# Automatic Criteria Weight Generation for Multi-criteria Decision Making Under Uncertainty

Mats Danielson[1,2] and Love Ekenberg[2,1(✉)]

[1] Department of Computer and Systems Sciences, Stockholm University, PO Box 7003, SE-164 07 Kista, Sweden
`mats.danielson@su.se`
[2] International Institute for Applied Systems Analysis, IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria
`ekenberg@iiasa.ac.at`

**Abstract.** Real-life decision situations almost invariably involve large uncertainties. In particular, there are several difficulties connected with the elicitation of probabilities, utilities, and criteria weights. In this article, we explore and test a robust multi-criteria weight generating method covering a broad set of decision situations, but which still is reasonably simple to use. We cover an important class of methods for criteria weight elicitation and propose the use of a reinterpretation of an efficient family (rank exponent) of methods for modelling and evaluating multi-criteria decision problems under uncertainty. We find that the rank exponent (RX) family generates the most efficient and robust weighs and works very well under different assumptions. Furthermore, it is stable under varying assumptions regarding the decision-makers' mindset and internal modelling. We also provide an example to show how the algorithm can be used in a decision-making context. It is exemplified with a problem of selecting strategies for combatting COVID-19.

**Keywords:** Multi-criteria decision analysis · Uncertain reasoning · Criteria weights · Criteria ranking · Rank order · Automatic weight generation · COVID-19

## 1 Introduction

A problem with many Multi-Criteria Decision Making (MCDM) models is that there is a lack of numerically precise information available in real life and it is hence difficult for a decision-maker to enter realistic input data into a model. There is, therefore, a perceived need for relaxing the demand for precise judgments to more realistically model decision problems. See, for instance, (Park 2004; Larsson et al. 2014) among others. Solutions to such problems are sometimes significantly hard to find and the results can be difficult to interpret. Quite well-known methods for approaching this problem are based on, e.g., sets of probability measures, upper and lower probabilities as well as interval probabilities and utilities (Coolen and Utkin 2008), fuzzy measures (Aven and Zio 2011; Shapiro

and Koissi 2015; Tang et al. 2018) as well as evidence and possibility theory, cf., e.g., (Dubois 2010; Dutta et al. 2018; Rohmer and Baudrit 2010) just to mention a few of them. Other approaches include second-order methods (Ekenberg et al. 2014; Danielson et al. 2007, 2019) and modifications of classical decision rules, cf., (Ahn and Park 2008; Sarabando and Dias 2009; Aguayo et al. 2014; Mateos et al. 2014). Regarding MCDM problems, Salo, Hämäläinen, and others have suggested methods for handling imprecise information, for instance, the PRIME method (Salo and Hämäläinen 2001) with various implementations thereof, see e.g. (Mustajoki et al. 2005b). Several other models are focussing on preference intensities, such as the MACBETH method (Bana e Costa et al. 2002), a variety of ROC approaches, such as (Sarabando and Dias 2010), or the Simos's method and variants thereof (Figueira and Roy 2002). Furthermore, there are smart swaps methods, such as (Mustajoki and Hämäläinen 2005a). Mixes of the above techniques are also common, as in Jiménez et al. (2006).

A major problem is combining interval and qualitative estimates without introducing evaluation measures like $\Gamma$-maximin or (Levi's) E-admissibility, cf., e.g., (Augustin et al. 2014). Greco et al. (2008) suggest the UTA[GMS] methodology for purposes similar to ours. By using an ordinal regression technique, they can form a representation based on a set of pairwise comparisons. This is generalised in Figueira et al. (2009) by introducing cardinalities for obtaining a class of total preference functions compatible with user assessments. However, this is less suitable for our purposes since it is unclear how interval constraints can be handled in combination with the extracted preference functions without encountering the computational difficulties discussed in, e.g., (Danielson and Ekenberg 2007). Also, structural constraints should be taken into consideration as discussed already in, e.g., (Ekenberg et al. 2005).

This paper will, more particularly, discuss a method for criteria weight elicitation that can be generally applied to any case where automatic weight generation is considered and with the property that weight functions can be elicited while preserving efficiency and correctness. Below we will provide a brief discussion of so-called surrogate weight methods and then propose a reinterpretation of the rank exponential method. Herein, we focus on ordinal information. In many circumstances, there is only ordinal information available which merits the investigation into ordinal weights. In (Danielson and Ekenberg 2017), it is investigated how much contribution cardinality brings over ordinality, where it is demonstrated that weights are much more insensitive to cardinality than values, which has implications for all ranking methods. We also provide experimental simulations and investigate some properties of the method. Thereafter, a problem of selecting a national strategy for handling the COVID-19 pandemic is discussed. The conclusion is that the method seems to be a very competitive candidate for weight elicitation and evaluations.

## 2   Rank Ordering Methods

Ordinal methods for generating weights, sometimes with some kind of further discrimination mechanism, constitute a quite commonly used approach to handle the difficulties in eliciting precise criteria weights from decision-makers, c.f., e.g., (Stewart 1993; Arbel and Vargas 1993; Barron and Barrett 1996a, 1996b; Katsikopoulos and Fasolo 2006). The decision-maker supplies ordinal information on importance, which subsequently is converted into numerical weights following the ordinal information. There have in

the literature been several suggestions of such methods, e.g., rank sum weights (RS), rank reciprocal weights (RR) (Stillwell et al. 1981), and centroid (ROC) weights (Barron 1992). Based on simulation experiments, Barron and Barrett (1996b) found ROC weights superior to RS and RR. Danielson and Ekenberg (2014 2016a, 2016b), applied in large-scale contexts, such as (Fasth et al. 2020, Komendantova et al. 2018, 2020), have also suggested a spectrum of methods and suggested some that are more robust than the earlier suggestions. In these experiments, surrogate weights as well as "true" reference weights are sampled from some underlying distribution. Then it is investigated how well the surrogate number results match the result of using the "true" results. The method is however dependent on the distribution used for generating the weight vectors.

RS is based on the idea that the rank order should be reflected directly in the weights. Given a simplex $S_w$ generated by $w_1 > w_2 > \ldots > w_N$, where $\Sigma w_i = 1$ and $0 \leq w_i$, assign an ordinal number to each item in the ranking, starting with the highest-ranked item as number 1. Let $i$ be the ranking number among $N$ items to rank. RS then becomes

$$w_i^{\text{RS}} = \frac{N + 1 - i}{\sum_{j=1}^{N}(N + 1 - j)} = \frac{2(N + 1 - i)}{N(N + 1)}$$

for all $i = 1, \ldots, N$.

RR has a similar design as RS but is based on the reciprocals (inverted numbers) of the rank order items. Assign an ordinal number to each item ranked, starting with the highest-ranked item (receiving number 1). Then assign the number $i$ to the i:th item in the ranking to obtain

$$w_i^{\text{RR}} = \frac{1/i}{\sum_{j=1}^{N}\frac{1}{j}}$$

ROC is a function based on the average of the corners in the polytope defined by the same simplex $S_w = w_1 > w_2 > \ldots > w_N$, $\Sigma w_i = 1$, and $0 \leq w_i$, where $w_i$ are variables representing the criteria weights. The ROC weights are given by

$$w_i^{\text{ROC}} = 1/N \sum_{j=i}^{N} \frac{1}{j}$$

for the ranking number $i$ among $N$ items to rank.

As a generalization to the RS method previously discussed, a rank exponent weight method was introduced by (Stillwell et al. 1981). In the original RS formula, they introduced the exponent $z < 1$ to yield the rank exponent (RX) weights given by

$$w_i^{\text{RX}(z)} = \frac{(N + 1 - i)^z}{\sum_{j=1}^{N}(N + 1 - j)^z}.$$

For $0 \leq z \leq 1$ the parameter $z$ mediates between the case of equal weights (no discrimination between the importance of criteria) and RS weights such that for $z = 0$ it in effect becomes equal weights and for $z = 1$ it instead becomes RS weights. Thus, for these values of the parameter $z$ the RX($z$) formula is the (exponential) combination of

equal and RS weights. In this paper, we suggest a reinterpretation of RX.[1] This has, to our knowledge, not been investigated before. Beyond $z = 1$ it becomes something else, a novel weighting scheme in its own right. Earlier, before the accessibility and use of simulations to evaluate different weights, parameters such as the $z$ parameter of RX was considered hard to estimate and thus less suitable for real-life decisions. In this work, we examine the potential of $RX(z)$ in detail and compare it to established state-of-the-art weights such as RS, RR, and ROC.

## 3   Assessing Automatically Generated Weights

There are basically two categories of elicitation models that are in use depending on the degrees of freedom (DoF) present when decision-makers assign their weights. In point allocation (PA), decision-makers are given point sums, e.g. 100, to distribute among $N$ criteria and there are consequently $N-1$ degrees of freedom. In the Direct Rating (DR) way of assigning weights, on the other hand, the decision-makers have no limitation on the point sum they are allowed to use, and thus a decision-maker may allocate as many points as desired. Only thereafter, the points are normalized, i.e., in DR there are $N$ degrees of freedom. Consequently, when generating weight vectors in an $N-1$ DoF model, they must sum to 100%, and when generating vectors for an $N$ DoF model, a vector is generated keeping components within [0%, 100%] which is thereafter normalised. Other distributions would of course at least theoretically be possible, but it is important to remember that the validation methods are strongly dependent on these assumptions and affect the validations. Different decision-makers use different mental strategies and models when weighting criteria. Thus, a reasonable weighting scheme must be able to perform well in both PA and DR cases, i.e. regardless of the degrees of freedom being $N-1$ or $N$ (Danielson and Ekenberg 2019).

### 3.1   Experimental Setup

The experiments below for an $N-1$ DoF model was based on a homogenous $N$-variate Dirichlet distribution generator, and a standard round-robin normalised random weight generator was used for the $N$ DoF experiments. We call the $N-1$ DoF model type of generator an *N–1-generator* and the $N$ DoF model type an *N-generator*. Details of the simulation generators are given in (Danielson and Ekenberg 2014).

The simulation experiment consisted of four numbers of criteria $N = \{3, 6, 9, 12\}$ and five numbers of alternatives $M = \{3, 6, 9, 12, 15\}$, i.e. a total of 20 simulation scenarios. These simulation sets were selected to cover the most common sizes of decision problems. The behaviour with large decision problems is not within the scope of this article. Each scenario was run 10 times with 10,000 trials for each of them yielding a total of 2,000,000 decision situations. Unscaled value vectors were generated uniformly, and no significant differences were observed with other value distributions. The results

---

[1] Stillwell et al. prescribed a very different use of the $z$ parameter, $0 < z < 1$, in which $z$ is the decision-maker's estimate of the largest weight. In their original version, $z$ is thus a required extra input parameter. In our reinterpretation, $z$ is a configuration parameter that can be determined beforehand and not required user input. We still chose to keep the RX name despite that.

of the simulations are shown in the tables below, where we show a subset of the results with chosen pairs $(N, M)$.

The "winner frequency" in the tables refers to the fraction of cases where the best alternative was correctly predicted. Other measurements include "podium frequency" where the three best alternatives are correctly predicted and "total frequency" where the positions of all alternatives are correctly predicted. The latter two measurements showed the same pattern across the weighting methods as the winners, and are thus not presented here since they would not add to the discussion [2].

The first set of tables shows the winner frequency for the $RX(z)$ family of methods and the second set of tables shows the winner frequency for the older ROC, RR, RS methods together with selected $RX(z)$ methods. Both sets of tables utilise the simulation methods $N-1$ DoF, $N$ DoF, and an equal combination of $N-1$ and $N$ DoF. All hit ratios in all tables are given in percent and are mean values of the 10 scenario runs. With hit ratios is meant the fraction of times that the correct winner is predicted.

The first set of studies concern the parameter $z$ of the $RX(z)$ method. Recall that $z = 1$ is the same as the RS method studied previously and which is used as one of the comparisons in the next set of tables. For values $0 \leq z \leq 1$, which is a combination of RS and equal weights, the algorithm underperforms compared to already known algorithms. This is easily understood since equal weights is a very weak weighting scheme as it does not take any information on the decision situation into account. Thus, this study focuses on parameters $z > 1$.

## 3.2  Results

In Table 1, using an $N-1$-generator, it can be seen that higher parameter values tend to outperform the others when looking at the winner. In Table 2, the frequencies have changed according to expectation since we employ a model with $N$ degrees of freedom. Now lower parameter values outperform higher (lower being closer to RS), but not at all by as much. In Table 3, the $N$ and $N-1$ DoF models are combined with equal emphasis on both. Now, we can see that in total medium-sized parameters generally perform the best. While (Stillwell et al. 1981) discussed $z < 1$, it is evident by examining the formula that it cannot outperform RS (which is $z = 1$) since it is the linear combination of RS and equal weights, the latter being the worst performer since it does not take any information into account. Thus, we did in this experiment vary $z$ from 1 (RS) and up in steps of 0.1 until the performance declined. The best performances for the different sizes were always found in the interval [1.1, 1.6]. Thus, it gives guidance to select the best $z$ given the problem size.

It is clear from the table that parameters z $\in$ [1.3, 1.5] are the best performers but that all of the range [1.2, 1.6] are performing well. Since we do not know exactly what goes on inside a particular decision-maker's head when giving input to a decision situation, it is not wise to rely on a weight function to perform well on only one side of the dimensionality spectrum above. Instead, we consider the mix of $N$ and $N-1$ dimensions to constitute the most valid measurement of a viable automatically generated weighting scheme.

---

[2] In a choice problem, which we discuss here, it is better to use "winner frequency" while in an ordering problem it would be more appropriate to use "total frequency".

**Table 1.** Hit ratio for predicting winners using an $N-1$-generator

| $N-1$ DoF | z = | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 |
|---|---|---|---|---|---|---|---|
| 3 criteria | 3 alternatives | 89.1 | 89.5 | 89.9 | 90.0 | 90.1 | 90.1 |
| 3 criteria | 15 alternatives | 77.0 | 77.4 | 77.8 | 78.3 | 78.5 | 78.6 |
| 6 criteria | 6 alternatives | 80.9 | 81.7 | 82.5 | 83.2 | 83.6 | 83.8 |
| 6 criteria | 12 alternatives | 76.7 | 77.7 | 78.8 | 79.4 | 79.8 | 80.2 |
| 9 criteria | 9 alternatives | 77.4 | 78.6 | 79.6 | 80.4 | 81.3 | 81.9 |
| 12 criteria | 6 alternatives | 79.1 | 80.1 | 81.2 | 82.3 | 83.0 | 83.6 |
| 12 criteria | 12 alternatives | 75.5 | 76.9 | 77.9 | 79.1 | 79.9 | 80.4 |

**Table 2.** Hit ratio for predicting winners using an $N$-generator

| $N$ DoF | z = | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 |
|---|---|---|---|---|---|---|---|
| 3 criteria | 3 alternatives | 89.2 | 89.0 | 88.8 | 88.5 | 88.3 | 87.9 |
| 3 criteria | 15 alternatives | 81.0 | 80.8 | 80.4 | 79.8 | 79.4 | 78.6 |
| 6 criteria | 6 alternatives | 87.0 | 86.7 | 86.3 | 85.7 | 84.9 | 84.1 |
| 6 criteria | 12 alternatives | 83.9 | 83.7 | 83.0 | 82.1 | 81.3 | 80.4 |
| 9 criteria | 9 alternatives | 86.9 | 86.5 | 85.7 | 84.9 | 83.7 | 82.6 |
| 12 criteria | 6 alternatives | 90.1 | 89.6 | 89.0 | 88.2 | 87.3 | 86.6 |
| 12 criteria | 12 alternatives | 87.2 | 86.8 | 86.1 | 85.1 | 83.8 | 82.6 |

**Table 3.** Hit ratio for predicting winners using $N$ and $N-1$ DoF generators combined

| Combined | z = | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 |
|---|---|---|---|---|---|---|---|
| 3 criteria | 3 alternatives | 89.2 | 89.3 | 89.4 | 89.3 | 89.2 | 89.0 |
| 3 criteria | 15 alternatives | 79.0 | 79.1 | 79.1 | 79.1 | 79.0 | 78.6 |
| 6 criteria | 6 alternatives | 84.0 | 84.2 | 84.4 | 84.5 | 84.3 | 84.0 |
| 6 criteria | 12 alternatives | 80.3 | 80.7 | 80.9 | 80.8 | 80.6 | 80.3 |
| 9 criteria | 9 alternatives | 82.2 | 82.6 | 82.7 | 82.7 | 82.5 | 82.3 |
| 12 criteria | 6 alternatives | 84.6 | 84.9 | 85.1 | 85.3 | 85.2 | 85.1 |
| 12 criteria | 12 alternatives | 81.4 | 81.9 | 82.0 | 82.1 | 81.9 | 81.5 |

But in line with that reasoning, we would also like to minimise the spread between the dimensions, i.e. having a generating function that differs less between both endpoints of the input dimensionality scale is preferred to one that has a larger spread. To that effect, in addition to studying the overall hit ratio, we also studied the spread of

the results from different dimensionalities. This is shown in Table 4 for the different $z$ parameters of RX($z$). Now a quite different picture emerges. While all parameters $z \in [1.2, 1.6]$ perform well overall, it is clear that higher $z$ keeps the spread down, especially for problems of larger size. Since this is a highly desirable property given that we don't know the thinking process of a particular decision-maker, we tend to favour higher $z$ parameters for their robustness as long as they do perform well overall. For comparisons with current well-known weight functions, we select both RX(1.5) and RX(1.6).

**Table 4.** Spread between hit ratio for predicting winners using $N$ and $N-1$ DoF generators

| Spread | $z =$ | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 |
|---|---|---|---|---|---|---|---|
| 3 criteria | 3 alternatives | 0.1 | 0.5 | 1.1 | 1.5 | 1.8 | 2.2 |
| 3 criteria | 15 alternatives | 4.0 | 3.4 | 2.6 | 1.5 | 0.9 | 0.0 |
| 6 criteria | 6 alternatives | 6.1 | 5.0 | 3.8 | 2.5 | 1.3 | 0.3 |
| 6 criteria | 12 alternatives | 7.2 | 6.0 | 4.2 | 2.7 | 1.5 | 0.2 |
| 9 criteria | 9 alternatives | 9.5 | 7.9 | 6.1 | 4.5 | 2.4 | 0.7 |
| 12 criteria | 6 alternatives | 11.0 | 9.5 | 7.8 | 5.9 | 4.3 | 3.0 |
| 12 criteria | 12 alternatives | 11.7 | 9.9 | 8.2 | 6.0 | 3.9 | 2.2 |

### 3.3   Comparing with Earlier State-of-the-Art Weights

In (Danielson and Ekenberg 2014), previous classic weighting functions were compared. Here, these results are repeated together with the new results for RX. The latter is represented by RX(1.5) and RX(1.6) which achieved the best results above. In Table 5, using an $N-1$-generator, it can be seen that ROC outperforms the other classical ones when looking at the winner. RR is better than RS (which is RX(1.0)). In Table 6, the frequencies have changed according to expectation since we employ a model with $N$ degrees of freedom. Now RS outperforms the others including RX while ROC and RR are far behind. In Table 7, the $N$ and $N-1$ DoF models are combined with equal emphasis on both. Now, we can see that in total RX generally performs the best.

It is clear from studying the resulting tables that the RX family of automatic weight functions easily outperform the more well-known functions, provided that it is possible to select the $z$ parameter in an informed manner. The picture becomes even clearer once the spread between different decision-maker ways of thinking is being taken into consideration.

None of the other studied state-of-the-art functions perform well under varying conditions, while the RX($z$) family is able to do so. Especially somewhat higher $z$ parameters perform very well, making parameter selection a trade-off between pure performance and robustness. Our suggestion is to use $z \in [1.5, 1.6]$ as the optimal compromise for the parameter. As was seen in Table 4, lower $z$-values lead to less robustness with respect to decision-maker styles of reasoning. With a higher parameter, the RX($z$) family by far outperforms the earlier known ROC, RS, and RR weighting schemes.

**Table 5.** Hit ratio for predicting winners using an $N-1$-generator

| $N-1$ DoF | | ROC | RS | RR | RX(1.5) | RX(1.6) |
|---|---|---|---|---|---|---|
| 3 criteria | 3 alternatives | 90.2 | 88.2 | 89.5 | 90.1 | 90.1 |
| 3 criteria | 15 alternatives | 79.1 | 76.6 | 76.5 | 78.5 | 78.6 |
| 6 criteria | 6 alternatives | 84.8 | 79.9 | 82.7 | 83.6 | 83.8 |
| 6 criteria | 12 alternatives | 81.3 | 75.6 | 78.2 | 79.8 | 80.2 |
| 9 criteria | 9 alternatives | 83.5 | 75.6 | 79.5 | 81.3 | 81.9 |
| 12 criteria | 6 alternatives | 86.4 | 77.8 | 80.8 | 83.0 | 83.6 |
| 12 criteria | 12 alternatives | 83.4 | 72.9 | 76.8 | 79.9 | 80.4 |

**Table 6.** Hit ratio for predicting winners using an $N$-generator

| $N$ DoF | | ROC | RS | RR | RX(1.5) | RX(1.6) |
|---|---|---|---|---|---|---|
| 3 criteria | 3 alternatives | 87.3 | 89.3 | 88.3 | 88.3 | 87.9 |
| 3 criteria | 15 alternatives | 77.9 | 81.1 | 79.1 | 79.4 | 78.6 |
| 6 criteria | 6 alternatives | 80.1 | 87.3 | 78.1 | 84.9 | 84.1 |
| 6 criteria | 12 alternatives | 76.4 | 84.3 | 74.3 | 81.3 | 80.4 |
| 9 criteria | 9 alternatives | 76.3 | 87.2 | 69.8 | 83.7 | 82.6 |
| 12 criteria | 6 alternatives | 77.5 | 90.3 | 67.8 | 87.3 | 86.6 |
| 12 criteria | 12 alternatives | 73.4 | 87.6 | 63.1 | 83.8 | 82.6 |

**Table 7.** Hit ratio for predicting winners using $N$ and $N-1$ DoF generators combined

| Combined | | ROC | RS | RR | RX(1.5) | RX(1.6) |
|---|---|---|---|---|---|---|
| 3 criteria | 3 alternatives | 88.8 | 88.8 | 88.9 | 89.2 | 89.0 |
| 3 criteria | 15 alternatives | 78.5 | 78.9 | 77.8 | 79.0 | 78.6 |
| 6 criteria | 6 alternatives | 82.5 | 83.6 | 80.4 | 84.3 | 84.0 |
| 6 criteria | 12 alternatives | 78.9 | 80.0 | 76.3 | 80.6 | 80.3 |
| 9 criteria | 9 alternatives | 79.9 | 81.4 | 74.7 | 82.5 | 82.3 |
| 12 criteria | 6 alternatives | 82.0 | 84.1 | 74.3 | 85.2 | 85.1 |
| 12 criteria | 12 alternatives | 78.4 | 80.3 | 70.0 | 81.9 | 81.5 |
| Mean | | 81.3 | 82.4 | 77.4 | 83.2 | 83.0 |

**Table 8.** Spread between hit ratio for predicting winners using $N$ and $N-1$ DoF generators

| Spread | | ROC | RS | RR | RX(1.5) | RX(1.6) |
|---|---|---|---|---|---|---|
| 3 criteria | 3 alternatives | 2.9 | 1.1 | 1.2 | 1.8 | 2.2 |
| 3 criteria | 15 alternatives | 1.2 | 4.5 | 2.6 | 0.9 | 0.0 |
| 6 criteria | 6 alternatives | 4.7 | 7.4 | 4.6 | 1.3 | 0.3 |
| 6 criteria | 12 alternatives | 4.9 | 8.7 | 3.9 | 1.5 | 0.2 |
| 9 criteria | 9 alternatives | 7.2 | 11.6 | 9.7 | 2.4 | 0.7 |
| 12 criteria | 6 alternatives | 8.9 | 12.5 | 13.0 | 4.3 | 3.0 |
| 12 criteria | 12 alternatives | 10.0 | 14.7 | 13.7 | 3.9 | 2.2 |
| Mean | | 5.7 | 8.6 | 7.0 | 2.3 | 1.2 |

## 4 Example

In the current outbreak of the COVID-19 pandemic, several nation-states seem to have been less than fully prepared. Where strategic plans existed, they were often either not complete or not followed. In some cases, the supply of resources was not sufficient to sustain the outbreak over time. Further, cognitive and behavioural biases seem to have played a significant role in the decision-making processes regarding which risk mitigation and management measures to implement. Many countries were to a large extent unprepared for a similar scenario to arise, despite the fact that predictions about a significant probability for a pandemic to occur in a foreseeable future, and national governments of several countries often acted in an uncoordinated manner, which have resulted in suboptimal responses from national bodies. The current discourse has had a strong emphasis on the number of direct fatalities, while there still is a multitude of relevant aspects of the current crisis. In this example, we briefly discuss how a more general framework, including epidemiological and socio-economic factors, could look like using a model for evaluating the qualitative and quantitative aspects involved.

A detailed account of all the relevant aspects is beyond the scope of this article and for demonstrational purposes only, we just use a few possible options and criteria for a national policy with four levels of restrictions suggested to be imposed on the population of a country affected by COVID-19.[3]

Some examples of possible mitigation strategies could then be:

1. An unmitigated response
2. Response by pharmaceutical measures and case isolation, public communication encouraging increased hygiene and personal protection.
3. 2 + additional personal protective measures and mild social distancing measures.

---

[3] A more complete discussion of possibilities for how to contain the virus spread from a policy point of view, while considering different societal and policy factors in a multi-stakeholder-multi-criteria context, as well as preferences amongst relevant stakeholder groups, is, without any mathematical details, provided in (Ekenberg et al. 2020).

4.  3 + self-selected social distancing and comprehensive contact tracing and publicly disclosed detailed location information of individuals that tested positive for COVID-19.

We use the following four criteria:

1.  Number of cases (including critical, severe and mild)
2.  Economic aspects
3.  Human rights violations
4.  Effects on education

The estimates (for demonstrational purpose only) on the values of each response level under each criterion are shown in Table 9 below.

**Table 9.** The valuation of strategies under the respective criteria

| Criterion/Measure | Cases (in 10s of thousands)[a] | Economy (GDP decline in %) | Human rights | Education (% of students having no access to online or any other schooling) |
|---|---|---|---|---|
| *Strategy 1* | 1415–1729 | 1–3 | Better than Str.2 | 0 |
| *Strategy 2* | 1412–1725 | 1–4 | Better than Str.3 | 1–5 |
| *Strategy 3* | 1334–1630 | 5–7 | Better than Str.4 | 10–20 |
| *Strategy 4* | 1217–1487 | 5–10 | | 10–30 |

[a]Epidemiological simulations and other estimates are from (Ekenberg et al. 2020). The estimates are based on the effects of various measures, which are sometimes overlapping due to the significant uncertainties involved and some significant similarities between them.

We need to calibrate the different scales since they are of very different characters and in this example, we assume that:

- The maximum difference between Str.1 and Str.4 in Cases is more important than the maximum difference of Str.1 and Str.4 in Economy.
- The maximum difference between Str.1 and Str.4 in Economy is more important than the maximum difference of Str.1 and Str.4 in Human rights.
- The maximum difference between Str.1 and Str.4 in Human rights is more important than the maximum difference between Str.1 and Str.4 in Education.

The resulting criteria ranking then becomes the following: The importance of Cases is higher than that of Economy, which in turn is more important than Human rights. Further,

Human rights is more important than Education. This ranking is then represented by the RX(1.5) weight generating algorithm. The weights (using the enumeration above) then become w(1) = 0.470, w(2) = 0.305, w(3) = 0.166, and w(4) = 0.059 respectively.

The generated weights together with estimates on the values of each response strategy can then be evaluated by solving successive optimisation problems using the program DecideIT which employs the RX weights together with algorithms from (Danielson et al. 2019). For the evaluation, belief distributions are generated from the input data (both weights and values) using the algorithms in the program. The value $V(S_i)$ for each strategy is then assessed as $V(S_i) = \sum w_i \cdot v_{ij}$ for all weight and value variables involved. The result can be seen in Fig. 1, where Str.1 is found to be the best option for a policy given the background information used herein. The strategy values $V(S_i)$ are seen at the top of the evaluation bars. The coloured parts are the contributions from each criterion.
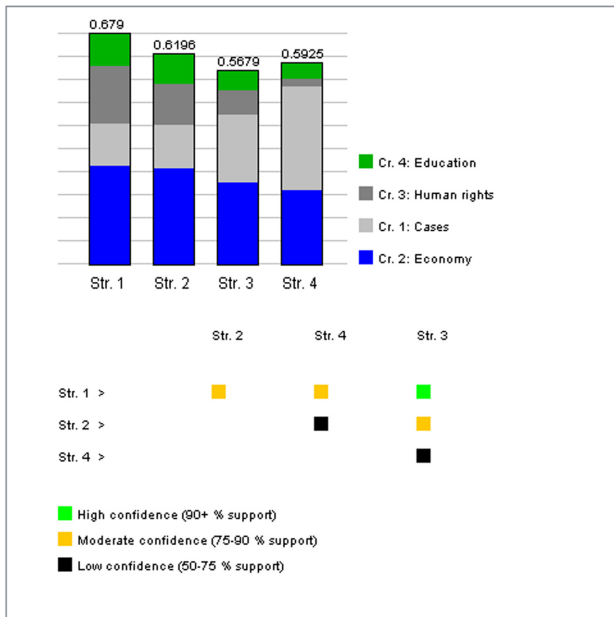


**Fig. 1.** The result of the analysis. (Color figure online)

Without going into the details, Fig. 1 shows that, given the background information, the higher the bars representing the strategies, the better the respective strategies are. We can also see the result's robustness by the colour markings. A green square means that there is a significant difference between the strategies and that there must be large changes in the input data for it to change. A yellow square means that there is still a significant difference, but that the result is more sensitive to input data. A black square means that there is no significant difference between the strategies. An extended explanation of the semantics regarding the bars and the colour markings are provided in (Danielson et al.

2019). Str.1 is thus the best option in this example. Furthermore, this result is quite robust. It is followed by Str.2, which is quite similar to Str.4, but better than Str.3.[4]

## 5   Conclusions

This paper aims to define and test a robust multi-criteria weight generating method covering a broad set of decision situations, but which still is reasonably simple to use. In the analyses, we have investigated the average hit rate in percent over the pairs ($N$, $M$) of number of criteria and alternatives. From Tables 7 and 8 concerning generated weight performances, RX(1.5) and RX(1.6) are found to be the best candidates for representing weights when searching an optimal alternative. The other weight generation methods are clearly inferior. In particular, ROC is heavily biased by an assumption that decision-makers have a decision process with $N-1$ degrees of freedom considering $N$ criteria, while a reasonable requirement on a robust a rank ordering method is that it should provide adequate alternative rankings under the varying assumptions that we have little real-life knowledge about. It is thus clear that the RX family of methods generates the most efficient and robust weighs and works very well on different problem sizes. Furthermore, it is stable under varying assumptions regarding the decision-makers' mindset and internal modelling. As further research, the obvious next step and extension to the observations in this paper is to find a configuration function that asserts different parameter values to problems of different sizes. This would further increase the efficiency of the RX family of automatic weight functions over its previous competitors.

## References

Aguayo, E.A., Mateos, A., Jiménez-Martín, A.: A new dominance intensity method to deal with ordinal information about a DM's preferences within MAVT. Knowl. Based Syst. **69**, 159–169 (2014)

Ahn, B.S., Park, K.S.: Comparing methods for multiattribute decision making with ordinal weights. Comput. Oper. Res. **35**(5), 1660–1670 (2008)

Arbel, A., Vargas, L.G.: Preference simulation and preference programming: robustness issues in priority derivation. Eur. J. Oper. Res. **69**, 200–209 (1993)

Augustin, T., Coolen, F.P., De Cooman, G., Troffaes, M.C. (eds.) Introduction to Imprecise Probabilities. Wiley Series in Probability and Statistics. Wiley, Hoboken (2014)

Aven, T., Zio, E.: Some considerations on the treatment of uncertainties in risk assessment for practical decision making. Reliabil. Eng. Syst. Saf. **96**, 64–74 (2011)

Bana e Costa, C.A., Correa, E.C., De Corte, J.M., Vansnick, J.C.: Facilitating bid evaluation in public call for tenders: a socio-technical approach. Omega **30**, 227–242 (2002)

---

[4] The result depends largely on the values in Table 9. These will not be discussed further here since the main purpose of the article is to showcase the weights.

Barron, F.H.: Selecting a best multiattribute alternative with partial information about attribute weights. Acta Physiol. (Oxf) **80**(1–3), 91–103 (1992)

Barron, F., Barrett, B.: The efficacy of SMARTER: simple multi-attribute rating technique extended to ranking. Acta Physiol. (Oxf) **93**(1–3), 23–36 (1996a)

Barron, F., Barrett, B.: Decision quality using ranked attribute weights. Manag. Sci. **42**(11), 1515–1523 (1996b)

Coolen, F.P., Utkin, L.V.: Imprecise reliability. In: Melnick, E.L., Everitt, B.S. (eds.) Encyclopedia of Quantitative Risk Analysis and Assessment (2008)

Danielson, M., Ekenberg, L.: Computing upper and lower bounds in interval decision trees. Eur. J. Oper. Res. **181**(2), 808–816 (2007)

Danielson, M., Ekenberg, L.: Rank ordering methods for multi-criteria decisions. In: Zaraté, P., Kersten, G.E., Hernández, J.E. (eds.) GDN 2014. LNBIP, vol. 180, pp. 128–135. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07179-4_14

Danielson, M., Ekenberg, L.: A robustness study of state-of-the-art surrogate weights for MCDM. Group Decis. Negot. **7** (2016a)

Danielson, M., Ekenberg, L.: The CAR method for using preference strength in multi-criteria decision making. Group Decis. Negot. **25**(4), 775–797 (2016b)

Danielson, M., Ekenberg, L.: Simplifying cardinal ranking in MCDM methods. In: Schoop, M., Kilgour, D.M. (eds.) Proceedings of the 17th International Conference on Group Decision and Negotiation, University of Hohenheim, Stuttgart (2017)

Danielson, M., Ekenberg, L.: An improvement to swing techniques for elicitation in MCDM methods. Knowl.-Based Syst. **168**, 70–79 (2019)

Danielson, M., Ekenberg, L., Larsson, A.: Distribution of belief in decision trees. Int. J. Approx. Reason. **46**(2), 387–407 (2007)

Danielson, M., Ekenberg, L., Larsson, A.: A second-order-based decision tool for evaluating decisions under conditions of severe uncertainty. Knowl.-Based Syst. **191** (2019)

Dubois, D.: Representation, propagation, and decision issues in risk analysis under incomplete probabilistic information. Risk Anal. **30**(3), 361–368 (2010)

Dutta, P.: Human health risk assessment under uncertain environment and its SWOT analysis. Open Publ. Health J. **11** (2018)

Ekenberg, L., Danielson, M., Larsson, A., Sundgren, D.: Second-order risk constraints in decision analysis. Axioms **3**, 31–45 (2014)

Ekenberg, L., Mihai, A., Fasth, T., Komendantova, N., Danielson, M.: Mitigating cognitive and behavioural biases during pandemics responses. Under review (2020)

Ekenberg, L., Thorbiörnson, J., Baidya, T.: Value differences using second order distributions. Int. J. Approx. Reason. **38**(1), 81–97 (2005)

Fasth, T., Bohman, S., Larsson, A., Ekenberg, L., Danielson, M.: Portfolio decision analysis for evaluating stakeholder conflicts in land use planning. Group Decis. Negot. **29**(2), 321–343 (2020)

Figueira, J.R., Greco, S., Słowiński, R.: Building a set of additive value functions representing a reference preorder and intensities of preference: GRIP method. Eur. J. Oper. Res. **195**(2), 460–486 (2009)

Figueira, J., Roy, B.: Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. Eur. J. Oper. Res. **139**, 317–326 (2002)

Greco, S., Mousseau, V., Słowiński, R.: Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. Eur. J. Oper. Res. **191**(2), 416–436 (2008)

Jiménez, A., Ríos-Insua, S., Mateos, A.: A generic multi-attribute analysis system. Comput. Oper. Res. **33**, 1081–1101 (2006)

Katsikopoulos, K., Fasolo, B.: New tools for decision analysis. IEEE Trans. Syst. Man Cybern. – Part A: Syst. Hum. **36**(5), 960–967 (2006)

Komendantova, N., Ekenberg, L., Marashdeh, L., Al Salaymeh, A., Danielson, M., Linnerooth-Bayer, J.: Are energy security concerns dominating environmental concerns? Evidence from stakeholder participation processes on energy transition in Jordan. Climate **6**(4) (2018)

Komendantova, N., et al.: Water-energy nexus: addressing stakeholders' preferences in Jordan. Sustainability (2020)

Larsson, A., Riabacke, M., Danielson M., Ekenberg, L.: Cardinal and rank ordering of criteria – addressing prescription within weight elicitation. Int. J. Inf. Technol. Decis. Mak. **13** (2014)

Mateos, A., Jiménez-Martín, A., Aguayo, E.A., Sabio, P.: Dominance intensity measuring methods in MCDM with ordinal relations regarding weights. Knowl. Based Syst. **70**, 26–32 (2014)

Mustajoki, J., Hämäläinen, R.: A preference programming approach to make the even swaps method even easier. Decis. Anal. **2**, 110–123 (2005)

Mustajoki, J., Hämäläinen, R., Salo, A.: Decision support by interval SMART/SWING - incorporating imprecision in the SMART and SWING methods. Decis. Sci. **36**(2), 317–339 (2005)

Park, K.S.: Mathematical programming models for characterizing dominance and potential optimality when multicriteria alternative values and weights are simultaneously incomplete. IEEE Trans. Syst. Man Cybern. - Part A: Syst. Hum. **34**(5), 601–614 (2004)

Rohmer, J., Baudrit, C.: The use of the possibility theory to investigate the epistemic uncertainties within scenario-based earthquake risk assessments. Nat. Hazards **56**, 613–632 (2010). Springer

Salo, A.A., Hämäläinen, R.P.: Preference Ratios in Multiattribute Evaluation (PRIME)—elicitation and decision procedures under incomplete information. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. **31**, 533–545 (2001)

Sarabando, P., Dias, L.: Multi-attribute choice with ordinal information: a comparison of different decision rules. IEEE Trans. Syst. Man Cybern. Part A **39**, 545–554 (2009)

Sarabando, P., Dias, L.: Simple procedures of choice in multicriteria problems without precise information about the alternatives' values. Comput. Oper. Res. **37**, 2239–2247 (2010)

Shapiro, A.F., Koissi, M.C.: Risk Assessment Applications of Fuzzy Logic. Casualty Actuarial Society, Canadian Institute of Actuaries, Society of Actuaries (2015)

Stewart, T.J.: Use of piecewise linear value functions in interactive multicriteria decision support: a Monte Carlo study. Manag. Sci. **39**(11), 1369–1381 (1993)

Stillwell, W., Seaver, D., Edwards, W.: A comparison of weight approximation techniques in multiattribute utility decision making. Org. Behav. Hum. Perform. **28**(1), 62–77 (1981)

Tang, M., Liao, H., Li, Z., Xu, Z.: Nature disaster risk evaluation with a group decision making method based on incomplete hesitant fuzzy linguistic preference relations. Int. J. Environ. Res. Publ. Health (2018)