

Trade-Offs for Ordinal Ranking Methods in Multi-criteria Decisions

Mats Danielson^{1,2} and Love Ekenberg^{1,2} 

¹ Department of Computer and Systems Sciences,
Stockholm University, Box 7003, 164 07 Kista, Sweden
mad@dsv.su.se, ekenberg@iiasa.ac.at

² International Institute of Applied Systems Analysis, IIASA,
Schlossplatz 1, 2361 Laxenburg, Austria

Abstract. Weight elicitation methods in multi-criteria decision analysis (MCDA) are often cognitively demanding, require too much precision and too much time and effort. Some of the issues may be remedied by connecting elicitation methods to an inference engine facilitating a quick and easy method for decision-makers to use weaker input statements, yet being able to utilize these statements in a method for decision evaluation. One important class of such methods ranks the criteria and converts the resulting ranking into numerical so called surrogate weights. We analyse the relevance of these methods and discuss how robust they are as candidates for modelling decision-makers and analysing multi-criteria decision problems under the perspectives of several stakeholders.

Keywords: Multi-criteria decision analysis · Criteria weights · Criteria ranking · Rank order

AQ1

1 Introduction

Regardless of the methods used for decision making, there exists a real problem in that numerically precise information is seldom available, and when it comes to providing reasonable weights in multi-criteria decision problems, there are severe difficulties due to decision-makers not seeming to have neither precise information at hand nor the required discrimination capacity. The same problem appears in group decision settings, where there is a desire to rank or in other ways compare the views or values of different participants or stakeholders and there are several approaches to this. For instance, Fig. 1 shows an implementation of the multi-criteria multi-stakeholder approach for using the CAR method from [4], based on rankings and imprecise information. It is developed for group decisions for infrastructure policy making in Swedish municipalities.

To somewhat circumvent these problems, some approaches utilise imprecise importance information to determine criteria weights and sometimes values of

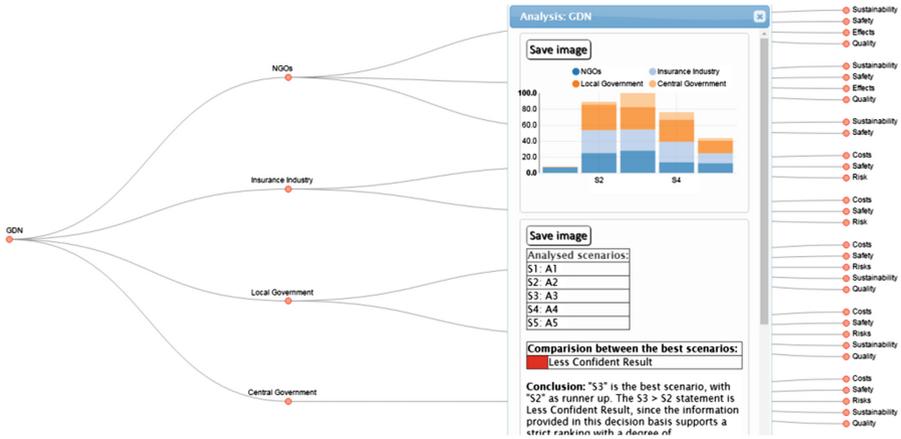


Fig. 1. The Group Decision tool Decision Wizard

alternatives.¹ Such methods were mostly assessed using case studies until [2] introduced a process utilising systematic simulations from criteria orderings. The basic idea is to generate surrogate weights from a criteria ordering and investigate how well the results of using these surrogates match the ordering provided by the decision-maker, while trying to determine which the ‘true’ weights of the decision-makers are.

The methodology is of course vulnerable since the validation result is heavily dependent on the model we have of the decision-maker’s mind set reflected in the distribution used for generating the weight vectors. To use these surrogate numbers, whichever way they are produced, they need to fulfil some robustness ideas, since we cannot for certain claim that we know exactly what decision-makers have in mind when stating such an ordering. This article discusses these robustness issues when translating orderings into surrogate numbers.

2 Rank Ordering Methods

Common for most elicitation methods is the assumption that all elicitation is made relative to a weight distribution held by the decision-maker.² One initial idea is to just skip the criteria elicitation and assign equal weights to every criterion. However, the information loss is then very large and it is most often worthwhile to at least rank the criteria, since rankings are (often) easier to provide than precise numbers. An ordering of the criteria is then achieved which can be handled in various ways. One such is to introduce so called surrogate weights, which are derived from the supposed ranking.

¹ In this paper the full features of the large variety of elicitation techniques will not be discussed. For more exhaustive discussion refer to [6].

² For various cognitive and methodological aspects of imprecision in decision making, see e.g., [3] and others by the same authors.

This technique is utilised in [2] and many others. In these classes of methods, the resulting ranking is converted into numerical weights by surrogate functions. Needless to say, for practical decision making, surrogate weights can seem as a peculiar way of motivating a method and the results of these kinds of methods should always be interpreted in the light of this. Nevertheless, some kind of absolute validation in this field is impossible and the surrogate methods are quite widely used and can be considered as some of several ways of trying to motivate the various generation methods suggested. The crucial issue then rather becomes how to set surrogate weights while losing as little information as possible and preserving the “correctness” when assigning the weights. In addition to surrogate weights (such as those discussed above), dominance procedures and classical methods have been discussed in various contexts. Dominance procedures are often versions of outranking, based on pairwise dominance. The classical methods consist of the well-known maximax, maximin, and minimax regret decision rules. Categories of weights other than surrogate weights are not considered any further in this paper because the discussed surrogate methods, as will be shown, are very efficient.

3 The RS, RR and ROC Methods

In the literature, various surrogate weight methods have been suggested. [8] discusses rank sum (RS) weights and rank reciprocal (RR) weights, which are alternatives to the quite popular ROC (rank order centroid) from [1]. The rank sum is based on the idea that the rank order should be reflected directly in the weight. Assume a simplex S_w generated by $w_1 > w_2 > \dots > w_N$, $\sum w_i = 1$ and $0 \leq w_i$. We will, unless otherwise stated, henceforth presume that decision problems are modelled as simplexes S_w . Assign an ordinal number to each item ranked, starting with the highest ranked item as number 1. Denote the ranking number i among N items to rank. Thus, a larger weight is assigned to lower ranking numbers.

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

Another idea discussed in [8] is rank reciprocal (RR) weights. They have a similar origin as the RS weights, but are based on the reciprocals (inverted numbers) of the rank order for each item ranked. These are obtained by assigning an ordinal number to each item ranked, starting with the highest ranked item as number 1. Thus, a larger weight is again assigned to lower ranking numbers.

$$w_i^{\text{RR}} = \frac{1/i}{\sum_{j=1}^N \frac{1}{j}}, \text{ where } i \text{ and } j \text{ is as above.} \quad (2)$$

ROC weights are the centroid components of the simplex S_w . That is, ROC is a function based on the average of the corners in the polytope defined by the simplex $S_w = w_1 > w_2 > \dots > w_N$, $\sum w_i = 1$, and $0 \leq w_i$.

$$w_i^{\text{ROC}} = 1/N \sum_{j=i}^N \frac{1}{j} \quad (3)$$

In this way, it resembles RR more than RS but is, particularly for lower dimensions, more extreme than both in the sense of weight distribution, especially the largest and smallest weights (Table 1).

Table 1. The RS weights up to ten dimensions

RS	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$
$N = 1$	1.0000									
$N = 2$	0.6667	0.3333								
$N = 3$	0.5000	0.3333	0.1667							
$N = 4$	0.4000	0.3000	0.2000	0.1000						
$N = 5$	0.3333	0.2667	0.2000	0.1333	0.0667					
$N = 6$	0.2857	0.2381	0.1905	0.1429	0.0952	0.0476				
$N = 7$	0.2500	0.2143	0.1786	0.1429	0.1071	0.0714	0.0357			
$N = 8$	0.2222	0.1944	0.1667	0.1389	0.1111	0.0833	0.0556	0.0278		
$N = 9$	0.2000	0.1778	0.1556	0.1333	0.1111	0.0889	0.0667	0.0444	0.0222	
$N = 10$	0.1818	0.1636	0.1455	0.1273	0.1091	0.0909	0.0727	0.0545	0.0364	0.0182

Of the three methods above, ROC has often been considered to be a quite reasonable candidate, despite that generated weights are sometimes perceived to be too sharp or discriminative, meaning that too large emphasis is put on the larger weights, i.e. on those criteria ranked highest up in the ranking order (Table 2).

When comparing the centroid weights of the ROC, RS, and RR, we can see that there are significant differences. To begin with, ROC is compared to RR in Table 4. RR shows a rather large similarity to ROC in terms of generating function. While it is true

Table 2. The RR weights up to ten dimensions

RR	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$
$N = 1$	1.0000									
$N = 2$	0.6667	0.3333								
$N = 3$	0.5455	0.2727	0.1818							
$N = 4$	0.4800	0.2400	0.1600	0.1200						
$N = 5$	0.4380	0.2190	0.1460	0.1095	0.0876					
$N = 6$	0.4082	0.2041	0.1361	0.1020	0.0816	0.0680				
$N = 7$	0.3857	0.1928	0.1286	0.0964	0.0771	0.0643	0.0551			
$N = 8$	0.3679	0.1840	0.1226	0.0920	0.0736	0.0613	0.0526	0.0460		
$N = 9$	0.3535	0.1767	0.1178	0.0884	0.0707	0.0589	0.0505	0.0442	0.0393	
$N = 10$	0.3414	0.1707	0.1138	0.0854	0.0683	0.0569	0.0488	0.0427	0.0379	0.0341

that ROC puts larger emphasis on the higher ranked criteria than the other methods for lower N , RR quickly takes over as the most heavy emphasize of the highest ranked criterion from $N = 6$ onwards (Table 3).

Table 3. the ROC weights up to 10 dimensions

ROC	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$
$N = 1$	1.0000									
$N = 2$	0.7500	0.2500								
$N = 3$	0.6111	0.2778	0.1111							
$N = 4$	0.5208	0.2708	0.1458	0.0625						
$N = 5$	0.4567	0.2567	0.1567	0.0900	0.0400					
$N = 6$	0.4083	0.2417	0.1583	0.1028	0.0611	0.0278				
$N = 7$	0.3704	0.2276	0.1561	0.1085	0.0728	0.0442	0.0204			
$N = 8$	0.3397	0.2147	0.1522	0.1106	0.0793	0.0543	0.0335	0.0156		
$N = 9$	0.3143	0.2032	0.1477	0.1106	0.0828	0.0606	0.0421	0.0262	0.0123	
$N = 10$	0.2929	0.1929	0.1429	0.1096	0.0846	0.0646	0.0479	0.0336	0.0211	0.0100

Table 4. ROC compared to RR

ROC-RR	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$
$N = 1$	0.0000									
$N = 2$	0.0833	-0.0833								
$N = 3$	0.0657	0.0051	-0.0707							
$N = 4$	0.0408	0.0308	-0.0142	-0.0575						
$N = 5$	0.0187	0.0377	0.0107	-0.0195	-0.0476					
$N = 6$	0.0002	0.0376	0.0223	0.0007	-0.0205	-0.0402				
$N = 7$	-0.0153	0.0347	0.0276	0.0121	-0.0043	-0.0201	-0.0347			
$N = 8$	-0.0282	0.0308	0.0296	0.0186	0.0057	-0.0070	-0.0191	-0.0304		
$N = 9$	-0.0392	0.0265	0.0298	0.0223	0.0122	0.0017	-0.0084	-0.0180	-0.0269	
$N = 10$	-0.0485	0.0222	0.0291	0.0242	0.0163	0.0077	-0.0009	-0.0091	-0.0168	-0.0241

RR also emphasizes the lowest ranked criteria, making it the weighing method that puts the most emphasis on the end-points of the ranking. The differences in the table (and in some of the subsequent ones) might not appear as too dramatic, but better precision of a method, even if it by some smaller percentage, is a way of fine tuning these originally quite rough methods based of a kind of statistical validity. Comparing ROC to RS in Table 5, it is evident that ROC emphasizes the higher ranked criteria at the expense of the middle and lower ranked ones. For the middle ranked criteria, this is the opposite of the RR method when compared to ROC.

Table 5. ROC compared to RS

ROC-RS	<i>i</i> = 1	<i>i</i> = 2	<i>i</i> = 3	<i>i</i> = 4	<i>i</i> = 5	<i>i</i> = 6	<i>i</i> = 7	<i>i</i> = 8	<i>i</i> = 9	<i>i</i> = 10
<i>N</i> = 1	0.0000									
<i>N</i> = 2	0.0833	-0.0833								
<i>N</i> = 3	0.1111	-0.0556	-0.0556							
<i>N</i> = 4	0.1208	-0.0292	-0.0542	-0.0375						
<i>N</i> = 5	0.1233	-0.0100	-0.0433	-0.0433	-0.0267					
<i>N</i> = 6	0.1226	0.0036	-0.0321	-0.0401	-0.0341	-0.0198				
<i>N</i> = 7	0.1204	0.0133	-0.0224	-0.0344	-0.0344	-0.0272	-0.0153			
<i>N</i> = 8	0.1175	0.0203	-0.0144	-0.0283	-0.0318	-0.0290	-0.0221	-0.0122		
<i>N</i> = 9	0.1143	0.0254	-0.0079	-0.0227	-0.0283	-0.0283	-0.0246	-0.0182	-0.0099	
<i>N</i> = 10	0.1111	0.0293	-0.0026	-0.0177	-0.0245	-0.0263	-0.0248	-0.0209	-0.0153	-0.0082

4 The SR Method

Not surprisingly, ROC, RS and RR perform well only for specific assumptions on decision-maker assignments of criteria weight preferences and since these weight models are in a sense opposites, it is interesting to see how extreme behaviours can be reduced. A natural candidate for this could be a linear combination of RS and RR. Since we have no reasons to expect anything else, we can, e.g., balance them equally in an additive combination of the Sum and the Reciprocal weight function that we call the SR weight method.

$$w_i^{SR} = \frac{1/i + \frac{N+1-i}{N}}{\sum_{j=1}^N (1/j + \frac{N+1-j}{N})}, \text{ i and j as above.} \tag{4}$$

Of course, other combinations of these would be thinkable, but the important observation is achieved by comparing SR with the others. The actual mix between the methods would affect the result according to its proportions. The details there are not crucial for our main point that all these results are a sensitive product of the underlying assumptions regarding the mind-settings of decision-makers. If this nevertheless would be important, the reasonable proportions must be elicited and fine-tuned with respect to the individual in question. A reasonable meaning of such a procedure escapes us and is in any case beyond the scope of this article, so the tables below show the equally proportional SR weights and its behaviour in relation to ROC and RS. Table 6 shows the weights w_i^{SR} for different numbers of criteria up to $N = 10$.

Table 6. The weights for SR

ROC-RS	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$
$N = 1$	0.0000									
$N = 2$	0.0833	-0.0833								
$N = 3$	0.1111	-0.0556	-0.0556							
$N = 4$	0.1208	-0.0292	-0.0542	-0.0375						
$N = 5$	0.1233	-0.0100	-0.0433	-0.0433	-0.0267					
$N = 6$	0.1226	0.0036	-0.0321	-0.0401	-0.0341	-0.0198				
$N = 7$	0.1204	0.0133	-0.0224	-0.0344	-0.0344	-0.0272	-0.0153			
$N = 8$	0.1175	0.0203	-0.0144	-0.0283	-0.0318	-0.0290	-0.0221	-0.0122		
$N = 9$	0.1143	0.0254	-0.0079	-0.0227	-0.0283	-0.0283	-0.0246	-0.0182	-0.0099	
$N = 10$	0.1111	0.0293	-0.0026	-0.0177	-0.0245	-0.0263	-0.0248	-0.0209	-0.0153	-0.0082

Table 7 demonstrates how the weights for this SR combination compare to ROC weights (RR is similar to ROC in this respect). Similarly, Table 8 demonstrates how the weights for this SR combination compare to RS weights. From Tables 7 and 8 we can, as expected, see that SR does indeed constitute a compromise that tries to compensate for shortcomings in the other methods but does not deviate too much from any of them.

Table 7. SR compared to ROC weights

SR-ROC	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$
$N = 1$	0.0000									
$N = 2$	-0.0833	0.0833								
$N = 3$	-0.0894	0.0266	0.0628							
$N = 4$	-0.0845	0.0019	0.0360	0.0466						
$N = 5$	-0.0781	-0.0106	0.0200	0.0330	0.0357					
$N = 6$	-0.0722	-0.0176	0.0097	0.0233	0.0285	0.0282				
$N = 7$	-0.0670	-0.0217	0.0028	0.0161	0.0226	0.0244	0.0229			
$N = 8$	-0.0626	-0.0242	-0.0021	0.0107	0.0177	0.0207	0.0209	0.0190		
$N = 9$	-0.0589	-0.0258	-0.0057	0.0065	0.0137	0.0174	0.0187	0.0181	0.0160	
$N = 10$	-0.0556	-0.0268	-0.0084	0.0031	0.0103	0.0145	0.0165	0.0168	0.0158	0.0137

Table 8. SR compared to RS weights

SR-RS	<i>i</i> = 1	<i>i</i> = 2	<i>i</i> = 3	<i>i</i> = 4	<i>i</i> = 5	<i>i</i> = 6	<i>i</i> = 7	<i>i</i> = 8	<i>i</i> = 9	<i>i</i> = 10
<i>N</i> = 1	0.0000									
<i>N</i> = 2	0.0000	0.0000								
<i>N</i> = 3	0.0217	-0.0290	0.0072							
<i>N</i> = 4	0.0364	-0.0273	-0.0182	0.0091						
<i>N</i> = 5	0.0452	-0.0206	-0.0233	-0.0103	0.0090					
<i>N</i> = 6	0.0504	-0.0140	-0.0224	-0.0168	-0.0056	0.0084				
<i>N</i> = 7	0.0534	-0.0084	-0.0197	-0.0183	-0.0118	-0.0028	0.0076			
<i>N</i> = 8	0.0549	-0.0039	-0.0166	-0.0177	-0.0141	-0.0083	-0.0011	0.0069		
<i>N</i> = 9	0.0555	-0.0004	-0.0136	-0.0162	-0.0146	-0.0108	-0.0058	-0.0001	0.0062	
<i>N</i> = 10	0.0555	0.0025	-0.0110	-0.0146	-0.0142	-0.0118	-0.0083	-0.0041	0.0005	0.0055

5 The EW, SIMOS and RE Methods

In the 1970s, the quite naïve idea of equal weights (EW) gained some recognition. The hope was that, given that elicitation is a hard problem, equal weights would perform as well as any other set of weights. As a generalization to RS as previously discussed, a rank exponent weight method was introduced by [8]. In the RS formula, introduce the exponent z to yield the rank exponent (RE) weights as

$$w_i^{RE} = \frac{(N + 1 - i)^z}{\sum_{j=1}^N (N + 1 - j)^z} \tag{5}$$

Thus, z mediates between equal weights and the RS weights and for $z = 0$, this becomes equal weights and for $z = 1$ it becomes RS weights. Thus, for $0 < z < 1$ it is the exponential combination of equal and RS weights. Beyond $z = 1$ it becomes a more extreme weighting scheme. This makes the z parameter of RE a bit hard to estimate and potentially less suitable in real-life decisions.

Another type of method is the SIMOS method, which has gained some interest in these contexts. It was proposed in [7] with the purpose of providing decision makers with a simple method, not requiring any former familiarity with decision analytical techniques and can easily express criteria hierarchies introducing some cardinality as well, if needed. It has been applied in a multitude of contexts and seems to have been comparatively well received by real-life decision-makers. The SIMOS method has, however, been criticised for not being robust when the preferences are changed and it has further some contra-intuitive features. [5] suggested a revised version, but introduces the severely complicating factor to correctly estimate a reliable and robust proportional factor z between the most and least important criteria. In this study, the

Simos weights are used ordinally by not using the blank cards or equal weights. The z factor is estimated at $N + 1$, with N being the number of criteria as usual.

6 Assessing Models for Surrogate Weights

Simulation studies have become a kind of de facto standard for comparing multi-criteria weights. The assumption is that there exist a set of ‘true’ weights in the decision-maker’s mind which are inaccessible in its pure form by any elicitation method. The modelling assumptions regarding decision-makers’ mind-sets above are then inherent in the generation of decision problem vectors by a random generator. When following an $N-1$ DoF model, a vector is generated in which the components sum to 100%. This simulation is based on a homogenous N -variate Dirichlet distribution generator. When following an N DoF model, a vector is generated without an initial joint restriction, only keeping components within [0%, 100%] implying N degrees of freedom, where these components subsequently are normalised so that their sum is 100%. We call the $N-1$ DoF model type of generator an $N-1$ -generator and the N DoF model type an N -generator. Depending on how we model the decision-maker’s weight assessments, the results then become very different: ROC weights in N dimensions coincide with the mass point for the vectors of the $N-1$ -generator over the polytope S_w . Similarly, RS weights are very close to the mass point of an N -generator over a polytope. In reality, though, we cannot know whether a specific decision-maker (or decision-makers in general) adhere more to $N-1$ or N DoF representations of their knowledge. Both as individuals and as a group, they might use either or be anywhere in between. A, in a reasonable sense, *robust* rank ordering mechanism must employ a surrogate weight function that at least handles both types of conceptualisation and anything in between.

The simulations were carried out with a varying number of criteria and alternatives. There were four numbers of criteria $N = \{3, 6, 9, 12\}$ and five numbers of alternatives $M = \{3, 6, 9, 12, 15\}$ creating a total of 20 simulation scenarios. Each scenario was run 10 times, each time with 10,000 trials, yielding a total of 2,000,000 decision situations generated. For this simulation, an N -variate joint Dirichlet distribution was employed to generate the random weight vectors for the $N-1$ DoF simulations and a standard round-robin normalised random weight generator for the N DoF simulations. Unscaled value vectors were generated uniformly, and no significant differences were observed with other value distributions.

The results of the simulations are shown in the tables below, where we show a subset of the results with chosen pairs (N, M) . There were three measures of success.³ The first is the hit ratio as in the previous studies (“winner”), the number of times the highest evaluated alternative using a particular method coincides with the true highest alternative. The second is the matching of the three highest ranked alternatives (“podium”), the number of times the three highest evaluated alternatives using a

³ Kendall’s tau was also computed and it does not deviate from the other findings in the tables. Thus, it is not shown in the tables.

particular method all coincide with the true three highest alternatives. This means that the internal order between these three alternatives is uninteresting. A third set generated is the matching of all ranked alternatives (“overall”), the number of times all evaluated alternatives using a particular method coincide with the true ranking of the alternatives. All three sets correlated strongly with each other and the latter two are not shown in this paper. The tables show the winner frequency for the six methods ROC, RE, RR, SIMOS, SR, and EW utilising the simulation methods $N-1$ DoF, N DoF and a 50% combination of $N-1$ DoF and N DoF. All hit ratios in all tables are given in per cent and are mean values of the 10 scenario runs. The standard deviations between sets of 10 runs were around 0.2–0.3 per cent. In Table 9, using an $N-1$ -generator, it can be seen that ROC not surprisingly outperforms the others, when looking at the winner, but with SR close behind. EW, likewise not surprisingly, is performing much worse than all the others.

Table 9. Using an $N-1$ -generator, it can be seen that ROC outperforms the others

N-1 DoF		ROC	RE	RR	Simos	SR	EW
3 criteria	3 alternatives	90.2	88.9	89.5	88.1	89.3	72.9
3 criteria	15 alternatives	79.1	77.2	76.5	76.2	76.9	56.5
6 criteria	6 alternatives	84.8	80.6	82.7	79.6	83.1	57.5
6 criteria	12 alternatives	81.3	76.9	78.2	75.5	78.9	50.0
9 criteria	9 alternatives	83.5	77.6	79.5	76.2	81.2	50.1
12 criteria	6 alternatives	86.4	78.5	80.8	77.1	84.1	54.2
12 criteria	12 alternatives	83.4	74.2	76.8	72.5	80.2	45.7

In Table 10 the frequencies have changed according to expectation since we employ a model with N degrees of freedom. Now the SIMOS model behaves very similar to RE and, as expected, outperforms the others, in particular when it comes to a larger number of criteria and alternatives. In Table 11, the N and $N-1$ DoF models are combined with equal emphasis on both. Now, we can see that in total RE and SR perform the best.

Table 10. For N degrees of freedom, RE and Simos are top of the form

N DoF		ROC	RE	RR	Simos	SR	EW
3 criteria	3 alternatives	87.3	89.3	88.3	89.2	89.1	78.3
3 criteria	15 alternatives	77.9	81.3	79.1	81.4	80.6	65.6
6 criteria	6 alternatives	80.1	87.3	78.1	87.4	85.1	67.4
6 criteria	12 alternatives	76.4	84.2	74.3	84.3	82.0	60.9
9 criteria	9 alternatives	76.3	87.0	69.8	87.2	83.0	62.3
12 criteria	6 alternatives	77.5	90.0	67.8	90.2	84.6	66.4
12 criteria	12 alternatives	73.4	87.4	63.1	87.7	81.7	58.7

Table 11. Winners when considering both DoF models

Combined		ROC	RE	RR	Simos	SR	EW
3 criteria	3 alternatives	88.8	89.1	88.9	88.7	89.2	75.6
3 criteria	15 alternatives	78.5	79.3	77.8	78.8	78.8	61.1
6 criteria	6 alternatives	82.5	84.0	80.4	83.5	84.1	62.5
6 criteria	12 alternatives	78.9	80.6	76.3	79.9	80.5	55.5
9 criteria	9 alternatives	79.9	82.3	74.7	81.7	82.1	56.2
12 criteria	6 alternatives	82.0	84.3	74.3	83.7	84.4	60.3
12 criteria	12 alternatives	78.4	80.8	70.0	80.1	81.0	52.2

Since we are looking for a surrogate model with both precision and robustness, we turn our attention to the spread between the methods' results under the two varying decision-maker assumptions regarding degrees of freedom when producing weights. Table 12 shows the spread as the absolute value of the difference between the frequencies (hit ratios) under the $N-1$ DoF model and the N DoF model.

Table 12. Spread as the absolute value of the difference between the frequencies (hit ratios) under the $N-1$ DoF model and the N DoF model for winners

Spread		ROC	RE	RR	Simos	SR	EW
3 criteria	3 alternatives	2.9	0.4	1.2	1.1	0.2	5.4
3 criteria	15 alternatives	1.2	4.1	2.6	5.2	3.7	9.1
6 criteria	6 alternatives	4.7	6.7	4.6	7.8	2.0	9.9
6 criteria	12 alternatives	4.9	7.3	3.9	8.8	3.1	10.9
9 criteria	9 alternatives	7.2	9.4	9.7	11.0	1.8	12.2
12 criteria	6 alternatives	8.9	11.5	13.0	13.1	0.5	12.2
12 criteria	12 alternatives	10.0	13.2	13.7	15.2	1.5	13.0

7 Concluding Remarks

The aim of this study has been to find reasonably robust multi-criteria weights that are able to cover a broad set of decision situations, but at the same time have a reasonably simple semantic regarding how they are generated. In this paper, we have considered $N = \{3, 6, 9, 12, 15\}$ and $M = \{3, 6, 9, 12\}$ and to summarise the analysis, we look at the average hit rate in per cent ("mean correct") over all the pairs (N, M) that we have reported in the tables above. Table 13 shows the conclusion of the performances. RE heads the table with SR not far behind. Further in Table 13, we can see the mean square spread between the different DoF. RE and SR are the best candidates when it comes to the mean value, followed by Simos. However, the robustness in the sense of mean square variation is significant. In that respect SR is clearly the most superior, rendering it the top ranked position. Since we appreciate both precision and robustness, the final score determining the most suitable surrogate weight method is the difference between the mean value of the "winners" and the mean value of the squared spread.

Table 13. Performance averages and mean square for the six methods

Conclusion	ROC	RE	RR	Simos	SR	EW
Mean correct (Table 11)	81.3	82.9	77.5	82.3	82.8	60.5
Rank from mean correct	4	1	5	3	2	6
Mean square spread (Table 12)	6.4	8.5	8.4	9.9	2.2	10.7
Rank from mean sq spread	2	4	3	5	1	6
Final score	74.9	74.4	69.1	72.4	80.7	49.8
Rank from final score	2	3	5	4	1	6

It is clear that SR is the preferred surrogate method of those investigated. In a further study, it would be interesting to study the RE method with a varying set of exponents to see if primarily the stability can be improved, in that case making it a candidate for real-life decision making tools.

References

1. Barron, F.H.: Selecting a best multiattribute alternative with partial information about attribute weights. *Acta Psychol.* **80**(1–3), 91–103 (1992)
2. Barron, F., Barrett, B.: Decision quality using ranked attribute weights. *Manag. Sci.* **42**(11), 1515–1523 (1996)
3. Danielson, M., Ekenberg, L., Larsson, A.: Distribution of belief in decision trees. *Int. J. Approx. Reason.* **46**(2), 387–407 (2007)
4. Danielson, M., Ekenberg, L.: The CAR method for using preference strength in multi-criteria decision making. *Group Decis. Negot.* **25**, 775–797 (2015)
5. 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)
6. Riabacke, M., Danielson, M., Ekenberg, L.: State-of-the-art in prescriptive weight elicitation. *Adv. Decis. Sci.* Article ID 276584, p. 24 (2012). doi:[10.1155/2012/276584](https://doi.org/10.1155/2012/276584)
7. Simos, J.: L'évaluation environnementale: Un processus cognitif neegociee. These de doctorat, DGF-EPFL, Lausanne (1990)
8. Stillwell, W., Seaver, D., Edwards, W.: A comparison of weight approximation techniques in multiattribute utility decision making. *Organ. Behav. Hum. Perform.* **28**(1), 62–77 (1981)