative importance of their swings from worst to best performances. As indicated earlier, the point allocation approach appears to be more cognitively demanding for decision makers as the sum of the weights has to meet the constraint represented by the original number of points available. No such constraint applies in the direct rating method and there is evidence that it more reliable than point allocation, at least where individual judgment are concerned (Bottomley and Doyle, 2013). Nevertheless, point allocation would have favoured ROC weights in the simulations.

Second, the simulation assumed that each of the probability distributions selected for the values and weights was equi-probable and that these distributions reflect the skewness of distributions of values and weights that would be found in real scenario planning exercises. However, without a representative survey of such exercises –assuming this would be practically possible - it is not possible to verify the validity of this assumption. All that we can conclude is that the MINIMOD method gave a robust performance over the many combinations of distributions that were simulated. A further limitation is that the distributions were generated independently so that the effect of correlations between values on different attributes was not investigated. Of course, many of the simulations would have contained such correlations by chance, especially where the number of options was small.

Third, the analysis described here only investigated the effectiveness of the methods when decisions of different ‘sizes’ were being made. There was no analysis of the effect of other characteristics of decisions on the efficacy of the approximation methods, such as different levels of skewness or the degree of nonlinearity of value functions. This was because the MINIMOD method is designed to minimize the amount of assessment that need to be made. If a pre-assessment was required before a decision was made to administer the method this would increase the demand placed on decision makers. In particular, it would clearly be nonsensical to require an assessment of exact values or weights before subsequently approximating them. Thus application of the method relies on post-modelling assessment and modification through Philips's 'requisite process' (Phillips, 1984) to ensure its reliability rather than pre-modelling assessment of its suitability.

Conclusions

The above results suggest that, despite its simplicity, the MINIMOD(RS) method provides a reliable approximation to SMART. Even when the method fails to identify the best option the utility loss tends to be low and managers always have the option of replacing scores for some attributes with 'rank scores' if they are uncomfortable with the indications of the method. As Phillips (1984) argues, the investigation of discrepancies between intuition and decision models can itself lead to new insights into the decision problem being faced and investigations like this are likely to be less burdensome when MINIMOD is being used. Because the method is simple and relatively quick to implement it is likely to appeal to busy managers who need to evaluate the performance of strategic options under a range of scenarios. This is particularly likely to be the case when the number of attributes and options is large so that the number of scores and weights that would need to be elicited for a full SMART model would be both demotivating and fatiguing. In addition the transparency of MINIMOD(RS) makes it suitable for use by diverse groups of stakeholders when they meet to form strategic planning teams (Cairns et al., 2016).

Of course, as Edwards and Barron (1994) pointed out, approximations like this may deny decision makers the benefits that come from the ‘deep soul searching’ and reflection that more advanced decision analysis methods may demand. These benefits may include deeper insights and a greater understanding of the decision problem. Nevertheless, there are likely to be many occasions where the MINIMOD(RS) method is likely to be useful and appropriate, not the least where the use of an alternative method would be either unacceptable to decision makers or infeasible because of time constraints.

References

Ahn, B. S., & Park, K. S. (2008). Comparing methods for multiattribute decision making with ordinal weights. *Computers & Operations Research*, 35, 1660-1670.

Aguayo, E. A., Mateos, A., & Jiménez, A. (2014). A new dominance intensity method to deal with ordinal information about a DM’s preferences within MAVT. *Knowledge-Based Systems*, 69, 159-169.

Arbel, A., & Vargas, L. G. (1993). Preference simulation and preference programming: robustness issues in priority derivation. *European Journal of Operational Research*, 69, 200-209.

Barron, F.H. & Barrett B.E. (1996). Decision quality using ranked attributes weights. *Management Science*, 42, 1515 - 1523.

Belton, V. and Gear, T., (1983). On a short-coming of Saaty's method of analytic hierarchies. Omega, 11, 228-230.

Belton, V. & Stewart, T.J. (2002). *Multiple Criteria Decision Analysis: An Integrated Approach*. Norwell, MA: Kluwer Academic Publishers.

Belton, V., Wright G., & Montibeller, G. (2008). When is swapping better than weighting? An evaluation of the even swaps method in comparison with multiattribute value analysis. *Working Paper, Dept. of Management Science, University of Strathclyde*.

Borcherding, K., Eppel, T. & von Winterfeldt, D. (1991). Comparison of weighting judgments in multiattribute utility measurement, *Management Science*, 37, 1603-1619.

Bottomley, P. A., & Doyle, J. R. (2013). Comparing the validity of numerical judgements elicited by direct rating and point allocation: Insights from objectively verifiable perceptual tasks. *European Journal of Operational Research*, 228, 148-157.

Cairns, G., Goodwin, P., & Wright, G. (2016). A decision-analysis-based framework for analysing stakeholder behaviour in scenario planning. *European Journal of Operational Research*, 249, 1050-1062.

Danielson, M., & Ekenberg, L. (2017). A robustness study of state-of-the-art surrogate weights for MCDM. *Group Decision and Negotiation*, 26, 677-691.

Doyle, J.R, Green, R.H, & Bottomley, P.A. (1997). Judging relative importance: direct rating and point allocation are not equivalent. *Organisational Behavior and Human Decision Processes*,70, 55-72.

Durbach, I. (2014). Outranking under uncertainty using scenarios. *European Journal of Operational Research*, 232, 98-108.

Durbach, I. (In Press). Scenario planning in the analytic hierarchy process. *Futures and Foresight Science.*

Durbach, I., & Stewart, T. J. (2003). Integrating scenario planning and goal programming. *Journal of Multi-Criteria Decision Analysis*, 12, 261-271.

Durbach, I. & Stewart, T.J. (2012). A comparison of simplified value function approaches for treating uncertainty in multi-criteria decision analysis. *Omega,* 40, 456-464.

Durbach, I. & Stewart, T.J. (2012). Modeling uncertainty in multi-criteria decision analysis. *European Journal of Operational Research*, 223, 1-14.

Edwards (1977). How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics* 7, 326-340.

Edwards, W. & Barron, H.F. (1994). SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organisational Behavior and Human Decision Processes*, 60, 306-325.

Eisenhardt, K. M., & Sull, D. N. (2001). Strategy as simple rules. *Harvard Business Review*, 79, 106-119.

Fasolo, B., McClelland, G. H., & Todd, P. M. (2007). Escaping the tyranny of choice: When fewer attributes make choice easier. *Marketing Theory*, 7, 13-26.

Goodwin, P. & Wright, G., (2001). Enhancing strategy evaluation in scenario planning: a role for decision analysis. *Journal of Management Studies*, 38**,** 1-16.

Goodwin, P. & Wright, G., (2014). *Decision Analysis for Management Judgment* (5th edn.). Chichester: Wiley.

Hammond, J.S., Keeney, R.L., & Raiffa, H. (1998). Even swaps: a rational method for making tradeoffs. *Harvard Business Review,* 76, 137-150.

Jia, J., Fischer, G. W., & Dyer, J. S. (1998). Attribute weighting methods and decision quality in the presence of response error: a simulation study. *Journal of Behavioral Decision Making*, 11, 85-105.

Jimenez, A., Mateos, A., & Sabio, P. (2013). Dominance intensity measure within fuzzy weight oriented MAUT: An application. *Omega*, 41, 397-405.

Keeney, R. L., & Raiffa, H. (1976). *Decision Analysis with Multiple Conflicting Objectives*. New York: Wiley.

Kwon I-W. (1978). *Statistical Decision Theory with Business and Economic Applications: A Bayesian Approach.* New York: Petrocelli/Charter.

Katsikopoulos, K. V., & Fasolo, B. (2006). New tools for decision analysts. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 36, 960-967.

MacGregor, D., Lichtenstein, S., & Slovic, P. (1988). Structuring knowledge retrieval: An analysis of decomposed quantitative judgments. *Organizational Behavior and Human Decision Processes*, 42, 303-323.

Montibeller, G., Gummer, H., & Tumidei, D. (2006). Combining scenario planning and multi-criteria decision analysis in practice. *Journal of Multi-Criteria Decision Analysis*, 14, 5-20.

Payne , J.W. (1976). Task complexity and contingent processing in decision making : an information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366-387.

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge: Cambridge University Press.

Phillips, L. D. (1984). A theory of requisite decision models. *Acta Psychologica*, 56, 29–48.

Ram, C., Montibeller, G., & Morton, A. (2011). Extending the use of scenario planning and MCDA for the evaluation of strategic options. *Journal of the Operational Research Society*, *62*(5), 817-829.

Roberts, R. & Goodwin, P. (2002). Weight approximations in multi-attribute decision models. *Journal of Multi-Criteria Decision Analysis*, 11, 291-203.

Saaty,T. (1994). *The Analytical Hierarchy Process.* New York: Wiley.

Salo, A. A., & Hämäläinen, R. P. (1997). On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-Criteria Decision Analysis*, 6, 309-319.

Salo, A.A. & Hämäläinen, R.P. (2001). Preference ratios in multiattribute evaluation (PRIME) – elicitation and decision procedures under incomplete information, *IEEE Transactions on Systems, Man, and Cybernetics Part A* 3, 533-545.

Salo, A.A. & Punkka, A. (2005). Rank inclusion in criteria hierarchies. *European Journal of Operational Researc*h, 163, 338-356

Schroeder, M. J., & Lambert, J. H. (2011). Scenario-based multiple criteria analysis for infrastructure policy impacts and planning. *Journal of Risk Research*, 14, 191-214.

Stewart, T. J. (1996). Robustness of additive value function methods in MCDM. *Journal of Multi-Criteria Decision Analysis*, 5, 301-309.

Stewart, T., French, S. & Rios, J. (2013). Integrating multicriteria decision analysis and scenario planning: review and extension. Omega, 41, 679-688.

Stillwell, W.G. Seaver, D.A, & Edwards, W. (1981). A comparison of weight approximation techniques in multiattribute utility decision making. *Organisational Behavior and Human Performance*. 28, 62-77.

Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, 79, 281–299.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

Van der Heijden, K. (1996). *Scenarios. The Art of Strategic Conversation.* Chichester: Wiley.

Vilkkumaa, E., Liesiö, J., Salo, A., & Ilmola-Sheppard, L. (2018). Scenario-based portfolio model for building robust and proactive strategies. *European Journal of Operational Research*, 266, 205-220.

Vose, D. (2000). *Risk Analysis. A Quantitative Guide* (2nd ed.). Chichester: Wiley.

Wang, Y. M., & Luo, Y. (2009). On rank reversal in decision analysis. *Mathematical and Computer Modelling*, 49, 1221-1229.

Wright, G. & Cairns, G. (2011). *Scenario Thinking. Practical Approaches to the Future.* Basingstoke: Palgrave Macmillan.

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting*, 25, 813-825.

		O_1	O_2	.	O_m
	w_1	v_{11}	v_{21}	.	v_{m1}
	w_2	v_{12}	v_{22}	.	v_{m2}
	.	.	.	.	.
	w_n	v_{1n}	v_{2n}	.	$v_{m\,n}$
Totals	1	$v(O_1)$	$v(O_2)$	.	$v(O_m)$

Table 1 Decision Matrix

	Attributes				
Strategy	**NPV ($m)**	**Market share**	**Environmental impact**	**Risk**	**Flexibility**
A	790	15%	Very High	High	Very Low
B	731	10%	High	Very High	Low
C	586	21%	Small	Medium	Medium
D	686	25%	None	Low	High
E	697	18%	Medium	Very Low	Very High

Table 2 Choice of strategies in a given scenario

	Attributes				
Strategy	**NPV ($m)**	**Market share**	**Environmental impact**	**Risk**	**Flexibility**
A	100	40	0	30	0
B	90	0	30	0	10
C	0	55	90	50	30
D	55	100	100	80	75
E	60	50	40	100	100

Table 3 Values assigned to strategies in SMART

Attribute	Swing	Weight	Normalised weight
NPV	\$586m to \$790m	100	0.31
Risk	Very High to Very Low	90	0.28
Flexibility	Very Low to Very High	85	0.26
Environmental impact	Very High to None	30	0.09
Market share	10% to 25%	20	0.06
	Sum	325	

Table 4 Weights assigned to strategies in SMART

Rank	Attribute	Swing	Rank sum weight
5	NPV	$586m to $790m	0.33
4	Risk	High to Very Low	0.27
3	Flexibility	Low to High	0.20
2	Environmental impact	High to Small	0.13
1	Market share	10% to 25%	0.07

Table 5 Weights assigned to strategies in MINIMOD

Strategy	NPV ($m)	Market share	Environmental impact	Risk	Flexibility
A	5	2	1	2	1
B	4	1	2	1	2
C	1	4	4	3	3
D	2	5	5	4	4
E	3	3	3	5	5

Table 6 Performance ranks of strategies on attributes

Positive skew	Uniform	Bell-shaped	Equal	Negative skew
0.13	0.11	0.10	1.00	0.11
0.14	0.33	0.40	1.00	0.22
0.18	0.56	0.55	1.00	0.90
0.75	0.78	0.40	1.00	0.95
1.00	1.00	1.00	1.00	1.00

Table 7 Possible distributions of non-normalized, unconstrained weights

<u>Figure Legends</u>

Figure 1 Sensitivity analysis

Figure 2 Hit rates

Figure 3   Mean percentage utility losses