Expert Systems With Applications 214 (2023) 119045

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A robust TOPSIS method for decision making problems with hierarchical and non-monotonic criteria

Salvatore Corrente [a,*], Menelaos Tasiou [b]

[a] Department of Economics and Business, University of Catania, Corso Italia 55, 95129, Catania, Italy
[b] Portsmouth Business School, University of Portsmouth, Richmond Building, Portland Street, Portsmouth, Hampshire PO1 3DE, UK

ARTICLE INFO

Keywords:
TOPSIS
Non-monotonic criteria
Qualitative criteria
Normalization
Hierarchy of criteria
Robustness concerns

ABSTRACT

This paper introduces an extension of a well-known Multiple Criteria Decision Aiding method, namely the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). Most of the TOPSIS applications assume that preferences are monotonic for each evaluation criterion and that qualitative scales are converted into quantitative ones before the method is applied. However, both assumptions have been subject of discussion and criticism in the literature. To this solution, this paper introduces a normalization technique based on simulations that permit taking into account non-monotonic preferences as well as qualitative criteria. An additional novelty lies in the integration of the Multiple Criteria Hierarchy Process, which extends the applicability of the method to problems in which criteria are hierarchically structured. To deal with robustness concerns, the Stochastic Multicriteria Acceptability Analysis will be used in the new proposal, giving information in statistical terms on the goodness of the considered alternatives. The new method has been applied to evaluate a set of banks listed in the LSE's FTSE350 Index.

1. Introduction

The 'Technique for Order Preference by Similarity to Ideal Solution' (TOPSIS) is a Multiple Criteria Decision Aiding (MCDA; Greco et al. (2016)) method proposed by Hwang and Yoon (1981) to deal with ranking problems. The logic behind this method is very close to the way people choose, as the best alternatives are those close to the positive ideal solution (PIS) whilst at the same time distant from the negative ideal solution (NIS) (Kim et al., 1997). This method is based on a very simple working theory, resulting in it being easily understood by decision makers (Jun et al., 2017). For this reason, as acknowledged by Behzadian et al. (2012), the application of TOPSIS in the field of MCDA is widespread.

Despite the extensive application of this method and its intuitive reasoning, several steps of the underlying algorithm have been criticized for a variety of reasons. These regard the normalization of the evaluations, the weighting of criteria, the norm used to compute the distance of each alternative from the PIS and NIS (Shih et al., 2007), and the way these distances are being put together to define the closeness index (Kuo, 2017). In this paper, we shall extensively deal with the normalization issue. Moreover, we shall propose an extension of the TOPSIS method that is able to take into account a hierarchically structured set of criteria, as well as to consider a plurality of weight vectors compatible with the information provided by the Decision Maker (DM). These are discussed in more detail forthwith.

Starting with the normalization step, this is used to express all performances in a unique common scale. Several normalization techniques have been applied, the most well-known being the vector normalization, the max-normalization, the min–max normalization and the sum-normalization (see, for example, Shih et al. (2007)). Arguably, all of them reach the final scope. However, the choice of the normalization procedure to be applied is quite important in TOPSIS, as it may radically alter the obtained results. What is more, some of these normalization techniques, in particular the vector normalization, are quite sensible to the unit scale of the criteria, so that, if the same criterion is expressed in different units, the application of the same normalization technique can produce different results (Opricovic & Tzeng, 2004). For this reason, many studies over the years have highlighted the impact of the different normalization techniques on the rankings produced by TOPSIS (see, for example, Acuña-Soto et al. (2018), Çelen (2014), Chatterjee and Chakraborty (2014), Milani et al. (2005) and Vafaei et al. (2018)).

Other two implicit assumptions of TOPSIS are that: (i) criteria are expressed in quantitative scales; (ii) the same criteria have a monotonic

* Corresponding author.
E-mail addresses: salvatore.corrente@unict.it (S. Corrente), Menelaos.Tasiou@port.ac.uk (M. Tasiou).

https://doi.org/10.1016/j.eswa.2022.119045
Received 19 April 2021; Received in revised form 7 May 2022; Accepted 12 October 2022
Available online 21 October 2022

(increasing or decreasing) direction of preference [1]. Indeed, in many real world decision making problems criteria are expressed in qualitative or ordinal scales and the preferences of the DM are non-monotonic with respect to the same evaluations. For example, in evaluating a car, the criterion ‘comfort’ can be expressed on a qualitative scale, the levels of which are e.g. *very low*, *low*, *medium*, *high* and *very high*; that is a simple monotonic type of evaluation. Now, consider evaluating a town for the summer holidays, on the basis of *mean temperature*; of course, the criterion is arguably non-monotonic. Although considering criteria expressed in qualitative terms or presenting a non-monotonic direction of preference is really important in MCDA (see e.g. Ghaderi et al. (2017) and Kadziński et al. (2020) for some recent papers related to the consideration of non-monotonic criteria in MCDA), very few contributions to TOPSIS deal with these two issues. On the one hand, the few papers related to the use of qualitative criteria in TOPSIS imply a transformation of the qualitative evaluations into quantitative ones (Shih et al., 2007). However, this transformation can, to some extent, be considered arbitrary since the difference among qualitative levels does not have a clear meaning (Roy, 2005). Moreover, the difference among two consecutive levels is not necessarily the same along the whole scale of the considered criterion — for example, in the case of the comfort criterion mentioned above, the difference between *very low* and *low* does not have to be necessarily equal to the difference between *high* and *very high*. On the other hand, the papers considering non-monotonic criteria in TOPSIS (Cables et al., 2016; Shih et al., 2007) are anyway subject to the normalization problem discussed above since a transformation of the original evaluation into a common scale has to be performed and, therefore, a normalization procedure has to be chosen.

In this paper, we shall propose to replace the normalizations that are generally used in TOPSIS with an alternative technique that has been presented by Angilella et al. (2015) for the Choquet integral preference model (Choquet, 1953; Grabisch, 1996). The considered normalization technique can be used indifferently to deal with criteria that are expressed in quantitative or qualitative scales, as well as criteria with a monotonic or non-monotonic direction of preference. It is based on a sampling procedure that only takes into account the ordinal information of the alternatives’ performances, not their intrinsic value. For this reason, we apply the normalization technique in an iterative way, thus simulating at each iteration a different normalization technique.

The other two issues taken into account in the paper are the hierarchical structure of criteria and the choice of their weights. Starting with the former, the original TOPSIS method provides recommendations by simultaneously taking into account all the considered criteria. However, in many MCDA problems, criteria are structured hierarchically. For this reason we will integrate the Multiple Criteria Hierarchy Process (MCHP; Corrente et al. (2012)) in the TOPSIS method. The new hierarchical version of the TOPSIS method will provide information not only at a comprehensive level (that is, considering all criteria at the same time as in the regular version of TOPSIS) but also at all levels of the hierarchy, thus focusing on a particular macrocriterion in which the DM is more interested. In particular, the integration of the MCHP in the TOPSIS method will permit to compute a ranking of the alternatives at each node of the hierarchy.

As in other MCDA approaches, TOPSIS also implies the definition of a weight vector denoting the importance of each criterion. Several methods have been applied to elicit weights in an objective (Hwang & Yoon, 1981) or subjective (Bhutia & Phipon, 2012; Yurdakul & IÇ, 2005) manner. However, the choice of the weight vector is an important step as the application of TOPSIS with different weight vectors could provide different results (Li et al., 2013; Olson, 2004). For this reason, in Okul et al. (2014) the SMAA-TOPSIS method has been proposed, integrating the Stochastic Multiobjective Acceptability Analysis (SMAA; Lahdelma et al. (1998)) in TOPSIS to provide robust recommendations by considering a plurality of weight vectors compatible with some preferences provided by the DM. Given the prominent benefits of SMAA-TOPSIS, in this paper, we shall also apply this method.

To sum up, we are proposing an extension of the TOPSIS method having a great relevance both from a theoretical as well as from a practical point of view. Indeed, the new proposal is able to overtake some drawbacks of the original method that have been criticized over the years, while being able to deal with many different issues of decision making problems, making it more applicable from its original version. The new method has the following important characteristics:

- *it iteratively applies a procedure to put all alternatives’ evaluations on the same scale*: the procedure permits to take easily into account criteria expressed on qualitative or ordinal scales as well as criteria having a non-monotonic direction of preference. Moreover, the application of the procedure in an iterative way permits to avoid the choice of a single normalization technique that, as acknowledged before, could be arbitrary and could affect the final results of the method;
- *it applies the MCHP to deal with problems in which criteria are hierarchically structured:* this application permits the DM to have a more in depth view of the problem under investigation. The original TOPSIS method provides a ranking of the alternatives at hand taking into account all criteria simultaneously. However, getting a ranking for each macro-criterion permits to have a more precise information on the strengths and weaknesses of each considered alternative;
- *it applies the SMAA methodology:* the new method produces results on the considered problem taking into account not only one weight vector but a plurality of weight vectors compatible with some preferences provided by the DM in indirect terms. In this way, it avoids the choice of a single weight vector and answers to robustness concerns.

The remaining of this paper is structured as follows: Section 2 provides the methodological preliminaries this proposal is based on. Section 3 presents the novel proposal and outlines its applicability on a step by step basis. Section 4 provides a simulated evaluation illustrating the applicability of the proposed method in evaluating a set of banks on the basis of financial and non-financial criteria, and Section 5 concludes this study.

2. Methodological background

2.1. TOPSIS

Let us consider a ranking problem in which alternatives of $A = \{a_1, \ldots, a_n\}$ are evaluated on criteria in $G = \{g_1, \ldots, g_m\}$. For brevity, the evaluation of alternative $a_i \in A$ on criterion $g_j \in G$ will be denoted by a_{ij} and we assume that $g_j : A \rightarrow \mathbb{R}$ for all $j = 1, \ldots, m$. However, this assumption will not be necessary to apply the TOPSIS method in our proposal as it has been introduced before and as it will be clarified in the next section. In our context, $g_j : A \rightarrow E_j$, where E_j is the scale of criterion g_j so that $E_j \subseteq \mathbb{R}$ iff g_j is expressed on a quantitative scale, while E_j is a set of qualitative levels iff g_j is expressed on a qualitative scale. A decision matrix will collect the evaluations of the alternatives in A on the criteria in G. The application of TOPSIS implies the following steps:

[1] A criterion has an increasing direction of preference if the higher the evaluation of an alternative is on that criterion, the better the alternative performs. Vice versa, a criterion has a decreasing direction of preference if the lower the evaluation of an alternative on that criterion is, the better the alternative is deemed to perform

1. *Normalization*[2]*:* for each criterion $g_j \in G$ and for each $a_i \in A$, the normalized value of a_{ij} is obtained as $z_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^{n} a_{ij}^2}}$,
2. *Weighting:* denoting by $\boldsymbol{w} = [w_1, \ldots, w_m]$ the vector composed of the importance of criteria in G, such that $w_j > 0$ for all $g_j \in G$ and $\sum_{j=1}^{m} w_j = 1$, for each $g_j \in G$ and for each $a_i \in A$; the normalized weighted value v_{ij} is obtained as $v_{ij} = z_{ij} \cdot w_j$,
3. *Positive Ideal Solution (PSI) and Negative Ideal Solution (NIS):* denoting by G_I and G_D the subsets of G composed of the increasing and decreasing criteria respectively, $PIS = A^+ = \left(v_1^+, \ldots, v_m^+\right)$ and $NIS = A^- = \left(v_1^-, \ldots, v_m^-\right)$ are computed so that

$$v_j^+ = \begin{cases} \max\limits_{i=1,\ldots,n} v_{ij} & \text{if} \quad g_j \in G_I, \\ \min\limits_{i=1,\ldots,n} v_{ij} & \text{if} \quad g_j \in G_D, \end{cases} \qquad \text{and}$$

$$v_j^- = \begin{cases} \min\limits_{i=1,\ldots,n} v_{ij} & \text{if} \quad g_j \in G_I, \\ \max\limits_{i=1,\ldots,n} v_{ij} & \text{if} \quad g_j \in G_D. \end{cases}$$

4. *Distance from PIS and NIS:* for each $a_i \in A$, the distance from the PIS and NIS denoted respectively by $d^+(a_i)$ and $d^-(a_i)$ are computed:

$$d^+(a_i) = \sqrt{\sum_{j=1}^{m} \left(v_{ij} - v_j^+\right)^2} \qquad \text{and} \qquad d^-(a_i) = \sqrt{\sum_{j=1}^{m} \left(v_{ij} - v_j^-\right)^2} \tag{1}$$

5. *Relative closeness to PIS and NIS and ranking:* for each $a_i \in A$ the relative closeness $C(a_i)$ to the PIS and NIS is computed as $C(a_i) = \frac{d^-(a_i)}{d^+(a_i)+d^-(a_i)}$. $C(a_i) \in [0,1]$, it is a decreasing function of $d^+(a_i)$ and an increasing function of $d^-(a_i)$. The alternatives are therefore ranked from the best to the worst with respect to decreasing values of $C(a_i)$.

2.2. Multiple Criteria Hierarchy Process

In real world applications, the evaluation criteria are not always at the same level; they are generally structured in a hierarchical manner. It is therefore possible to highlight the *root* criterion, i.e. the objective of the decision making problem, some *macro*criteria descending from the root criterion, and the hierarchy keeps unfolding accordingly until the very bottom of the hierarchy, where the elementary criteria are located, i.e. those criteria on which the alternatives are evaluated. The Multiple Criteria Hierarchy Process (MCHP; Corrente et al. (2012)) is a methodology that was recently introduced in MCDA to deal with problems in which criteria have the above mentioned hierarchical structure. The use of the MCHP permits to decompose the problem into sub-problems that separately take into account different aspects the DM is more interested in, thus avoiding to take into account all elementary criteria simultaneously.

From a formal point of view, G is the set composed of all criteria in the hierarchy; I_G is the set of the indices of criteria in G; $G_E \subseteq G$ is the set of elementary criteria — that is, the criteria at the bottom of the hierarchy and based on which the performances of the alternatives are provided; $E_G \subseteq I_G$ is the set of indices of the elementary criteria; g_0 is the root criterion; $g_{\mathbf{t}}$ with $\mathbf{t} \in E_G$ is an elementary criterion, while $g_{\mathbf{r}}$, with $\mathbf{r} \in I_G \setminus \{E_G\}$, is a non-elementary criterion; given a non-elementary criterion $g_{\mathbf{r}}$, its subcriteria at the level immediately below are denoted by $g_{(\mathbf{r},1)}, \ldots, g_{(\mathbf{r},n(\mathbf{r}))}$, while $E\left(g_{\mathbf{r}}\right) \subseteq E_G$ is the set of indices of the elementary criteria descending from $g_{\mathbf{r}}$.

The application of the MCHP to a ranking problem gives a complete order of the alternatives on each non-elementary criterion. Given $g_{\mathbf{r}}$, $\mathbf{r} \in I_G \setminus E_G$, a preference $\succ_{\mathbf{r}}$ and an indifference $\sim_{\mathbf{r}}$ relations are defined so that $a_i \succ_{\mathbf{r}} a_k$ iff a_i is preferred to a_k on $g_{\mathbf{r}}$ and $a_i \sim_{\mathbf{r}} a_k$ iff a_i is indifferent to a_k on $g_{\mathbf{r}}$.

2.3. Stochastic Multicriteria Acceptability Analysis

As it has been described in Section 2.1, the application of the TOPSIS methodology implies the knowledge of the weight vector $[w_1, \ldots, w_m]$. In order to elicit the values composing this vector, a direct or an indirect technique can be used in MCDA (Jacquet-Lagrèze & Siskos, 2001). In the direct case, the DM is able to provide exact values for the parameters of the model (the weights in our context), while in the indirect one, the DM provides some preference information from which values of the parameters compatible with such preferences can be inferred. Generally, the indirect technique is mainly used as it requires less cognitive effort on behalf of the DM. Two families of methodologies implementing the indirect technique are mostly used in MCDA; these are, the Robust Ordinal Regression (ROR; Greco et al. (2008) and Corrente, Greco, Kadziński et al. (2013)) and the Stochastic Multicriteria Acceptability Analysis (SMAA; Lahdelma et al. (1998) and Pelissari et al. (2020)). Both explore the whole set of parameters compatible with the preferences provided by the DM, even if accomplished in different ways. In this paper we shall apply the SMAA methodology and, for this reason, we shall briefly recall its main concepts and notation taking into account a ranking problem considering the MCHP framework discussed in the previous sub-section.

The use of SMAA is based on the knowledge of two distributions f_χ and f_W on the evaluation space χ and on the weight space W, respectively. In this paper, we shall assume that the performances of the considered alternatives are fixed and, consequently, we shall only take into account the weight variability. For each $a_i \in A$, for each non-elementary criterion $g_{\mathbf{r}}$, and for each $w \in W$ following the distribution f_W, a rank function is defined so that

$$rank_{\mathbf{r}}(a_i, w) = 1 + \sum_{k=1}^{n} \rho(a_k \succ_{\mathbf{r}} a_i) \tag{2}$$

where $\rho(true) = 1$ and $\rho(false) = 0$. Moreover, for each $g_{\mathbf{r}}$, $\mathbf{r} \in I_G \setminus E_G$, and for each rank position $s \in \{1, \ldots, n\}$, the set $W_{\mathbf{r}}^s(a_i) \subseteq W$ comprised of all the weight vectors for which the considered preference model gives a_i the position s with respect to $g_{\mathbf{r}}$ is computed:

$$W_{\mathbf{r}}^s(a_i) = \left\{ w \in W \; : \; rank_{\mathbf{r}}(a_i, w) = s \right\}. \tag{3}$$

The extension of SMAA to the MCHP thus permits to obtain for each non-elementary criterion $g_{\mathbf{r}}$ the following indices:

- *rank acceptability index,* $b_{\mathbf{r}}^s(a_i)$: it gives the frequency with which a_i fills the position s in the ranking obtained on $g_{\mathbf{r}}$ and it is computed as

$$b_{\mathbf{r}}^s(a_i) = \int_{w \in W_{\mathbf{r}}^s(a_i)} f_W(w)\, dw. \tag{4}$$

  $b_{\mathbf{r}}^s(a_i) \in [0,1]$ and the best alternatives are those presenting higher values of $b_{\mathbf{r}}^s(a_i)$ for the first positions, and low values of $b_{\mathbf{r}}^s(a_i)$ for the last positions in the ranking;
- *pairwise winning index,* $p_{\mathbf{r}}(a_i, a_k)$ (Leskinen et al., 2006): it gives the frequency with which alternative a_i is preferred to alternative a_k on $g_{\mathbf{r}}$ and it is computed as

$$p_{\mathbf{r}}(a_i, a_k) = \int_{w \in W \,:\, a_i \succ_{\mathbf{r}} a_k} f_W(w)\, dw. \tag{5}$$

The multidimensional integrals used to compute the rank acceptability indices and the pairwise winning indices are approximated by Monte Carlo simulations. The rank acceptability indices as well as the pairwise winning indices provide robust recommendations taking into account

[2] We will hereby consider the vector normalization just for explanatory purposes. However, in the following we shall not use any normalization technique often considered in TOPSIS.

the plurality of models compatible with the preferences provided by the DM. However, they do not give a single representative ranking of the alternatives at hand. For this reason, several procedures have been proposed to summarize the data provided by these two indices (see, for example, Kadziński and Michalski (2016) and Vetschera (2017)). Following Kadziński and Michalski (2016), in this paper, we shall aggregate the rank acceptability indices values computing the expected raking of each alternative a_i on each non-elementary criterion $g_{\mathbf{r}}$ on the basis of the following index:

$$ER_{\mathbf{r}}(a_i) = -\sum_{s=1}^{n} s \cdot b^s_{\mathbf{r}}(a_i). \tag{6}$$

$ER_{\mathbf{r}}(a_i)$ is expressed as a sum of the $b^s_{\mathbf{r}}(a_i)$ having corresponding weight $-s$. Of course, the higher the value $ER_{\mathbf{r}}(a_i)$, the better the alternative a_i performs on criterion $g_{\mathbf{r}}$. The computation of $ER_{\mathbf{r}}(a_i)$ will provide a total order of the considered alternatives on $g_{\mathbf{r}}$.

3. The new proposal

3.1. The normalization step

As described in Section 2.1, the first step in the application of TOPSIS is the normalization of the evaluations. However, three different issues are associated with this step:

1. Alternative normalization schemes generally provide alternative recommendations on the problem at hand. Even if normalization is one important aspect in several MCDA problems, the recommendations provided by the method could be sensible to the performed normalization. In the TOPSIS method, different normalization techniques can be used and the most applied ones are the vector normalization (described in the previous section), the linear max–min normalization, the linear max normalization and the linear sum normalization. Several studies delve into the effects of normalization in TOPSIS (see, for example, Çelen (2014)),
2. As stated by Hwang and Yoon (1981, p.130), *"TOPSIS assumes that each attribute in the decision matrix takes either monotonically increasing or monotonically decreasing utility"*. Despite the fact that non-monotonic criteria are generally used in real world decision making problems (several authors worked on this topic in other MCDA methods see, for example, Despotis and Zopounidis (1995), Ghaderi et al. (2017) and Kadziński et al. (2020)), in our opinion, this problem has not been adequately considered under the TOPSIS framework,
3. All the normalization techniques mentioned in step 1 can be used in the case that alternatives' evaluations are discrete. This means that if the evaluations are expressed in a qualitative scale, for example *bad*, *medium*, *good*, etc, none of the considered normalization techniques in the TOPSIS method can be used without an *a priori* translation of the qualitative evaluations into quantitative ones (Shih et al., 2007). However, this translation can present additional problems because, on one hand, the difference between qualitative levels has not a real meaning (Roy, 2005), while, on the other hand, it assumes that the difference between successive grades is always the same,[3] which implies more than what the qualitative scales actually allow to do (Cinelli et al., 2020).

To solve the three above mentioned problems, we shall propose to apply a normalization technique presented by Angilella et al. (2015) for the application of the Choquet integral preference model (Choquet, 1953). Let us recall the steps of the considered normalization technique for each criterion $g_j \in G$:

- Order the alternatives in A from the best to the worst (with the possibility of some ex aequo) with respect to the preferences of the DM on criterion g_j, that is, $a_{(1)}, a_{(2)}, \ldots, a_{(n)}$, where $(\cdot)$ is a permutation of the indices of the alternatives so that $a_{(i)} \succ_j a_{(i+1)}$ ($a_{(i)}$ is preferred to $a_{(i+1)}$ on criterion g_j) for $i = 1, \ldots, n-1$; in the description we shall assume that there are not indifferent alternatives on g_j but, of course, this is not an assumption of the method,
- Sample n different real numbers $x_1, x_2, \ldots, x_n$ in the interval $[0,1]$ and reorder them in a decreasing way, $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ where $(\cdot)$ is a permutation of the indices $\{1,\ldots,n\}$ so that $x_{(1)} > x_{(2)} > \cdots > x_{(n)}$,
- Assign $x_{(i)}$ to $a_{(i)}$ for all $i = 1, \ldots, n$, that is, assign the greatest value $\left(x_{(1)}\right)$ to the preferred alternative $\left(a_{(1)}\right)$, the second greatest value $\left(x_{(2)}\right)$ to the second preferred alternative $\left(a_{(2)}\right)$, and so on until the lowest value $\left(x_{(n)}\right)$ is assigned to the worst alternative $\left(a_{(n)}\right)$.

Our claim is that the considered procedure can represent both monotonic and not-monotonic preferences of the DM as shown in the following example.

Let us consider 5 alternatives $a_1, \ldots, a_5$ evaluated on criterion g_j as follows: $g_j(a_1) = 1$, $g_j(a_2) = 2$, $g_j(a_3) = 3$, $g_j(a_4) = 4$ and $g_j(a_5) = 5$ and the following three cases:

Case (1) the DM orders the alternatives from the best to the worst as $a_5 \succ_j a_4 \succ_j a_3 \succ_j a_2 \succ_j a_1$; that is $a_{(1)} = a_5$, $a_{(2)} = a_4$, $a_{(3)} = a_3$, $a_{(4)} = a_2$ and $a_{(5)} = a_1$,

Case (2) the DM orders the alternatives from the best to the worst as $a_1 \succ_j a_2 \succ_j a_3 \succ_j a_4 \succ_j a_5$; that is $a_{(i)} = a_i$, for all $i = 1, \ldots, 5$,

Case (3) the DM orders the alternatives from the best to the worst as $a_3 \succ_j a_4 \succ_j a_2 \succ_j a_5 \succ_j a_1$; that is $a_{(1)} = a_3$, $a_{(2)} = a_4$, $a_{(3)} = a_2$, $a_{(4)} = a_5$ and $a_{(5)} = a_1$.

Let us use the procedure described above to normalize the evaluations of the 5 alternatives on g_j. Since there is no indifference between consecutive alternatives in the three cases, we have to sample 5 real values in the interval $[0,1]$. Let us suppose the sampled values are the following: $x_1 = 0.8147$, $x_2 = 0.9058$, $x_3 = 0.127$, $x_4 = 0.9134$ and $x_5 = 0.6324$.

Ordering these values in a decreasing manner we have $x_{(1)} = 0.9134 = x_4$, $x_{(2)} = 0.9058 = x_2$, $x_{(3)} = 0.8147 = x_1$, $x_{(4)} = 0.6324 = x_5$ and $x_{(5)} = 0.127 = x_3$.

The values are therefore assigned to the five alternatives as given in Table 1 and as shown in Figs. 1–3.

Table 1
Normalization of the evaluations of the five alternatives on g_j in the considered cases.

Evaluation	Normalized value		
	Case 1	Case 2	Case 3
$g_j(a_1) = 1$	0.127	0.9134	0.127
$g_j(a_2) = 2$	0.6324	0.9058	0.8147
$g_j(a_3) = 3$	0.8147	0.8147	0.9134
$g_j(a_4) = 4$	0.9058	0.6424	0.9058
$g_j(a_5) = 5$	0.9134	0.127	0.6424

As it is evident from the three figures, on the one hand, the DM's preferences in case 1 are increasing with the values of g_j, while the DM's preferences in case 2 are decreasing with the values of g_j. On the other hand, the DM's preferences in case 3 are non-monotonic. In particular, they are increasing in the interval $[1,3]$, while they are decreasing in the interval $[3,5]$ with the value 3 being the most preferred on g_j for the DM.

Let us observe that the normalization procedure defined above is independent on the numerical value representing the evaluations of the

[3] For example, the difference between *bad* and *medium* is the same as the difference between *medium* and *good*.

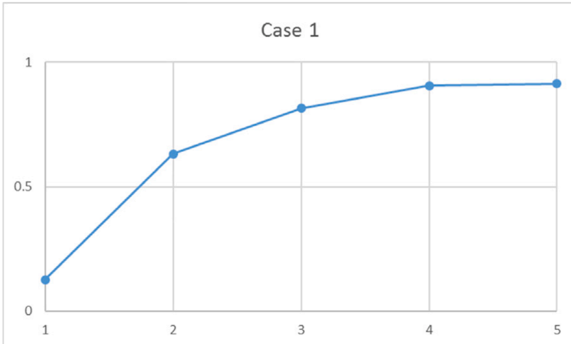

Fig. 1. Monotonic increasing prefs.

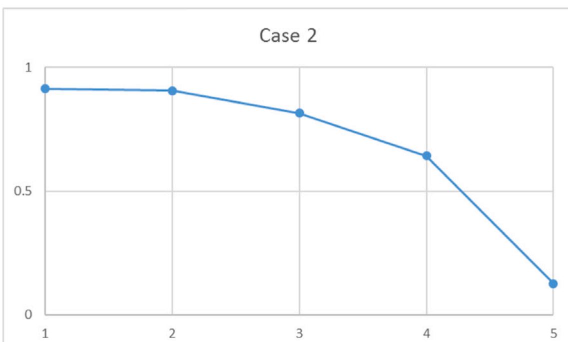

Fig. 2. Monotonic decreasing prefs.

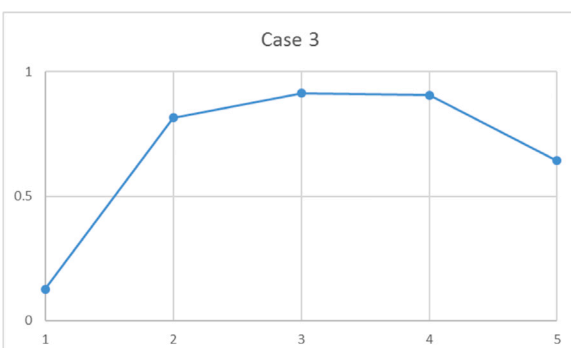

Fig. 3. Non-monotonic prefs.

alternatives on the considered criteria, but it depends on the ranking of the alternatives on the same criterion with respect to the DM's preferences. For this reason, the procedure can be used efficiently to deal also with decision making problems in which the evaluations are not expressed quantitatively but qualitatively, provided that the DM is able to preferentially order these evaluations from the best to the worst.

Once the normalization procedure is applied for each criterion, the evaluations of the alternatives on the considered criteria are then expressed on the same scale and the first step of the TOPSIS application can therefore be considered done.

We shall conclude this section by underlying two things. First, the normalization procedure presented above is sensible to the sampled numbers x_i and does not take into account the difference in the evaluations of the alternatives, but only the preference order of the alternatives on the considered criterion. This could be considered a drawback of the method. However, indeed, we do not claim that we have to work with a single normalized matrix but we would like to apply the SMAA methodology recalled in Section 2.3 to provide robust recommendations on the problem at hand. As it will be evident from the summary of the new procedure given in Section 3.3, we shall build several normalization matrices using the procedure above and, therefore, we shall apply TOPSIS once for each sampled normalization matrix. At the end, by applying SMAA, we shall provide information in statistical terms on the goodness of the considered alternatives.

Another important aspect is that a different normalization can provide different results on the TOPSIS application (see, for example, Çelen (2014), Chatterjee and Chakraborty (2014), Milani et al. (2005) and Vafaei et al. (2018)). For this reason, in our opinion, the choice of only one normalization can be considered arbitrary. The normalization procedure described above – in conjunction with the SMAA application – avoids taking into account only one normalization technique since each sampled normalization matrix corresponds, in some sense, to a different normalization of the performance matrix. We think that this is an added value and a beneficial aspect of our proposal.

3.2. A hierarchical TOPSIS method

In this section, we shall introduce the integration of the TOPSIS method with the MCHP methodology to deal with decision making problems structured in a hierarchical way. For such a reason, we shall use the notation and methodology introduced in Section 2.2. Let us recall once again that, in the case that a hierarchical structure is considered, the evaluations of the alternatives are given on the elementary criteria only, i.e. those at the bottom of the hierarchical structure.

For each non-elementary criterion $g_{\mathbf{r}}$, let us describe how the TOPSIS steps have to be modified to provide a complete ranking of the alternatives on $g_{\mathbf{r}}$ only.

- *Normalization:* Let us use the procedure described in the previous section, i.e. for each alternative $a_i \in A$ and for each elementary criterion $g_{\mathbf{t}}$, the values $z_{i\mathbf{t}} = \frac{a_{i\mathbf{t}}}{\sqrt{\sum_{i=1}^{n} a_{i\mathbf{t}}^2}}$ are now replaced by the normalized values z_{ij}, obtained by applying the introduced normalization procedure;
- *Weighting:* Let us assume that a weight vector is obtained using an indirect preference information provided by the DM and that the weighted normalized values are therefore computed as follows: for each $a_i \in A$ and for each $g_{\mathbf{t}} \in G_E$, $v_{i\mathbf{t}} = z_{i\mathbf{t}} \cdot w_{\mathbf{t}}$;
- *Partial positive ideal and negative ideal solutions:* the partial positive ideal solution $A_{\mathbf{r}}^{+} = \left(v_{\mathbf{t}_1}^{+}, \ldots, v_{\mathbf{t}_{|E(g_{\mathbf{r}})|}}^{+}\right)$ and the partial negative ideal solution $A_{\mathbf{r}}^{-} = \left(v_{\mathbf{t}_1}^{-}, \ldots, v_{\mathbf{t}_{|E(g_{\mathbf{r}})|}}^{-}\right)$ are computed. In defining $A_{\mathbf{r}}^{+}$ and $A_{\mathbf{r}}^{-}$, we shall take into account only the elementary criteria descending from $g_{\mathbf{r}}$. Let us note here that, in order to define such solutions, we do not have to distinguish between elementary criteria having an increasing or a decreasing direction of preference. Indeed, using the normalization procedure described above, we can also take into account elementary criteria having a non-monotonic direction of preference. Moreover, the most preferred alternative on a criterion $g_{\mathbf{t}}$ will have the greatest normalized value on $g_{\mathbf{t}}$ and vice versa. Consequently, $v_{\mathbf{t}}^{+} = \max_{i=1,\ldots,n} v_{i\mathbf{t}}$ and $v_{\mathbf{t}}^{-} = \min_{i=1,\ldots,n} v_{i\mathbf{t}}$ for each $g_{\mathbf{t}} \in G_E$,
- *Distance to partial positive ideal and negative ideal solutions*: for each $a_i \in A$, its distance $d_{\mathbf{r}}^{+}(a_i)$ from the partial positive ideal solution $A_{\mathbf{r}}^{+}$ and its distance $d_{\mathbf{r}}^{-}(a_i)$ from the partial negative ideal solution $A_{\mathbf{r}}^{-}$ are computed as

$$d_{\mathbf{r}}^{+}(a_i) = \sqrt{\sum_{\mathbf{t} \in E(g_{\mathbf{r}})} \left(v_{i\mathbf{t}} - v_{\mathbf{t}}^{+}\right)^2} \quad \text{and} \quad d_{\mathbf{r}}^{-}(a_i) = \sqrt{\sum_{\mathbf{t} \in E(g_{\mathbf{r}})} \left(v_{i\mathbf{t}} - v_{\mathbf{t}}^{-}\right)^2}, \tag{7}$$

- *Relative closeness to partial positive ideal and negative ideal solutions and ranking:* for each alternative $a_i \in A$ the partial relative closeness is computed as $C_{\mathbf{r}}(a_i) = \frac{d_{\mathbf{r}}^{-}(a_i)}{d_{\mathbf{r}}^{+}(a_i)+d_{\mathbf{r}}^{-}(a_i)}$. This gives an estimate of the closeness of a_i to the positive ideal and negative ideal solutions taking into account only criterion $g_{\mathbf{r}}$.

To sum up, the extension of the TOPSIS method to the MCHP methodology is reduced to the application of TOPSIS not only at a comprehensive level, that is, considering the whole set of elementary criteria simultaneously, but to the application of the method once for each non-elementary criterion $g_{\mathbf{r}}$ taking into account only the elementary criteria descending from $g_{\mathbf{r}}$.

3.3. The MCHP-TOPSIS-SMAA method

In this section we shall describe in detail the main steps of the MCHP-TOPSIS-SMAA method summarizing what has been explained in the previous sections. To make such a description clearer, Fig. 4 shows a flow chart of the proposed method.

Step (0) The analyst, together with the help of the DM, has to structure the set of criteria in a hierarchical way following the MCHP principles. The root criterion will represent the main objective of the problem; elementary criteria are located at the bottom of the hierarchy and the performance of the alternatives is provided only on the basis of these criteria. Macro-criteria are placed between the root criterion $g_{\mathbf{0}}$ and the elementary ones to denote a particular aspect of the problem,

Step (1) In this step, the DM is invited to explicitly state her/his preferences. At first, to perform the normalization following what has been presented in Section 3.1, for each elementary criterion the DM has to order the alternatives from the best to the worst. Then, if (s)he is willing to do so, (s)he can provide information on the comparison between elementary criteria or macro-criteria in terms of their importance. For example, (s)he can state that elementary criterion $g_{\mathbf{t}_1}$ is more important than elementary criterion $g_{\mathbf{t}_2}$ or that macro-criterion $g_{\mathbf{r}_1}$ is more important than macro-criterion $g_{\mathbf{r}_2}$ and so on. This preference information will be translated into linear constraints used to restrict the whole space of weights vectors; that is

$$W=\left\{\left(w_{\mathbf{t}_1},\ldots,w_{\mathbf{t}_{|E_G|}}\right):w_{\mathbf{t}}>0\;\forall\mathbf{t}\in E_G\text{ and }\sum_{\mathbf{t}\in E_G}w_{\mathbf{t}}=1\right\}. \tag{8}$$

For example, the preference of $g_{\mathbf{t}_1}$ over $g_{\mathbf{t}_2}$ will be translated into the constraint $w_{\mathbf{t}_1}\geqslant w_{\mathbf{t}_2}+\varepsilon$,[4] while the preference of $g_{\mathbf{r}_1}$ over $g_{\mathbf{r}_2}$ will be translated into the constraint $w_{\mathbf{r}_1}\geqslant w_{\mathbf{r}_2}+\varepsilon$, where, following Corrente, Greco and Słowiński (2013), the weight of a non-elementary criterion $g_{\mathbf{r}}$, that is, $w_{\mathbf{r}}$ is the sum of the weights of the elementary criteria descending from it. Formally, $w_{\mathbf{r}}=\sum_{\mathbf{t}\in E(g_{\mathbf{r}})}w_{\mathbf{t}}$. In this way, the provided preferences are all expressed in terms of the weights of only the elementary criteria,

Step (2) The analyst has to check if the preferences provided by the DM are consistent, that is, if there exists at least one weight vector $\left(w_{\mathbf{t}_1},\ldots,w_{\mathbf{t}_{|E_G|}}\right)$ compatible with such preferences. For this reason, one has to solve the following LP problem

$$\varepsilon^*=\max\varepsilon,\ \text{subject to}$$
$$\left.\begin{array}{l}w_{\mathbf{t}}>0\;\forall\mathbf{t}\in E_G,\\ \sum_{\mathbf{t}\in E_G}w_{\mathbf{t}}=1,\\ E^{Pref}\end{array}\right\}E^{DM}$$

where E^{Pref} is the set of constraints translating the preferences of the DM. If E^{DM} is feasible and $\varepsilon^*>0$, then there exists at least one weight vector compatible with the preferences provided by the DM, while, in the opposite case (E^{DM} is infeasible or $\varepsilon^*\leqslant 0$), there does not appear any compatible weight vector to exist and, therefore, the cause of the infeasibility of E^{DM} should be investigated and solved by using, for example, one of the methods proposed in Mousseau et al. (2003).

In case that the preferences are consistent or after the same preferences have been revised to make them consistent, the number of performed iterations is initialized to 1,

Step (3) Build a normalization matrix as described in Section 3.1 and sample a weight vector compatible with the preferences provided by the DM. Let us observe that the set of constraints translating the preferences of the DM define a convex set and one can sample several vectors from this set by using, for example, the Hit-And-Run method (Smith, 1984; Tervonen et al., 2013),

Step (4) For each non-elementary criterion $g_{\mathbf{r}}$, apply the MCHP-TOPSIS method using the built normalized matrix and the sampled weight vector in step (3). Compute, therefore, the partial ranking of the alternatives at hand not only considering the root criterion $g_{\mathbf{0}}$ but all non-elementary criteria $g_{\mathbf{r}}$ in the hierarchy as described in Section 3.2. Store the obtained rankings,

Step (5) Check if the maximum number of iterations has been reached. If yes, pass on to the following step, otherwise go to step (3),

Step (6) Apply the SMAA methodology computing, for each non-elementary criterion $g_{\mathbf{r}}$, the rank acceptability index of each alternative a_i for each rank position s $\left(b^s_{\mathbf{r}}(a_i)\right)$ and the pairwise winning index of each ordered pair of alternatives $\left(p_{\mathbf{r}}(a_i,a_k)\right)$.

Let us observe that the idea of integrating the SMAA methodology into the TOPSIS method has been proposed by Okul et al. (2014). However, in that paper the authors take into account the variability of the alternatives' evaluations and the variability of the weights of criteria applying the vector normalization and comparing the results of the application of SMAA to both TOPSIS and the weighted sum preference models. In this paper, we are instead interested in proposing a different normalization step and, therefore, we apply the SMAA methodology to get robust recommendations varying the normalized matrices and the weights vectors compatible with some preferences provided by the DM.

4. Bank overall evaluation: an illustrative study

4.1. Context and selection of indicators

This section illustrates the merits of the proposed approach in a simulated evaluation of a sample of banks primary listed on the London Stock Exchange (LSE) in 2020. The evaluation will assess the banks in question in terms of their performance on a set of financial and non-financial attributes, the latter often found missing from evaluation frameworks despite their evolving recognition (Deloitte, 2019) for a more sustainable and inclusive landscape in the corporate world (Tett, 2019).

Starting with the financial characteristics, these are based on the 'CAMEL' framework, introduced in 1979 by US regulators to assess the financial health of banks around five key areas, namely the *capital adequacy* (C), *asset quality* (A), *management* (M), *earnings* (E) and *liquidity* (L). The framework is an integral part of the assessment process carried out by central banks and regulatory bodies, as it identifies banks' strengths and weaknesses (Gaganis et al., 2021; Ravisankar & Ravi, 2010; Reddy & Ravi, 2013). Both the actual criteria of the framework and the outcome of the evaluation are not disclosed publicly, with

[4] Let us observe that ε is an auxiliary variable introduced to transform the strict inequality constraints in weak ones.

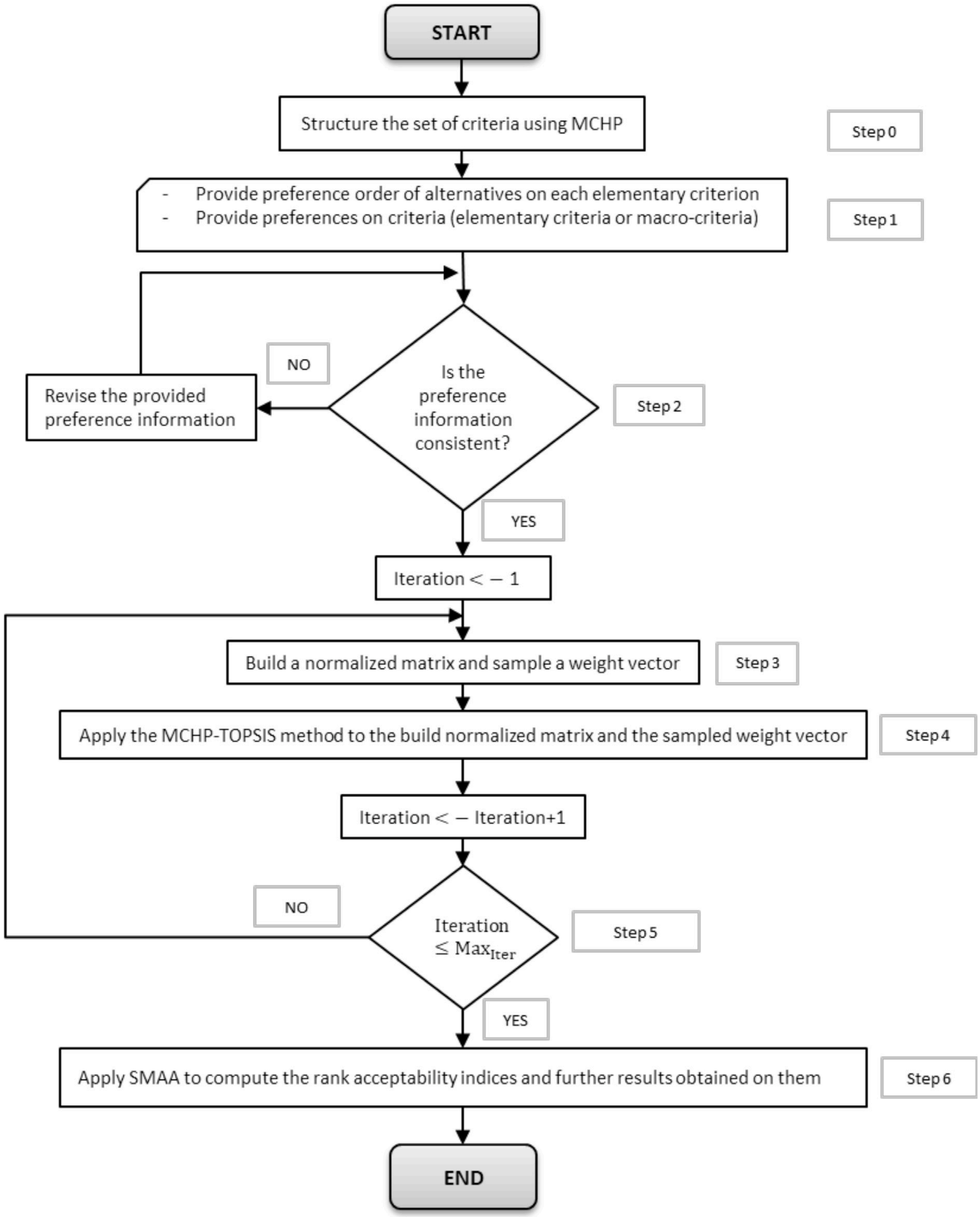

Fig. 4. Flow chart of the proposed method.

only the bank executives being aware of the latter. Regardless, prior literature (among others, see e.g. Berger et al. (2000, 2001), Cole and White (2012) and Doumpos and Zopounidis (2010)) suggests the use of a range of ratios to proxy the five areas of this framework.

One of the most comprehensive list of criteria is disclosed in the study of Doumpos and Zopounidis (2010, Table 1, p.58). The authors present a decision support system that evaluates the population of Greek banks on the basis of 26 criteria selected and parameterized in cooperation with analysts from the Bank of Greece. We closely follow their framework for this illustrative study. Although, due to lack of granular data and such detailed information from a central bank, we end up with a smaller set of 16 indicators. We should note that our only difference in the selected criteria appears in the proxies for their elementary criteria 'Man4' and 'Man5'. In particular, regarding the 'Man4' criterion (which Doumpos and Zopounidis (2010) describe as 'Top management competencies, qualifications and continuity'), we only have information on the top management's qualifications (educational data). Instead of using the whole top management team's qualifications in a form of a composite index, we only use the CEO's qualification; that is primarily due to the illustration of the capacity of the normalization technique to factor in different types of data, such as an ordinal scale.[5] Finally, the 'Man5' criterion (which Doumpos and Zopounidis (2010) describe as 'Managers' experience and competence) is here proxied by experience as in the number of Boards CEOs have sat on (Man5.1), and competence with the network size (number of overlaps through employment, other activities, and education) of the CEO (Man5.2).[6] To avoid 'double-counting', these two indicators share the weight of the principle criterion they are trying to proxy, i.e. 'Man5'. Data are obtained from the Standard & Poor's (S&P) Market Intelligence and the BoardEx databases. The list of criteria, their description and sources are displayed in Table 2, and descriptive statistics are given in Table 3.

Evaluation according to financial characteristics is undoubtedly important as financial institutions entail the element of systemic risk (Chen & Cheng, 2013), the close monitoring of which has been signified after the global financial crisis. However, an evaluation could well be augmented to include other, non-financial attributes that portray also the sustainable aspect of entities. These generally involve socio-economic characteristics that usually come under the umbrella commonly referred to as 'ESG' (Environmental, Social and Governance). Discussions around it have intensified in recent years as banks, regulatory bodies, businesses, NGOs, and governments become more attuned to evaluating investments and corporate performance through the lens of ESG impact (Deloitte, 2019), whilst socially responsible investments are notoriously growing in recent decades (Bilbao-Terol et al., 2014). When it comes to banks, ESG may be also particularly important given the downturn of their public trust following the global financial crisis. As Del Andeson (2017) adds, "[...] investors feel the banking industry's reputation has been tarnished in recent years following the many high-profile breakdowns in governance and breaches of public trust. [...] we do not believe that these past transgressions should be overwhelming factors in the forward-looking ESG assessment of individual banks, particularly for banks that performed well through the financial crisis or banks that have wholly revamped their management teams and governance processes".

Despite a large number of ESG-involving evaluations in the academic arena (among others, see Cai et al. (2011), Cheung et al. (2010), Hillman and Keim (2001), Oliveira et al. (2019), Ortas et al. (2015) and Puggioni and Stefanou (2019)), there is no single framework outlining key indicators to evaluate a set of entities on.[7] For this simulated evaluation, we rely on the ASSET4 ESG evaluation framework provided by EIKON, hosted by Thomson Reuters' Refinitiv database. In particular, the database collects data points on over 400 metrics from companies' annual reports, 178 of which are chosen after consultation with analysts and clientele to form the three major categories ('E', 'S' and 'G'). The bottom layer of the 178 indicators is essentially aggregated into pillars,[8] the score of which varies in the $[0, 100]$ range, and is used as an indicator to proxy a specific aspect. Due to lack of access on the bottom layer 178 scores, we use the pillar ranking scores Thomson Reuters share in their ESG ASSET4 database. These are listed in Table 2, yet we abridge the indicators'[9] in the following.

Starting with the environmental dimension, the three key composite indicators score an entity's efforts to reduce the resources it uses (resource reduction), reduce its emissions (emission reduction) and being innovative with respect to environmentally related issues (environmental innovation). These three composite metrics essentially contain aspects such as whether an entity has policies and committees dedicated to reduce their environmental resources and emissions and encourage environmental innovation (e.g. through turning to, promoting or investing in clean energy solution products). Looking at the social dimension of ESG, an entity's score is judged on four pillars, including *workforce quality and diversity* (e.g. staff training, equal opportunities between genders, welcoming employees' disabilities, flexible working schemes etc.), *human rights* (e.g. policies and actions ranging from respecting universal human rights set by the UN to respecting freedom of speech etc.), *community* (e.g. good citizenship in many forms, including fair competition, awards and recognition from local surroundings to supporting the community), and *product responsibility* (e.g. policies about protecting customer health and safety, personal data, and general integrity and privacy). Last subdimension is the corporate governance with three pillars: *board governance and structure* (e.g. policies ensuring accurate reporting, ratios of internal to independent committee members, number of meetings, % of females on board, CEO duality and compensation linked to executives' targets), *shareholders' rights* (e.g. anti-takeover defences, voting rights, election of board members with majority, as well as policies for minority shareholder protection etc.) and, finally, *CSR strategy* (e.g. having a committee dedicated to CSR actions, publishing related reports and having external auditors).

4.2. Sample and parameter modelling

Our sample consists of 8 out of the 9 UK banks listed on the FTSE350 Banks Index of the LSE. The sample is purely conditioned on data availability on ESG attributes. The reason for turning our focus to the UK market is twofold. First, we would like to avoid comparing entities spanning across different countries due to potentially different regulatory standards in the set of non-financial criteria driving the differences in performance. Second, our subscription to the provider of a set of data we rely on for this analysis (BoardEx database) is primarily covering the UK. Banks in our sample will be evaluated on a set of criteria (26 in total) in a Monte Carlo simulation environment involving 100,000 iterations.

Consistent with the step count presented in Section 3.3 and illustrated in Fig. 4, step 0 entails structuring the set of criteria. In more detail, the set of criteria chosen for this evaluation are hierarchically structured in 2 levels. The first consists of the highest categorization between financial and non-financial attributes. Financial attributes are split in five categories (each proxying a different aspect of the CAMEL framework), whilst non-financial ones are split in three categories (each proxying the different aspect of the ESG framework). The hierarchy is visually displayed in Fig. 5.

[5] We treat the education of the CEO as an ordinal category similar to Papadimitri et al. (2020), which categorizes the levels of education in undergraduate, postgraduate and doctoral.

[6] Due to missing data on the actual years CEOs sat in each board, we only provide the ordinal information (i.e. the number of boards) aside of the years they have been serving for. Moreover, as 'competence' is difficult to capture in principle, we use network size as proxy for CEOs' capacity to use their networks to efficiently gather and control private information facilitating value-creation knowledge in several decision making situations, from acquisitions (El-Khatib et al., 2015) to fostering innovation from within and/or venturing in new domains (Cao et al., 2006; Smith et al., 2005).

[7] The International Organization for Standardization (ISO) along with the Global Reporting Initiative (GRI) distribute a document (see ISO (2010)) that draws upon international consensus from the broadest view of stakeholder groups, in order to ensure that the concept of evaluation in these aspects is gradually standardized. According to the GRI, 75% of the largest 250 companies in the globe use the GRI principles (GRI, 2019) already. This is a deeply welcoming fact, and a great initiative from a reporting perspective, although still far from an apt evaluation framework.

[8] The pillar score is essentially calculated by Thomson Reuters as an equally-weighted average of each indicator's rank score, i.e. $score = \frac{\text{no. of companies with a worst value} + \frac{\text{no. of companies with the same value}}{2}}{\text{no. of companies with a value}}$.

[9] The full breakdown of all indicators used in the composite metric are confidential and only visible to the subscribers of the database. Therefore, we only provide a summary of the thematic areas captured by those indicators here.

Table 2
Evaluation criteria.

Level 1	Level 2	Abbreviation	Short description	Source	Orientation
Financial	*Capital adequacy*	Cap1	Capital adequacy ratio	Market intelligence	Max
		Cap2	TIER II capital/TIER I	Market intelligence	Min
	Asset quality	Ass1	Risk-weighted assets/Assets	Market intelligence	Min
		Ass2	(Non performing loans - Provisions)/Loans	Market intelligence	Min
		Ass4	(Non performing loans/2 - Provisions)/Equity	Market intelligence	Min
	Managerial quality	Man1	Operating expenses/Operating income	Market intelligence	Min
		Man2	Staff cost/Assets	Market intelligence	Min
		Man4	CEO's level of education (1 = BA, 2 = MA/MBA, 3 = PhD)	BoardEx	Max
		Man5.1	CEO's prior managerial experience (boards sat on)	BoardEx	Max
		Man5.2	CEO's network size	BoardEx	Max
	Earnings quality	Ear1	Net income/Assets	Market Intelligence	Max
		Ear2	Net income/Equity	Market intelligence	Max
		Ear3	Interest revenue/Assets	Market intelligence	Max
		Ear4	Other operating revenue/Assets	Market intelligence	Max
	Liquidity	Liq1	Cash/Assets	Market intelligence	Non-mon
		Liq2	(Loans - Provisions)/Deposits	Market intelligence	Non-mon
Non-financial	Corporate governance	Gov1	CSR strategy score	Asset4 - EIKON	Max
		Gov2	Shareholders' rights score	Asset4 - EIKON	Max
		Gov3	Board governance & structure score	Asset4 - EIKON	Max
	Social responsibility	Soc1	Product responsibility Score	Asset4 - EIKON	Max
		Soc2	Community score	Asset4 - EIKON	Max
		Soc3	Human rights score	Asset4 - EIKON	Max
		Soc4	Workforce quality & diversity score	Asset4 - EIKON	Max
	Environmental responsibility	Env1	Environmental innovation score	Asset4 - EIKON	Max
		Env2	Emission reduction score	Asset4 - EIKON	Max
		Env3	Resource reduction score	Asset4 - EIKON	Max

Notes: Abbreviation (column) of CAMEL criteria as in Doumpos and Zopounidis (2010) for consistency. The only difference with the authors' study in this list is located in criterion 'Man5', which we proxy with Man5.1 and 5.2 (both randomly sharing the weight assigned to Man5). Orientation shows whether that criterion is to be maximized (max) or minimized (min). Criteria 'Liq1' and 'Liq2' are non-monotonic. In particular, in its normalization, Liq1 has been set to be maximally preferred in the $[10, 12]$ interval (equal preference for any value in this interval) whilst Liq2 reaches its maximum score in the $[70, 80]$ interval (any value in the interval is equally preferred).

Table 3
Descriptive statistics.

Level 1	Level 2	Criterion	Min	Q1	Mean	Median	Q3	Max	Std.
Financial	Capital adequacy	Cap1	15.00	16.42	19.51	20.36	21.89	23.41	3.05
		Cap2	17.74	20.51	23.96	21.92	26.45	36.40	5.75
	Asset quality	Ass1	25.87	27.44	44.15	35.66	50.12	83.38	21.36
		Ass2	0.36	1.13	2.39	2.11	3.24	5.29	1.55
		Ass4	1.32	1.86	6.51	5.15	8.30	18.46	5.56
	Managerial quality	Man1	36.52	61.66	65.42	65.53	71.79	86.79	13.81
		Man2	0.62	0.67	1.13	0.77	1.15	3.01	0.76
		Man4	1.00	1.00	1.63	1.50	2.00	3.00	0.70
		Man5.1	4.00	4.75	6.63	7.00	8.00	10.00	2.06
		Man5.2	439.00	1393.50	2878.25	2221.50	4106.75	6397.00	2015.18
	Earnings quality	Ear1	0.14	0.21	0.92	0.42	0.95	3.37	1.10
		Ear2	−11.36	4.57	7.16	5.57	9.91	25.47	9.71
		Ear3	1.34	1.87	3.37	2.15	3.16	9.45	2.68
		Ear4	0.00	0.04	0.24	0.07	0.25	0.96	0.34
	Liquidity	Liq1	8.10	11.10	13.53	13.58	16.68	17.81	3.31
		Liq2	64.27	74.70	93.33	86.91	114.74	131.88	22.94
Non-financial	Corporate governance	Cg1	22.18	53.84	70.53	77.69	93.86	98.47	25.65
		Cg2	7.29	19.69	56.46	65.86	91.35	95.01	36.70
		Cg3	28.85	68.29	74.08	79.80	88.24	91.69	19.67
	Social responsibility	Soc1	4.80	33.57	53.58	56.37	81.53	96.53	32.18
		Soc2	6.94	26.46	60.64	72.83	93.07	99.27	36.40
		Soc3	34.84	81.06	85.10	96.77	97.44	99.44	20.89
		Soc4	72.11	74.56	80.58	78.07	86.05	92.66	7.22
	Environmental responsibility	Env1	30.00	76.31	78.23	84.52	87.47	95.97	19.71
		Env2	4.27	65.40	69.75	74.66	87.10	96.05	27.27
		Env3	27.18	58.35	74.96	82.34	94.72	99.60	24.19

Notes: Abbreviation (column) of CAMEL criteria as in Doumpos and Zopounidis (2010) for consistency. The only difference with the authors' study in this list is located in criterion 'Man5', which we proxy with Man5.1 and 5.2 (both randomly sharing the weight assigned to Man5). Cross sectional data regard the fiscal year end 2018.

Step 1 involves elicitation of preferences from the DM and steps 2–3 check the compatibility of expressed preferences and the count of simulations. As we lack such type of information from an interaction with a DM, preferences here are simulated in the notion of SMAA-TOPSIS (Okul et al., 2014). In more detail, weights are drawn unconditionally from a uniform distribution[10] hierarchically. That is,

[10] The procedure we follow is detailed in the online supplementary appendix in the study of Doumpos et al. (2016).

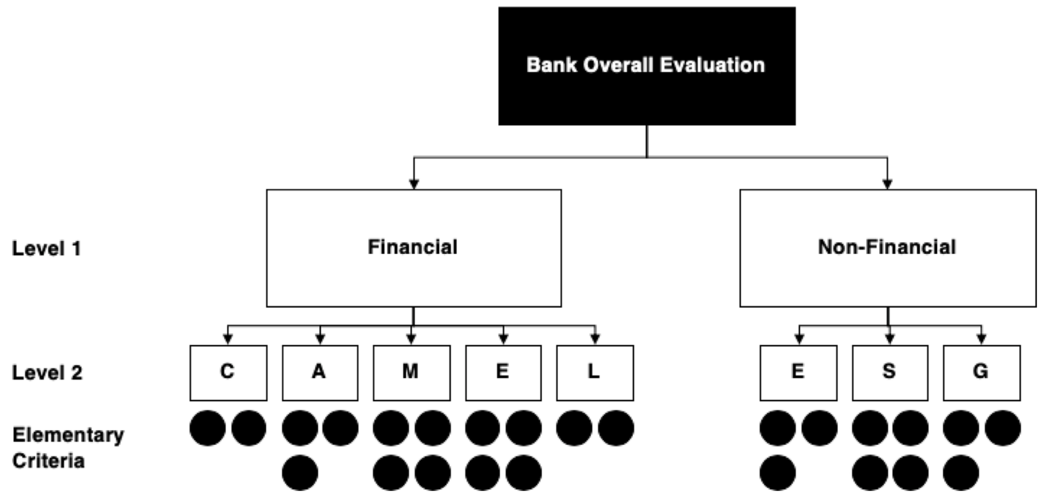

Fig. 5. Hierarchical representation of the evaluation.

Table 4
Overall results — Rank Acceptability Indices (RAI).

Bank/Rank Acceptability Index (%)	#1	#2	#3	#4	#5	#6	#7	#8	Exp. Class. (c_i)	Rank
HSBC Holdings Plc	44.42	20.74	12.41	14.55	6.03	1.22	0.58	0.06	2.23	1
Lloyds Banking Group Plc	19.14	33.64	19.73	12.01	6.56	4.31	3.56	1.06	2.86	2
Royal Bank of Scotland Group Plc	14.76	16.83	17.22	23.58	15.10	5.87	5.59	1.05	3.53	3
Standard Chartered Plc	14.13	12.14	26.64	16.73	10.36	6.16	7.43	6.41	3.77	4
Barclays Plc	3.77	7.40	10.38	15.23	35.31	17.90	7.39	2.64	4.67	5
Close Brothers Group Plc	3.27	5.45	6.63	7.59	15.34	35.51	17.98	8.25	5.46	6
Metro Bank Plc	0.40	3.12	5.37	7.71	7.85	17.72	28.15	29.69	6.34	7
Bank of Georgia Group PLC	0.13	0.68	1.62	2.62	3.47	11.32	29.33	50.84	7.14	8

in each iteration a set of weights declares the importance between financial and non-financial attributes (Level 1), followed by a set of weights declaring the importance between the five dimensions of financial attributes (CAMEL) and another set of weights declaring the importance between the three non-financial attributes (ESG) (Level 2). Then, in the same iteration, the importance of each Level 2 dimension is further split uniformly in the elementary criteria. Additionally, in each iteration, the normalization procedure (as described in Section 3.1) is also randomly chosen via a 'trigger' mechanism, identical to that used in uncertainty analysis of composite indicators (see Saisana et al. (2005)). Let us note here that, the space of preferences (either weight or normalization technique selection) could be conditioned if any type of information is elicited from the DM, such as assurance regions, inequality relationships between dimensions at any level etc. These are commonly used to bind optimized weights in DEA models (see Allen et al. (1997) and Allen and Thanassoulis (2004)). Finally, steps 5 and 6 regard the computational procedure of the proposed approach.

4.3. Results & discussion

This section abridges the results and illustrates some of the main insights that can be made by using this approach. To conserve space, we only discuss a subset of the results and we refer the interested reader to the full set provided in the online supplementary material.

Starting with an overall overview taking into account all criteria combined, Table 4 shows the RAIs for each bank i, and an expected overall classification using the listed probabilities (p_i) and the ordinal rankings ($r = 1, \ldots, 8$) as metaweights similar to Lahdelma and Salminen (2001, p.449), i.e. $c_i = \sum_{r=1}^{8} p_r r$ (see also Kadziński and Michalski (2016)). Seemingly, HSBC comes overall first, which is something confirmed in 44.42% of the simulated scenarios. A more detailed overview of the results of the simulations is given in Table 5, showing the PWIs. HSBC seems to beat its second – in the overall classification – counterpart, Lloyds Bank, in approximately 61% of the simulated scenarios.

The added benefit of the MCHP over the overall evaluation is its capacity to show where a unit falls short (finding the 'bottleneck') when evaluated against its counterparts. This is particularly magnified when the evaluation is comprised of several indicators that span across several dimensions. For instance, PWIs and RAIs presented in Tables 4 and 5 are available for every node of the hierarchy to facilitate questions the DM might have on particular performance overviews. To conserve space, we do not report these results in a tabular form, but we provide all results in the online supplementary material.

A way to abridge the evaluation (in an ordinal form) and get a first overview of bottleneck performance is by looking at the expected classification at each node of the hierarchy. This is in the notion of a scorecard used to break down composite indicators – being used frequently for business intelligence purposes (Barone et al., 2011) – albeit scorecards denote evaluation in the elementary set of indicators single and not in a node of interest. Fig. 6 shows the ranking of each bank (obtained from their expected classification similar to Table 4) coloured in an RGB gamut visually illustrating the state of a bank's performance. Also displayed in the figure are the average weights (in the 100,000 simulations) per Level 1 and 2 dimensions.

This is an intuitive way to look at macro dimensions and subdimensions that a firm in question may work on to improve its standing. For instance, taking the example of Barclays Plc, it comes overall fifth, and looking at its position against Standard Chartered Plc (overall fourth), a bottleneck in Barclays can be attributed in the 'Earnings' subdimension of the Financial attributes, and in the 'Governance' subdimension of the non-financial attributes. These are areas the DM could look into first, in order to improve the performance of the bank in question, *ceteris paribus*. For instance, if one looks at the PWIs for these two Level 2 dimensions (i.e. 'Earnings' and 'Governance' – PWIs available in the online appendix), Standard Chartered is preferred to Barclays in 80.98% of scenarios in the Earnings subdimension and 80.14% of scenarios in the Governance subdimension. Therefore, the possibilities for improvement in the space of preference simulated in this evaluation

Table 5
Overall results — Pairwise Winning Indices (PWI).

	Bank/Pairwise Winning Index (%)	B1	B2	B3	B4	B5	B6	B7	B8
B1	HSBC Holdings Plc	0	61.16	68.76	70.43	87.24	94.33	97.67	97.14
B2	Lloyds Banking Group Plc	38.84	0	61.57	68.59	79.44	83.95	86.31	95.64
B3	Royal Bank of Scotland Group Plc	31.24	38.43	0	46.97	70.38	80.05	84.61	95.64
B4	Standard Chartered Plc	29.57	31.41	53.03	0	64.99	74.44	79.65	89.60
B5	Barclays Plc	12.76	20.56	29.62	35.01	0	66.14	79.03	89.51
B6	Close Brothers Group Plc	5.67	16.05	19.95	25.56	33.86	0	70.30	82.65
B7	Metro Bank Plc	2.33	13.69	15.39	20.35	20.97	29.70	0	63.90
B8	Bank of Georgia Group PLC	2.86	4.36	4.37	10.40	10.49	17.35	36.10	0

		Level 1	Level 2					Level 1	Level 2		
Bank	Overall	Financial	C	A	M	E	L	Non-Financial	E	S	G
HSBC Holdings Plc	1	1	2	1	3	5	1	1	1	3	2
Lloyds Banking Group Plc	2	4	4	6	1	3	6	3	2	4	7
Royal Bank of Scotland Group Plc	3	2	1	4	2	6	3	5	5	5	5
Standard Chartered Plc	4	7	7	7	5	4	8	2	6	1	1
Barclays Plc	5	5	3	3	6	8	4	4	3	2	6
Close Brothers Group Plc	6	3	5	5	7	1	5	6	7	7	4
Metro Bank Plc	7	6	6	2	8	7	2	7	8	6	3
Bank of Georgia Group PLC	8	8	8	8	4	2	7	8	4	8	8
Avg. Weight (Level 1) - Σ of Level 2		49.97%						50.03%			
Avg. Weight (Level 2) - Σ of elementary			9.98%	10.08%	9.92%	10.06%	9.93%		16.68%	16.66%	16.70%

Fig. 6. Hierarchical scorecard.

are almost identical for both macro-criteria. Consequently, the DM can freely decide which dimension is worth improving more based on the costs associated to achieve these improvements.

Finally, delving more into the abridged results, there appears to be no statistically significant correlation between rankings in terms of financial and non-financial attributes (Pearson's correlation = 0.45, $p = 0.26$). Indeed, apart from HSBC, which is ranked consistently first at the 1st level of hierarchy, as well as Bank of Georgia, which is ranked consistently eighth, the relationship between the rest of the banks' rankings at Level 1 is more vague, with some banks performing better in the financial aspects and some in the non-financial ones. Interestingly, a banks' Total Assets is highly correlated with its non-financial ranking (−0.86, Pearson's correlation, significant at the 5% level; $p = 0.013$) yet the correlation between Total Assets and financial rankings is lost in statistical terms (−0.61, $p = 0.11$). Although a conjecture at this point, this may suggest that larger banks in our sample are not necessarily overall financially sound – in terms of the financial framework selected – yet they may have the capacity and/or resources to invest in their sustainable, non-financial aspects.

5. Conclusions

The Technique for Order Preferences by Similarity to Ideal Solutions (TOPSIS; Hwang and Yoon (1981)) is among the most well known and applied Multiple Criteria Decision Aiding (MCDA) methods. Citing Behzadian et al. (2012, pp.13052), "*To apply this technique, attribute values must be numeric, monotonically increasing or decreasing, and have commensurable units*". In this paper, we propose to embed into TOPSIS a normalization procedure proposed by Angilella et al. (2015) for the Choquet integral preference model (Choquet, 1953; Grabisch, 1996). The normalization procedure puts the performance of the alternatives in a [0,1] scale, by taking into account only the ranking of the performances with respect to the preferences of the Decision Maker (DM). The proposed normalization procedure resolves the three mentioned requirements of the TOPSIS method since (i) it gives the possibility to deal with criteria performance of which is expressed on a qualitative scale (not only ordinal); (ii) permits to take into account non-monotonic preferences of the DM that can be present in real world applications; (iii) it renders the performance on the considered criteria commensurable since all of them are expressed in a [0,1] scale.

Since the applied normalization technique produces a single normalization performance matrix and, as acknowledged by many studies, using different normalization techniques can affect the produced TOPSIS rankings (Opricovic & Tzeng, 2004), we applied the Stochastic Multicriteria Acceptability Analysis (SMAA; Lahdelma et al. (1998) and Pelissari et al. (2020)) which provides more robust recommendations. Indeed, by applying SMAA, we took into account several normalization matrices giving therefore a view of the frequency with which an alternative reaches a certain position, or the frequency with which an alternative is preferred to another one. Additionally, the application of SMAA gives the opportunity to consider a plurality of weight vectors, instead of just a single vector.

Finally, in order to deal with problems in which criteria are structured in a hierarchical manner, we integrated the Multiple Criteria Hierarchy Process (MCHP; Corrente et al. (2012)) concepts into the TOPSIS method. In this way, the DM is provided not only with a comprehensive ranking taking into account all criteria simultaneously, but also one ranking for each node of the hierarchy thus getting a more analytical knowledge of the strong and weak points of each alternative.

The comprehensive hierarchical and robust TOPSIS method has been applied to a simulated evaluation of a set of banks listed on the LSE's FTSE350 Banks Index, providing a ranking of the banks in question in terms of their overall financial strength, as this has been captured on the basis of financial and non-financial attributes related to sustainable aspects.

We think that the new proposal has merits both from the theoretical as well as from the practical point of view. Indeed, on the one hand, it overtakes some drawbacks of the original method that have been criticized over the years in literature, while, on the other hand, the main characteristics of the method mentioned above make it more applicable in practice since it is able to deal with many issues of real world decision making problems.

Several directions of research can be envisaged: (i) studying how the proposed normalization technique affects the rank reversal of TOPSIS method (Garćia-Cascales & Lamata, 2012); (ii) extending the TOPSIS method to take into account possible interactions between criteria as in many well known MCDA methods (Grabisch & Labreuche, 2016); (iii) studying the link between the ranking of alternatives at a certain node $g_{\mathbf{r}}$ and the ranking of the same alternatives on the criteria immediately below it, that is $g_{(\mathbf{r},1)} \ldots , g_{(\mathbf{r},n(\mathbf{r}))}$, and applying alternative fuzzy variants dealing with special type of multiattribute decision making problems (Mohammed, 2020; Venkatesh et al., 2019; Verma et al., 2022; Yu et al., 2019); finally, (iv) uncertainties are not explicitly considered and uncertainties can be good or bad for a system (Wang & Wu, 2021). Capturing uncertainties in the framework is another interesting research topic.

CRediT authorship contribution statement

Salvatore Corrente: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Menelaos Tasiou:** Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Salvatore Corrente wishes to acknowledge the support of the Starting Grant 2020 from the University of Catania, Italy and the support of the Ministero dell'Istruzione, dell'Universitá e della Ricerca (MIUR), PRIN 2017, Italy, project "Multiple Criteria Decision Analysis and Multiple Criteria Decision Theory", grant 2017CY2NCA.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eswa.2022.119045.

References

Acuña-Soto, C., Liern, V., & Pérez-Gladish, B. (2018). Normalization in TOPSIS-based approaches with data of different nature: application to the ranking of mathematical videos. *Annals of Operations Research*, 1–29.

Allen, R., Athanassopoulos, A., Dyson, R., & Thanassoulis, E. (1997). Weights restrictions and value judgements in Data Envelopment Analysis: Evolution, development and future directions. *Annals of Operations Research*, *73*, 13–34.

Allen, R., & Thanassoulis, E. (2004). Improving envelopment in data envelopment analysis. *European Journal of Operational Research*, *154*(2), 363–379.

Angilella, S., Corrente, S., & Greco, S. (2015). Stochastic multiobjective acceptability analysis for the Choquet integral preference model and the scale construction problem. *European Journal of Operational Research*, *240*, 172–182.

Barone, D., Jiang, L., Amyot, D., & Mylopoulos, J. (2011). Composite indicators for business intelligence. In *International conference on conceptual modeling* (pp. 448–458). Springer.

Behzadian, M., Khanmohammadi Otaghsara, S., Yazdani, M., & Ignatius, J. (2012). A state-of the-art survey of TOPSIS applications. *Expert Systems with Applications*, *39*(17), 13051–13069.

Berger, A., Davies, S., & Flannery, M. (2000). Comparing market and supervisory assessments of bank performance: who knows what when? *Journal of Money, Credit and Banking*, 641–667.

Berger, A., Kyle, M., & Scalise, J. (2001). Did US bank supervisors get tougher during the credit crunch? Did they get easier during the banking boom? Did it matter to bank lending? In *Prudential supervision: What works and what doesn't* (pp. 301–356). University of Chicago Press.

Bhutia, P., & Phipon, R. (2012). Appication of AHP and TOPSIS method for supplier selection problem. *IOSR Journal of Engineering*, *2*, 43–50.

Bilbao-Terol, A., Arenas-Parra, M., Cañal-Fernández, V., & Antomil-Ibias, J. (2014). Using TOPSIS for assessing the sustainability of government bond funds. *Omega*, *49*, 1–17.

Cables, E., Lamata, M., & Verdegay, J. (2016). RIM-reference ideal method in multicriteria decision making. *Information Sciences*, *337–338*, 1–10.

Cai, Y., Jo, H., & Pan, C. (2011). Vice or virtue? The impact of corporate social responsibility on executive compensation. *Journal of Business Ethics*, *104*(2), 159–173.

Cao, Q., Maruping, L. M., & Takeuchi, R. (2006). Disentangling the effects of CEO turnover and succession on organizational capabilities: A social network perspective. *Organization Science*, *17*(5), 563–576.

Çelen, A. (2014). Comparative Analysis of Normalization Procedures in TOPSIS Method: With an Application to Turkish Deposit Banking Market. *Informatica*, *25*(2), 185–208.

Chatterjee, P., & Chakraborty, S. (2014). Investigating the effect of normalization norms in flexible manufacturing system selection using multi-criteria decision-making methods. *Journal of Engineering Science and Technology Review*, *7*(3), 141–150.

Chen, Y.-S., & Cheng, C.-H. (2013). Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry. *Knowledge-Based Systems*, *39*, 224–239.

Cheung, Y., Tan, W., Ahn, H.-J., & Zhang, Z. (2010). Does corporate social responsibility matter in Asian emerging markets? *Journal of Business Ethics*, *92*(3), 401–413.

Choquet, G. (1953). Theory of capacities. *Annales de I'Institut Fourier*, *5*(54), 131–295.

Cinelli, M., Kadzinski, M., Gonzalez, M., & Słowiński, R. (2020). How to Support the Application of Multiple Criteria Decision Analysis? Let Us Start with a Comprehensive Taxonomy. *Omega*, *96*, Article 102261.

Cole, R., & White, L. (2012). Déjà vu all over again: The causes of US commercial bank failures this time around. *Journal of Financial Services Research*, *42*(1–2), 5–29.

Corrente, S., Greco, S., Kadziński, M., & Słowiński, R. (2013). Robust ordinal regression in preference learning and ranking. *Machine Learning*, *93*, 381–422.

Corrente, S., Greco, S., & Słowiński, R. (2012). Multiple criteria hierarchy process in robust ordinal regression. *Decision Support Systems*, *53*(3), 660–674.

Corrente, S., Greco, S., & Słowiński, R. (2013). Multiple criteria hierarchy process with ELECTRE and PROMETHEE. *Omega*, *41*, 820–846.

Del Andeson, C. (2017). ESG in Action: Evaluating Global Financials. Avalaible at: https://www.pimco.co.uk/en-gb/insights/viewpoints/esg-in-action-evaluating-global-financials/, Accessed 9 September 2019.

Deloitte (2019). *Sustainable finance: Embrasing ESG in the financial service industry*: *Technical Report*, Avalaible at: https://www2.deloitte.com/content/dam/Deloitte/sg/Documents/risk/sea-risk-sustainable-finance.pdf, Accessed 9 September 2019.

Despotis, D., & Zopounidis, C. (1995). Building Additive Utilities in the Presence of Non-Monotonic Preferences. In P. Pardalos, Y. Siskos, & C. Zopounidis (Eds.), *Advances in multicriteria analysis. nonconvex optimization and its applications* (pp. 101–114). Boston, MA: Springer.

Doumpos, M., Gaganis, C., & Pasiouras, F. (2016). Bank diversification and overall financial strength: International evidence. *Financial Markets, Institutions & Instruments*, *25*(3), 169–213.

Doumpos, M., & Zopounidis, C. (2010). A multicriteria decision support system for bank rating. *Decision Support Systems*, *50*(1), 55–63.

El-Khatib, R., Fogel, K., & Jandik, T. (2015). CEO network centrality and merger performance. *Journal of Financial Economics*, *116*(2), 349–382.

Gaganis, C., Papadimitri, P., & Tasiou, M. (2021). A multicriteria decision support tool for modelling bank credit ratings. *Annals of Operations Research*, *306*, 27–56.

Garćia-Cascales, M., & Lamata, M. (2012). On rank reversal and TOPSIS method. *Mathematical and Computer Modelling*, *56*, 123–132.

Ghaderi, M., Ruiz, F., & Agell, N. (2017). A linear programming approach for learning non-monotonic additive value functions in multiple criteria decision aiding. *European Journal of Operational Research*, *259*(3), 1073–1084.

Grabisch, M. (1996). The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, *89*(3), 445–456.

Grabisch, M., & Labreuche, C. (2016). Fuzzy measures and integrals in MCDA. In S. Greco, M. Ehrgott, & J. Figueira (Eds.), *Multiple criteria decision analysis: State of the art surveys* (pp. 553–603). New York: Springer-Verlag.

Greco, S., Ehrgott, M., & Figueira, J. (2016). *Multiple criteria decision analysis: State of the art surveys*. New York: Springer-Verlag.

Greco, S., Mousseau, V., & Słowiński, R. (2008). Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research*, *191*(2), 416–436.

GRI (2019). The GRI Standards: the global standards for sustainability reporting. Available at: https://www.globalreporting.org/standards/media/2458/gri_standards_brochure.pdf.

Hillman, A., & Keim, G. (2001). Shareholder value, stakeholder management, and social issues: What's the bottom line? *Strategic Management Journal*, *22*(2), 125–139.

Hwang, C., & Yoon, K. (1981). *Multiple attribute decision making*. New York: Springer Verlag.

ISO (2010). GRI G4 Guidelines and ISO 26000:2010. How to use the GRI G4 Guidelines and ISO 26000 in conjunction. Available at: https://www.iso.org/files/live/sites/isoorg/files/archive/pdf/en/iso-gri-26000_2014-01-28.pdf.

Jacquet-Lagrèze, E., & Siskos, Y. (2001). Preference disaggregation: 20 years of MCDA experience. *European Journal of Operational Research*, *130*(2), 233–245.

Jun, W., Lingyu, T., Yuyan, L., & Peng, G. (2017). A weighted EMD-based prediction model based on TOPSIS and feed forward neural network for noised time series. *Knowledge-Based Systems*, *132*, 167–178.

Kadziński, M., Martyn, K., Cinelli, M., Słowiński, R., Corrente, S., & Greco, S. (2020). Preference disaggregation for multiple criteria sorting with partial monotonicity constraints: Application to exposure management of nanomaterials. *International Journal of Approximate Reasoning*, *117*, 60–80.

Kadziński, M., & Michalski, M. (2016). Scoring procedures for multiple criteria decision aiding with robust and stochastic ordinal regression. *Computers & Operations Research*, *71*, 54–70.

Kim, G., Park, C., & Yoon, P. (1997). Identifying investment opportunities for advanced manufacturing systems with comparative-integrated performance measurement. *International Journal of Production Economics*, *50*, 23–33.

Kuo, T. (2017). A modified TOPSIS with a different ranking index. *European Journal of Operational Research*, *260*, 152–160.

Lahdelma, R., Hokkanen, J., & Salminen, P. (1998). SMAA - Stochastic multiobjective acceptability analysis. *European Journal of Operational Research*, *106*(1), 137–143.

Lahdelma, R., & Salminen, P. (2001). SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. *Operations Research*, *49*(3), 444–454.

Leskinen, P., Viitanen, J., Kangas, A., & Kangas, J. (2006). Alternatives to incorporate uncertainty and risk attitude in multicriteria evaluation of forest plans. *Forest Science*, *52*(3), 304–312.

Li, P., Qian, H., Wu, J., & Chen, J. (2013). Sensitivity analysis of TOPSIS method in water quality assessment: I. Sensitivity to the parameter weights. *Environmental Monitoring and Assessment*, *185*(3), 2453–2461.

Milani, A., Shanian, A., Madoliat, R., & Nemes, J. (2005). The effect of normalization norms in multiple attribute decision making models: A case study in gear material selection. *Structural and Multidisciplinary Optimization*, *29*(4), 312–318.

Mohammed, A. (2020). Towards a sustainable assessment of suppliers: an integrated fuzzy TOPSIS-possibilistic multi-objective approach. *Annals of Operations Research*, *293*, 639–668.

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 Journal of Operational Research*, *147*(1), 72–93.

Okul, D., Gencer, C., & Aydogan, E. (2014). A method based on SMAA-TOPSIS for stochastic multi-criteria decision making and a real-world application. *International Journal of Information Technology and Decision Making*, *13*(5), 957–978.

Oliveira, R., Zanella, A., & Camanho, A. (2019). The assessment of corporate social responsibility: The construction of an industry ranking and identification of potential for improvement. *European Journal of Operational Research*, *278*(2), 498–513.

Olson, D. (2004). Comparison of weights in TOPSIS models. *Mathematical and Computer Modelling*, *40*, 721–727.

Opricovic, S., & Tzeng, G. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *European Journal of Operational Research*, *156*(2), 445–455.

Ortas, E., Gallego-Alvarez, I., & Álvarez Etxeberria, I. (2015). Financial factors influencing the quality of corporate social responsibility and environmental management disclosure: A quantile regression approach. *Corporate Social Responsibility and Environmental Management*, *22*(6), 362–380.

Papadimitri, P., Pasiouras, F., Tasiou, M., & Ventouri, A. (2020). The effects of board of directors' education on firms' credit ratings. *Journal of Business Research*, *116*, 294–313.

Pelissari, R., Oliveira, M., Ben Amor, S., Kandakoglu, A., & Helleno, A. (2020). SMAA methods and their applications: a literature review and future research directions. *Annals of Operations Research*, *293*, 433–493.

Puggioni, D., & Stefanou, S. (2019). The value of being socially responsible: A primal-dual approach. *European Journal of Operational Research*, *276*(3), 1090–1103.

Ravisankar, P., & Ravi, V. (2010). Financial distress prediction in banks using Group Method of Data Handling neural network, counter propagation neural network and fuzzy ARTMAP. *Knowledge-Based Systems*, *23*(8), 823–831.

Reddy, K., & Ravi, V. (2013). Differential evolution trained kernel principal component WNN and kernel binary quantile regression: Application to banking. *Knowledge-Based Systems*, *39*, 45–56.

Roy, B. (2005). Paradigm and challenges. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), *Multiple criteria decision analysis: State of the art surveys* (pp. 3–24). Berlin: Springer.

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(2), 307–323.

Shih, H., Shyur, H., & Lee, E. (2007). An extension of TOPSIS for group decision making. *Mathematical and Computer Modelling*, *45*, 801–813.

Smith, R. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, *32*, 1296–1308.

Smith, K., Collins, C., & Clark, K. (2005). Existing knowledge, knowledge creation capability, and the rate of new product introduction in high-technology firms. *Academy of Management Journal*, *48*(2), 346–357.

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 Journal of Operational Research*, *224*, 552–559.

Tett, G. (2019). Moral Money: Bridging the yawning information gap on ESG investing. *Financial Times*, Retrieved from: https://www.ft.com/content/e278933c-91d0-11e9-b7ea-60e35ef678d2.

Vafaei, N., Ribeiro, R., & Camarinha-Matos, L. (2018). Data normalisation techniques in decision making: case study with TOPSIS method. *International Journal of Information and Decision Sciences*, *10*(1), 19–38.

Venkatesh, V., Zhang, A., Deakins, E., Luthra, S., & Mangla, S. (2019). A fuzzy AHP-TOPSIS approach to supply partner selection in continuous aid humanitarian supply chains. *Annals of Operations Research*, *283*(1), 1517–1550.

Verma, S., Mehlawat, M., & Mahajan, D. (2022). Software component evaluation and selection using TOPSIS and fuzzy interactive approach under multiple applications development. *Annals of Operations Research*, *312*, 441–471.

Vetschera, R. (2017). Deriving rankings from incomplete preference information: A comparison of different approaches. *European Journal of Operational Research*, *258*(1), 244–253.

Wang, W., & Wu, Y. (2021). Is uncertainty always bad for the performance of transportation systems? *Communications in Transportation Research*, *1*, Article 100021.

Yu, B., Cai, M., & Li, Q. (2019). A λ-rough set model and its applications with TOPSIS method to decision making. *Knowledge-Based Systems*, *165*, 420–431.

Yurdakul, M., & IÇ, Y. (2005). Development of a performance measurement model for manufacturing companies using the AHP and TOPSIS approaches. *International Journal of Productions Research*, *43*(21), 4609–4641.