Computers & Operations Research 157 (2023) 106268

Contents lists available at ScienceDirect

Computers and Operations Research

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

Scoring from pairwise winning indices

Sally Giuseppe Arcidiacono[1], Salvatore Corrente *,[1], Salvatore Greco[1]

Department of Economics and Business, University of Catania, Corso Italia, 55, 95129 Catania, Italy

ARTICLE INFO

Keywords:
Scoring method
Pairwise Winning Indices
Stochastic Multicriteria Acceptability Analysis
Explainability
Simulations

ABSTRACT

The pairwise winning indices, computed in the Stochastic Multicriteria Acceptability Analysis, give the probability with which an alternative is preferred to another. They are computed taking into account all the instances of the assumed preference model compatible with the preference information provided by the Decision Maker mainly, but not exclusively, in terms of pairwise preference comparisons of reference alternatives. In this paper we present a new scoring method assigning a value to each alternative summarizing the results of the pairwise winning indices. Several procedures of this type have been provided in literature. However, our method, expressing the score in terms of a representative additive value function, permits to disaggregate the overall evaluation of each alternative in the sum of contributions of considered criteria. This permits not only to rank the alternatives but also to explain the reasons for which an alternative obtains its evaluation and, consequently, fills a certain ranking position. We also present a probabilistic model underlying our methodology. This probabilistic model is based on a simple piecewise linear approximation of the cumulative normal distribution, which allows the use of linear programming. To prove the efficiency of the method in representing the preferences of the Decision Maker, we performed an extensive set of simulations varying the number of alternatives and criteria. The results of the simulations, analyzed from a statistical point of view, prove the reliability of our procedure. The applicability of the method to decision making problems is explained by means of a case study related to the evaluation of financial funds.

1. Introduction

Given a set of alternatives $A = \{a, b, \dots\}$ evaluated on a coherent family of criteria (Roy, 1996) $G = \{g_1, \dots, g_m\}$, choice, ranking and sorting are the typical decision problems handled through Multiple Criteria Decision Making (MCDM) methods (Greco et al., 2016; Keeney and Raiffa, 1976). In this paper we are interested in ranking decision problems in which alternatives have to be ordered from the best to the worst taking into account the preferences of the Decision Maker (DM).[2]

Several methods aiming to produce such a ranking have been proposed in literature and they mainly differ (i) with respect to the form the DM's preference information are articulated and, (ii) in the procedures used to get the final ranking. With respect to the first point we distinguish between direct and indirect preference information. In case of direct preference information, the DM is asked to provide directly values of the parameters involved in the decision model used to produce the ranking, while, in case of indirect preference information, the DM is asked to provide some information in terms of preference comparison between alternatives or comparison of criteria with respect to their importance; this information is then used to infer parameters of the assumed preference model so that the application of the model with the inferred parameters restores the preferences expressed by the DM. To figure out the difference between direct and indirect preference information, consider a decision problem in which there are two criteria g_1 and g_2 providing evaluations in the interval [0,1]. Let us also assume that we want to represent the DM's preference assigning to each alternative a an overall evaluation $U(a)$ in terms of an overall value function formulated as a weighted sum, that is, $U(a) = w_1 \cdot g_1(a) + w_2 \cdot g_2(a)$, with $w_1, w_2 \geqslant 0$ and $w_1 + w_2 = 1$. In case of a direct preference information, the DM is asked to provide the values of w_1 and w_2. For example, the DM could say that $w_1 = 0.2, w_2 = 0.8$. In case of an indirect preference, the DM could be asked to say if between two alternatives a and b, with $g_1(a) = 0.5, g_2(a) = 0.6, g_1(b) = 0.4$ and $g_2(b) = 0.9$, he/she prefers one of the two or if they are indifferent. If

* Corresponding author.
E-mail addresses: s.arcidiacono@unict.it (S.G. Arcidiacono), salvatore.corrente@unict.it (S. Corrente), salgreco@unict.it (S. Greco).

[1] All authors have worked equally for all the parts of the paper.

[2] We agree that different actors can participate in a decision aiding process, such as decision maker, analyst, experts, stakeholders and so on. This distinction is for sure important in a fruitful development of the decision process (see e.g. chapter 2 in Roy (1996)). However, in this paper, we consider a specific methodology and for the sake of clarity we believe that it is better to abstract from the different types of actors participating to the decision process. In this perspective we prefer to consider the simplest case in which there is a single decision maker.

https://doi.org/10.1016/j.cor.2023.106268
Received 25 March 2022; Received in revised form 26 April 2023; Accepted 28 April 2023
Available online 3 May 2023
0305-0548/

the DM is answering that the two alternatives are indifferent, we would conclude that

$$U(a) = U(b) \Leftrightarrow w_1 \cdot g_1(a) + w_2 \cdot g_2(a)$$
$$= w_1 \cdot g_2(b) + w_2 \cdot g_2(b) \Leftrightarrow w_1 \cdot 0.5 + w_2 \cdot 0.6 = w_1 \cdot 0.4 + w_2 \cdot 0.9$$

from which $0.1 \cdot w_1 = 0.3 \cdot w_2$ and, consequently, $w_1 = 0.75$ and $w_2 = 0.25$.

If, instead, the DM says that a is preferred to b, then we have to conclude that

$$U(a) > U(b) \Leftrightarrow w_1 \cdot g_1(a) + w_2 \cdot g_2(a)$$
$$> w_1 \cdot g_2(b) + w_2 \cdot g_2(b) \Leftrightarrow w_1 \cdot 0.5 + w_2 \cdot 0.6 > w_1 \cdot 0.4 + w_2 \cdot 0.9$$

from which we get $0.1 \cdot w_1 > 0.3 \cdot w_2 \Leftrightarrow 0.1 \cdot w_1 > 0.3 \cdot (1 - w_1) \Leftrightarrow 0.4 \cdot w_1 > 0.3$ and, consequently, $w_1 > 0.75$ and $w_2 < 0.25$. Conversely, if the DM says that b is preferred to a, we would get that $w_1 < 0.75$ and $w_2 > 0.25$.

In general, the indirect preference information should be favorite because of the lower cognitive effort asked to the DM in providing it. The approach applying this way of asking preferences to the DM is known as preference disaggregation or ordinal regression (Jacquet-Lagreze and Siskos, 1982, 2001).

As explained above, in preference disaggregation one aims to infer an instance of the assumed preference model compatible with the preferences provided by the DM. However, many of these instances can exist. Even if all of them give the same recommendation on the comparison of alternatives on which the DM expressed his/her preferences, they can provide different recommendations comparing other alternatives. For such a reason, giving a recommendation using only one of these instances can be considered arbitrary to some extent. Robust Ordinal Regression (ROR) (Corrente et al., 2013; Greco et al., 2008) and Stochastic Multicriteria Acceptability Analysis (SMAA) (Lahdelma et al., 1998; Pelissari et al., 2020) are two families of MCDM methods aiming to provide robust recommendations on the problem at hand. Indeed, they take into account all the instances of the preference model compatible with the preference information given by the DM. On the one hand, ROR builds necessary and possible preference relations for which a is necessarily preferred to b iff a is at least as good as b for all compatible models, while a is possibly preferred to b iff a is at least as good as b for at least one compatible model. On the other hand, SMAA gives recommendations in statistical terms producing mainly two indices: (i) the Rank Acceptability Index, $b^k(a)$, being the probability with which an alternative a fills the position k in the final ranking, (ii) the Pairwise Winning Index (PWI), $p(a,b)$, being the probability with which a is at least as good as b. In this paper we are interested in the SMAA methodology and, in particular, in the aggregation of the PWIs to get a complete ranking of the alternatives under consideration. Several methods have been proposed in literature to deal with this problem. In the following, without any ambition to be exhaustive, we shall review some of them. The first methods appear in social choice theory (Arrow et al., 2010). For example, Dodgson (1876) proposes a ranking method such that best ranked alternatives are the closest to be Condorcet winner (Condorcet, 1785). This is the alternative $a \in A$ (if it exists) such that, with respect to any other alternative $b \in A$ there is a majority of compatible value functions for which a is at least as good as b (that is, the alternatives $a \in A$ for which $p(a,b) > 0.5$ for all $b \in A$). Another very well-known ranking algorithm relating social choice with probability values $p(a,b)$ is the Simpson procedure (Simpson, 1969). It ranks alternatives in A assigning to each $a \in A$ a score being its minimal $p(a,b)$ over all b in A. In the perspective of applying some social choice ranking procedure to PWIs, Kadziński and Michalski (2016) present nine methods summarizing the PWIs by sum, min and max operators. In a more computational oriented approach, Vetschera (2017) presents different ranking methods based on the solution of specific Mixed Integer Linear Programming (MILP) problems.

Even if, as observed above, several methods have been proposed in the past to build a complete ranking of the alternatives summarizing the information provided by the PWIs, none of them provides any easily understandable explanation of this ranking. We then aim to fill this gap proposing a method that, on the basis of the PWIs supplied by the SMAA methodology gives a ranking of the considered alternatives and explains it through an additive value function. In this way, differently from the other methods proposed up to now in literature to obtain a final ranking from the PWIs of SMAA, the DM gets also a scoring for the considered alternatives. Moreover, using the value function provided by our procedure, he/she can investigate on the reasons for which an alternative gets a certain ranking position. Indeed, since the ranking is provided by means of an additive value function, the DM can look at the contribution given by each criterion to the global value permitting to identify criteria being weak and strong points for the considered alternatives. From a methodological point of view, as it will be clear later, the computation of the additive value function involves only to solve an LP problem.

Let us also observe that, in line with what has been proposed for ROR in Kadziński et al. (2012b), the value function provided by our method can be seen as representative of the many value functions compatible with the preferences expressed by the DM. Therefore, our approach can be interpreted also as a procedure to construct a representative value function for SMAA methodology. We also present a probabilistic model permitting to interpret the scores as mean values of probabilistic distributions of the overall evaluations that can be assigned to the considered alternatives. In particular, we assume that the probability distributions of alternative evaluations have independent normal distributions with a common standard deviation σ. To handle these probability distributions through linear programming, we introduce a simple piecewise linear approximation of the cumulative normal distribution, which we believe has an independent interest that goes beyond the specific method we propose.

From a preference learning perspective (Fürnkranz and Hüllermeier, 2010), we proved the efficiency of our method not only in explaining the preferences of the DM but also in predicting his/her comprehensive preferences starting from some available information. To do this, we performed an extensive set of simulations considering different numbers of alternatives (n) and criteria (m). For each (n, m) configuration, we assumed the existence of an artificial DM. Its preferences, obtained through a value function that is unknown to the method we use to produce the representative value function, have to be discovered on the basis of some pairwise comparisons of alternatives it provided. To evaluate the performances of our method with respect to this objective, we compute the Kendall-Tau correlation coefficient (Kendall, 1938) between the ranking of the alternatives produced by the artificial DM and the one given by our procedure. We compared our method to other sixteen methods known in literature aiming to summarize the information contained in the PWIs. The results prove that there is not a method being the best for each (n, m) configuration and the difference in terms of Kendall-Tau between the values obtained by our proposal and those obtained by the best method in each configuration is not statistically significant with respect to a Kolmogorov–Smirnov test with 5% significance level (Massey, 1951). This proves that our method is good not only for its capacity to *explain* the preferences of the DM but also to *learn* preferences with reliable results.

Finally, to show how the method can support decision making in a real world problem, we applied it to a financial context in which seven funds have to be overall evaluated taking into account their performances on five criteria.

The paper is structured as follows. In the next section, we present the new method as well as some extensions; a comprehensive comparison between our proposal and other methods presented in literature to deal with the same problem is performed in Section 3, while the method is applied to a real world financial problem in Section 4; finally, some conclusions and further directions of research are given in Section 5.

2. Giving a score to the alternatives on the basis of the Pairwise Winning Indices

2.1. Methodological background

In the following, we shall briefly recall the main concepts of the preference disaggregation and SMAA.

2.1.1. Additive value functions and indirect preference information

Without loss of generality, let us assume that all criteria are of the gain type (the greater the evaluation of an alternative $a \in A$ on criterion $g_j \in G$, the more a is preferred on g_j). $g_j(a)$ denotes the evaluation of a on g_j. Let us also suppose that the model assumed to represent the preferences of the DM is an additive value function $U : A \rightarrow [0,1]$ of the following type

$$U(a) = U\left(g_1(a), \ldots, g_m(a)\right) = \sum_{j=1}^{m} u_j\left(g_j(a)\right). \tag{1}$$

In Eq. (1), $u_j : X_j \rightarrow [0,1]$, for all $g_j \in G$, is the marginal value function related to g_j, while $X_j = \left\{x_j^0, \ldots, x_j^{n_j}\right\}$ is the set of evaluations taken by alternatives in A on g_j such that $x_j^k < x_j^{k+1}$ for all $k = 0, \ldots, n_j - 1$. Moreover, u_j is non-decreasing in X_j for all $g_j \in G$.

To build an additive value function, the marginal value functions should be defined. Under the preference disaggregation approach, this is done by taking into account some preference information provided by the DM in terms of comparisons between alternatives of the following type:

- a is preferred to b (denoted by $a \succ_{DM} b$) translated into the constraint $U(a) > U(b)$,
- a is at least as good as b ($a \succsim_{DM} b$) translated into the constraint $U(a) \geqslant U(b)$,
- a is indifferent to b ($a \sim_{DM} b$) translated into the constraint $U(a) = U(b)$.[3]

2.1.2. Checking for the existence of a compatible value function

A *compatible value function* is an additive value function of type (1) compatible with the preferences provided by the DM and, therefore, satisfying the following set of constraints:

$$\left.\begin{array}{l}
U(a) = \sum_{j=1}^{m} u_j\left(x_j^k\right), \text{ where } x_j^k \in X_j : x_j^k = g_j(a),\ \forall g_j \in G, \forall a \in A,\\
U(a) > U(b), \text{ iff } a \succ_{DM} b,\\
U(a) \geqslant U(b), \text{ iff } a \succsim_{DM} b,\\
U(a) = U(b), \text{ if } a \sim_{DM} b,\\
u_j\left(x_j^{k+1}\right) \geqslant u_j\left(x_j^k\right),\ \forall g_j \in G \text{ and } k = 0, \ldots, n_j - 1,\\
u_j\left(x_j^0\right) = 0,\ \forall g_j \in G,\\
\sum_{j=1}^{m} u_j\left(x_j^{n_j}\right) = 1.
\end{array}\right\} E^{DM}$$

To check for the existence of at least one compatible value function one needs to solve the following LP problem in the unknown variables $u_j(x_j^k), g_j \in G, k = 1, \ldots, n_j$, and ε,

$$\varepsilon^* = \max \varepsilon \text{ subject to}$$

$$\left.\begin{array}{l}
U(a) = \sum_{j=1}^{m} u_j\left(x_j^k\right), \text{ where } x_j^k \in X_j : x_j^k = g_j(a),\ \forall g_j \in G, \forall a \in A,\\
U(a) \geqslant U(b) + \varepsilon, \text{ iff } a \succ_{DM} b,\\
U(a) \geqslant U(b), \text{ iff } a \succsim_{DM} b,\\
U(a) = U(b), \text{ if } a \sim_{DM} b,\\
u_j\left(x_j^{k+1}\right) \geqslant u_j\left(x_j^k\right),\ \forall g_j \in G \text{ and } k = 0, \ldots, n_j - 1,\\
u_j\left(x_j^0\right) = 0,\ \forall g_j \in G,\\
\sum_{j=1}^{m} u_j\left(x_j^{n_j}\right) = 1
\end{array}\right\} E^{DM'}$$

where ε is an auxiliary variable used to convert the strict inequalities ($U(a) > U(b)$) in weak ones ($U(a) \geqslant U(b) + \varepsilon$). If $E^{DM'}$ is feasible and $\varepsilon^* > 0$, then, there exists at least one compatible value function. We shall call such a function the *Most Discriminant Value Function*. In the opposite case ($E^{DM'}$ is infeasible or $\varepsilon^* \leqslant 0$), then, it does not exist any compatible value function and the causes can be investigated by means of the methods presented in Mousseau et al. (2003).

2.1.3. The SMAA methodology

Let us assume that $E^{DM'}$ is feasible and $\varepsilon^* > 0$. Therefore, there exists at least one compatible value function. In this case, in general, as already observed in the introduction, infinitely many compatible value functions exist. The SMAA methodology aims to give a recommendation on the problem at hand by taking into account a well distributed sampling of them. Since the constraints in $E^{DM'}$ define a convex set of parameters, some compatible value functions can be sampled by using, for example, the Hit-And-Run (HAR) algorithm (Smith, 1984; Tervonen et al., 2013; Van Valkenhoef et al., 2014). Let us denote by $\mathcal{U}$ the set of sampled compatible value functions. Each value function in $\mathcal{U}$ will give a certain recommendation on each pair of alternatives $(a,b) \in A \times A$. For this reason, for each $(a,b) \in A \times A$, two different subsets of $\mathcal{U}$ can be defined:

$$\mathcal{U}_{a>b} = \{U \in \mathcal{U} : U(a) > U(b)\}; \quad \mathcal{U}_{a\sim b} = \{U \in \mathcal{U} : U(a) = U(b)\}. \tag{2}$$

For each $(a,b) \in A \times A$, the PWI of the pair (a,b), denoted by $p(a,b)$, can then be computed in the following way:

$$p(a,b) = \frac{|\mathcal{U}_{a>b}| + \frac{1}{2}|\mathcal{U}_{a\sim b}|}{|\mathcal{U}|}. \tag{3}$$

Let us observe that other definitions of the PWI have been provided in literature. For example, in Kadziński and Michalski (2016), $p(a,b) = \frac{|\mathcal{U}_{a\succsim b}|}{|\mathcal{U}|}$ where $\mathcal{U}_{a\succsim b} = \{U \in \mathcal{U} : U(a) \geqslant U(b)\}$ and, therefore, $\left|\mathcal{U}_{a\succsim b}\right| = \left|\mathcal{U}_{a>b}\right| + \left|\mathcal{U}_{a\sim b}\right|$, while in Vetschera (2017) $p(a,b) = \frac{|\mathcal{U}_{a>b}|}{|\mathcal{U}|}$. In the following, without loss of generality, we shall consider the PWI defined in Eq. (3).

2.2. Proposed methodology

The idea under the construction of a compatible value function able to represent in the best way the PWIs is the following:

> "*Given $a,b \in A$, if $p(a,b) \geqslant 0.5$, that is, for at least 50% of the sampled compatible value functions a is at least as good as b, then, the greater $p(a,b)$, the larger should be the difference between the utilities of a and b*".

Given in different terms, we aim to build a compatible value function U such that the difference $U(a) - U(b)$ increases with $p(a,b)$. Formally, this requirement is translated into the following constraint:

$$\text{if } p(a,b) \geqslant 0.5, \text{ then } U(a) - U(b) \geqslant \eta\left(p(a,b) - 0.5\right) \tag{4}$$

[3] The indifference between alternatives a and b can also be translated into a different constraint using an auxiliary threshold as described in Branke et al. (2017). However, the way the indifference relation is translated does not affect the following description and, for this reason, we shall assume that $a \sim_{DM} b$ iff $U(a) = U(b)$.

where $\eta \in \mathbb{R}^+$.

Let us observe that if $\eta > 0$, the constraint (4) perfectly represents the eventual preferences provided by the DM on some pairs of alternatives. Indeed,

- if $a \succsim_{DM} b$, then $p(a,b) \geqslant 0.5$ and, therefore, $U(a) - U(b) \geqslant \eta \cdot (p(a,b) - 0.5) \geqslant 0$, from which it follows that $U(a) \geqslant U(b)$,
- if $a \succ_{DM} b$, then $p(a,b) = 1$ and, therefore, $U(a) - U(b) \geqslant 0.5 \cdot \eta > 0$ from which it follows that $U(a) > U(b)$,
- if $a \sim_{DM} b$, then $p(a,b) = 0.5 = p(b,a)$ and, therefore, on the one hand, $U(a) - U(b) \geqslant 0$ and, on the other hand, $U(b) - U(a) \geqslant 0$, so that we get $U(a) = U(b)$.

2.2.1. Checking for a compatible scoring function

To check for an additive value function having all the characteristics mentioned above (we shall call it *compatible scoring function*), the following LP problem, denoted by LP_0, should be solved:

$$\eta^* = \max \eta, \text{ subject to,}$$

$$\left.\begin{array}{l} U(a) = \sum_{j=1}^{m} u_j\left(x_j^k\right), \text{ where } x_j^k \in X_j : x_j^k = g_j(a),\ \forall g_j \in G, \forall a \in A, \\ U(a) - U(b) \geqslant \eta \cdot (p(a,b) - 0.5),\ \forall (a,b) \in A \times A : p(a,b) \geqslant 0.5, \\ u_j\left(x_j^{k+1}\right) \geqslant u_j\left(x_j^k\right),\ \forall g_j \in G \text{ and } k = 0, \dots, n_j - 1, \\ u_j\left(x_j^0\right) = 0,\ \forall g_j \in G, \\ \sum_{j=1}^{m} u_j\left(x_j^{n_j}\right) = 1. \end{array}\right\} E^{SF}$$

If E^{SF} is feasible and $\eta^* > 0$, then there is at least one compatible scoring function. In the opposite case (E^{SF} is infeasible or $\eta^* \leqslant 0$), then there is not any compatible value function satisfying all the constraints in E^{SF} with a positive value of η. The causes of this infeasibility could be detected by using one of the approaches proposed in Mousseau et al. (2003). In the first case (E^{SF} is feasible and $\eta^* > 0$), the compatible scoring function obtained by solving LP_0 can then be used to assign a value to each alternative in A producing, therefore, a complete ranking of the alternatives at hand. Observe that the value assigned to the alternatives can be considered a score obtained taking into account the whole set of compatible value functions that have been sampled to compute the PWIs. In this perspective the obtained compatible scoring function can be seen as a representative value function (Kadziński et al., 2012a) built on the basis of the principle "one for all, all for one". Indeed, on the one hand, all compatible value functions concur to define the value function U corresponding to η^*, while, on the other hand, U represents the whole set of the compatible value functions.

Let us observe that even if E^{SF} is feasible but $\eta^* \leqslant 0$, the compatible scoring function obtained solving LP_0 can be used to assign a value to each alternative and, then, ranking completely all alternatives under consideration. In this case, there will be at least one pair of alternatives $(a,b) \in A \times A$ for which $p(a,b) \geqslant 0.5$ and nevertheless $U(a) < U(b)$. However, maximizing η, that is, in this case, minimizing its absolute value since it is negative, the optimization problem gives a value function U which minimizes the deviation from the preferences represented by the PWIs.

2.2.2. A probabilistic model underlying the proposed methodology

In this section we present a probabilistic model supporting the proposed methodology. We assume that each alternative a has an overall value $\widetilde{U}(a)$ being a random variable normally distributed with mean $U(a)$ and standard deviation σ being equal for all alternatives. We assume also that the random variables $\widetilde{U}(a), a \in A$, are independent. Under these hypotheses, for all $a, b \in A$, the random variable $\widetilde{U}(a) - \widetilde{U}(b)$ has a normal distribution with mean $U(a) - U(b)$ and standard deviation $\sqrt{2}\sigma$. Therefore, considering the cumulative normal distribution, the probability that a is preferred to b is given by

$$P(a,b) = Prob(\widetilde{U}(a) - \widetilde{U}(b) > 0) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{U(a)-U(b)}{\sqrt{2}\sigma}}^{+\infty} e^{-t^2/2}\, dt$$

and, for the symmetry of the normal distribution,

$$P(a,b) = Prob(\widetilde{U}(a) - \widetilde{U}(b) > 0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{U(a)-U(b)}{\sqrt{2}\sigma}} e^{-t^2/2}\, dt.$$

Of course, this formulation of the probability of preference of a over b cannot be treated using linear programming. We then reformulate the probability $P(a,b)$ approximating for $x \geqslant 0$ the cumulative normal distribution $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$ with the function $\hat{\Phi}(x) = min\left\{\frac{1}{2} + \lambda x, 1\right\}$.

Considering the values taken by $\Phi(x)$ and $\hat{\Phi}(x)$ for $x = 0.5, 0.501, 0.502, \dots, 1$, we computed the value of the parameter λ minimizing the maximum error $|\Phi(x) - \hat{\Phi}(x)|$. This value is $\hat{\lambda} = 0.301$. In correspondence of $\lambda = \hat{\lambda}$, the maximum error, the average absolute error and the average quadratic error take a value of 0.048, 0.030 and 0.033, respectively. In simple words, we can say that taking $\lambda = 0.301$, approximating $\Phi(x)$ by $\hat{\Phi}(x)$ we get an average error of around 3% with a maximum error not greater than 5%. Considering the natural imprecision and variability of human preferences, this error seems acceptable for the type of quantity represented by the probability $P(a,b)$.

Approximating $P(a,b) = \Phi\left(\frac{U(a)-U(b)}{\sqrt{2}\sigma}\right)$ with $\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(a)-U(b)}{\sqrt{2}\sigma}\right)$ and considering $\lambda = \hat{\lambda}$, we have that

$$\hat{P}(a,b) = min\left\{\frac{1}{2} + \hat{\lambda}\frac{U(a) - U(b)}{\sqrt{2}\sigma}, 1\right\}.$$

The following result holds.

Proposition 2.1. *For $U(a) \geqslant U(b)$,*

$$\hat{P}(a,b) \geqslant p(a,b) \Leftrightarrow \frac{1}{2} + \hat{\lambda}\frac{U(a) - U(b)}{\sqrt{2}\sigma} \geqslant p(a,b).$$

Proof. First, let us prove that $\hat{P}(a,b) \geqslant p(a,b)$ implies $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant p(a,b)$. Two cases are possible:

- $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} < 1$: in this case, $\hat{P}(a,b) = \frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma}$, so that $\hat{P}(a,b) \geqslant p(a,b)$ implies $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant p(a,b)$;
- $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant 1$: in this case, $\hat{P}(a,b) = 1 \geqslant p(a,b)$, and, then, $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant p(a,b)$.

Let us prove now that $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant p(a,b)$ implies $\hat{P}(a,b) \geqslant p(a,b)$. We have two cases:

- if $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} < 1$, then $\hat{P}(a,b) = \frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma}$, so that, $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant p(a,b)$ implies $\hat{P}(a,b) \geqslant p(a,b)$;
- if $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant 1$, then $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant \hat{P}(a,b) = 1 \geqslant p(a,b)$. □

Observing that

$$\frac{1}{2} + \hat{\lambda}\frac{U(a) - U(b)}{\sqrt{2}\sigma} \geqslant p(a,b) \Leftrightarrow U(a) - U(b) \geqslant \frac{\sqrt{2}\sigma}{\hat{\lambda}}(p(a,b) - 0.5),$$

in case $\eta^* \geqslant 0$, we can reformulate problem LP_0 as follows

$$\left.\begin{array}{l}\sigma^* = \max \sigma, \text{ subject to,}\\ U(a) = \sum_{j=1}^{m} u_j\left(x_j^k\right), \text{ where } x_j^k \in X_j : x_j^k = g_j(a), \ \forall g_j \in G, \forall a \in A,\\ U(a) - U(b) \geqslant \eta \cdot (p(a,b) - 0.5), \ \forall (a,b) \in A \times A : p(a,b) \geqslant 0.5,\\ u_j\left(x_j^{k+1}\right) \geqslant u_j\left(x_j^k\right), \ \forall g_j \in G \text{ and } k = 0, \dots, n_j - 1,\\ u_j\left(x_j^0\right) = 0, \ \forall g_j \in G,\\ \eta = \frac{\sqrt{2}\sigma}{\hat{\lambda}},\\ \sum_{j=1}^{m} u_j\left(x_j^{n_j}\right) = 1. \end{array}\right\}$$

Taking into consideration above Proposition 2.1, the constraint $\eta = \frac{\sqrt{2}\sigma}{\hat{\lambda}}$ permits to reformulate the constraints

$$U(a) - U(b) \geqslant \eta \cdot (p(a,b) - 0.5), \ \forall (a,b) \in A \times A : p(a,b) \geqslant 0.5$$

as

$$\hat{P}(a,b) \geqslant p(a,b), \ \forall (a,b) \in A \times A : p(a,b) \geqslant 0.5.$$

In view of this, the problem LP_0 gives the compatible value function U maximizing the standard deviation σ for which, for all a, b such that $p(a,b) \geqslant 0.5$, the estimated probability $\hat{P}(a,b)$ is not smaller than the observed probability $p(a,b)$. Remembering that $\hat{P}(a,b)$ is a non-increasing function of σ, maximizing σ under the condition that $\hat{P}(a,b) \geqslant p(a,b)$ amounts to minimize the maximal absolute error $|\hat{P}(a,b) - p(a,b)|$ for all a, b for which $p(a,b) \geqslant 0.5$, that is to find the minimal ε such that for all a, b for which $p(a,b) \geqslant 0.5$, $|\hat{P}(a,b) - p(a,b)| \leqslant \varepsilon$. In simple words, maximizing σ permits to identify the compatible value function U related to the probability values $\hat{P}(a,b)$ that best approximates the pairwise winning indices $p(a,b)$.

If the solution of problem LP_0 gives $\eta^* < 0$, of course the constraint $\eta = \frac{\sqrt{2}\sigma}{\hat{\lambda}}$ is no more acceptable, because we would have $\sigma < 0$ which is absurd. In this case, we can replace the constraint $\eta = \frac{\sqrt{2}\sigma}{\hat{\lambda}}$ with the constraint $-\eta = \frac{\sqrt{2}\sigma}{\hat{\lambda}}$. Moreover, we can approximate the cumulative normal distribution $\Phi(x)$ for $x \leqslant 0$ with the function $\hat{\Phi}(x) = max\left\{\frac{1}{2} + \lambda x, 0\right\}$. This minimizes the maximum error with respect to $\Phi(x)$ again for $\lambda = \hat{\lambda} = 0.301$ with the same maximum error, average absolute error and average quadratic error taken by $\hat{\Phi}(x)$ for $x \geqslant 0$. Using $\hat{\Phi}(x)$ for $x \leqslant 0$, we can approximate $P(a,b) = \Phi\left(\frac{U(a)-U(b)}{\sqrt{2}\sigma}\right)$ with $\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(a)-U(b)}{\sqrt{2}\sigma}\right)$ also in case $U(a) - U(b) \leqslant 0$. Let us observe that putting together the definition of $\hat{\Phi}(x)$ for $x \leqslant 0$ and $x \geqslant 0$, we get

$$\hat{\Phi}(x) = \begin{cases} 0 & \text{if} \quad x \leqslant -\frac{1}{2\hat{\lambda}},\\ \frac{1}{2} + \hat{\lambda}x & \text{if} \quad -\frac{1}{2\hat{\lambda}} < x < \frac{1}{2\hat{\lambda}},\\ 1 & \text{if} \quad x \geqslant \frac{1}{2\hat{\lambda}} \end{cases} \tag{5}$$

which corresponds to a uniform distribution having the probability density function

$$f(x) = \begin{cases} \hat{\lambda} & \text{if } -\frac{1}{2\hat{\lambda}} \leqslant x \leqslant \frac{1}{2\hat{\lambda}}\\ 0 & \text{otherwise.} \end{cases}$$

Fig. 1 compares the cumulative normal distribution $\Phi(x)$ and its approximation $\hat{\Phi}(x)$ (see Eq. (5)) by their graphs.

Taking into account the constraint $-\eta = \frac{\sqrt{2}\sigma}{\hat{\lambda}}$ and that $p(a,b) + p(b,a) = 1$, from the constraint

$$U(a) - U(b) \geqslant \eta \cdot (p(a,b) - 0.5), \ \forall (a,b) \in A \times A : p(a,b) \geqslant 0.5$$

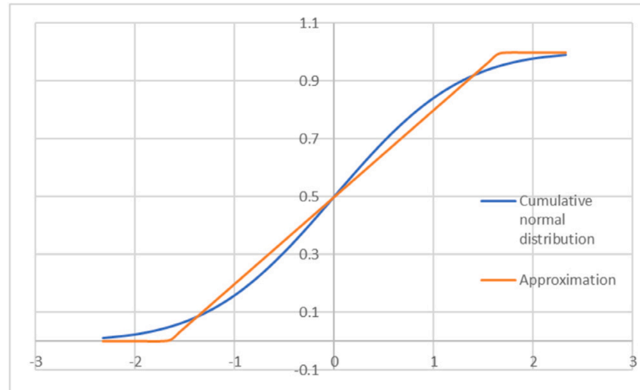


Fig. 1. Approximation of the cumulative normal distribution by means of the function defined by Eq. (5).

we get

$$\frac{1}{2} + \hat{\lambda}\frac{U(a) - U(b)}{\sqrt{2}\sigma} \geqslant p(b,a), \ \forall (a,b) \in A \times A : p(a,b) \geqslant 0.5.$$

Proposition 2.2. *For all $a, b \in A$,*

$$\hat{P}(a,b) + \hat{P}(b,a) = 1.$$

Proof. Three cases are possible:

- $\frac{U(a)-U(b)}{\sqrt{2}\sigma} \leqslant -\frac{1}{2\hat{\lambda}}$: in this case $\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(a)-U(b)}{\sqrt{2}\sigma}\right) = 0$ and, as $\frac{U(b)-U(a)}{\sqrt{2}\sigma} \geqslant \frac{1}{2\hat{\lambda}}$, $\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(b)-U(a)}{\sqrt{2}\sigma}\right) = 1$, so that $\hat{P}(a,b) + \hat{P}(b,a) = 1$;
- $-\frac{1}{2\hat{\lambda}} < \frac{U(a)-U(b)}{\sqrt{2}\sigma} < \frac{1}{2\hat{\lambda}}$: in this case $\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(a)-U(b)}{\sqrt{2}\sigma}\right) = \frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma}$ and, as also $-\frac{1}{2\hat{\lambda}} < \frac{U(b)-U(a)}{\sqrt{2}\sigma} < \frac{1}{2\hat{\lambda}}$, $\hat{P}(b,a) = \hat{\Phi}\left(\frac{U(b)-U(a)}{\sqrt{2}\sigma}\right) = \frac{1}{2} + \hat{\lambda}\frac{U(b)-U(a)}{\sqrt{2}\sigma}$, so that $\hat{P}(a,b) + \hat{P}(b,a) = 1$;
- $\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant \frac{1}{2\hat{\lambda}}$: in this case $\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(a)-U(b)}{\sqrt{2}\sigma}\right) = 1$ and, as $\frac{U(b)-U(a)}{\sqrt{2}\sigma} \leqslant -\frac{1}{2\hat{\lambda}}$, $\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(b)-U(a)}{\sqrt{2}\sigma}\right) = 0$, so that $\hat{P}(a,b) + \hat{P}(b,a) = 1$. □

Proposition 2.3. *For all $a, b \in A$,*

$$\frac{1}{2} + \hat{\lambda}\frac{U(a) - U(b)}{\sqrt{2}\sigma} \geqslant p(b,a) \Rightarrow [\hat{P}(a,b) \geqslant p(b,a) \textit{ and } \hat{P}(b,a) \leqslant p(a,b)].$$

Proof. Three cases are possible:

- $U(a) \geqslant U(b)$ and $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant 1$: in this case, we have
$$\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(a) - U(b)}{\sqrt{2}\sigma}\right) = 1 \geqslant p(b,a);$$
- $U(a) \geqslant U(b)$ and $\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} < 1$: in this case, we have
$$\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(a) - U(b)}{\sqrt{2}\sigma}\right) = \frac{1}{2} + \hat{\lambda}\frac{U(a) - U(b)}{\sqrt{2}\sigma} \geqslant p(b,a);$$
- $U(a) < U(b)$: in this case, since
$$\frac{1}{2} + \hat{\lambda}\frac{U(a) - U(b)}{\sqrt{2}\sigma} \geqslant p(b,a) \geqslant 0$$

we have

$$\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(a)-U(b)}{\sqrt{2}\sigma}\right) = \frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant p(b,a).$$

Consequently, we get

$$\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant p(b,a) \Rightarrow \hat{P}(a,b) \geqslant p(b,a).$$

Remembering that $\hat{P}(a,b)+\hat{P}(b,a)=1$ (by Proposition 2.2) and $p(a,b)+p(b,a)=1$, we get that $\hat{P}(a,b) \geqslant p(b,a)$ is equivalent to $\hat{P}(b,a) \leqslant p(a,b)$ and, consequently, we obtain also

$$\frac{1}{2} + \hat{\lambda}\frac{U(a)-U(b)}{\sqrt{2}\sigma} \geqslant p(b,a) \Rightarrow \hat{P}(b,a) \leqslant p(a,b). \quad \square$$

Taking into account all the above considerations, in case $\eta^* < 0$, LP_0 problem can be reformulated as follows

$$\sigma^* = \min \sigma, \text{ subject to},$$

$$\left.\begin{array}{l} U(a) = \sum_{j=1}^{m} u_j\left(x_j^k\right), \text{ where } x_j^k \in X_j : x_j^k = g_j(a),\ \forall g_j \in G, \forall a \in A, \\ U(a) - U(b) \geqslant \eta \cdot (p(a,b) - 0.5),\ \forall (a,b) \in A \times A : p(a,b) \geqslant 0.5, \\ u_j\left(x_j^{k+1}\right) \geqslant u_j\left(x_j^k\right),\ \forall g_j \in G \text{ and } k = 0,\ldots,n_j - 1, \\ u_j\left(x_j^0\right) = 0,\ \forall g_j \in G, \\ \eta = -\frac{\sqrt{2}\sigma}{\hat{\lambda}}, \\ \sum_{j=1}^{m} u_j\left(x_j^{n_j}\right) = 1. \end{array}\right\}$$

In particular, the constraints

$$U(a) - U(b) \geqslant \eta \cdot (p(a,b) - 0.5),\ \forall (a,b) \in A \times A :\ p(a,b) \geqslant 0.5$$

can be interpreted as follows: for all $(a,b) \in A \times A$ for which $p(a,b) \geqslant 0.5$

- $\hat{P}(a,b) \geqslant 0.5$ if $U(a) \geqslant U(b)$, because $\hat{P}(a,b) = \hat{\Phi}\left(\frac{U(a)-U(b)}{\sqrt{2}\sigma}\right) \geqslant 0.5$,
- $\hat{P}(a,b) \geqslant p(b,a)$ if $U(a) \leqslant U(b)$, due to Proposition 2.3.

In case $p(a,b) \geqslant 0.5$ but $U(a) \leqslant U(b)$, condition $\hat{P}(a,b) \geqslant p(b,a)$ says that in a certain form the estimated probabilities $\hat{P}(\cdot,\cdot)$ continues to represent the preferences conveyed by the pairwise comparison indices $p(\cdot,\cdot)$. Indeed, even if $\hat{P}(a,b) < 0.5$ while $p(a,b) \geqslant 0.5$, $\hat{P}(a,b) \geqslant p(b,a)$ implies $\hat{P}(b,a) \leqslant p(a,b)$. Consequently, if the DM determines his/her final preferences taking into consideration the estimated probability $\hat{P}(b,a) > 0.5$ (for which it is more probable that b is preferred to a) and the observed probability $p(a,b) > 0.5$ (for which it is more probable that a is preferred to b), $p(a,b) > 0.5$ should prevail over $\hat{P}(b,a) > 0.5$, because $\hat{P}(b,a)$ is not greater than $p(a,b)$.

2.3. Some extensions of the scoring method

Let us assume that E^{SF} is feasible and $\eta^* > 0$. This means that at least one compatible scoring function exists. However, many of them could exist. Let us denote by $\mathcal{U}^{SF}$ the set of additive value functions satisfying all constraints in E^{SF} with $\eta = \eta^*$, that is, all value functions maximally discriminating between the alternatives $(a,b) \in A \times A$ on the basis of the corresponding PWIs. From now on, we shall call such functions *maximally discriminant compatible scoring functions*. As already observed in the introduction, the previous scoring procedures based on PWIs compute a value for each alternative just to obtain a complete ranking of them. However, in some methods, this number has not a particular meaning and, therefore, it can be completely useless for the DM. In our approach, the complete ranking of the alternatives under examination is produced on the basis of the construction of a compatible scoring function. This function, on the one hand, assigns a score to each alternative and, on the other hand, *explains* the reasons for which an alternative has received a particular ranking position. This explanation is provided defining the contribution of each criterion $g_j \in G$ to the overall evaluation through the values taken by the corresponding marginal value function u_j. For this reason, from the explainability point of view (see e.g. Arrieta et al., 2020), it is important checking if among the maximally discriminant compatible scoring functions in $\mathcal{U}^{SF}$, there exists at least one having one of the following characteristics: (i) all criteria $g_j \in G$ give a contribution to the compatible scoring function U; (ii) for each $g_j \in G$ the marginal value function u_j is strictly monotone in X_j.

2.3.1. Maximally discriminant compatible scoring functions with non-null contribution of all criteria

In this case, one is looking for a maximally discriminant compatible scoring function $U \in \mathcal{U}^{SF}$ such that $u_j\left(x_j^{n_j}\right) > 0$ for each $g_j \in G$. Indeed, in an additive value function, $u_j\left(x_j^{n_j}\right)$ can be considered, in some way, as the "importance" of criterion g_j being the marginal value assigned to the greatest performance on g_j (that is $x_j^{n_j}$) by the marginal value function u_j (Jacquet-Lagreze and Siskos, 1982). To this aim, the following LP problem, denoted by LP_1, has to be solved:

$$h^* = \max\ h, \text{ subject to},$$

$$\left.\begin{array}{l} \eta = \eta^*, \\ u_j\left(x_j^{n_j}\right) \geqslant h,\ \ \forall g_j \in G, \\ E^{SF}. \end{array}\right\} E^{SF}_{AllContr}$$

If $E^{SF}_{AllContr}$ is feasible and $h^* > 0$, then, there is at least one maximally discriminant compatible scoring function $U \in \mathcal{U}^{SF}$ in which all criteria contribute to U. In the opposite case, in all maximally discriminant compatible scoring functions, $u_j\left(x_j^{n_j}\right) = 0$ for at least one $g_j \in G$ and, therefore, g_j is not contributing to the global value assigned to the alternatives by U. Let us denote by $\mathcal{U}^{SF}_{AllContr}$ the subset of $\mathcal{U}^{SF}$ composed of all maximally discriminant compatible scoring functions in which all criteria contribute to the global value of each alternative.

2.3.2. Strictly monotone maximally discriminant compatible scoring functions

In this case, one is looking for a maximally discriminant compatible scoring function $U \in \mathcal{U}^{SF}$ where, for each $g_j \in G$, $u_j\left(x_j^k\right) < u_j\left(x_j^{k+1}\right)$ for all $k = 0,\ldots,n_j - 1$. To check for the existence of a maximally discriminant compatible scoring function having the mentioned characteristic, one has to solve the following LP problem denoted by LP_2:

$$\gamma^* = \max\ \gamma, \text{ subject to},$$

$$\left.\begin{array}{l} \eta = \eta^*, \\ u_j\left(x_j^{k+1}\right) \geqslant u_j\left(x_j^k\right) + \gamma,\ \ \forall g_j \in G \text{ and } k = 0,\ldots,n_j - 1, \\ E^{SF}. \end{array}\right\} E^{SF}_{AllInc}$$

If E^{SF}_{AllInc} is feasible and $\gamma^* > 0$, then there exists at least one maximally discriminant compatible scoring function in $\mathcal{U}^{SF}$ for which all marginal value functions are increasing, while, this is not the case if E^{SF}_{AllInc} is infeasible or $\gamma^* \leqslant 0$. Let us denote by $\mathcal{U}^{SF}_{AllInc}$ the subset of $\mathcal{U}^{SF}$ composed of the maximally discriminant compatible scoring functions with $\eta = \eta^*$ such that all marginal value functions are increasing.

Note 2.1. $\mathcal{U}^{SF}_{AllInc} \subseteq \mathcal{U}^{SF}_{AllContr}$ and, therefore, if $\mathcal{U}^{SF}_{AllInc} \neq \emptyset$ then $\mathcal{U}^{SF}_{AllContr} \neq \emptyset$, while the opposite is not true. This means that the existence of a maximally discriminant compatible scoring function in which all marginal functions contribute to the global value of an alternative does not imply the existence of a maximally discriminant compatible scoring function in which all marginal value functions are increasing.

2.3.3. Specific classes of maximally discriminant compatible scoring functions

As already underlined above, the sets $\mathcal{U}^{SF}$, $\mathcal{U}^{SF}_{AllContr}$ or $\mathcal{U}^{SF}_{AllInc}$ could be composed of more than one maximally discriminant compatible scoring function. Each of them assigns a different value to each alternative. Therefore, taking into consideration robustness concerns (Roy, 2010), could be interesting to consider a well-distributed sample of maximally discriminant compatible scoring functions. In the following, we shall describe in detail how to compute such a well-distributed sample of maximally discriminant compatible scoring functions in $\mathcal{U}^{SF}$. The same procedure can be easily adapted to compute a well-distributed sample of maximally discriminant compatible scoring functions in $\mathcal{U}^{SF}_{AllContr}$ or $\mathcal{U}^{SF}_{AllInc}$.

Before going ahead, let us observe that an additive value function U, as the one in Eq. (1), is uniquely defined by the marginal values assigned to the evaluations x_j^k from marginal value functions u_j, that is, $u_j\left(x_j^k\right)$ for all $g_j \in G$ and for all $k = 0,\dots,n_j$. For the sake of simplicity, denoting by u_j^k the values $u_j\left(x_j^k\right)$, an additive value function U can also be represented by the vector $U = \left[u_j^k\right]_{\substack{g_j\in G\\ k=0,\dots,n_j}}$.

Let us assume that $\mathcal{U}^{SF} \neq \emptyset$ and that $U^1 = \left[u_j^{k,1}\right]$ is the maximally discriminant compatible scoring function obtained solving LP_0. Another maximally discriminant compatible scoring function $U \in \mathcal{U}^{SF}$ is different from U^1 iff $u_j^k \neq u_j^{k,1}$ (being equivalent to "$u_j^k > u_j^{k,1}$ or $u_j^k < u_j^{k,1}$") for at least one criterion $g_j \in G$ and for at least one $k \in \{0,\dots,n_j\}$. Therefore, the existence of a maximally discriminant compatible scoring function, "sufficiently" different from U^1, can be checked by solving the following MILP problem that we shall denote by $MILP-1$:

$$\delta_2^* = \max\delta \ \text{ subject to,}$$

$$\left.\begin{array}{l}
\eta = \eta^*,\\
E^{SF},\\
\delta \geqslant \delta_{min},\\
\left.\begin{array}{l}
u_j^k \geqslant u_j^{k,1} + \delta - M y_{j,1}^{k,1}, \ \ k=0,\dots,n_j,\\
u_j^k + \delta \leqslant u_j^{k,1} + M y_{j,2}^{k,1}, \ \ k=0,\dots,n_j,\\
y_{j,1}^{k,1}, y_{j,2}^{k,1} \in \{0,1\},\\
\sum\limits_{j=1}^{m}\sum\limits_{k=0}^{n_j}\left[y_{j,1}^{k,1}+y_{j,2}^{k,1}\right] \leqslant 2\cdot\sum\limits_{g_j\in G}\left(n_j+1\right)-1.
\end{array}\right\} E_1
\end{array}\right\}$$

In the constraints above we have that:

- M is a big positive number;
- δ is the minimal difference between the marginal value attached to x_j^k (where $g_j \in G$ and $k = 0,\dots,n_j$) by the maximally discriminant compatible scoring functions U^1 and U that needs to be maximized to ensure a sufficient difference of U^1 from U. To ensure that this difference is not too low, we fix a lower bound for this variable being δ_{min}.[4] Of course, the choice of δ_{min} will influence the computation of the other maximally discriminant compatible scoring functions obtained in addition to U^1 in the well-diversified sample we are looking for;
- $u_j^k \geqslant u_j^{k,1} + \delta - M y_{j,1}^{k,1}$ translates the constraint $u_j^k > u_j^{k,1}$. In particular, if $y_{j,1}^{k,1}=0$, then the constraint is reduced to $u_j^k \geqslant u_j^{k,1}+\delta$ and, therefore, $u_j^k > u_j^{k,1}$;
- $u_j^k + \delta \leqslant u_j^{k,1} + M y_{j,2}^{k,1}$ translates the constraint $u_j^k < u_j^{k,1}$. In particular, if $y_{j,2}^{k,1}=0$, then the constraint is reduced to $u_j^k+\delta \leqslant u_j^{k,1}$ and, therefore, $u_j^k < u_j^{k,1}$;
- the constraint $\sum\limits_{j=1}^{m}\sum\limits_{k=0}^{n_j}\left[y_{j,1}^{k,1}+y_{j,2}^{k,1}\right] \leqslant 2\cdot\sum\limits_{g_j\in G}\left(n_j+1\right)-1$ is used to impose that at least one binary variable is equal to 0 and, consequently, at least one u_j^k is such that $u_j^k \neq u_j^{k,1}$. Indeed, each maximally discriminant compatible scoring function is defined by n_j+1 values for each $g_j \in G$ and, therefore, by $\sum\limits_{g_j\in G}\left(n_j+1\right)$ values.

Solving $MILP-1$, two cases can occur:

case (1) $MILP-1$ is feasible: there is at least one maximally discriminant compatible scoring function in $\mathcal{U}^{SF}$ different from U^1 and the marginal values u_j^k obtained solving $MILP-1$ define such a function. We shall denote by $U^2 = \left[u_j^{k,2}\right]$ the marginal values defining the new maximally discriminant compatible scoring function;

case (2) $MILP-1$ is infeasible: the maximally discriminant compatible scoring function obtained solving LP_0 is unique. Therefore, considering δ_{min}, the well-diversified sample of maximally discriminant compatible scoring functions we are looking for contains only U^1. Of course, this is true for the fixed value δ_{min} because reducing the value of δ_{min} could bring to the discovery of other maximally discriminant compatible scoring functions sufficiently different from U^1 in the well-diversified sample we are looking for.

In case (1) one can proceed in a iterative way to find a sample of well distributed maximally discriminant compatible scoring functions in $\mathcal{U}^{SF}$ with the considered δ_{min} value. Let us assume that t functions $U^r = \left[u_j^{k,r}\right]$, $r = 1,\dots,t$, in $\mathcal{U}^{SF}$ have already been computed; the $(t+1)$th function $U = \left[u_j^k\right]$ in $\mathcal{U}^{SF}$, if it exists, is obtained solving the following problem denoted by $MILP-t$,

$$\delta_{t+1}^* = \max\delta \ \text{ subject to,}$$

$$\left.\begin{array}{l}
\eta = \eta^*,\\
E^{SF},\\
\delta \geqslant \delta_{min},\\
E_1 \cup E_2 \cup \dots \cup E_t
\end{array}\right\}$$

where $E_r, r = 1,\dots,t$, is the following set of constraints

$$\left.\begin{array}{l}
u_j^k \geqslant u_j^{k,r} + \delta - M y_{j,1}^{k,r},\\
u_j^k + \delta \leqslant u_j^{k,r} + M y_{j,2}^{k,r},\\
y_{j,1}^{k,r}, y_{j,2}^{k,r} \in \{0,1\},\\
\sum\limits_{j=1}^{m}\sum\limits_{k=0}^{n_j}\left[y_{j,1}^{k,r}+y_{j,2}^{k,r}\right] \leqslant 2\cdot\sum\limits_{g_j\in G}\left(n_j+1\right)-1.
\end{array}\right\} E_r$$

Let us observe that constraints in E_r avoid that function U is the same as U^r, $r = 1,\dots,t$.

If $MILP-t$ is infeasible, then, there are only t maximally discriminant compatible scoring functions, that is, $U^1,\dots,U^t$. In the opposite case ($MILP-t$ is feasible), the function $U = \left[u_j^k\right]$ obtained solving the MILP problem is the $(t+1)$th maximally discriminant compatible scoring function, that is, $U = U^{t+1}$.

Let us conclude this section observing that the sample of maximally discriminant compatible scoring functions obtained, can be used, if necessary, as a starting point for a more detailed exploration going beyond

[4] For example, in the case study presented in Section 4, we fix $\delta_{min} = 0.1$.

the value functions in the sample. In fact, each convex combination $\overline{U}$ of maximally discriminant compatible scoring functions $U^1, \ldots, U^t$, that is,

$$\overline{U} = \lambda_1 U^1 + \cdots + \lambda_t U^t, \text{ with } \lambda_r \geqslant 0, \text{ for all } r \in \{1, \ldots, t\}, \text{ and } \sum_{r=1}^{t} \lambda_r = 1,$$

is, in turn, another maximally discriminant compatible scoring function. Indeed, as $U^1, \ldots, U^t$ satisfy constraints from $E^{SF} \cup \{\eta = \eta^*\}$, also their convex combination $\overline{U}$ satisfies the same constraints. This means that starting from the obtained sample of maximally discriminant compatible scoring functions, others can be easily and meaningfully obtained. Let us observe that all of them provide the same alternatives ranking since they have to satisfy the same constraints in E^{SF}. However, this ranking can be explained in different ways since the contribution of each criterion changes from one maximally discriminant compatible scoring function in the sample to another. In this perspective, different value functions supply different interpretations to the score. Of course, among the different explanations provided by the different value functions, there can be someone which is more convincing or more acceptable for the DM. In this respect, the contribution in terms of decision aiding supplied by the plurality of value functions has to be seen in terms of possibility for the DM to select the interpretation that is more convenient for him/her. This is of fundamental importance in a constructive approach of decision aiding in which concepts, models, procedures and results are considered as suitable tools for developing convictions (Roy, 1993).

3. Comparison with other scoring methods based on PWIs

In this section we test our scoring procedure with respect to its capacity to predict preferences on the basis of some comparisons provided by the DM. In this perspective, our procedure will be compared with other sixteen methods. They represent the state of the art in literature to obtain a ranking on the basis of the knowledge of the PWIs represented in the matrix $PWM = [p(a,b)]$. First of all, let us briefly review the sixteen methods with which we shall compare our scoring procedure. The first three methods (Vetschera, 2017) aim to define a complete order (that is a complete, asymmetric and transitive binary relation) that optimally represents the PWIs. In particular the complete order on A is represented by the 0–1 variables $y_{ab}, a, b \in A$, such that if $y_{ab} = 1$, then a is preferred to b, while this not the case if $y_{ab} = 0$. The properties of completeness and asymmetry are ensured by the constraints $y_{a,b} + y_{b,a} = 1,\ a, b \in A, a \neq b$. The transitivity is instead obtained through the constraints $y_{a,b} \geqslant y_{a,c} + y_{c,b} - 1.5$, for all $a, b, c \in A$ with $c \in A \setminus \{a, b\}$. Considering three different goodness indicators, the three following methods $\mathbf{M_1}, \mathbf{M_2}$ and $\mathbf{M_3}$ are then obtained.

$\mathbf{M_1}$:

$$\max \sum_{\substack{(a,b)\in A\times A,\\ a\neq b}} p(a,b) y_{ab}, \text{ subject to}$$

$$\left.\begin{array}{l} y_{ab} + y_{ba} = 1, \\ y_{ab} \geqslant y_{ac} + y_{cb} - 1.5,\ \forall c \in A \setminus \{a,b\}, \\ y_{ab} \in \{0,1\} \end{array}\right\} \forall (a,b) \in A \times A,\ a \neq b$$

$\mathbf{M_2}$:

$$\max \sum_{\substack{(a,b)\in A\times A,\\ a\neq b}} \log(p(a,b))\, y_{ab}, \text{ subject to}$$

$$\left.\begin{array}{l} y_{ab} + y_{ba} = 1, \\ y_{ab} \geqslant y_{ac} + y_{cb} - 1.5,\ \forall c \in A \setminus \{a,b\}, \\ y_{ab} \in \{0,1\} \end{array}\right\} \forall (a,b) \in A \times A,\ a \neq b$$

$\mathbf{M_3}$:

$$\max f_{MM}$$

$$\left.\begin{array}{l} f_{MM} \leqslant p(a,b) + (1 - y_{ab}), \\ y_{ab} + y_{ba} = 1, \\ y_{ab} \geqslant y_{ac} + y_{cb} - 1.5,\ \forall c \in A \setminus \{a,b\}, \\ y_{ab} \in \{0,1\}. \end{array}\right\} \forall (a,b) \in A \times A,\ a \neq b$$

The following methods $\mathbf{M_4} - \mathbf{M_{16}}$ rank the alternatives from A according to the increasing order of the values assigned to alternatives a from A by the following ranking functions:

$\mathbf{M_4}$: *The positive outranking index*:

$$PosOI(a, A, PWM) = \frac{1}{|A| - 1} \sum_{\substack{b\in A\setminus\{a\}:\\ p(a,b)\geqslant 0.5}} p(a,b);$$

$\mathbf{M_5}$: *Max in favor*:

$$MF(a, A, PWM) = \max_{b\in A\setminus\{a\}} p(a,b);$$

$\mathbf{M_6}$: *Min in favor*[5]:

$$mF(a, A, PWM) = \min_{b\in A\setminus\{a\}} p(a,b);$$

$\mathbf{M_7}$: *Sum in favor*:

$$SF(a, A, PWM) = \sum_{b\in A\setminus\{a\}} p(a,b);$$

$\mathbf{M_8}$: *Max against*:

$$MA(a, A, PWM) = -\max_{b\in A\setminus\{a\}} p(b,a);$$

$\mathbf{M_9}$: *Min against*:

$$mA(a, A, PWM) = -\min_{b\in A\setminus\{a\}} p(b,a);$$

$\mathbf{M_{10}}$: *Sum against*:

$$SA(a, A, PWM) = -\sum_{b\in A\setminus\{a\}} p(b,a);$$

$\mathbf{M_{11}}$: *Max difference*:

$$MD(a, A, PWM) = \max_{b\in A\setminus\{a\}} [p(a,b) - p(b,a)];$$

$\mathbf{M_{12}}$: *Min difference*:

$$mD(a, A, PWM) = \min_{b\in A\setminus\{a\}} [p(a,b) - p(b,a)];$$

$\mathbf{M_{13}}$: *Sum of differences*:

$$SD(a, A, PWM) = \sum_{b\in A\setminus\{a\}} [p(a,b) - p(b,a)];$$

$\mathbf{M_{14}}$: *Copeland score*:

$$Cop(a, A, PWM) = \sum_{\substack{b\in A\\ b\neq a}} r_{ab}, \quad \text{where } r_{ab} = \begin{cases} 1, & \text{if } p(a,b) \geqslant 0.5 \\ -1, & \text{otherwise.} \end{cases}$$

[5] The function is also known as Simpson score as presented in Leskinen et al. (2006).

M$_{15}$: *Most Discriminant Score*: It is the score assigned to alternatives in A by the most discriminant among the value functions compatible with the preferences given by the DM. From a computational point of view it is obtained solving the following LP problem

$$\max \varepsilon, \text{ subject to } E^{DM'}$$

that has been presented in Section 2.1.2. Of course, the most discriminant value function is of the same type of the value function used to represent the preferences given by the DM (weighted sum or general additive value function in our context)[6];

M$_{16}$: *Barycenter Score:* It is the score assigned to alternatives in A by the barycenter of the set of sampled value functions compatible with the preferences given by the DM. From a computational point of view it is obtained averaging, component by component, all the sampled value functions compatible with the preferences given by the DM and on the basis of which the PWIs have been computed (see Section 2.1.3). For example, let us suppose that the sampled value functions are weighted sums represented by the weight vectors $(w_1^t, \ldots, w_m^t)$, with $t = 1, \ldots, |\mathcal{U}|$, with $\mathcal{U}$ being the set of sampled value functions. Then, the barycenter of $\mathcal{U}$ is a weighted sum represented by the weight vector $\left(w_1^b, \ldots, w_m^b\right)$ where $w_j^b = \frac{1}{|\mathcal{U}|}\sum_{t=1}^{|\mathcal{U}|} w_j^t$ for all $j = 1, \ldots, m$.

Methods **M$_5$** - **M$_{13}$** have been presented in Kadziński and Michalski (2016), while methods **M$_4$** and **M$_{14}$** are reported in Leskinen et al. (2006).

All methods presented above, as well as our scoring procedure, produce a complete ranking of the alternatives at hand. Therefore, in the following, we shall denote by $Ranking_{\mathbf{M}}$ the ranking obtained by the method **M** with $\mathbf{M} \in \{\mathbf{M_1}, \ldots, \mathbf{M}_{16}, \mathbf{ScPr}_{AVF}, \mathbf{ScPr}_{WS}\}$. On the one hand, $\mathbf{ScPr}_{AVF}$ is the scoring procedure based on the solution of the LP_0 problem presented in Section 2.2 and aiming to summarize the PWIs by an additive value function (see Eq. (1)). On the other hand, $\mathbf{ScPr}_{WS}$ is the scoring procedure based on the solution of the same LP_0 problem but in which PWIs are summarized by a weighted sum, that is, $U(a) = \sum_{j=1}^{m} w_j \cdot g_j(a)$. In this case, the scoring function is obtained solving LP_0 but replacing E^{SF} with the following set of constraints:

$$\left.\begin{array}{l} U(a) = \sum_{j=1}^{m} w_j \cdot g_j(a),\ \forall a \in A, \\ U(a) - U(b) \geqslant \eta\,(p(a,b) - 0.5)\,,\ \forall (a,b) \in A \times A \,:\, p(a,b) \geqslant 0.5, \\ w_j \geqslant 0,\ \forall j = 1, \ldots, m, \\ \sum_{j=1}^{m} w_j = 1. \end{array}\right\} E_{WS}^{SF}$$

As observed in the previous section, let us observe that the number of parameters defining an additive value function is $\sum_{j=1}^{m}\left(n_j + 1\right)$ (one for each criterion and for each possible performance on the same criterion), while, the number of parameters defining a weighted sum is m (only one for each criterion).

3.1. Simulations details

In order to compare $\mathbf{ScPr}_{AVF}$ and $\mathbf{ScPr}_{WS}$ with the sixteen methods $\mathbf{M_1} - \mathbf{M_{16}}$ reviewed above, we shall perform an extensive set of simulations considering different decision problems. In each simulation we assume the existence of an "artificial" DM whose preferences are represented by a random generated value function. This value function will rank order n alternatives evaluated on m criteria where $n \in \{6, 9, 12, 15\}$ and $m \in \{3, 5, 7\}$. To provide more robust conclusions (see Dede et al. (2022, 2016) and Zio and Pedroni (2012)), for each $(n, m) \in \{6, 9, 12, 15\} \times \{3, 5, 7\}$, 10,000 independent runs will be done. Algorithm 1 presents the steps that have to be performed in each of the considered runs. These steps are described in the following lines:

1: A performance matrix $PM = \left[pm_{ij}\right]_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$ composed of n rows and m columns is built. The ith row of the matrix $\left(pm_{i1}, \ldots, pm_{im}\right)$ is a vector of m values taken randomly in a uniform way in the $[0, 1]$ interval. They represent the evaluations of alternative $a_i \in A$ on criteria $g_j \in G$, that is, $pm_{ij} = g_j(a_i)$. Moreover, the performance matrix is built so that the alternatives (having as evaluations the values in the rows of the performance matrix) are non-dominated[7];

2: Let us assume that the artificial DM's value function is a weighted sum such that for each alternative $a_i \in A$

$$WS_i = WS(pm_{i1}, \ldots, pm_{im}) = w_1 pm_{i1} + \cdots + w_m pm_{im} = \sum_{j=1}^{m} w_j pm_{ij}$$

where $w_j \geqslant 0$ for all $j = 1, \ldots, m$ and $\sum_{j=1}^{m} w_j = 1$. To simulate the DM's value function, we then sample m non-negative values $w_1^{DM}, \ldots, w_m^{DM}$ such that their sum is 1 following the procedure proposed by Rubinstein (1982);[8]

3: Apply the artificial DM's value function defined by the vector $\left(w_1^{DM}, \ldots, w_m^{DM}\right)$ to compute the weighted sum of each alternative. On the basis of the values assigned to all alternatives compute their ranking and denote it by $Ranking_{\mathbf{DM}}$;

4: The artificial DM's preference information is provided using the procedure proposed in Vetschera (2017): we sample an alternative $a_i \in A$ and, then, we compare a_i with all the other alternatives a_{i_1}, that is, $a_{i_1} \in A \setminus \{a_i\}$. If $WS_i > WS_{i_1}$, then $a_i \succ_{DM} a_{i_1}$; if $WS_i < WS_{i_1}$ then $a_{i_1} \succ_{DM} a_i$; finally, if $WS_i = WS_{i_1}$, then $a_i \sim_{DM} a_{i_1}$;

5: Sample 10,000 value functions compatible with the preferences provided by the artificial DM. In this case, we assume that a compatible value function is identified by a vector of weights $(w_1, \ldots, w_m)$ so that the following set of constraints is satisfied:

$$\left.\begin{array}{l} w_1 pm_{i1} + \cdots + w_m pm_{im} \geqslant w_1 pm_{i_1 1} + \cdots + w_m pm_{i_1 m} + \varepsilon, \ \text{ if } \ a_i \succ_{DM} a_{i_1}, \\ w_1 pm_{i1} + \cdots + w_m pm_{im} = w_1 pm_{i_1 1} + \cdots + w_m pm_{i_1 m}, \ \text{ if } \ a_i \sim_{DM} a_{i_1}, \\ w_j \geqslant 0, \ \text{ for all } j = 1, \ldots, m, \\ \sum_{j=1}^{m} w_j = 1, \\ \varepsilon > 0. \end{array}\right\} E_{WS}^{DM}$$

Since the artificial DM's value function is a weighted sum and the value function used to approximate its preferences is a weighted sum as well, there exists at least one compatible value function. Therefore, we can sample 10,000 compatible value functions (for example using the HAR method) from the space defined by the constraints in E_{WS}^{DM}. Denoting by $\mathcal{U}$ the set of

[6] In fact, there could exist more than one value function satisfying the set of constraints $E^{DM'}$ with $\varepsilon = \varepsilon^*$. In the following, for the sake of simplicity, we consider only the first solution supplied by the solver.

[7] For each $a_{i_1}, a_{i_2} \in A$, $\exists g_{j_1}, g_{j_2} \in G$ such that $pm_{i_1 j_1} > pm_{i_2 j_1}$ and $pm_{i_2 j_2} < pm_{i_1 j_2}$.

[8] Take randomly $m-1$ values $v_1, \ldots, v_{m-1}$ in the interval $[0, 1]$ and reorder the values in the set $\left\{0, v_1, \ldots, v_{m-1}, 1\right\}$ in a non-decreasing way so that $0 = v_{(0)} \leqslant v_{(1)} \leqslant v_{(2)} \leqslant \cdots \leqslant v_{(m-1)} \leqslant v_{(m)} = 1$. For each $j = 1, \ldots, m$, put $w_j = v_{(j)} - v_{(j-1)}$.

Algorithm 1 Single run steps

repeat
1: Generate a performance matrix of n alternatives and m criteria
2: Build the DM's value function
3: Compute the ranking of alternatives at hand by using the artificial DM's value function and denote it by $Ranking_{\mathbf{DM}}$
4: Elicit the artificial DM's preferences
5: Sample 10,000 value functions compatible with the artificial DM's preferences and compute the PWIs
6: Apply method $\mathbf{M}$ with $\mathbf{M} \in \{\mathbf{M_1}, \ldots, \mathbf{M}_{16}, \mathbf{ScPr}_{AVF}, \mathbf{ScPr}_{WS}\}$ to get the ranking of the alternatives at hand and denote this ranking by $Ranking_{\mathbf{M}}$
7: Compute the Kendall-Tau correlation coefficient between $Ranking_{\mathbf{DM}}$ and $Ranking_{\mathbf{M}}$ for all $\mathbf{M} \in \{\mathbf{M_1}, \ldots, \mathbf{M}_{16}, \mathbf{ScPr}_{AVF}, \mathbf{ScPr}_{WS}\}$
until 10,000 runs have not been performed
8: Compute statistics on the obtained results

Table 1
Average (over 10,000 independent runs) Kendall-Tau correlation coefficient between $Ranking_{\mathbf{DM}}$ and $Ranking_{\mathbf{M}}$.

	(n,m)											
	(6,3)	(6,5)	(6,7)	(9,3)	(9,5)	(9,7)	(12,3)	(12,5)	(12,7)	(15,3)	(15,5)	(15,7)
$\mathbf{M}_1$	0.8447	0.7990	0.7791	0.8542	0.7944	0.7642	0.8661	0.8060	0.7700	0.8758	0.8170	0.7787
$\mathbf{M}_2$	0.8447	0.7990	0.7791	0.8542	0.7944	0.7642	0.8660	0.8061	0.7700	0.8759	0.8169	0.7786
$\mathbf{M}_3$	0.8447	0.7990	0.7791	0.8542	0.7944	0.7642	0.8661	0.8059	0.7701	0.8759	0.8169	0.7786
$\mathbf{M}_4$	0.8446	0.7988	0.7790	0.8540	0.7945	0.7643	0.8662	0.8060	0.7701	0.8758	0.8171	0.7788
$\mathbf{M}_5 - \mathbf{M}_9 - \mathbf{M}_{11}$	0.6262	0.6567	0.6599	0.5360	0.5918	0.6123	0.4768	0.5461	0.5762	0.4323	0.5084	0.5466
$\mathbf{M}_6 - \mathbf{M}_8 - \mathbf{M}_{12}$	0.6457	0.6590	0.6634	0.5699	0.5973	0.6094	0.5142	0.5558	0.5729	0.4752	0.5167	0.5411
$\mathbf{M}_7 - \mathbf{M}_{10} - \mathbf{M}_{13}$	0.8431	0.7977	0.7789	0.8526	0.7941	0.7640	0.8654	0.8058	0.7698	0.8748	0.8168	0.7792
$\mathbf{M}_{14}$	0.8447	0.7992	0.7796	***0.8544***	***0.7950***	***0.7648***	0.8664	0.8065	***0.7708***	0.8762	0.8175	***0.7796***
$\mathbf{M}_{15}$	0.8202	0.7745	0.7592	0.8260	0.7708	0.7438	0.8401	0.7830	0.7493	0.8508	0.7936	0.7561
$\mathbf{M}_{16}$	***0.8450***	***0.7993***	***0.7797***	0.8543	0.7946	0.7641	***0.8670***	***0.8067***	0.7705	***0.8767***	***0.8176***	0.7795
$\mathbf{ScPr}_{WS}$	0.8447	0.7982	0.7780	0.8542	0.7943	0.7627	0.8661	0.8059	0.7693	0.8759	0.8169	0.7787
$\mathbf{ScPr}_{AVF}$	0.8438	0.7958	0.7753	0.8505	0.7881	0.7533	0.8606	0.7937	0.7542	0.8678	0.8040	0.7606
$dev\left(\mathbf{ScPr}_{WS}\right)$	0.0406%	0.1324%	0.2259%	0.0251%	0.0890%	0.2743%	0.0991%	0.1010%	0.1980%	0.0855%	0.0795%	0.1204%
$dev\left(\mathbf{ScPr}_{AVF}\right)$	0.1512%	0.4342%	0.5653%	0.4556%	0.8784%	1.5122%	0.7394%	1.6157%	2.1538%	1.0056%	1.6621%	2.4427%

sampled compatible value functions we have $|\mathcal{U}| = 10{,}000$. Computing the sets $\mathcal{U}_{a \succ b}$ and $\mathcal{U}_{a \sim b}$ for each $(a,b) \in A \times A$ with $a \neq b$ as in Eq. (2), we then can compute the PWIs $p(a,b)$ as defined in Eq. (3);

6: Apply the sixteen methods presented above as well as the scoring procedure we are proposing to compute the alternatives' ranking $Ranking_{\mathbf{M}}$ with $\mathbf{M} \in \{\mathbf{M}_1, \ldots, \mathbf{M}_{16}, \mathbf{ScPr}_{AVF}, \mathbf{ScPr}_{WS}\}$;

7: For each $\mathbf{M} \in \{\mathbf{M}_1, \ldots, \mathbf{M}_{16}, \mathbf{ScPr}_{AVF}, \mathbf{ScPr}_{WS}\}$ compute the Kendall-Tau (Kendall, 1938) between $Ranking_{\mathbf{M}}$ and $Ranking_{\mathbf{DM}}$, denoted as $\tau(\mathbf{M}, \mathbf{DM})$. This value ranges in $[-1, 1]$ and the greater the value, the more correlated are the two rankings. In particular, a Kendall-Tau equal to 1 implies perfect correlation between the two rankings, while, a Kendall-Tau equal to -1 implies inverse correlation between them;

8: For each $\mathbf{M} \in \{\mathbf{M}_1, \ldots, \mathbf{M}_{16}, \mathbf{ScPr}_{AVF}, \mathbf{ScPr}_{WS}\}$ we compute the average Kendall-Tau between $Ranking_{\mathbf{M}}$ and $Ranking_{\mathbf{DM}}$ over the 10,000 performed independent runs. We denote it as $\overline{\tau}(\mathbf{M}, \mathbf{DM})$. Moreover, we perform two different versions of the 2-sample Kolmogorov–Smirnov test (Massey, 1951) between each pair of methods at the 5% significance level. In the first version (see Table 2), we test the *null hypothesis* of equality between the cumulative distribution functions of the Kendall-Tau coefficients of two methods, say F_1 and F_2, versus the *alternative hypothesis* that the cumulative distribution functions are different. In the second version (see Table 3), we test the *null hypothesis* that F_1 is greater than or equal to F_2 versus the *alternative hypothesis* that F_1 is smaller than F_2, meaning that the method related to F_1 is better than the one related to F_2.

For each (n,m) configuration, we highlight "*the best method*" between $\mathbf{M}_1$, …, $\mathbf{M}_{16}$, $\mathbf{ScPr}_{AVF}$, $\mathbf{ScPr}_{WS}$ in terms of their capability of replying $Ranking_{DM}$. This is the one having the maximum average Kendall-Tau, that is, the method $\overline{\mathbf{M}} \in \{\mathbf{M}_1, \ldots, \mathbf{M}_{16}, \mathbf{ScPr}_{AVF}, \mathbf{ScPr}_{WS}\}$ such that

$$\overline{\tau}\left(\overline{\mathbf{M}}, \mathbf{DM}\right) = \overline{\tau}_{Max} = \max_{\mathbf{M} \in \{\mathbf{M}_1, \ldots, \mathbf{M}_{16}, \mathbf{ScPr}_{AVF}, \mathbf{ScPr}_{WS}\}} \{\overline{\tau}(\mathbf{M}, \mathbf{DM})\}.$$

For example, looking at the data in Table 1, $\mathbf{M}_{16}$ is the best method for the $(6,3)$ configuration since the average Kendall-Tau coefficient between $Ranking_{\mathbf{M}_{16}}$ and $Ranking_{DM}$ is 0.8450.

As one can see from the data in the table, there is not any method being the best in all configurations. More in detail, $\mathbf{M}_{16}$ (based on the barycenter of the set of sampled value functions) is the best method in seven out of the twelve configurations, while $\mathbf{M}_{14}$ is the best method in the remaining five configurations. Let us observe that, because $p(a,b) + p(b,a) = 1$ for all $a, b \in A$, it is easy to prove the equivalence of some of considered scoring procedures.

Proposition 3.1. *Given a set of alternatives A and the PWIs matrix PWM the following holds:*

1. *$\mathbf{M}_5$, $\mathbf{M}_9$ and $\mathbf{M}_{11}$ provide the same ranking of alternatives from A,*
2. *$\mathbf{M}_6$, $\mathbf{M}_8$ and $\mathbf{M}_{12}$ provide the same ranking of alternatives from A,*
3. *$\mathbf{M}_7$, $\mathbf{M}_{10}$ and $\mathbf{M}_{13}$ provide the same ranking of alternatives from A.*

Proof. For all $a \in A$:

1. $(\mathbf{M}_5)$ $MF(a, A, PWM) = \max_{b \in A \setminus \{a\}} p(a,b) = \max_{b \in A \setminus \{a\}} [1 - p(b,a)] = 1 - \min_{b \in A \setminus \{a\}} p(b,a) = 1 + mA(a, A, PWM)$ $(\mathbf{M}_9)$
$(\mathbf{M}_{11})$ $MD(a, A, PWM) = \max_{b \in A \setminus \{a\}} [p(a,b) - p(b,a)] = \max_{b \in A \setminus \{a\}} [1 - 2p(b,a)] = 1 - 2 \min_{b \in A \setminus \{a\}} p(b,a) = 1 + 2 \cdot mA(a, A, PWM)$ $(\mathbf{M}_9)$,
2. $(\mathbf{M}_6)$ $mF(a, A, PWM) = \min_{b \in A \setminus \{a\}} p(a,b) = \min_{b \in A \setminus \{a\}} [1 - p(b,a)] = 1 - \max_{b \in A \setminus \{a\}} p(b,a) = 1 + MA(a, A, PWM)$ $(\mathbf{M}_8)$

Table 2
First version of the 2-sample Kolmogorov–Smirnov test at 5% significance level.

h/p-value	(n,m)											
	(6,3)	(6,5)	(6,7)	(9,3)	(9,5)	(9,7)	(12,3)	(12,5)	(12,7)	(15,3)	(15,5)	(15,7)
$\mathbf{ScPr}_{WS}-\overline{\mathbf{M}}$	0/0.9992	0/0.9999	0/0.9999	0/1	0/0.9999	0/0.9779	0/0.9994	0/0.9985	0/0.9975	0/1	0/0.9975	0/0.9948
$\mathbf{ScPr}_{AVF}-\overline{\mathbf{M}}$	0/0.9992	0/0.9985	0/0.9967	0/0.9975	0/0.7091	0/0.1966	0/0.2351	1/0.0210	1/0.0015	0/0.1483	1/0.0022	1/0

$(\mathbf{M}_{12})$ $mD(a,A,PWM) = \min_{b\in A\setminus\{a\}}[p(a,b)-p(b,a)] = \min_{b\in A\setminus\{a\}}[1-2p(b,a)] = 1-2\max_{b\in A\setminus\{a\}} p(b,a) = 1+2\cdot MA(a,A,PWM)$ $(\mathbf{M}_8)$,

3. $(\mathbf{M}_7)$ $SF(a,A,PWM) = \sum_{b\in A\setminus\{a\}} p(a,b) = \sum_{b\in A\setminus\{a\}}[1-p(b,a)] = |A|-1-\sum_{b\in A\setminus\{a\}} p(b,a) = |A|-1+SA(a,A,PWM)$ $(\mathbf{M}_{10})$
$(\mathbf{M}_{13})$ $SD(a,A,PWM) = \sum_{b\in A\setminus\{a\}}[p(a,b)-p(b,a)] = \sum_{b\in A\setminus\{a\}}[1-2p(b,a)] = |A|-1-2\sum_{b\in A\setminus\{a\}} p(b,a) = |A|-1+2\cdot SA(a,A,PWM)$ $(\mathbf{M}_{10})$. □

In consequence of Proposition 3.1, in Table 1, we grouped the data corresponding to these three triplets of methods.

To have an estimate of how $\mathbf{ScPr}_{WS}$ and $\mathbf{ScPr}_{AVF}$ behave with respect to the best method $\overline{\mathbf{M}}$ in each configuration, we computed their deviation from it. This is a "normalized distance" of $\overline{\tau}(\mathbf{ScPr}_{WS},\mathbf{DM})$ and $\overline{\tau}(\mathbf{ScPr}_{AVF},\mathbf{DM})$ from $\overline{\tau}_{Max}$ obtained as follows:

$$dev(\mathbf{ScPr}_{WS}) = \frac{\overline{\tau}(\mathbf{ScPr}_{WS},\mathbf{DM})-\overline{\tau}_{Max}}{\overline{\tau}_{Max}} \text{ and } dev(\mathbf{ScPr}_{AVF}) = \frac{\overline{\tau}(\mathbf{ScPr}_{AVF},\mathbf{DM})-\overline{\tau}_{Max}}{\overline{\tau}_{Max}}. \quad (6)$$

The data in Table 1 underline that this error is in the interval $[0.0406\%, 0.2743\%]$ for $\mathbf{ScPr}_{WS}$ and in the interval $[0.1512\%, 2.4427\%]$ for $\mathbf{ScPr}_{AVF}$. To check if the difference between the distributions of values $\tau(\mathbf{ScPr}_{WS},\mathbf{DM})$ and $\tau(\overline{\mathbf{M}},\mathbf{DM})$ and, analogously, between the distributions of values $\tau(\mathbf{ScPr}_{AVF},\mathbf{DM})$ and $\tau(\overline{\mathbf{M}},\mathbf{DM})$ are significant from the statistical point of view, we performed the 2-sample Kolmogorov–Smirnov test with 5% significance level. In Table 2, we report the p-value and the indicator $h\in\{0,1\}$ of the performed test for the null hypothesis that the cumulative distribution functions of the Kendall-Tau coefficients are equal ($h=0$ means that the null hypothesis is not rejected, while, $h=1$ means the opposite).

Looking at Table 2, one can observe the following:

- For all (n,m) configurations, the difference between the distributions of values $\tau(\mathbf{ScPr}_{WS},\mathbf{DM})$ and $\tau(\overline{\mathbf{M}},\mathbf{DM})$ is not significant from the statistical point of view. This means that even if our scoring procedure is not the best in any configuration, it obtains an approximation of the DM's ranking of the alternatives at hand at least as good as the one produced by the best method $\overline{\mathbf{M}}$;
- For all (n,m) configurations apart from $(12,5)$, $(12,7)$, $(15,5)$ and $(15,7)$ the difference between the distributions of values $\tau(\mathbf{ScPr}_{AVF},\mathbf{DM})$ and $\tau(\overline{\mathbf{M}},\mathbf{DM})$ is not significant from the statistical point of view. In particular, the difference between them for the configuration (12,5) is also not significant if the test is done at the 2% significance level. In these four configurations, we checked the alternative hypothesis that one of the cumulative distribution functions is smaller than the other. The results are shown in Table 3 ($h=1$ means that the null hypothesis is rejected in favor of the alternative hypothesis, while, $h=0$ means the opposite)
One can see that, on the one hand, for configurations $(12,5)$ and $(15,5)$, the cumulative distribution function corresponding to $\mathbf{M}_{16}$ is smaller than the cumulative distribution function corresponding to $\mathbf{ScPr}_{AVF}$ ($\mathbf{M}_{16}$ is therefore better than $\mathbf{ScPr}_{AVF}$), while, on the other hand, for configurations $(12,7)$ and $(15,7)$, the cumulative distribution function corresponding to $\mathbf{M}_{14}$ is smaller than the cumulative distribution function corresponding to $\mathbf{ScPr}_{AVF}$ ($\mathbf{M}_{14}$ is therefore better than $\mathbf{ScPr}_{AVF}$).

We would like to conclude this section by commenting the statistical tests in the following way:

1. Even if $\mathbf{ScPr}_{WS}$ is not the best method for any (n,m) configuration, the difference between the cumulative distribution functions of values $\tau(\mathbf{ScPr}_{WS},\mathbf{DM})$ and $\tau(\overline{\mathbf{M}},\mathbf{DM})$ is not significant from the statistical point of view. Moreover, even if $\mathbf{ScPr}_{WS}$, $\mathbf{M}_{14}$ and $\mathbf{M}_{16}$ are able to produce a ranking of the alternatives under consideration summarizing the PWIs, our method presents an added value with respect to any of them. First of all, $\mathbf{ScPr}_{WS}$ produces a ranking based on the score assigned to each alternative by a weighted sum which is able to explain the contribution given by each criterion to the overall evaluation of each alternative. This is not the case for $\mathbf{M}_{14}$ where the score assigned to each alternative a is equal to the difference between the number of alternatives in A to which a is at least as good in at least 50% of the cases and the number of alternatives to which a is at least as good in less than 50% of the cases. With respect to $\mathbf{M}_{16}$, the added value of our method is due to the consideration of the whole set of value functions compatible with the preferences given by the DM and not only one of them. Indeed, the ranking produced by $\mathbf{M}_{16}$ is finally related to only one specific compatible value function, even if, in some form, this value function depends on all the compatible value functions because it is the barycenter of all them (or, more precisely, from a computational point of view, of all the compatible value functions in the considered sample). Instead, through $\mathbf{ScPr}_{WS}$ one constructs a value function that represents the ranking orders of all the compatible value functions as summarized by the PWIs, which is something more than simply selecting one among the many compatible value functions;
2. The difference between the distributions of values $\tau(\mathbf{ScPr}_{AVF},\mathbf{DM})$ and $\tau(\overline{\mathbf{M}},\mathbf{DM})$ is not statistically significant in eight of the twelve considered configurations. Moreover, in the four configurations in which the difference is significant from the statistical point of view, the error of $Ranking_{\mathbf{ScPr}_{AVF}}$ with respect to the best method, that is, $dev(\mathbf{ScPr}_{AVF})$, varies between 1.6157% (for configuration $(12,5)$) and 2.4427% (for configuration $(15,7)$), which seems quite acceptable. Therefore, we can conclude that $Ranking_{\mathbf{ScPr}_{AVF}}$ gives, in any case, a good approximation of $Ranking_{DM}$. Furthermore, as already observed for $\mathbf{ScPr}_{WS}$ in the previous item, differently from $\mathbf{M}_{14}$ and $\mathbf{M}_{16}$, $\mathbf{ScPr}_{AVF}$ is able to explain the ranking obtained by each alternative. The non optimal performance of $\mathbf{ScPr}_{AVF}$ with respect to $\mathbf{ScPr}_{WS}$ is due to the larger number of parameters that need to be estimated (35 in $\mathbf{ScPr}_{AVF}$ vs. 5 in $\mathbf{ScPr}_{WS}$). The lack in the performance of $\mathbf{ScPr}_{AVF}$ is nevertheless counterbalanced by its more detailed and analytic capacity to explain the ranking of alternatives. Indeed, while $\mathbf{ScPr}_{WS}$ assigns to each criterion a weight that holds for all values taken by the criterion itself, $\mathbf{ScPr}_{AVF}$ assigns a specific marginal value to each performance of the criterion. As will be shown in the next section, this can give more insight into the problem at hand and, therefore, can

Table 3
Second version of the 2-sample Kolmogorov–Smirnov test at 5% significance level.

(12,5)			(12,7)			(15,5)			(15,7)		
h/p-value	$\mathbf{M}_{16}$	$\mathbf{ScPr}_{AVF}$	h/p-value	$\mathbf{M}_{14}$	$\mathbf{ScPr}_{AVF}$	h/p-value	$\mathbf{M}_{16}$	$\mathbf{ScPr}_{AVF}$	h/p-value	$\mathbf{M}_{14}$	$\mathbf{ScPr}_{AVF}$
$\mathbf{M}_{16}$	[redacted]	1/0.0105	$\mathbf{M}_{14}$	[redacted]	1/0.0008	$\mathbf{M}_{16}$	[redacted]	1/0.0011	$\mathbf{M}_{14}$	[redacted]	1/0
$\mathbf{ScPr}_{AVF}$	0/1	[redacted]	$\mathbf{ScPr}_{AVF}$	0/1	[redacted]	$\mathbf{ScPr}_{AVF}$	0/0.9975	[redacted]	$\mathbf{ScPr}_{AVF}$	0/1	[redacted]

Table 4
Evaluation of the 7 funds on the five considered criteria.

		$g_1(\cdot)$	$g_2(\cdot)$	$g_3(\cdot)$	$g_4(\cdot)$	$g_5(\cdot)$
a_1	Allianz Multipartner Multi20	0.0403	0,0010	−0.0155	−0.0030	−0,0010
a_2	Amundi Bilanciato Euro C	0.0257	0.0004	−0.0103	−0.0014	−0.0008
a_3	Arca Te - Titoli Esteri	0.0322	0.0009	−0.0133	−0.0022	−0.0011
a_4	Bancoposta Mix 2 A Cap	0.0193	0.0003	−0.0080	−0.0011	−0.0006
a_5	Etica Rendita Bilanciata I	0.0334	0.0005	−0.0150	−0.0009	−0.0013
a_6	Eurizon Pir Italia 30 I	0.0219	0.0003	−0.0088	−0.0011	−0.0007
a_7	Pramerica Global Multiasset 30	−0.0018	0.0000	0.0007	−0.0010	0.0006

Table 5
Normalized values of the 7 funds on the five considered criteria.

		$\overline{g}_1(\cdot)$	$\overline{g}_2(\cdot)$	$\overline{g}_3(\cdot)$	$\overline{g}_4(\cdot)$	$\overline{g}_5(\cdot)$
a_1	Allianz Multipartner Multi20	0.6940	0.7349	0.3370	0.1917	0.4171
a_2	Amundi Bilanciato Euro C	0.5157	0.4869	0.4924	0.5268	0.4651
a_3	Arca Te - Titoli Esteri	0.5943	0.6939	0.4013	0.3492	0.3923
a_4	Bancoposta Mix 2 A Cap	0.4370	0.3980	0.5596	0.5880	0.5381
a_5	Etica Rendita Bilanciata I	0.6102	0.4950	0.3522	0.6333	0.3455
a_6	Eurizon Pir Italia 30 I	0.4694	0.4342	0.5357	0.5926	0.4933
a_7	Pramerica Global Multiasset 30	0.1793	0.2571	0.8219	0.6184	0.8487

be beneficial for the DM in terms of a better understanding of the considered decision problem.

4. Case study

In this section, we shall apply the method described in Section 2 showing its main characteristics. For this reason, we shall consider a financial problem in which the returns of seven funds are evaluated with respect to five performance measures (for a standard survey on performance measures see, for example, Amenc and Le Sourd (2005) and Bacon (2012)). The considered funds belong to the sub-class of the balanced bond funds which may invest in stocks a proportion of their assets between 10% and 50%. The historical data used to estimate the relevant statistics are the daily logarithmic return of the interval between the 01/01/2018–01/01/2021 (784 data points). The considered performance measures are:

- g_1: the Sharpe Ratio (SR; Sharpe, 1998b), defined as the ratio between the expected return and the standard deviation. It is often chosen by practitioners to rank managed portfolios and belongs to the class of reward-to-variability measures;
- g_2: the Treynor Ratio (TR; Scholz and Wilkens, 2005), defined as the ratio between the expected return and the systematic risk sensitivity with respect to a benchmark (here we consider the MSCI World index Bacmann and Scholz, 2003). It assumes the form of a reward-to-variability measure but it derives from the Capital Asset Pricing Model (CAPM) portfolio theory;
- g_3: the Average Value-at-Risk Ratio (AVaRR; Rockafellar and Uryasev, 2000), defined as a relative performance measure of the same class of SR and TR. It looks at the average amount of potential losses suffered by the portfolio manager;
- g_4: the Jensen Alpha (JA; Jensen, 1968), derived from the CAPM theory, defined as the excess return of a fund over the theoretical benchmark (MSCI World index);
- g_5: the Morningstar Risk-Adjusted Return (MRAR; Sharpe, 1998a), derived within the Expected Utility theory, defined as the annualized geometric average of the returns.

The evaluations of the alternatives on the considered criteria are given in Table 4.

The evaluations of the considered alternatives on the criteria at hand will be aggregated by the weighted sum (7)

$$WS(a) = WS\left(g_1(a), \dots, g_m(a)\right) = \sum_{j=1}^{m} w_j \cdot g_j(a) \tag{7}$$

where w_j are the weights of criteria g_j and they are such that $w_j \geqslant 0$ for all $g_j \in G$ and $\sum_{j=1}^{m} w_j = 1$.

Table 6
Pairwise winning indices of the considered funds.

$p(\cdot,\cdot)$	a_1	a_2	a_3	a_4	a_5	a_6	a_7
a_1	0.50	0.4166	0.4261	0.4082	0.5237	0.4070	0.3673
a_2	0.5834	0.50	0.5928	0.4074	0.5577	0.3673	0.3681
a_3	0.5739	0.4072	0.50	0.4050	0.5482	0.3969	0.3729
a_4	0.5918	0.5926	0.5950	0.50	0.5595	0.4570	0.3586
a_5	0.4763	0.4423	0.4519	0.4405	0.50	0.4145	0.3933
a_6	0.5930	0.6327	0.6031	0.5431	0.5855	0.50	0.3816
a_7	0.6327	0.6319	0.6271	0.6414	0.6067	0.6184	0.50

However, the use of a weighted sum implies that the evaluations are expressed on the same scale. For such a reason, before applying the weighted sum, we used a standardization technique proposed in Greco et al. (2018) transforming the evaluation of each $a \in A$ on $g_j \in G$, that is $g_j(a)$, into the standardized value $\overline{g}_j(a)$ as follows:

$$\overline{g}_j(a) = \begin{cases} 0 & \text{if} \quad g_j(a) \leqslant M_j - 3s_j, \\ 0.5 + \frac{g_j^z(a)}{6} & \text{if} \quad M_j - 3s_j < g_j(a) < M_j + 3s_j, \\ 1 & \text{if} \quad g_j(a) \geqslant M_j + s_j \end{cases}$$

if g_j has an increasing direction of preference or

$$\overline{g}_j(a) = \begin{cases} 0 & \text{if} \quad g_j(a) \geqslant M_j + 3s_j, \\ 0.5 - \frac{g_j^z(a)}{6} & \text{if} \quad M_j - 3s_j < g_j(a) < M_j + 3s_j, \\ 1 & \text{if} \quad g_j(a) \leqslant M_j - s_j \end{cases}$$

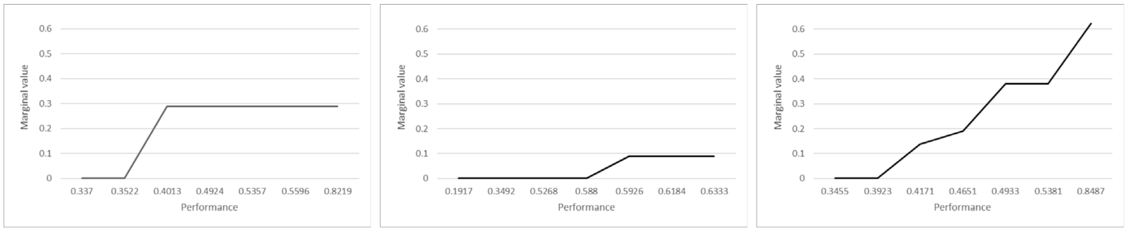
if g_j has a decreasing direction of preference. In both cases, $M_j = \frac{1}{|A|}\sum_{a \in A} g_j(a)$, $s_j = \sqrt{\frac{\sum_{a\in A}(g_j(a) - M_j)^2}{|A|}}$, and $g_j^z(a) = \frac{g_j(a) - M_j}{s_j}$. With this procedure, observing that the five considered criteria have an increasing direction of preference, the values in Table 4 are transformed in those given in Table 5.

For the sake of simplicity let us assume that there is not any preference information provided by the DM (see the next section for a sensitivity analysis on the number of preferences given by the DM) and, consequently, the space from which the compatible models (the weighted sum in our case) have to be sampled is the following:

$$W = \{(w_1, \dots, w_5) \in \mathbb{R}^5 : w_j \geqslant 0, \ \forall j = 1, \dots, 5, \text{ and } \sum_{j=1}^{5} w_j = 1\}. \tag{8}$$

Sampling 100,000 weight vectors $(w_1, \dots, w_5)$ from the space W and computing the value assigned by the weighted sum to the seven funds for each of the 100,000 weight vectors, we obtain the PWIs shown in Table 6 being the basis of our scoring procedure.

Following the notation introduced in Section 2, since on each criterion all alternatives have different evaluations, $n_j = 6$, for all



(a) AVaRR ($\overline{g}_3$) (b) JA ($\overline{g}_4$) (c) MRAR ($\overline{g}_5$)

Fig. 2. Marginal value functions for criteria AVaRR, JA and MRAR obtained solving LP_0.

Table 7
The additive value function obtained solving the LP_0 problem.

SR ($\overline{g}_1$)		TR ($\overline{g}_2$)		AVaRR ($\overline{g}_3$)		JA ($\overline{g}_4$)		MRAR ($\overline{g}_5$)	
$x_1^0 = 0.1793$	$u_1^{0,1} = 0$	$x_2^0 = 0.2571$	$u_2^{0,1} = 0$	$x_3^0 = 0.337$	$u_3^{0,1} = 0$	$x_4^0 = 0.1917$	$u_4^{0,1} = 0$	$x_5^0 = 0.3455$	$u_5^{0,1} = 0$
$x_1^1 = 0.437$	$u_1^{1,1} = 0$	$x_2^1 = 0.398$	$u_2^{1,1} = 0$	$x_3^1 = 0.3522$	$u_3^{1,1} = 0$	$x_4^1 = 0.3492$	$u_4^{1,1} = 0$	$x_5^1 = 0.3923$	$u_5^{1,1} = 0$
$x_1^2 = 0.4694$	$u_1^{2,1} = 0$	$x_2^2 = 0.4342$	$u_2^{2,1} = 0$	$x_3^2 = 0.4013$	$u_3^{2,1} = 0.2885$	$x_4^2 = 0.5268$	$u_4^{2,1} = 0$	$x_5^2 = 0.4171$	$u_5^{2,1} = 0.137$
$x_1^3 = 0.5157$	$u_1^{3,1} = 0$	$x_2^3 = 0.4869$	$u_2^{3,1} = 0$	$x_3^3 = 0.4924$	$u_3^{3,1} = 0.2885$	$x_4^3 = 0.588$	$u_4^{3,1} = 0$	$x_5^3 = 0.4651$	$u_5^{3,1} = 0.1904$
$x_1^4 = 0.5943$	$u_1^{4,1} = 0$	$x_2^4 = 0.495$	$u_2^{4,1} = 0$	$x_3^4 = 0.5357$	$u_3^{4,1} = 0.2885$	$x_4^4 = 0.5926$	$u_4^{4,1} = 0.0883$	$x_5^4 = 0.4933$	$u_5^{4,1} = 0.3803$
$x_1^5 = 0.6102$	$u_1^{5,1} = 0$	$x_2^5 = 0.6939$	$u_2^{5,1} = 0$	$x_3^5 = 0.5596$	$u_3^{5,1} = 0.2885$	$x_4^5 = 0.6184$	$u_4^{5,1} = 0.0883$	$x_5^5 = 0.5381$	$u_5^{5,1} = 0.3803$
$x_1^6 = 0.694$	$u_1^{6,1} = 0$	$x_2^6 = 0.7349$	$u_2^{6,1} = 0$	$x_3^6 = 0.8219$	$u_3^{6,1} = 0.2885$	$x_4^6 = 0.6333$	$u_4^{6,1} = 0.0883$	$x_5^6 = 0.8487$	$u_5^{6,1} = 0.6232$

Table 8
Global utility of the seven funds applying the maximally discriminant compatible scoring function U^1 obtained solving LP_0.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7
$U^1(\cdot)$	0.137	0.4789	0.2885	0.6688	0.0883	0.7571	1

$j = 1, \ldots, 5$. Consequently, an additive value function is defined by five marginal value functions $u_j(\cdot)$ assigning the seven different evaluations $u_j^k = u_j\left(x_j^k\right)$, $k = 0, \ldots, 6$. In this way, with a slight abuse of notation, the value function U can be identified by the vector of marginal values u_j^k assigned by it, that is, $U = \left[u_j^k\right]_{\substack{j=1,\ldots,5\\k=0,\ldots,6}}$.

Solving the LP_0 problem presented in Section 2.2.1, we find that E^{SF} is feasible and $\eta^* = 2.0513$. Therefore, at least one compatible scoring function exists and the one obtained solving LP_0, denoted by U^1, is given in Table 7, while the respective marginal value functions are shown in Fig. 2.

As one can see from the table, the greatest values on criteria SR and TR have a null marginal value meaning that the obtained maximally discriminant compatible scoring function does not assign any "importance" to these criteria. Consequently, they do not contribute to the global value obtained by each alternative. Going to the other three criteria, on the one hand, the greatest maximal shares correspond to MRAR (0.6232), followed by AVaRR (0.2885), while, on the other hand, the least maximal share is due to JA (0.0833). Looking at the changes in the marginal values, they are different for the three criteria. In particular, the greatest variation is observed for MRAR in passing from 0.5381 ($u_5(0.5381) = 0.3803$) to 0.8487 ($u_5(0.8487) = 0.6232$). Going to AVaRR and JA, the corresponding marginal value functions present only a single change in passing, on the one hand, from 0.3522 ($u_3(0.3522) = 0$) to 0.4013 ($u_3(0.4013) = 0.2885$) and, on the other hand, in passing from 0.588 ($u_4(0.588) = 0$) to 0.5926 ($u_4(0.5926) = 0.0883$). This means that changes in the performances on these two criteria affect very marginally the global value of the considered funds.

The maximally discriminant compatible scoring function obtained solving LP_0 and shown in Table 7 assigns a unique value to each fund as shown in Table 8. The ranking of the considered funds on the basis of this function is the following:

$$a_7 \succ a_6 \succ a_4 \succ a_2 \succ a_3 \succ a_1 \succ a_5.$$

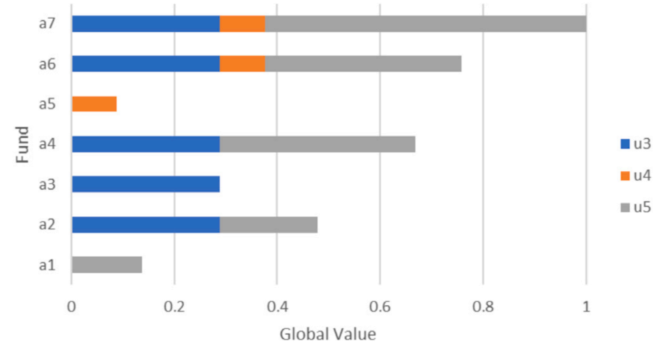
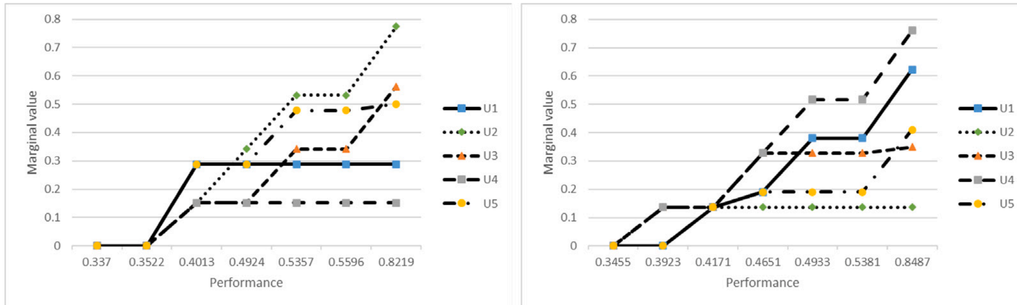


Fig. 3. Global value assigned to each fund by the scoring function obtained solving LP_0 and shown in Table 7.

To better explain why each fund fills a certain position in the provided ranking we examine the global score assigned to them. In particular, we examine the contribution given to this global value by the criteria under consideration as shown in Fig. 3. For example, even if AVaRR and JA contribute in the same way to the global value of a_6 and a_7, the better global value obtained by a_7 can be explained by its better performance on MRAR ($\overline{g}_5(a_7) = 0.8487$ and $\overline{g}_5(a_6) = 0.4933$) to which corresponds a difference in the marginal value of 0.2429 ($u_5(0.8487) - u_5(0.4933) = 0.6232 - 0.3803 = 0.2429$). Analogously, even if a_5 is better than a_1 on AVaRR and JA, this is not enough to compensate the greater value obtained by a_1 on MRAR. Indeed, for both funds only one criterion gives a marginal contribution to their global value (MRAR for a_1 and JA for a_5) since all the other performances on the remaining criteria are associated with a null marginal value. However, the marginal value given by MRAR to a_1 (0.137) is greater than the one given by JA to a_5 (0.0833) that presents the greatest performance on this criterion (0.6333). Analogous considerations can be done for the funds put in the middle of the obtained ranking, that is, a_2, a_3 and a_4.

4.1. Sets of diversified maximally discriminant compatible scoring functions

As already described above, solving the LP_0 problem we found that $\mathcal{U}^{SF} \neq \emptyset$ since $\eta^* > 0$ and, therefore, there exists at least one compatible scoring function able to summarize the information of the



(a) AVaRR ($\overline{g}_3$) (b) MRAR ($\overline{g}_5$)

Fig. 4. Marginal value functions for criteria AVaRR and MRAR for five maximally discriminant compatible value functions being the most distant among them.

PWIs shown in Table 6. Moreover, we observed that for the function U^1 obtained solving the LP problem and shown in Table 7, only the last three criteria (AVaRR, JA and MRAR) are giving a contribution to the global value of each fund while the marginal value attached to the best performances on SR and TR, that is 0.694 and 0.7349, respectively, is zero. For this reason, following the procedure shown in Section 2, one can wonder if there is a maximally discriminant compatible scoring function such that all criteria give a contribution to the global value of the funds at hand. Solving LP_1, we find that $E^{SF}_{AllContr}$ is feasible and $h^* = 0$. This means that each function in $\mathcal{U}^{SF}$ is such that at least one marginal value function gives a null contribution to the global value assigned to the alternatives and, consequently, $\mathcal{U}^{SF}_{AllContr} = \emptyset$. Because of the Note 2.1, we find that also $\mathcal{U}^{SF}_{AllInc} = \emptyset$ and, therefore, there is not any maximally discriminant compatible scoring function such that all marginal value functions are monotone in their domain.

Since $\mathcal{U}^{SF} \neq \emptyset$, one can wonder if there exists another maximally discriminant compatible scoring function "sufficiently" different from U^1. For this reason, we iteratively solved the MILP problems described in Section 2.3.3 considering $\delta_{min} = 0.1$. In this way, in addition to the function found solving LP_0, we get a sample of twenty well-diversified maximally discriminant compatible scoring functions. An interesting aspect is that for all these functions, the marginal value corresponding to the greatest performance on criteria SR and TR is zero meaning that both criteria have a null impact on the global value of the alternatives. Moreover, the marginal value function for criterion JA is the same for the twenty functions and, in particular, is the one shown in Fig. 2(b). Consequently, the twenty considered maximally discriminant compatible scoring functions in the sample, together with the one obtained solving LP_0, differ only for the marginal value functions of AVaRR and MRAR. For this reason, in Fig. 4 we have shown the marginal value functions with respect to these two criteria for LP_0 and other four "most distant functions" among the other twenty obtained by the procedure described in Section 2.3. The four functions are chosen in the following way. Let us denote by $\mathcal{U}_{\mathcal{F}}$ the set composed of the twenty maximally discriminant compatible scoring functions in the sample and by U^1 the function obtained solving LP_0. At first, let us select the maximally discriminant compatible scoring function in $\mathcal{U}_{\mathcal{F}}$ being the farthest from U^1, that is $U^2 \in \mathcal{U}_{\mathcal{F}}$ such that $d(U^2, U^1) = \max_{U \in \mathcal{U}_{\mathcal{F}}} d(U, U^1)$, where $d(U, U^1) = \sqrt{\sum_{j=1}^{5}\sum_{k=0}^{6}\left[u_j\left(x_j^k\right) - u_j^1\left(x_j^k\right)\right]^2}$, that is, the Euclidean vector-to-vector distance between U and U^1. Let us denote by $\mathcal{CF}$ the set composed of the chosen functions up to now, that is, $\mathcal{CF} = \{U^1, U^2\}$. After that, let us add to $\mathcal{CF}$ the function $U^k \in \mathcal{U}_{\mathcal{F}}$ presenting the maximum $\min_{U \in \mathcal{CF}}\{d(U^k, U)\}$, that is, the function presenting the maximal minimal distance from the functions that have already been included in $\mathcal{CF}$. The procedure continues then iteratively until $|\mathcal{CF}| = 5$.

Looking at the marginal value functions corresponding to AVaRR (Fig. 4(a)) and MRAR (Fig. 4(b)), one can observe that the five considered maximally discriminant compatible scoring functions are quite different with respect to the importance assigned to two mentioned criteria, that is, to the marginal value assigned to the greatest performance on the considered criterion. For example, in U^2 and U^4, the two marginal value functions contribute in a completely different way to the global value of the seven alternatives. On the one hand, AVaRR slightly contributes to U^4 since, for this function, $u_3^4(0.8219) = 0.1515$, while it has a very great importance in U^2 since $u_3^2(0.8219) = 0.7747$. On the other hand, the opposite behavior can be observed for U^2 and U^4 for criterion MRAR. Indeed, this criterion slightly contributes to the global value of the alternatives in U^2 since $u_5^2(0.8487) = 0.1369$, while it contributes in a considerable way in U^4 since $u_5^4(0.8487) = 0.7601$. This sheds light on the importance of taking into account not only one maximally discriminant compatible scoring function but all the maximally discriminant compatible scoring functions in the well distributed sample obtained through the iterative procedure described in Section 2.3.3.

Table 9
Vectors of weights representing the user's value functions.

	w_1	w_2	w_3	w_4	w_5
$\mathbf{w}_1$	0.2	0.2	0.2	0.2	0.2
$\mathbf{w}_2$	0.4567	0.2567	0.1567	0.09	0.04
$\mathbf{w}_3$	0.2567	0.1567	0.09	0.04	0.4567
$\mathbf{w}_4$	0.1567	0.09	0.04	0.4567	0.2567
$\mathbf{w}_5$	0.09	0.04	0.4567	0.2567	0.1567
$\mathbf{w}_6$	0.04	0.4567	0.2567	0.1567	0.09

4.2. Sensitivity analysis with respect to preference information

In this section we shall study how the scoring procedure is sensible to the amount of preference information provided by the DM (Dede et al., 2011, 2021; Puppo et al., 2021; Zio and Pedroni, 2012). To this aim, we considered the same problem presented in Section 4. We assume the existence of an artificial DM which ranks the seven alternatives at hand. The preferences of the artificial DM are represented by a weighted sum and we considered the six different weight vectors shown in Table 9.

The six weight vectors in the Table are well-distributed in the polyhedron W. Apart from $\mathbf{w}_1$ weighting equally the five criteria, following Paelinck (1974) (see also Corrente et al., 2014), $\mathbf{w}_1, \ldots, \mathbf{w}_5$, are the barycenters of the following subsets of W, respectively:

$$W_2 = \{(w_1, \ldots, w_5) \in W : w_1 \geqslant w_2 \geqslant w_3 \geqslant w_4 \geqslant w_5\}$$
$$W_3 = \{(w_1, \ldots, w_5) \in W : w_5 \geqslant w_1 \geqslant w_2 \geqslant w_3 \geqslant w_4\}$$
$$W_4 = \{(w_1, \ldots, w_5) \in W : w_4 \geqslant w_5 \geqslant w_1 \geqslant w_2 \geqslant w_3\}$$
$$W_5 = \{(w_1, \ldots, w_5) \in W : w_3 \geqslant w_4 \geqslant w_5 \geqslant w_1 \geqslant w_2\}$$
$$W_6 = \{(w_1, \ldots, w_5) \in W : w_2 \geqslant w_3 \geqslant w_4 \geqslant w_5 \geqslant w_1\}$$

Algorithm 2 Steps in the sensitivity analysis

For each artificial DM's value function represented by the weight vector $\mathbf{w}_j$, $j = 1, \ldots, 6$, compute the ranking of alternatives at hand by using the artificial DM's value function and denote it by $Ranking_{\mathbf{DM}}$

repeat

1: Elicit artificial DM's preference information in terms of k pairwise comparisons of alternatives, with $k = 1, \ldots, 6$,

2: Sample s value functions compatible with the artificial DM's preference information and compute the PWIs

3: Apply $\mathbf{ScPr}_{AVF}$ to get the ranking of the alternatives at hand and denote this ranking by $Ranking_{\mathbf{ScPr}_{AVF}}$

4: Compute the Kendall-Tau correlation coefficient between $Ranking_{\mathbf{DM}}$ and $Ranking_{\mathbf{ScPr}_{AVF}}$

until t runs have not been performed

5: Compute statistics on the obtained results

For each of the considered weight vectors, we evaluate how the recommendations obtained by $ScPr_{AVF}$ change with the number of pairwise comparisons provided by the DM. Algorithm 2 presents the steps that have to be performed for each weight vector.

These steps are described in the following lines:

1: Artificial DM's preferences are expressed by pairwise comparisons of alternatives. $k = 1$ means that the DM provides one pairwise comparison; $k = 2$ means that the DM provides two pairwise comparisons and so on; the maximum value we assume for k is 6 since $n - 1$ is the minimum number of pairwise comparisons sufficient to define a complete ranking of the n alternatives,[9]

2: To avoid to sample every time a different set of value functions compatible with the k pairwise comparisons, we use the following procedure: (i) at first, 1,000,000 value functions are sampled from the space W defined in Eq. (8). Let us denote the set composed of these functions by $\mathcal{U}$; (ii) we choose in $\mathcal{U}$ only the value functions compatible with the k pairwise comparisons elicited in step 2:. Let us denote the set composed of these value functions by $\mathcal{U}_k$. The PWIs are then computed on the basis of the value functions in $\mathcal{U}_k$,

3: $\mathbf{ScPr}_{AVF}$ described in Section 2.2.1 is applied. The computed value function is a general additive value function as in Eq. (1). This value function is then used to rank the alternatives at hand. The obtained ranking is denoted by $Rank_{\mathbf{ScPr}_{AVF}}$,

4: The Kendall-Tau coefficient between $Ranking_{DM}$ and $Rank_{\mathbf{ScPr}_{AVF}}$ is then computed.

Let us observe that the number of runs t that has been done for each value of k is $\binom{21}{k}$. Indeed, considering 7 alternatives it is possible to perform $\binom{7}{2} = 21$ pairwise comparisons between the alternatives under consideration. Therefore, the number of ways k pairs can be chosen from the 21 considered is $\binom{21}{k}$. This means that for $k = 1$, we perform 21 different runs, every time, considering one of the twenty one possible pairs; for $k = 2$, we perform $\binom{21}{2} = 210$ different runs, and so on, until for $k = 6$, we perform $\binom{21}{6} = 54{,}264$ different runs considering, every time, six different pairwise comparisons among the possible twenty one. In Table 10 we report, for each k and for each DM's value function, the average Kendall-Tau coefficient over the $t = \binom{21}{k}$ considered runs between $Ranking_{DM}$ and $Ranking_{\mathbf{ScPr}_{AVF}}$.

As one can see from the values in the Table, for each of the six considered DM's value functions, the average Kendall-Tau coefficient increases with k. Even in cases $\mathbf{w}_3$ and $\mathbf{w}_6$ in which the average Kendall-Tau coefficient for $k = 1$ is quite low,[10] increasing the number of pairwise comparisons implies an increasing of the average Kendall-Tau coefficient. This shows that the capacity of $\mathbf{ScPr}_{AVF}$ to reply the DM's ranking increases with the number of provided pairwise comparisons.

Table 10
Average Kendall-Tau coefficient between $Ranking_{DM}$ and $Ranking_{\mathbf{ScPr}_{AVF}}$.

# pairs	$\mathbf{w}_1$	$\mathbf{w}_2$	$\mathbf{w}_3$	$\mathbf{w}_4$	$\mathbf{w}_5$	$\mathbf{w}_6$
1	0.7732	0.8685	0.1474	0.6780	0.9410	0.1565
2	0.8463	0.9814	0.3887	0.7760	0.9651	0.3633
3	0.8595	0.9933	0.5631	0.7963	0.9727	0.5797
4	0.8722	0.9971	0.6667	0.8079	0.9782	0.7096
5	0.8828	0.9988	0.7318	0.8189	0.9832	0.7998
6	0.8916	0.9996	0.7759	0.8303	0.9854	0.8418

5. Conclusions

In this paper we have proposed a new scoring procedure. The procedure assigns a value to each alternative under consideration summarizing the Pairwise Winning Indices (PWIs) provided by the Stochastic Multicriteria Acceptability Analysis (SMAA; Lahdelma et al., 1998; Pelissari et al., 2020). The method builds an additive value function compatible with the PWIs that assigns a score to each alternative and that, for this reason, is called *compatible scoring function*. The idea under the proposal is that the difference between the scores assigned to two alternatives a and b by the compatible scoring function should be proportional to the difference $p(a, b) - 0.5$, that is, to the excess over 0.5 of the probability with which a is considered at least as good as b. Consequently, the greater $p(a, b)$, the larger the difference of the scores attributed to a and b. The compatible scoring function is obtained solving a simple LP problem. In case more than one compatible scoring function maximally discriminating among the alternatives exists, we show an iterative procedure aiming to find a well-diversified sample of them. Moreover, some LP problems are presented to discover, among the maximally discriminant compatible scoring functions, some of them presenting particular characteristics: (i) functions for which all criteria contribute to the global score assigned to the alternatives from the built function, or (ii) functions for which the marginal value functions are strictly monotone.

Several other methods summarizing the information contained in the PWIs have been proposed in literature before. However, differently from our proposal, they aim only to rank order all alternatives from the best to the worst. Instead, our proposal has the advantage that the score assigned to each alternative is based on the construction of an additive value function. Through its marginal value functions, it allows the DM to get some explanations on the reasons for which an alternative fills a given position and obtains its specific score. In particular, the ranking position and the score of each alternative can be explained in terms of the contributions given by each criterion to the global value assigned to the considered alternative. This is a great advance from the decision aiding point of view since the scoring function we are proposing answers to the explainability concerns being nowadays very relevant for any decision aiding method (see, e.g. Arrieta et al., 2020).

[9] It is enough asking the DM to compare the preferred alternative with the second preferred one; to compare the second preferred to the third preferred, and so on, until the $(n-1)$th preferred is compared to the nth.

[10] This means that, one pairwise comparison is not enough to restore the DM's alternatives ranking.

To justify the proposed scoring procedure, we presented also a probabilistic model based on the assumption that the values that can be assigned to each alternative have a normal distribution with the assigned score as mean and with a common standard deviation. To take into account this distribution in the linear programming model defined to assign the score, we proposed a simple piecewise linear approximation of the cumulative normal distribution, which we believe has an independent interest that goes beyond the proposed method. Moreover, to prove that the new proposal, beyond explaining the rank position and the score of the alternatives, is efficient in predicting the preferences of the DM, we performed a large set of computational experiments. We compared our scoring procedure to other sixteen methods that have been proposed in literature and that represent the state of the art in this field. We simulated an artificial DM in problems composed of n alternatives and m criteria trying to replicate the ranking produced by the artificial DM itself. We considered 6, 9, 12 and 15 alternatives and 3, 5 and 7 criteria. For robustness reasons, for each (n, m) configuration, we performed 10,000 independent runs applying the sixteen mentioned methods and our scoring procedure. To check how efficient the methods are in replicating the ranking of the artificial DM, we computed the Kendall-Tau between the preference ranking of the artificial DM and the ranking produced by each considered method. These Kendall-Tau values are then averaged over the 10,000 independent runs.

The results show that even if our proposal is not getting the best value (the maximum average Kendall-Tau) for any of the considered configurations, the "deviation" from the best average Kendall-Tau value is always lower than 2.5%. To check if the difference between the Kendall-Tau values of our proposal and the ones of the best method for each (n, m) configuration is significant from the statistical point of view, we performed a Kolmogorov–Smirnov test with 5% significance level. The test shows that in the considered configurations the difference is never statistically significant with very few exceptions. This means that our scoring procedure is able to reproduce the preferences of the artificial DM and, at the same time, differently from all the other methods, it is able to give an explanation of the reasons giving to the alternatives a certain rank position.

Finally, we have shown how to apply the new scoring procedure to a financial problem in which seven funds are evaluated with respect to five different criteria underlying the potentialities of the proposed method in explaining how the criteria contribute to the global value assigned to the alternatives. Moreover, we checked how the recommendations given by our scoring procedure are sensible to the number of preferences introduced by the DM. Assuming the existence of an artificial DM, we have shown that the similarity between the ranking produced by our scoring procedure and the one produced by the artificial DM increases in average with the number of pairwise comparisons produced by his/her.

As further directions of research we plan to apply the new proposal to some real world decision problems to which SMAA has been applied and for which a final ranking of the alternatives under consideration has to be produced. Moreover, how to extend the scoring procedure to summarize the PWIs obtained in a problem presenting a hierarchical structure of criteria (see, for example, Corrente et al., 2017) deserves to be investigated.

Data availability

No data was used for the research described in the article.

Acknowledgments

The authors wish to acknowledge the support of the Ministero dell'Istruzione, dell'Universitá e della Ricerca (MIUR) - PRIN 2017, project "Multiple Criteria Decision Analysis and Multiple Criteria Decision Theory", grant 2017CY2NCA. Moreover, Salvatore Corrente wishes to acknowledge the support of the STARTING GRANT of the University of Catania.

References

Amenc, N., Le Sourd, V., 2005. Portfolio Theory and Performance Analysis. John Wiley & Sons.
Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115.
Arrow, K., Sen, A., Suzumura, K., 2010. Handbook of Social Choice and Welfare, Vol. 2. Elsevier.
Bacmann, J., Scholz, S., 2003. Alternative performance measures for hedge funds. AIMA J. 1 (1), 1–9.
Bacon, C., 2012. Practical Risk-Adjusted Performance Measurement. John Wiley & Sons.
Branke, J., Corrente, S., Greco, S., Gutjahr, W., 2017. Efficient pairwise preference elicitation allowing for indifference. Comput. Oper. Res. 88, 175–186.
Condorcet, J.-M., 1785. Essai sur l'Application de l'Analyse á la Probabibité des Décisions Rendues á la Pluralitédes Voix. Imprimerie Royale, Paris.
Corrente, S., Figueira, J., Greco, S., 2014. The SMAA-PROMETHEE method. European J. Oper. Res. 239 (2), 514–522.
Corrente, S., Figueira, J., Greco, S., Słowiński, R., 2017. A robust ranking method extending ELECTRE III to hierarchy of interacting criteria, imprecise weights and stochastic analysis. Omega 73, 1–17.
Corrente, S., Greco, S., Kadziński, M., Słowiński, R., 2013. Robust Ordinal Regression in preference learning and ranking. Mach. Learn. 93, 381–422.
Dede, G., Kamalakis, T., Anagnostopoulos, D., 2022. A framework of incorporating confidence levels to deal with uncertainty in pairwise comparisons. CEJOR Cent. Eur. J. Oper. Res. 30 (3), 1051–1069.
Dede, G., Kamalakis, T., Sphicopoulos, T., 2016. Theoretical estimation of the probability of weight rank reversal in pairwise comparisons. European J. Oper. Res. 252 (2), 587–600.
Dede, G., Kamalakis, T., Varoutas, D., 2011. Evaluation of optical wireless technologies in home networking: An analytical hierarchy process approach. J. Opt. Commun. Netw. 3 (11), 850–859.
Dede, G., Mitropoulou, P., Nikolaidou, M., Kamalakis, T., Michalakelis, C., 2021. Safety requirements for symbiotic human–robot collaboration systems in smart factories: a pairwise comparison approach to explore requirements dependencies. Requir. Eng. 26 (1), 115–141.
Dodgson, C., 1876. A Method of Taking Votes on More than Two Issues. Clarendon Press, Oxford.
Fürnkranz, J., Hüllermeier, E. (Eds.), 2010. Preference Learning. Springer, Berlin.
Greco, S., Ehrgott, M., Figueira, J., 2016. Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, Berlin.
Greco, S., Ishizaka, A., Matarazzo, B., Torrisi, G., 2018. Stochastic multi-attribute acceptability analysis (SMAA): an application to the ranking of Italian regions. Reg. Stud. 52 (4), 585–600.
Greco, S., Mousseau, V., Słowiński, R., 2008. Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. European J. Oper. Res. 191 (2), 416–436.
Jacquet-Lagreze, E., Siskos, Y., 1982. Assessing a set of additive utility functions for multicriteria decision-making, the UTA method. European J. Oper. Res. 10 (2), 151–164.
Jacquet-Lagreze, E., Siskos, Y., 2001. Preference disaggregation: 20 years of MCDA experience. European J. Oper. Res. 130 (2), 233–245.
Jensen, M., 1968. The performance of mutual funds in the period 1945–1964. J. Finance 23 (2), 389–416.
Kadziński, M., Greco, S., Słowiński, R., 2012a. Selection of a representative set of parameters for robust ordinal regression outranking methods. Comput. Oper. Res. 39 (11), 2500–2519.
Kadziński, M., Greco, S., Słowiński, R., 2012b. Selection of a representative value function in robust multiple criteria ranking and choice. Eur. J. Oper. Res. 217 (3), 541–553.
Kadziński, M., Michalski, M., 2016. Scoring procedures for multiple criteria decision aiding with robust and stochastic ordinal regression. Comput. Oper. Res. 71, 54–70.
Keeney, R., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. J. Wiley, New York.
Kendall, M., 1938. A new measure of rank correlation. Biometrika 30 (1/2), 81–93.
Lahdelma, R., Hokkanen, J., Salminen, P., 1998. SMAA - Stochastic multiobjective acceptability analysis. European J. Oper. Res. 106 (1), 137–143.
Leskinen, P., Viitanen, J., Kangas, A., Kangas, J., 2006. Alternatives to incorporate uncertainty and risk attitude in multicriteria evaluation of forest plans. For. Sci. 52 (3), 304–312.
Massey, Jr., F., 1951. The Kolmogorov-Smirnov test for goodness of fit. J. Amer. Statist. Assoc. 46 (253), 68–78.
Mousseau, V., Figueira, J., Dias, L., Gomes da Silva, C., Climaco, J., 2003. Resolving inconsistencies among constraints on the parameters of an MCDA model. European J. Oper. Res. 147 (1), 72–93.

Paelinck, J., 1974. Qualitative multiple criteria analysis, environmental protection and multiregional development. Pap. Reg. Sci. Assoc. 36, 59–74.

Pelissari, R., Oliveira, M., Ben Amor, S., Kandakoglu, A., Helleno, A., 2020. SMAA methods and their applications: a literature review and future research directions. Ann. Oper. Res. 293, 433–493.

Puppo, L., Pedroni, N., Bersano, A., Di Maio, F., Bertani, C., Zio, E., 2021. Failure identification in a nuclear passive safety system by Monte Carlo simulation with adaptive Kriging. Nucl. Eng. Des. 380, 111308.

Rockafellar, R., Uryasev, S., 2000. Optimization of conditional value-at-risk. J. Risk 2, 21–42.

Roy, B., 1993. Decision science or decision-aid science? European J. Oper. Res. 66 (2), 184–203.

Roy, B., 1996. Multicriteria Methodology for Decision Aiding. Kluwer Academic Publishers, Dordrecht.

Roy, B., 2010. Robustness in operational research and decision aiding: A multi-faceted issue. European J. Oper. Res. 200 (3), 629–638.

Rubinstein, R., 1982. Generating random vectors uniformly distributed inside and on the surface of different regions. European J. Oper. Res. 10 (2), 205–209.

Scholz, H., Wilkens, M., 2005. Investor-specific performance measurement: A justification of Sharpe Ratio and Treynor Ratio. Int. J. Finance 17 (4), 3671.

Sharpe, W., 1998a. Morningstar's risk-adjusted ratings. Financ. Anal. J. 54 (4), 21–33.

Sharpe, W., 1998b. The Sharpe ratio. In: Streetwise–The Best of the Journal of Portfolio Management. pp. 169–185.

Simpson, P., 1969. On defining areas of voter choice: Professor Tullock on stable voting. Q. J. Econ. 83 (3), 478–490.

Smith, R., 1984. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. Oper. Res. 32, 1296–1308.

Tervonen, T., Van Valkenhoef, G., Bastürk, N., Postmus, D., 2013. Hit-And-Run enables efficient weight generation for simulation-based multiple criteria decision analysis. European J. Oper. Res. 224, 552–559.

Van Valkenhoef, G., Tervonen, T., Postmus, D., 2014. Notes on "Hit-And-Run enables efficient weight generation for simulation-based multiple criteria decision analysis". European J. Oper. Res. 239, 865–867.

Vetschera, R., 2017. Deriving rankings from incomplete preference information: A comparison of different approaches. European J. Oper. Res. 258, 244–253.

Zio, E., Pedroni, N., 2012. Monte Carlo simulation-based sensitivity analysis of the model of a thermal-hydraulic passive system. Reliab. Eng. Syst. Saf. 107, 90–106.