A simulation study on the probabilities of rank reversal, tie making, and tie breaking for multiple criteria decision making methods

Lisheng Jiang[1,2], Huchang Liao *[1], and Bernard De Baets[2]

[1]Business School, Sichuan University, Chengdu 610064, China

[2]KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure links 653, B-9000 Ghent, Belgium

Abstract

Multiple criteria decision making (MCDM) methods can be affected by preference reversal, meaning that the order of two alternatives is reversed after adding or deleting another alternative. Here, we focus on methods that produce rankings with ties (*i.e.*, weak orders). In this context, one usually talks about rank reversal. Existing rank reversal probability simulation experiments are subject to improvement on the following points: 1) the small number of MCDM methods included, 2) the unclear relation between the rank reversal probability and the rank of the deleted alternative, and 3) the lack of consideration of ties. In this paper, considering both strict preferences and ties, we distinguish two new phenomena: tie breaking, *i.e.*, the shift from a tie to a strict preference, and tie making, *i.e.*, the shift from a strict preference to a tie. To investigate the probabilities of rank reversal, tie making, and tie breaking, a simulation study involving thirty versions of twelve MCDM methods and six simulation factors is set up. Results demonstrate that for MCDM methods using pairwise comparison data, deleting an alternative ranked first or last leads to smaller probabilities than deleting an alternative in the middle, while the opposite holds for the methods using evaluation data under criteria. Four findings and three suggestions are given to help decision makers select MCDM methods to use.

Keywords: Multiple criteria decision making; rank reversal; rank-based probability; tie breaking; tie making

*Corresponding author. Email: liaohuchang@163.com.

1 Introduction

Multiple criteria decision making (MCDM) methods assist decision makers in their selection of alternatives on the basis of the data provided. It has been observed that these methods can be affected by preference reversal, meaning that the order of two alternatives is reversed after adding or deleting another alternative. This phenomenon was first observed in the Analytic Hierarchy Process (AHP, Belton and Gear, 1983), and was later studied in the Preference Ranking Organization METHod of Enrichment Evaluations (PROMETHEE, De Keyser and Peeters, 1996), the series of ELECTRE methods (an abbreviation in French for elimination and choice translating reality, Wang and Triantaphyllou, 2008; Liu and Ma, 2021), the Technique for Order of Preferences by Similarity to Ideal Solution (TOPSIS, Garcia-Cascales and Lamata, 2012; Mousavi-Nasab and Sotoudeh-Anvari, 2018), and Data Envelopment Analysis (DEA, Soltanifar and Shahghobadi, 2014). In this paper, we focus on MCDM methods that produce rankings with ties (*i.e.*, weak orders), meaning that no incomparability is considered. In this context, one usually talks about rank reversal, although the term is also used to refer to preference reversal in general.

Many studies have looked into rank reversal from three perspectives: the reasons for rank reversal, the estimation of rank reversal probabilities, and the elimination of rank reversal, all of which are discussed in Section 2. For estimating rank reversal probabilities, simulation experiments are the only approach that is available for all MCDM methods. The most recent simulation study on rank reversal probabilities involving more than one MCDM method dates back to 1998. Zanakis et al. (1998) performed a simulation experiment to estimate the rank reversal probabilities for five kinds of MCDM methods, *i.e.*, the ELECTRE methods, AHP, TOPSIS, Simple Additive Weighting (SAW), and Multiplicative Exponential Weighting (MEW). Although Zanakis et al. (1998) provided some helpful advice for selecting MCDM methods, their study did not include certain common MCDM methods. Since 1998, also some significant improvements to MCDM methods have been proposed[1], the rank reversal probabilities of which still remain unstudied (Cinelli et al., 2020), and a thorough comparative study of different MCDM methods would be welcome (Bhatt et al., 2023). Furthermore, when estimating the rank reversal probability produced by deleting an alternative, an interesting question has not been addressed: does deleting alternatives at different ranks result in

[1]The MCDM methods commonly used are PROMETHEE II (Brans and Vincke, 1985) and TODIM (an abbreviation in Portuguese for interactive and multiple criteria decision making, Gomes and Lima, 1991). Important improvements to MCDM methods are the Best-Worst Method (BWM, Rezaei, 2015, 2016), the Gained and Lost Dominance Score (GLDS, Wu and Liao, 2019), the Double Normalization-based Multiple Aggregations (DNMA, Liao and Wu, 2020), and Quasi-Statistical Processing (QSP, Tomashevskii and Tomashevskii, 2021).

the same rank reversal probability? Last but not least, in the study of rank reversal, only strict preferences have been considered. Besides strict preferences, also ties are common in real-world situations as well as in the results of MCDM methods. Although evidence indicates that ties can be translated into strict preferences in real life (Corbin and Marley, 1974), such a shift has not been studied in MCDM methods.

To address the three aforementioned issues, we set up a simulation study involving thirty versions of twelve different MCDM methods that produce rankings with ties. However, the ELECTRE methods are not included here because they might result in incomparabilities. In this paper, not only rank reversal is studied, but also the shift between ties and strict preferences. Also, to investigate the relation between the rank of the alternative deleted and the rank reversal probability, the rank reversal probabilities caused by deleting an alternative at different ranks are computed independently.

The contributions and innovations of this paper are summarized as follows.

1) This paper considers a broader range of MCDM methods, as well as more simulation environments and parameters, than Zanakis et al. (1998). The thirty versions of twelve different MCDM methods used in this paper represent various decision making ideas and significant improvements to MCDM methods. This work fills the gaps in the research on the rank reversal probabilities of the Best-Worst Method (BWM, Rezaei, 2015, 2016), the Gained and Lost Dominance Score (GLDS, Wu and Liao, 2019), the Double Normalization-based Multiple Aggregations (DNMA, Liao and Wu, 2020), and Quasi-Statistical Processing (QSP, Tomashevskii and Tomashevskii, 2021).

2) This paper introduces two new phenomena: tie breaking, which represents the shift from a tie to a strict preference, and tie making, which represents the shift from a strict preference to a tie. Besides the rank reversal probability, this paper investigates the probabilities of tie making and tie breaking as well.

3) This paper investigates the effects of six simulation factors (*i.e.*, the number of alternatives, the number of criteria, the information consistency, the ratio scale maximum, the data distribution type, and the criterion weight type) on the probabilities of rank reversal, tie breaking, and tie making by performing regressions. Furthermore, this study evaluates the rank-based probabilities of rank reversal, tie making, and tie breaking by deleting the alternative at each rank independently.

This paper is organized as follows. Section 2 reviews the research on rank reversal and introduces the tie making and tie breaking phenomena. Section 3 discusses the design of the simulation experiments. The experimental results are given in Section 4. Finally, Section 5 offers some suggestions and the paper closes with conclusions in Section 6.

2 Literature review and research problem identification

As far as we know, Huntington (1938), who inductively generalized a postulate of relevancy recognized as a basic principle for ranking sports teams, was the first to study the influence of one alternative on the dominance relation between two other alternatives. The relevancy postulate states that when ranking two teams, the performance of a third team should not be taken into account. Later, in the field of social choice, the research of Huntington (1938) was discussed in the context of the axiom of independence of irrelevant alternatives which says that social preferences between two social states depend only on the individual preferences for these two social states, and not on the individual preferences for other social states (Arrow, 1951). Recently, Kondratev et al. (2023) studied the methods to avoid rank reversal for the scoring rules that are widely used to rank athletes in sports and candidates in elections.

In MCDM, Belton and Gear (1983) discovered that AHP might reverse the strict preference between alternatives after adding an alternative[2] This paper does not consider the latter cases., which was considered a shortcoming of AHP. Many scholars have studied the rank reversal of AHP since then. Vargas (1994), Millet and Saaty (2000), and Saaty (2006) explained the rank reversal phenomenon in AHP and pointed out that it is not reasonable for rank reversal to occur in AHP when alternatives are independent of each other (*e.g.*, when the axiom of independence of irrelevant alternatives is satisfied); however, it is reasonable when alternatives are dependent on each other (*e.g.*, the case of context-dependent preferences, Tversky and Simonson, 1993).

In the case that alternatives are independent, several studies have been done on AHP to 1) uncover the causes of rank reversal, 2) quantify the rank reversal probability, and 3) modify AHP to avoid rank reversal. In the studies on finding the reasons for rank reversal, five factors leading to rank reversal are considered: the method used for measuring preference intensities (Saaty, 1987), the size of the ratio scale (Schoner et al., 1992), the weighted normalization operations (Schenkerman, 1994), the inconsistency of the pairwise comparison information (Leskinen and Kangas, 2005; Tu and Wu, 2023), and the changes in local priorities (Wang and Elhag, 2006; Tu and Wu, 2023). As follow-up research to Leskinen and Kangas (2005), Faramondi et al. (2023) quantified the smallest degree of

[2] In this paper, when talking about rank reversal, we mean that the MCDM method is fixed while the set of alternatives is changed. Besides, rank reversal can also be caused by using similar but different methods, using different scales, or modifying pairwise comparisons when the set of alternatives is fixed. For example, AHP with right eigenvector and AHP with left eigenvector can produce different rankings (Johnson et al., 1979). Also, for AHP with right eigenvector, increasing the value of a pairwise comparison can result in a counter-intuitive rank reversal (Csató and Petróczy, 2021).

uncertainty that is able to result in rank reversal in AHP. Alvarez et al. (2021) formulated AHP as a system of linear equations and found that a high condition number and deficient rank might result in rank reversal in AHP. For estimating the rank reversal probability of AHP, point estimates and confidence intervals (Stam and Silva, 1997), simulation experiments (Dede et al., 2015), and the use of the multivariate normal cumulative distribution function (Dede et al., 2016) are three approaches. The rank reversal probability of AHP calculated by means of the multivariate normal cumulative distribution function is almost identical to the probability obtained from simulations (Dede et al., 2016). In terms of modifying AHP to avoid rank reversal, some studies attempted to reduce rank reversal by finding local stability intervals (Aguarón and Moreno-Jiménez, 2000), estimating weight errors (Tomashevskii, 2015), or using weight modification functions (Tomashevskii, 2018). Combinations of AHP and other methods, such as DEA (Ramanathan, 2006; Wang et al. 2008; Mirhedayatian and Farzipoor Saen, 2011), have also been used to reduce the rank reversal probability. In addition, a feedback system called Sinarchy (Leung and Cao, 2001) was built to minimize the rank reversal probability in AHP by promoting information exchange among decision makers.

Other MCDM methods have been investigated for rank reversal as well. For finding the causes of rank reversal, Liu and Ma (2021) used directed graphs to study the rank reversal caused by the distillation process in ELECTRE II and provided necessary and sufficient conditions for rank reversal to occur in ELECTRE II. In estimating the rank reversal probability, Zanakis et al. (1998) performed a simulation study to compare the rank reversal probabilities of five MCDM methods so as to provide suggestions for decision makers to select an MCDM method. Wang and Triantaphyllou (2008) used simulation experiments to estimate the rank reversal probabilities of ELECTRE II and III. For reducing the rank reversal probabilities of MCDM methods, Aires and Ferreira (2019) used the max-min standardization technique to avoid rank reversal in TOPSIS.

Referring to the notion of rank reversal given by Liu and Ma (2021), we refine the mathematical definition of rank reversal. We restrict to the case of MCDM methods resulting in a weak order, *i.e.*, a ranking with ties.

Definition 1 ***(Rank reversal)***. *Consider an MCDM method that always results in a weak order. When applying this method to a set of alternatives $\Phi = \{A_1, A_2, \ldots, A_n\}$, the strict preference relation is the antisymmetric relation $\mathbf{P}_\Phi = \{(A_i, A_j) \in \Phi^2 \mid A_i \text{ is strictly preferred to } A_j\}$. Let $\bar{\Phi}$ be the set of alternatives after adding an alternative to Φ or deleting an alternative from Φ. Rank reversal occurs if $\exists A_i, A_j \in \Phi \cap \bar{\Phi}$, $(A_i, A_j) \in \mathbf{P}_\Phi$, while $(A_j, A_i) \in \mathbf{P}_{\bar{\Phi}}$.*

Similarly, we can define tie making and tie breaking as follows.

Definition 2 *(**Tie making and tie breaking**). Consider an MCDM method that always results in a weak order. When applying this method to a set of alternatives $\Phi = \{A_1, A_2, \ldots, A_n\}$, the indifference relation is the symmetric relation $\mathbf{I}_\Phi = \{(A_i, A_j) \in \Phi^2 \mid A_i$ and A_j are indifferent$\}$. Let $\bar{\Phi}$ be the set of alternatives after adding an alternative to Φ or deleting an alternative from Φ. Tie making occurs if $\exists A_i, A_j \in \Phi \cap \bar{\Phi}$, $(A_i, A_j) \in \mathbf{P}_\Phi$, while $(A_i, A_j) \in \mathbf{I}_{\bar{\Phi}}$. Tie breaking occurs if $\exists A_i, A_j \in \Phi \cap \bar{\Phi}$, $(A_i, A_j) \in \mathbf{I}_\Phi$, while $(A_i, A_j) \in \mathbf{P}_{\bar{\Phi}}$.*

After reviewing the research on rank reversal, we find three points that require further discussion.

(1) For some MCDM methods, simulation experiments are the only means to compute rank reversal probabilities. Most simulation experiments on rank reversal only consider one kind of MCDM method, such as AHP (Dede et al., 2015) and the ELECTRE methods (Wang and Triantaphyllou, 2008). Zanakis et al. (1998) presented a simulation experiment on the rank reversal probabilities of five different MCDM methods. However, some MCDM methods, such as PROMETHEE II (Brans and Vincke, 1985) and TODIM (an abbreviation in Portuguese for interactive and multiple criteria decision making, Gomes and Lima, 1991), were not taken into account by Zanakis et al. (1998). After 1998, several improvements to MCDM methods, such as BWM (Rezaei, 2015, 2016) and QSP (Tomashevskii and Tomashevskii, 2021), were proposed, the rank reversal probabilities of which remain unstudied.

(2) As formalized in Definition 1, both adding or deleting an alternative might lead to rank reversal. Are the rank reversal probabilities the same when the alternative is deleted at different ranks?

(3) The research on rank reversal is an investigation into the reversal of the strict preference between alternatives when the set of alternatives changes. Besides strict preference, most MCDM methods can also produce ties. Could the strict preference between two alternatives be changed to a tie after adding or deleting an alternative? If the answer to this question is affirmative, then what is the probability? As far as we know, few studies have investigated these questions.

3 Design of simulation experiments

Given that simulation studies are the only means to calculate the rank reversal probabilities of all MCDM methods, we set up a simulation study to address the aforementioned three questions. This section introduces the simulation experiments of this paper from three points of view: the selection of MCDM methods, the design of simulation environments and parameters, and the simulation experiment steps. As far as we know, Zanakis et al. (1998) involved the largest variety of MCDM methods among the studies of ranking reversal simulations. To show the appropriateness and advantages of this study, we compare the simulation experiments of this paper to the simulation experiments of Zanakis et al. (1998).

3.1 Selection of MCDM methods

In this paper, we only consider individual MCDM methods and not combinations thereof (*e.g.*, a combination of AHP and TOPSIS is not considered). We focus on MCDM methods producing rankings with ties. Hence, methods resulting in incomparabilities, like the ELECTRE methods, are not considered. Also, as it is difficult to interact with MCDM methods in simulations, methods that need the opinion of a decision maker to drive the decision process, such as the dominance-based rough set approach (Cinelli et al., 2020), are also not considered.

Popular MCDM methods that are often discussed in the rank reversal research are AHP, the ELECTRE methods, TOPSIS, PROMETHEE, and TODIM (Aires and Ferreira, 2018). Here, we first divide the methods into two categories according to the input data type and then introduce several versions of these methods as well as some further improvements.

For MCDM methods using pairwise comparison data as input, this paper includes three versions of AHP, *i.e.*, AHP with right eigenvector (AHP_RE, Saaty, 1977), AHP with linear goal programming (AHP_LGP, Bryson, 1995), and AHP with logarithmic least squares method[3] (AHP_LLSM, Crawford and Williams, 1985; Csató, 2019). Besides, we take into account some improved versions. QSP (Tomashevskii and Tomashevskii, 2021) uses a statistical process to analyze the weights of alternatives calculated by AHP_RE to avoid rank reversal. Although it is said that QSP can eliminate rank reversal, it has not validated on large-scale data. Therefore, QSP is examined in this paper to verify its ability to avoid rank reversal. Compared to AHP_LGP, BWM (Rezaei, 2015, 2016) needs less data to obtain the weights of alternatives, reducing the difficulty for decision makers to provide pairwise comparison data. As want to know whether a reduction in the amount of data could affect the rank reversal probability, BWM is also included in this paper.

The MCDM methods that use evaluation data under criteria as input are divided into three subgroups according to different decision making ideas, as indicated below.

(1) In the first subgroup, MCDM methods rank alternatives according to the dominance degrees of alternatives. PROMETHEE II (Brans and Vincke, 1985) is a well-known and widely used method. It is based on a preference function that can be selected from the following five options: the usual function, the quasi-function, the linear preference function, the level function, and the Gaussian

[3] Actually, AHP_RE and AHP_LLSM are equivalent in case of three alternatives (Crawford and Williams, 1985).

function (for more information on these five functions, we refer to Brans and Vincke, 1985). These five preference functions correspond to five versions of PROMETHEE II denoted by PROMETHEE II_T with $\mathrm{T} \in \{\mathrm{U}, \mathrm{Q}, \mathrm{LP}, \mathrm{L}, \mathrm{G}\}$. In this study, these five versions of PROMETHEE II are referred to as classical methods. TODIM (Gomes and Lima, 1991; Leoneti and Gomes, 2021) and GLDS (Wu and Liao, 2019) take into account human psychological factors, such as regret in GLDS and loss aversion in TODIM, when computing dominance degrees. For TODIM, we need to select one standardization technique from four options: max linear, max-min, sum linear, or vector standardization. Also, we need to select one Phi function from three different types. Therefore, TODIM has $4 \times 3 = 12$ versions, which are denoted by TODIM_ST_TYPE where $\mathrm{ST} \in \{\mathrm{Max}, \mathrm{Maxmin}, \mathrm{Sum}, \mathrm{Vector}\}$ and $\mathrm{TYPE} \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$. The standardization techniques and the Phi functions can be found in Leoneti and Gomes (2021). This paper includes GLDS and twelve versions of TODIM as improvements to PROMETHEE II.

(2) In the second subgroup of MCDM methods, a weighted approach is used to rank the alternatives. Two classical weighted methods are SWM and MEW. Recently, Cinelli et al. (2020) stated that for weighted methods, it is important to consider compensation between criteria. As DNMA (Liao and Wu, 2020) considers complete compensatory, non-compensatory, and incomplete compensatory criteria, it is used as an update of SWM and MEW.

(3) In the third subgroup, TOPSIS (Chen and Hwang, 1963) is a classical method based on the positive and negative ideal solutions. Similar to TODIM, TOPSIS requires one standardization technique, resulting in four versions referred to as TOPSIS_ST where $\mathrm{ST} \in \{\mathrm{Max}, \mathrm{Maxmin}, \mathrm{Sum}, \mathrm{Vector}\}$.

As shown in Table 1, Zanakis et al. (1998) considered more MCDM methods than Dede et al. (2015) and Wang and Triantaphyllou (2008), but they did not consider the popular methods PROMETHEE II and TODIM. Compared to Zanakis et al. (1998), this paper not only includes all versions of popular MCDM methods (except the ELECTRE methods), but also takes into account recent improvements to the classical methods, such as DNMA and GLDS. The MCDM methods considered in this study are extensively used and reflect a variety of decision-making ideas as well as significant improvements to MCDM methods.

3.2 Simulation environments and parameters

Given that there are two kinds of input data, *i.e.*, pairwise comparison data and evaluation data, this paper includes two simulation studies according to the type of input data.

The first simulation study, referred to as the simulations of comparison information (CIS), uses pairwise comparison data as input. The pairwise comparison data are drawn from a ratio scale

Table 1: List of MCDM methods used in four studies.

Method	Type of input	Idea for decision making	Included in Paper 1?	Included in Paper 2?	Included in Paper 3?	Included in this paper?
AHP_RE	Pairwise comparison data	Right eigenvector	✓	✓	✗	✓
QSP	Pairwise comparison data	Right eigenvector	✗	✗	✗	✓
AHP_LLSM	Pairwise comparison data	Geometric mean	✗	✓	✗	✓
AHP_LGP	Pairwise comparison data	Goal programming	✗	✗	✗	✓
BWM	Pairwise comparison data	Goal programming	✗	✗	✗	✓
PROMETHEE II	Evaluation data	Dominance degree	✗	✗	✗	All five versions
TODIM	Evaluation data	Dominance degree	✗	✗	✗	All twelve versions
GLDS	Evaluation data	Dominance degree	✗	✗	✗	✓
ELECTRE	Evaluation data	Dominance degree	✓	✗	✓	✗
SWM	Evaluation data	Weight	✓	✗	✗	✓
MEW	Evaluation data	Weight	✓	✗	✗	✓
DNMA	Evaluation data	Weight	✗	✗	✗	✓
TOPSIS	Evaluation data	Ideal solution	Only one version	✗	✗	All four versions

Note. Paper 1 is Zanakis et al. (1998). Paper 2 is Dede et al. (2015). Paper 3 is Wang and Triantaphyllou (2008).

$S = \{\frac{1}{M}, \frac{1}{M-1}, \ldots, \frac{1}{2}, 1, 2, \ldots, M-1, M\}$ where M is the ratio scale maximum. The comparison information in CIS can be of two types, *i.e.*, consistent data and inconsistent data, corresponding to two simulation environments. Saaty (1977) provided a method for judging the consistency of pairwise comparison data. Each environment has two simulation parameters: the ratio scale maximum and the number of alternatives. In this paper, the number of alternatives is set to 3, 5, 7, 9, 11, 13, and 15 in CIS. In Zanakis et al. (1998), the values were 3, 5, 7, and 9. Due to the small number of alternatives, the rank reversal probability did not stabilize as the number of alternatives increased in Zanakis et al. (1998). This study expands the number of alternatives to investigate how the rank reversal probability changes with an increasing number of alternatives. The ratio scale maximum, as suggested by Saaty (1977), is set to 5, 7, and 9.

The simulations of evaluation information (EIS), which is the second simulation study, use evaluation data under criteria as input. EIS has four simulation environments:

(1) The first environment considers evaluation data sampled from a standard normal distribution and evaluation-independent criterion weights that are sampled from a uniform distribution on $[0, 1]$.

(2) The second environment considers evaluation data sampled from a standard normal distribution and evaluation-dependent criterion weights that are generated by the method presented in Diakoulaki et al. (1995) from evaluation data (see Appendix A).

(3) The third environment considers evaluation data sampled from a uniform distribution on $[0, 1]$ and evaluation-independent criterion weights sampled from a uniform distribution on $[0, 1]$.

(4) The fourth environment considers evaluation data sampled from a uniform distribution on $[0, 1]$ and evaluation-dependent criterion weights.

Each environment has two simulation parameters: the number of alternatives and the number of criteria. The number of alternatives is set to 3, 5, 7, 9, 11, 13, and 15. The number of criteria is set to 2, 3, 4, 5, 10, 15, 20, and 30. The criteria in this paper are benefit criteria, meaning that the higher the evaluations of an alternative are, the better the alternative is.

Two points need to be stated in EIS. The first point is about the selection of the data distribution. Zanakis et al. (1998) only considered the uniform distribution, thus generating positive evaluation data with equal probability only. In real problems, evaluation data might be not only positive but also negative, and the probabilities might be different as well. Hence, we use the standard normal distribution to create both positive and negative evaluation data with different probabilities. Here, we do not consider skewed distributions or heavy-tailed distributions because they have too many parameters to facilitate the comparison of results. Therefore, only the uniform distribution on [0,1] and the standard normal distribution are considered. The second point is about the interactions between criteria. In practical problems, there may be interactions between criteria. For example, if a student does well in mathematics, it is highly likely that he/she will also do well in physics. Here, there is an interaction between math and physics. Some special tools, such as the Choquet integral (Choquet, 1954), have been used in conjunction with MCDM methods to address issues where criteria interact. However, some methods in Table 1, such as DNMA, have not been combined with these special tools, making the study of these interactions in the context of rank reversal difficult. Therefore, the criteria are considered as being non-interactive in this paper.

Table 2 summarizes the environments and parameters considered in this study. Compared with Zanakis et al. (1998), this study considers more parameters and environments. For example, criterion weights can be directly generated from a distribution as well as derived from evaluation data. Zanakis et al. (1998) did not consider the second situation, while this study considers both cases.

3.3 Simulation steps

Step 1. Create original input data. For CIS, there are 7 values for the number of alternatives $\times$ 3 values for the ratio scale maximum $\times$ 2 consistency levels (*i.e.*, consistent and inconsistent) = 42 simulations. Each simulation consists of 1000 replicates, *i.e.*, creating 1000 matrices of pairwise comparison data denoted by D_C. The method used to obtain the pairwise comparison data is given in Appendix B. For EIS, there are 7 values for the number of alternatives $\times$ 8 values for the number of

Table 2: Comparisons of simulation environments and parameters.

		Zanakis et al. (1998)	This paper
Parameters	Number of alternatives	3, 5, 7, 9	3, 5, 7, 9, 11, 13, 15
	Number of criteria	5, 10, 15, 20	2, 3, 4, 5, 10, 15, 20, 30
	Ratio scale maximum	Not considered	5, 7, 9
Environments	Consistency	Not considered	Considered
	Distribution type of evaluation	Uniform distribution	Standard normal distribution Uniform distribution
	Criterion weight type	Equal weights Uniform distribution U-shaped distribution	Uniform distribution Evaluation-dependent weights

Table 3: Values of thresholds or parameters of MCDM methods in this paper.

Method	Threshold/Parameters	Values
PROMETHEE II	Indifference threshold	0.1, 0.3, 0.5, 0.7, 0.9 and smaller than the value of the preference threshold
	Preference threshold	0.1, 0.3, 0.5, 0.7, 0.9
	Parameter in Gaussian function	1, 2, 3
TODIM	Parameter in Type III Phi function	1, 2, 3

criteria $\times$ 2 distribution types (*i.e.*, uniform and standard normal distributions) $\times$ 2 criterion weight types (*i.e.*, evaluation-dependent and evaluation-independent weights) = 224 simulations. Each simulation again consists of 1000 replicates, *i.e.*, creating 1000 matrices of evaluation data denoted by D_E. Here, we separately create data under different criteria to make criteria non-interactive. The 1000 vectors of evaluation-independent weights generated from the uniform distribution on $[0, 1]$ are denoted by W_i. The 1000 vectors of evaluation-dependent weights are denoted by W_d and are calculated from D_E.

Step 2. Calculate original rankings. The MCDM methods in Table 1 are used to obtain the ranking (with ties) of alternatives from D_C or D_E. In some MCDM methods, there might be thresholds or parameters that need to be determined by decision makers. The values of thresholds or parameters used in this paper are listed in Table 3. In each simulation, we obtain 1000 original rankings from D_C (resp. D_E) denoted by R_C (resp. R_E). In this paper, scores are kept to four decimals. Two alternatives are judged tied if their scores are the same up to four decimals.

Step 3. Add an alternative. For CIS, we add an alternative to D_C with values randomly sampled from the ratio scale to create D_C^r. In the simulations with consistent information, the pairwise comparison data should be consistent before and after adding an alternative. Therefore, if D_C is consistent, then D_C^r is also consistent. For EIS, we also add an alternative to D_E with random evaluations, creating D_E^r. The distribution type of the new alternative's evaluations is in line with the simulation environment. After adding an alternative, the evaluation data are changed. Therefore, the evaluation-dependent weights W_d should be recalculated, and the 1000 new weight vectors are given in W_d^r corresponding to D_E^r.

Step 4. Delete an alternative. For CIS, we independently delete an alternative from D_C to obtain several data matrices D_C^δ, where $\delta = 1, 2, \ldots, n$ and n is the original number of alternatives. For EIS, we independently delete an alternative from D_E to create several data matrices D_E^δ, where $\delta = 1, 2, \ldots, n$ and n is the original number of alternatives. After deleting an alternative, the evaluation data are changed. Therefore, the evaluation-dependent weights W_d should be recalculated, and the 1000 new weight vectors are given in W_d^δ corresponding to D_E^δ where $\delta = 1, 2, \ldots, n$.

Step 5. Calculate the rankings after adding or deleting an alternative. Step 5 is similar to Step 2. In each simulation, we can derive rankings with ties R_C^r, R_C^δ, R_E^r, and R_E^δ from D_C^r, D_C^δ, D_E^r, and D_E^δ, respectively.

Step 6. Calculate probabilities. In this step, we show the process of CIS as an example for analyzing the results. The same process is done in EIS. For CIS, we can compare R_C^r with R_C to count the replicates showing rank reversal, those showing tie breaking, and those showing tie making when adding an alternative. The number of replicates showing rank reversal can be identified according to Definition 1. The rank reversal probability is calculated as

$$P_1 = \frac{\nu_1}{\nu_2}, \tag{1}$$

where ν_1 is the number of replicates showing rank reversal, and ν_2 is the number of replicates containing at least one strict preference relation. For instance, in every replicate, there are $\frac{n(n-1)}{2}$ pairs of alternatives that are common to R_C^r and R_C. According to Definition 1, we can count the number of pairs of alternatives showing rank reversal by comparing the rankings in R_C^r and R_C. A replicate shows rank reversal when the number of pairs of alternatives showing rank reversal is not 0. Similarly, according to Definition 2, we can calculate the tie making probability as

$$P_2 = \frac{\nu_3}{\nu_2}, \tag{2}$$

where ν_3 is the number of replicates showing tie making, while the tie breaking probability is calculated as

$$P_3 = \frac{\nu_4}{\nu_5}, \tag{3}$$

where ν_4 is the number of replicates showing tie breaking and ν_5 is the number of replicates containing at least one tie. Referring to Definition 2, tie breaking presupposes the existence of a tie, so the denominator of Eq. (3) is the number of replicates containing at least one tie.

In the case of deleting an alternative, we further calculate the above three kinds of probabilities when deleting the alternative at different ranks to investigate the relation between the probabilities and the ranks of the alternative being deleted (with the rank referring to the original ranking).

After obtaining the probabilities, we do step-wise regressions (Draper and Smith, 1998) with the simulation factors as independent variables and the probabilities as the dependent variable to investigate the effects of simulation environments and parameters on the probabilities.

Things worth noting in Step 6 are shown below.

(1) When the MCDM method considered includes thresholds, we calculate the average of the probabilities of rank reversal, tie making, and tie breaking over different thresholds.

(2) By comparing every R_C^δ (resp. R_E^δ), $\delta = 1, 2, \ldots, n$, with R_C (resp. R_E), we can obtain the rank-based probabilities of rank reversal, tie making and tie breaking. When two alternatives are tied, we first calculate the probabilities obtained by deleting the two alternatives separately. Then, the rank-based probabilities at the ranks of the two alternatives are defined as the average of the two probabilities. For instance, suppose that there are four alternatives and that the rank reversal probabilities when separately deleting the alternatives from the first rank to the last rank are 0.3, 0.4, 0.6, and 0.5. The rank-based probability of rank reversal at the first rank is 0.3. The rank-based probabilities of rank reversal at the second and third ranks are $\frac{0.4+0.6}{2} = 0.5$. The rank-based probability of rank reversal at the last rank is 0.5.

(3) In the regression analyses, we use the average of rank-based probabilities at different ranks to represent the probability caused by deleting an alternative.

4 Experimental results

The probabilities and rank-based probabilities of rank reversal, tie making, and tie breaking are shown in the figures of the supplementary material. In Section 4.1, we compare the probabilities of rank reversal, tie making, and tie breaking for different MCDM methods and provide some summaries. In Section 4.2, we discuss the influence of the simulation environments and parameters on the three kinds of probabilities. In Section 4.3, we study the relations between the ranks of alternatives and the probabilities, *i.e.*, the rank-based probabilities.

4.1 Comparing the probabilities of different MCDM methods

Comparisons in CIS.

Usually, QSP, an improvement to AHP_RE, has lower probabilities of rank reversal, tie making, and tie breaking than AHP_RE, AHP_LGP, AHP_LLSM, and BWM. BWM using incomplete pairwise comparison data has higher probabilities of rank reversal, tie making, and tie breaking than the other four methods. AHP_LGP and AHP_LLSM have almost the same probabilities of rank reversal and tie breaking, but the tie making probability of AHP_LGP is higher than that of AHP_LLSM if inconsistent pairwise comparison data is used. We notice that if the data is consistent, then the probabilities of rank reversal, tie making, and tie breaking are all 0 for AHP_RE, AHP_LGP, and AHP_LLSM.

Comparisons in EIS.

For TODIM and TOPSIS, the sum linear standardization technique results in a higher rank reversal probability than the other three standardization techniques. We also find that DNMA, a decision making method considering compensation among criteria, has a higher rank reversal probability than SWM and MEW, which means that considering compensation among criteria can increase the rank reversal probability. As for the tie making probability, PROMETHEE II_U has a higher tie making probability than the other methods in EIS. An explanation for this observation is that the small output range of the usual function, *i.e.*, {0, 0.5, 1}, makes it easy to produce the same dominance degree for alternatives, leading to high tie making probabilities. As for the tie breaking probability, PROMETHEE II_U has a lower tie breaking probability than the other methods, which means that the tie relation produced by PROMETHEE II_U is stable if the set of alternatives is changed. On the contrary, the tie relations produced by other methods, such as TOPSIS and TODIM, are easily broken.

4.2 Discussion on the effects of simulation environments and parameters

4.2.1 *Discussion on the simulations of comparison information*

We discuss the effects of four simulation environments and parameters, *i.e.*, the number of alternatives, the ratio scale maximum, the information consistency, and the operation of deleting or adding alternatives, on the probabilities of rank reversal, tie making, and tie breaking for AHP_RE, AHP_LGP, AHP_LLSM, QSP, and BWM . The regression model used in this section is

$$\mathit{Probability} = \mathit{Constant} + \theta_1 \times n + \theta_2 \times M + \theta_3 \times \mathit{consistency} + \theta_4 \times \mathit{operation}\,, \tag{4}$$

where n is the number of alternatives and M is the ratio scale maximum. The values of consistency are "0" and "1": "1" means the data is consistent, while "0" means not. The values of operation are "0" and "1": "0" represents the operation of deleting an alternative, while "1" represents the operation of adding an alternative. θ_1, θ_2, θ_3, and θ_4 are the coefficients of the number of alternatives, ratio scale maximum, consistency, and operation, respectively.

Summary for rank reversal probability.

Table 4 shows the regression results for the rank reversal probabilities. The ratio scale maximum and the operation of adding or deleting alternatives do not affect the rank reversal probability. For AHP_RE, consistent information results in lower rank reversal probabilities than inconsistent information. For AHP_LGP, AHP_LLSM, and BWM (resp. QSP), a rise in the number of alternatives increases (resp. decreases) the rank reversal probabilities, and consistent information results in lower (resp. higher) rank reversal probabilities than inconsistent information.

Table 4: Regression results for rank reversal probabilities in CIS.

Method	θ_1	θ_2	θ_3	θ_4	P value
AHP_RE	0	0	-0.7384	0	0
AHP_LGP	0.0242	0	-0.7149	0	0
AHP_LLSM	0.0278	0	-0.8143	0	0
BWM	0.029	0	-0.0412	0	0
QSP	-0.0241	0	0.1136	0	0

Note. For all numbers in the table, we retain four decimal places. For all tables in Section 4.2, we consider the model valid when the P value is smaller than or equal to 0.05. In a valid model, when the coefficient is higher than 0, the corresponding parameter has a significant and positive effect on the probability, while when the coefficient is smaller than 0, the effect is significant and negative. When the coefficient is 0, the corresponding parameter has no significant effect on the probability. In this paper, the word "significant" is omitted.

Summary for tie making probability.

Table 5 shows the regression results for the tie making probabilities. For AHP_LGP and BWM, a rise in the number of alternatives increases the tie making probabilities, while the opposite holds for AHP_LLSM and QSP. The ratio scale maximum only influences the tie making probability of AHP_LLSM, the rise of which decreases the tie making probability. For all five methods, consistent information results in lower tie making probabilities than inconsistent information. For AHP_LLSM, the operation of adding an alternative results in lower tie making probabilities than the operation of deleting an alternative, while the opposite holds for QSP.

Summary for tie breaking probability.

Table 5: Regression results for tie making probabilities in CIS.

Method	θ_1	θ_2	θ_3	θ_4	P value
AHP_RE	0	0	-0.0204	0	0
AHP_LGP	0.0357	0	-0.6774	0	0
AHP_LLSM	-0.0008	-0.0044	-0.0237	-0.0072	0
BWM	0.0533	0	-0.4462	0	0
QSP	-0.0098	0	-0.0635	0.0373	0

Table 6: Regression results for tie breaking probabilities in CIS.

Method	θ_1	θ_2	θ_3	θ_4	P value
AHP_RE	0.0064	0	-0.9607	0.0302	0
AHP_LGP	0.0207	0	-0.7727	0	0
AHP_LLSM	0.0065	0	-0.9202	0	0
BWM	0.0352	0	-0.4909	-0.0685	0
QSP	-0.0232	0	-0.168	-0.146	0

Table 6 shows the regression results for the tie breaking probabilities. For QSP, a rise in the number of alternatives decreases the tie breaking probabilities, while the opposite is true for the other four methods. The ratio scale maximum has no effect on the tie breaking probabilities of all five methods. Also, consistent information results in lower tie breaking probabilities than inconsistent information. For AHP_RE, the operation of adding an alternative has a higher tie breaking probability than the operation of deleting an alternative, while the opposite holds for BWM and QSP.

4.2.2 *Discussion on the simulation of evaluation information*

We discuss the effects of five simulation environments and parameters, *i.e.*, the number of criteria, the number of alternatives, data distribution type, criterion weight type, and the operations of deleting or adding an alternative, on the probabilities of rank reversal, tie making, and tie breaking for GLDS, DNMA, SWM, MEW, five versions of PROMETHEE II, twelve versions of TODIM, and four versions of TOPSIS. The regression model used in this section is

$$Probability = Constant + \theta_1 \times n + \theta_5 \times N_c + \theta_6 \times \rho_1 + \theta_7 \times \rho_2 + \theta_4 \times operation\,, \tag{5}$$

where N_c is the number of criteria. ρ_1 represents the data distribution type, the values of which are "1" and "0", referring to uniformly distributed data and data sampled from the standard normal distribution, respectively. ρ_2 is the criterion weight type: it takes value "0" when the weights are evaluation-independent, and the value "1" when the weights are evaluation-dependent. θ_5, θ_6, and θ_7 are the coefficients of the number of criteria, data distribution type, and criterion weight type, respectively.

Summary for rank reversal probability.

Table 7 shows the regression results for the rank reversal probabilities. The effects of the five simulation environments and parameters are summarized as follows.

Effects of the number of alternatives. For all methods in EIS, the rank reversal probabilities go up if the number of alternatives rises.

Effects of the number of criteria. For all methods in EIS, the rank reversal probabilities increase if the number of criteria rises.

Effects of data distribution type. For PROMETHEE II_U, GLDS, SWM, and MEW, the data distribution type does not affect the rank reversal probabilities. For PROMETHEE II_G and TOPSIS_Vector, data sampled from the uniform distribution results in higher rank reversal probabilities than data sampled from the standard normal distribution, while the opposite holds for the remaining nineteen methods/versions.

Effects of criterion weight type. For the five versions of PROMETHEE II, the criterion weight type has no influence on the rank reversal probabilities. For the other methods/versions, the evaluation-dependent weights lead to higher rank reversal probabilities than the evaluation-independent weights.

Effects of operation. For SWM and MEW, the operations do not affect the rank reversal probabilities. For the other methods/versions, the operation of adding an alternative leads to higher rank reversal probabilities than the operation of deleting an alternative.

Summary for tie making probability.

Table 8 shows the regression results for the tie making probabilities. The effects of the five simulation environments and parameters are summarized as follows.

Effects of the number of alternatives. For TOPSIS_Maxmin, the tie making probability is not affected by the number of alternatives. As the number of alternatives increases, the tie making probabilities of PROMETHEE II_LP, TODIM_Max_III, and three versions of TODIM_Maxmin decrease, while the opposite holds for the remaining nineteen methods/versions.

Effects of the number of criteria. For the five versions of PROMETHEE II, the tie making probability decreases if the number of criteria rises. The number of criteria has no influence on the tie making probabilities for TODIM_Max_III and three versions of TODIM_Maxmin. A rise in the number of criteria increases the tie making probabilities of the remaining sixteen methods/versions.

Table 7: Regression results for rank reversal probabilities in EIS.

Method	θ_1	θ_5	θ_6	θ_7	θ_4	P value	Method	θ_1	θ_5	θ_6	θ_7	θ_4	P value
PROMETHEE II_U	0.0293	0.0141	0	0	0.4509	0	TODIM_Sum_III	0.0329	0.0064	-0.417	0.0941	0.0407	0
PROMETHEE II_Q	0.0435	0.0095	-0.0968	0	0.2501	0	TODIM_Vector_I	0.0311	0.0055	-0.1337	0.1463	0.028	0
PROMETHEE II_LP	0.0453	0.0072	-0.0616	0	0.2395	0	TODIM_Vector_II	0.0388	0.0035	-0.0734	0.1128	0.0303	0
PROMETHEE II_L	0.0453	0.0085	-0.0822	0	0.246	0	TODIM_Vector_III	0.0381	0.0037	-0.0712	0.1128	0.0309	0
PROMETHEE II_G	0.0444	0.0081	0.049	0	0.2079	0	GLDS	0.0375	0.0118	0	0.0932	0.0478	0
TODIM_Max_I	0.0267	0.009	-0.1321	0.1656	0.0313	0	SWM	0.0175	0.0021	0	0.3873	0	0
TODIM_Max_II	0.04	0.0041	-0.0777	0.1071	0.0384	0	MEW	0.0183	0.0021	0	0.3458	0	0
TODIM_Max_III	0.0362	0.0051	-0.0947	0.113	0.0358	0	DNMA	0.0326	0.0113	-0.1118	0.2495	0.0349	0
TODIM_Maxmin_I	0.0309	0.0081	-0.0405	0.1499	0.0373	0	TOPSIS_Max	0.0222	0.0118	-0.1011	0.3031	0.0319	0
TODIM_Maxmin_II	0.0443	0.0038	-0.0133	0.0972	0.0484	0	TOPSIS_Maxmin	0.0191	0.0107	-0.0402	0.3817	0.0304	0
TODIM_Maxmin_III	0.0404	0.0045	-0.026	0.1049	0.0446	0	TOPSIS_Sum	0.0348	0.0095	-0.2954	0.0905	0.0506	0
TODIM_Sum_I	0.0297	0.006	-0.5305	0.1052	0.04	0	TOPSIS_Vector	0.0366	0.0101	0.0226	0.1748	0.0426	0
TODIM_Sum_II	0.0297	0.0069	-0.3759	0.1059	0.0378	0							

Effects of data distribution type. Data sampled from the standard normal distribution results in smaller tie making probabilities than data sampled from the uniform distribution in SWM, MEW, DNMA, PROMETHEE II_Q, PROMETHEE II_L, PROMETHEE II_G, TODIM_Max_I, three versions of TODIM_Sum, TODIM_Vector_I, TODIM_Vector_III, TOPSIS_Max, and TOPSIS_Sum, while the opposite holds for TOPSIS_Vector. The data distribution type does not affect the tie making probabilities of the other ten methods/versions.

Effects of criterion weight type. The criterion weight type has no influence on the tie making probabilities of TODIM_Max_I, TODIM_Vector_II, and five versions of PROMETHEE II. For the remaining eighteen methods/versions, the evaluation-dependent weights lead to higher tie making probabilities than the evaluation-independent weights.

Effects of operation. For PROMETHEE II_LP, TODIM_Max_III, TOPSIS_Maxmin, and three versions of TODIM_Maxmin, the operation of deleting an alternative generates higher tie making probabilities than the operation of adding an alternative, while the opposite is true for GLDS, DNMA, PROMETHEE II_U, TODIM_Vector_I, TOPSIS_Max, TOPSIS_Sum, and TOPSIS_Vector. The operation does not affect the tie making probabilities of the remaining twelve methods/versions.

Summary for tie breaking probability.

Table 9 shows the regression results for the tie breaking probabilities. The *P* value is "NaN" when the coefficients of the parameters are all 0. The effects are summarized as follows.

Effects of the number of alternatives. A rise in the number of alternatives decreases the tie breaking probabilities of GLDS, SWM, DNMA, TOPSIS_Max, TOPSIS_Maxmin, and TOPSIS_Sum, while the opposite is true for PROMETHEE II_U. The tie breaking probabilities of the remaining eighteen methods/versions are not affected by the number of alternatives.

Table 8: Regression results for tie making probabilities in EIS.

Method	θ_1	θ_5	θ_6	θ_7	θ_4	P value	Method	θ_1	θ_5	θ_6	θ_7	θ_4	P value
PROMETHEE II_U	0.0551	-0.008	0	0	0.0703	0	TODIM_Sum_III	0.00005	0.00002	0.0003	0.0001	0	0
PROMETHEE II_Q	0.013	-0.0052	0.0674	0	0	0	TODIM_Vector_I	0.0001	0.00004	0.0005	0.0002	0.0002	0
PROMETHEE II_LP	-0.0002	-0.00008	0	0	-0.0013	0	TODIM_Vector_II	0.00004	0.00002	0	0	0	0
PROMETHEE II_L	0.0073	-0.0032	0.0395	0	0	0	TODIM_Vector_III	0.00004	0.00002	0.0001	0.00009	0	0
PROMETHEE II_G	0.0006	-0.0001	0.0041	0	0	0	GLDS	0.0085	0.0009	0	0.0075	0.0127	0
TODIM_Max_I	0.00005	0.00002	0.0003	0	0	0	SWM	0.0007	0.0002	0.0042	0.0078	0	0
TODIM_Max_II	0.00003	0.00001	0	0.00002	0	0	MEW	0.0005	0.0001	0.0048	0.005	0	0
TODIM_Max_III	-0.0051	0	0	0.0273	-0.0273	0	DNMA	0.0011	0.00009	0.002	0.001	0.0014	0
TODIM_Maxmin_I	-0.005	0	0	0.0271	-0.0269	0	TOPSIS_Max	0.0022	0.0006	0.0016	0.0062	0.0025	0
TODIM_Maxmin_II	-0.0051	0	0	0.0273	-0.0273	0	TOPSIS_Maxmin	0	0.0006	0	0.02	-0.0108	0
TODIM_Maxmin_III	-0.0051	0	0	0.0273	-0.0273	0	TOPSIS_Sum	0.0018	0.0003	0.0053	0.002	0.0018	0
TODIM_Sum_I	0.0003	0.00006	0.0023	0.0003	0	0	TOPSIS_Vector	0.0026	0.0007	-0.0016	0.0052	0.0029	0
TODIM_Sum_II	0.00005	0.00002	0.0002	0.0002	0	0							

Effects of the number of criteria. A rise in the number of criteria decreases the tie breaking probability of MEW, while the opposite is true for PROMETHEE II_U, PROMETHEE II_Q, PROMETHEE II_L, PROMETHEE II_G, GLDS, and four versions of TOPSIS. The tie breaking probabilities of the remaining fifteen methods/versions are not affected by the number of criteria.

Effects of data distribution type. For MEW, data sampled from the standard normal distribution results in smaller tie breaking probabilities than data sampled from the uniform distribution, while the opposite is true for PROMETHEE II_Q, PROMETHEE II_L, PROMETHEE II_G, and TOPSIS_Vector. The data distribution type does not affect the tie breaking probabilities of the other twenty methods/versions.

Effects of criterion weight type. For SWM, MEW, DNMA, TOPSIS_Max, and TOPSIS_Maxmin, the evaluation-dependent weights lead to higher tie breaking probabilities than the evaluation-independent weights. The criterion weight type has no influence on the tie breaking probabilities of the other tewnty methods/versions.

Effects of operation. For GLDS, deleting an alternative leads to higher tie breaking probabilities than adding an alternative, while the opposite is true for TODIM_Max_I, DNMA, and five versions of PROMETHEE II. These operations do not affect the tie breaking probabilities of the other seventeen methods/versions.

4.2.3 *Conclusions for the effects of simulation environments and parameters*

In CIS, consistent information does not result in higher probabilities of rank reversal, tie making, and tie breaking than inconsistent information in most cases. A rise in the ratio scale maximum does not increase the three kinds of probabilities. For all MCDM methods in EIS, we find that the evaluation-independent weights do not increase the probabilities of rank reversal, tie making, and tie

Table 9: Regression results for tie breaking probabilities in EIS.

Method	θ_1	θ_5	θ_6	θ_7	θ_4	P value	Method	θ_1	θ_5	θ_6	θ_7	θ_4	P value
PROMETHEE II_U	0.006	0.0067	0	0	0.1285	0	TODIM_Sum_III	0	0	0	0	0	NaN
PROMETHEE II_Q	0	0.0059	-0.0465	0	0.048	0	TODIM_Vector_I	0	0	0	0	0	NaN
PROMETHEE II_LP	0	0	0	0	0.6822	0	TODIM_Vector_II	0	0	0	0	0	NaN
PROMETHEE II_L	0	0.0053	-0.0402	0	0.0438	0	TODIM_Vector_III	0	0	0	0	0	NaN
PROMETHEE II_G	0	0.0016	-0.0154	0	0.0193	0	GLDS	-0.0015	0.0004	0	0	-0.0047	0
TODIM_Max_I	0	0	0	0	0.0017	0	SWM	-0.0007	0	0	0.9932	0	0
TODIM_Max_II	0	0	0	0	0	NaN	MEW	0	-0.0017	0.0873	0.9842	0	0
TODIM_Max_III	0	0	0	0	0	NaN	DNMA	-0.0007	0	0	0.0047	0.0029	0.0022
TODIM_Maxmin_I	0	0	0	0	0	NaN	TOPSIS_Max	-0.0122	0.0107	0	0.2927	0	0
TODIM_Maxmin_II	0	0	0	0	0	NaN	TOPSIS_Maxmin	-0.0122	0.0108	0	0.29	0	0
TODIM_Maxmin_III	0	0	0	0	0	NaN	TOPSIS_Sum	-0.0008	0.0002	-0.005	0	0	0
TODIM_Sum_I	0	0	0	0	0	NaN	TOPSIS_Vector	0	0.001	0	0	0	0
TODIM_Sum_II	0	0	0	0	0	NaN							

breaking.

4.3 Discussions on the rank-based probabilities

In this section, we show only two figures: a figure about the rank-based rank reversal probabilities for CIS using inconsistent data and a figure about the rank-based rank reversal probabilities of TODIM methods with Type I Phi function since they clearly show how the rank-based probabilities change as the rank increases. The figures about the rank-based probabilities of other methods are given in the supplementary material.

4.3.1 *Discussions on the rank-based rank reversal probability*

In Figure 1, we can find some "$\wedge$"-shaped curves for AHP_RE, AHP_LGP, AHP_LLSM, BWM, and QSP, meaning that deleting the alternative ranked first or last usually results in smaller rank reversal probabilities than deleting an alternative in the middle when the data is inconsistent. The "$\wedge$"-shaped curve feature diminishes as the number of alternatives increases and eventually disappears. In the supplementary material, only QSP has "$\wedge$"-shaped curves when the data is consistent.

When evaluation data is sampled from the uniform distribution on $[0, 1]$ and the criterion weights are evaluation-dependent, the rank-based rank reversal probabilities of the TODIM methods with Type I Phi function are shown in Figure 2 in which each row represents one standardization technique. We can find "$\vee$"-shaped curves in Figure 2, meaning that the deletion of the alternative ranked first or last mostly results in higher rank reversal probabilities than the deletion of an alternative in the middle. Referring to the supplementary material, we can find that, regardless of the type of

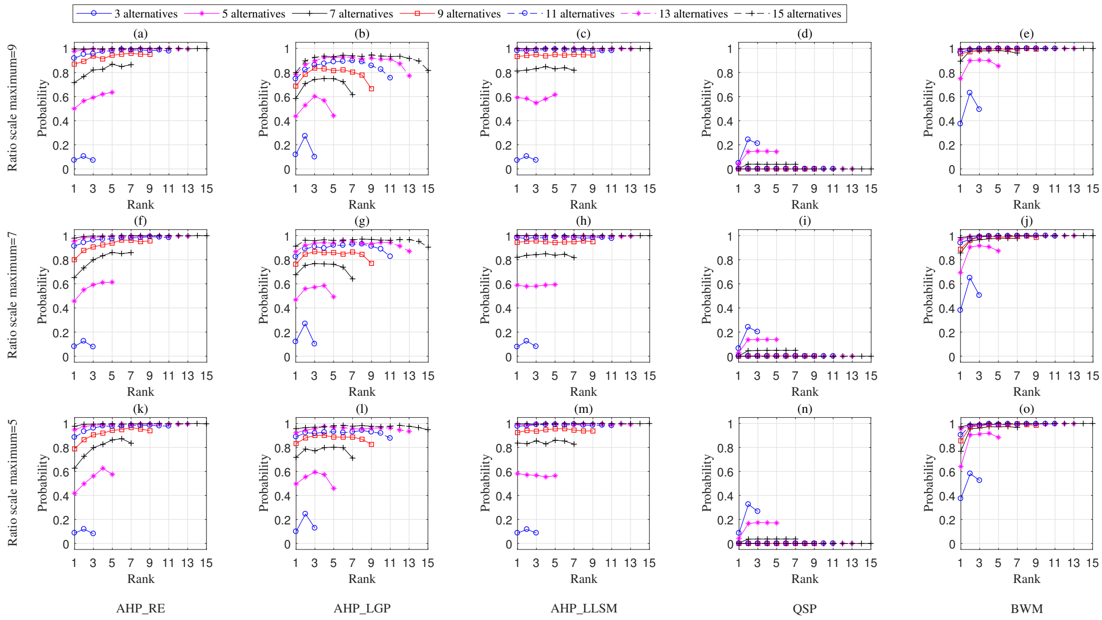

Figure 1: Rank-based rank reversal probabilities for CIS with inconsistent data.

data distribution, all methods/versions except SWM and MEW have "∨"-shaped curves when the weights are evaluation-independent. All methods/versions have "∨"-shaped curves when the weights are evaluation-dependent. In EIS, the "∨"-shaped curve feature also diminishes as the number of alternatives increases.

To summarize, for the rank-based rank reversal probability, the MCDM methods in EIS have a higher probability if the alternative ranked first or last is deleted than if the alternative in the middle is deleted, while the MCDM methods in CIS have the opposite feature.

4.3.2 *Discussions on the rank-based tie making probability*

The findings in this and the next section are all summarized from the figures in the supplementary material.

If the pairwise comparison data is consistent, then deleting the first or last alternative results in smaller tie making probabilities than deleting an alternative in the middle for BWM. For BWM and AHP_LGP, in the simulations with inconsistent information, the tie making probabilities at the first or last rank are smaller than the tie making probabilities at the middle ranks. AHP_RE, AHP_LLSM, and QSP do not show the "∧"-shaped curve feature in all simulations.

For PROMETHEE II_U, deleting the alternative at the first or last rank results in higher tie making probabilities than deleting an alternative in the middle for all data distribution types (*i.e.*, standard normal distribution and uniform distribution) and all types of weights (*i.e.*, evaluation-

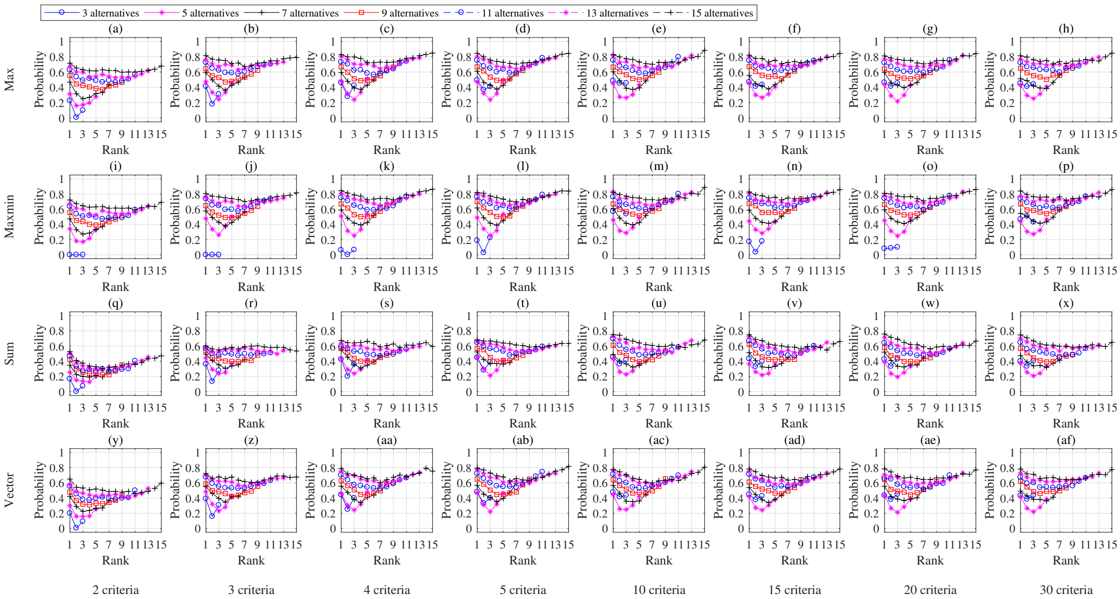

Figure 2: Rank-based rank reversal probabilities of TODIM methods with Type I Phi function when data is sampled from the uniform distribution on $[0, 1]$ and the criterion weights are evaluation-dependent.

dependent and evaluation-independent). For the other methods/versions in EIS, the tie making probabilities are almost the same for all ranks.

4.3.3 *Discussions on the rank-based tie breaking probability*

Only for BWM, deleting the alternative at the first or last rank results in smaller tie breaking probabilities than deleting an alternative in the middle if the pairwise comparison data is consistent. If the data is inconsistent, then all five methods in CIS have "∧"-shaped curves.

For some methods/versions, such as TOPSIS and TODIM, if the number of criteria is small, then the tie breaking probabilities fluctuate violently with increasing ranks. For any method in EIS, if the number of criteria is large, then the tie breaking probabilities are usually almost the same for all ranks. Therefore, a high number of criteria can help to eliminate the differences in the tie breaking probabilities at different ranks.

4.3.4 *Conclusions for the rank-based probabilities*

If an alternative is deleted from the set of alternatives, then the alternative's rank influences the probabilities of rank reversal, tie making, and tie breaking. Such influence on the three probabilities is associated with two factors.

(1) The first factor is the input data type. In most circumstances, deleting the first or last alternative results in smaller probabilities than deleting an alternative in the middle if the input data is pairwise comparison data. The opposite holds for evaluation data.

(2) The second factor is the number of alternatives. As the number of alternatives increases, the influence of the rank of the alternative deleted on the probabilities weakens, and the probabilities tend to be the same at all ranks.

5 Findings and suggestions

5.1 Findings

Finding 1. The influences of simulation environments and parameters on the probabilities of rank reversal, tie making, and tie breaking for CIS (resp. EIS) are summarised in Table 10 (resp. Table 11).

Table 10: Summary of the effects of simulation environments and parameters in CIS

	Rank reversal probability				Tie making probability				Tie breaking probability			
	NoA	RSM	Cons	Oper	NoA	RSM	Cons	Oper	Noa	RSM	Cons	Oper
AHP_RE	$\oslash$	$\oslash$	$\ominus$	$\ominus$	$\oslash$	$\oslash$	$\ominus$	$\oslash$	$\oplus$	$\oslash$	$\ominus$	$\oplus$
AHP_LGP	$\oplus$	$\oslash$	$\ominus$	$\oslash$	$\oplus$	$\oslash$	$\ominus$	$\oslash$	$\oplus$	$\oslash$	$\ominus$	$\oslash$
AHP_LLSM	$\oplus$	$\oslash$	$\ominus$	$\oslash$	$\ominus$	$\oslash$	$\ominus$	$\ominus$	$\oplus$	$\oslash$	$\ominus$	$\oslash$
BWM	$\oplus$	$\oslash$	$\ominus$	$\oslash$	$\oplus$	$\oslash$	$\ominus$	$\ominus$	$\oplus$	$\oslash$	$\ominus$	$\ominus$
QSP	$\ominus$	$\oslash$	$\oplus$	$\oslash$	$\ominus$	$\oslash$	$\ominus$	$\oplus$	$\ominus$	$\oslash$	$\ominus$	$\ominus$

Note. In Table 10, "NoA", "RSM", "Con", and "Oper" are abbreviations for "Number of alternatives", "Ratio scale maximum", "Consistency", and "Operation", respectively. "$\oplus$" (resp. "$\ominus$") represents "Positive relation" (resp. "Negative relation"). "$\oslash$" represents "No relation".

Table 11: Summary of the effects of simulation environments and parameters in EIS

	Rank reversal probability					Tie making probability					Tie breaking probability				
	NoA	NoC	DDT	WT	Oper	NoA	NoC	DDT	WT	Oper	NoA	NoC	DDT	WT	Oper
PROMETHEE II_U	$\oplus$	$\oplus$	$\oslash$	$\oslash$	$\oplus$	$\oplus$	$\ominus$	$\oslash$	$\oslash$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oslash$	$\oplus$
PROMETHEE II_Q	$\oplus$	$\oplus$	$\ominus$	$\oslash$	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oplus$	$\ominus$	$\oslash$	$\oplus$
PROMETHEE II_LP	$\oplus$	$\oplus$	$\ominus$	$\oslash$	$\oplus$	$\ominus$	$\ominus$	$\oslash$	$\oslash$	$\ominus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oplus$
PROMETHEE II_L	$\oplus$	$\oplus$	$\ominus$	$\oslash$	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oplus$	$\ominus$	$\oslash$	$\oplus$
PROMENTHEE II_G	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oplus$	$\ominus$	$\oslash$	$\oplus$
TODIM_Max_I	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oplus$
TODIM_Max_II	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
TODIM_Max_III	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\ominus$	$\oslash$	$\oslash$	$\oplus$	$\ominus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
TODIM_Maxmin_I	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\ominus$	$\oslash$	$\oslash$	$\oplus$	$\ominus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
TODIM_Maxmin_II	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\ominus$	$\oslash$	$\oslash$	$\oplus$	$\ominus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
TODIM_Maxmin_III	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\ominus$	$\oslash$	$\oslash$	$\oplus$	$\ominus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
TODIM_Sum_I	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
TODIM_Sum_II	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
TODIM_Sum_III	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
TODIM_Vector_I	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
TODIM_Vector_II	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
TODIM_Vector_III	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$	$\oslash$
GLDS	$\oplus$	$\oplus$	$\oslash$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oslash$	$\oslash$	$\ominus$
SWM	$\oplus$	$\oplus$	$\oslash$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\ominus$	$\oslash$	$\oslash$	$\oplus$	$\oslash$
MEW	$\oplus$	$\oplus$	$\oslash$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oslash$	$\oslash$	$\ominus$	$\oplus$	$\oplus$	$\oslash$
DNMA	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\ominus$	$\oslash$	$\oslash$	$\oplus$	$\ominus$
TOPSIS_Max	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oslash$	$\oplus$	$\oslash$
TOPSIS_Maxmin	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oslash$	$\oplus$	$\ominus$	$\ominus$	$\oplus$	$\oslash$	$\oplus$	$\oslash$
TOPSIS_Sum	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\ominus$	$\oslash$	$\oslash$
TOPSIS_Vector	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus$	$\ominus$	$\oplus$	$\oplus$	$\oslash$	$\oplus$	$\oslash$	$\oslash$	$\oslash$

Note. In Table 11, "NoA", "NoC", "DDT", "WT", and "Oper" are abbreviations for "Number of alternatives", "Number of criteria", "Data distribution type", "Weight type", and "Operation", respectively. "$\oplus$" (resp. "$\ominus$") represents "Positive relation" (resp. "Negative relation"). "$\oslash$" represents "No relation".

Finding 2. In the case of deleting an alternative, we find that the rank of the alternative deleted affects the probabilities of rank reversal, tie making, and tie breaking. This rank effect can be reduced by increasing the number of alternatives and/or criteria.

Finding 3. In most cases, inconsistent pairwise comparison information results in higher probabilities of rank reversal, tie making, and tie breaking than consistent data. BWM, which uses less pairwise comparison data than AHP_RE, AHP_LGP, and AHP_LLSM, has higher probabilities of rank reversal, tie making, and tie breaking than other methods in CIS.

Finding 4. The use of evaluation-independent weights does not increase the three kinds of probabilities. The consideration of compensation among criteria in MCDM methods can increase the rank reversal probability. When considering evaluation-dependent weights or compensation among criteria, the rank of an alternative not only depends on the evaluations of the alternative itself, but also on those of other alternatives, which strengthens the connection between the rank of the alternative and the set of alternatives. As a result, the rank of the alternative is easily affected by the change of the set of alternatives.

5.2 Suggestions

Suggestion 1. If decision makers want to get a total order of alternatives (without ties) from evaluation data under criteria, PROMETHEE II is not recommended because ties have a higher chance of appearing in PROMETHEE II's results than in the results of other MCDM methods.

Suggestion 2. When employing MCDM methods using pairwise comparison data, the information is recommended to be complete and consistent. When employing MCDM methods using evaluation data, evaluation-independent weights rather than evaluation-dependent weights are recommended.

Suggestion 3. In many real-world decision making problems, such as supply chain selection in public tenders (Schotanus et al., 2022), decision makers tend to use MCDM methods with low rank reversal probabilities. We notice that some improvements to MCDM methods, such as QSP, reduce the rank reversal probability, while others, such as BWM, increase the rank reversal probability. We suggest that when researchers propose new MCDM methods or improve existing MCDM methods, they should consider rank reversal, tie making, and tie breaking to avoid a situation where the performance of MCDM methods is improved at the expense of greatly increasing the probabilities of rank reversal, tie making, and tie breaking.

6 Conclusions

In this paper, we investigated the probabilities of rank reversal, tie making, and tie breaking for thirty versions of twelve MCDM methods. We discussed the effects of six simulation factors, *i.e.*, type of data distribution, weight type, number of criteria, number of alternatives, ratio scale maximum, and information consistency, on the probabilities of rank reversal, tie making, and tie breaking. We also reported a new factor affecting the probabilities, *i.e.*, the rank of the alternative deleted. Four findings and three suggestions were given to help decision makers select and use MCDM methods.

The attitudes of decision makers to rank reversal are diverse. Some decision makers oppose rank reversal, while others can tolerate it. When selecting MCDM methods for real problems, decision makers can consider their attitude to rank reversal and select MCDM methods with reference to the rank reversal probabilities of MCDM methods. At this point, many studies, such as Cinelli et al. (2014), Schotanus et al. (2022), as well as the present paper, can provide useful information. However, screening and comparing MCDM methods are currently time-consuming and difficult. One way to address this issue is to improve existing methods or propose new methods to enable decision makers to control the probability of rank reversal with some parameters. In this way, decision makers do not need to go through the trouble of choosing the MCDM method. They only need to adjust the parameters. For example, if decision makers want to know the rankings of alternatives in both situations of high and low rank reversal probabilities, they do not need to find two different methods; they merely adjust the parameters to get two rankings. The idea of MCDM methods with controllable rank reversal is crucial and meaningful, which should be a subject of future study.

There are several limitations to this paper as well. For example, we did not consider interactions among criteria. We considered just one method to calculate the evaluation-dependent weights. The impact of various methods to determine criterion weights on the probabilities is a subject of future research. In addition, some MCDM methods, such as the ELECTRE methods, may produce incomparabilities. A further direction is to study shifts between strict preference, indifference and incomparability. Furthermore, the method to generate inconsistent pairwise comparison information in Appendix B can lead to a high level of inconsistency. Considering the influence of inconsistency levels on the three kinds of probabilities can be a subject of future research.

Acknowledgements

Many thanks to the editor and reviewers for their helpful comments. The work was supported by the National Natural Science Foundation of China (71971145, 72171158, 72371173), the State Scholarship Fund of China Scholarship Council (202106240077), and the Fundamental Research Funds for the Central Universities (2023ZY-SX016).

References

Aguarón, J., & Moreno-Jiménez, J. M. (2000). Local stability intervals in the analytic hierarchy process. *European Journal of Operational Research*, *125*(1), 113–132. https://doi.org/10.1016/S0377-2217(99)00204-0

Aires, R. F. D. F., & Ferreira, L. (2018). The rank reversal problem in multi-criteria decision making: A literature review. *Pesquisa Operacional*, *38*, 331–362. https://doi.org/10.1590/0101-7438.2018.038.02.0331

Aires, R. F. D. F., & Ferreira, L. (2019). A new approach to avoid rank reversal cases in the TOPSIS method. *Computers & Industrial Engineering*, *132*, 84–97. https://doi.org//10.1016/j.cie.2019.04.023

Alvarez, G. B., de Almeida, R. G., Hernandez, C. T., & de Sousa, P. A. P. (2021). Some mathematical comments about the analytic hierarchy process: Part I-theoretical analysis. *Journal of Multi-Criteria Decision Analysis*, *28*(5-6), 296–310. https://doi.org/10.1002/mcda.1762

Arrow, K. J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Belton, V., & Gear, T. (1983). On a short-coming of Saaty's method of analytic hierarchies. *Omega*, *11*(3), 228–230. https://doi.org/10.1016/0305-0483(83)90047-6

Bhatt, D., Naqvi, S., & Gunasekaran, A. (2023). Prescriptive analytics applications in sustainable operations research: Conceptual framework and future research challenges. *Annals of Operations Research*, *in press*. https://doi.org/10.1007/s10479-023-05251-3

Brans, J.-P., & Vincke, P. (1985). Note—a preference ranking organisation method. *Management Science*, *31*(6), 647–656. https://doi.org/10.1287/mnsc.31.6.647

Bryson, N. (1995). A goal programming method for generating priority vectors. *Journal of the Operational Research Society*, *46*(5), 641–648. https://doi.org/10.1057/jors.1995.88

Chen, S. J., & Hwang, C. L. (1963). *Fuzzy Multiple Attribute Decision Making: Methods and Applications*. Springer Verlag, Berlin.

Choquet, G. (1954). Theory of Capacities. *Annales de l'institut Fourier*, *5*, 131–295. https://doi.org/10.5802/aif.53

Cinelli, M., Coles, S. R., & Kirwan, K. (2014). Analysis of the potentials of multi-criteria decision analysis methods to conduct sustainability assessment. *Ecological Indicators*, *46*, 138–148. https://doi.org/10.1016/j.ecolind.2014.06.011

Cinelli, M., Kadziński, 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*, 102261. https://doi.org/10.1016/j.omega.2020.102261

Corbin, R., & Marley, A. (1974). Random utility models with equality: An apparent, but not actual, generalization of random utility models. *Journal of Mathematical Psychology*, *11*(3), 274–293. https://doi.org/10.1016/0022-2496(74)90023-6

Crawford, G., & Williams, C. (1985). A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, *29*, 387–405. https://doi.org/10.1016/0022-2496(85)90002-1

Csató, L. (2019). A characterization of the logarithmic least squares method. *European Journal of Operational Research*, *276*(1), 212–216. https://doi.org/10.1016/j.ejor.2018.12.046

Csató, L. (2024). Right-left asymmetry of the eigenvector method: A simulation study. *European Journal of Operational Research*, *313*, 708–717. https://doi.org/10.1016/j.ejor.2023.09.022

Csató, L., & Petróczy, D. G. (2021). On the monotonicity of the eigenvector method. *European Journal of Operational Research*, *292*, 230–237. https://doi.org/10.1016/j.ejor.2020.10.020

De Keyser, W., & Peeters, P. (1996). A note on the use of PROMETHEE multi-criteria methods. *European Journal of Operational Research*, *89*(3), 457–461. https://doi.org/10.1016/0377-2217(94)00307-6

Dede, G., Kamalakis, T., & Sphicopoulos, T. (2015). Convergence properties and practical estimation of the probability of rank reversal in pairwise comparisons for multi-criteria decision making problems. *European Journal of Operational Research*, *241*(2), 458–468. https://doi.org/10.1016/j.ejor.2014.08.037

Dede, G., Kamalakis, T., & Sphicopoulos, T. (2016). Theoretical estimation of the probability of weight rank reversal in pairwise comparisons. *European Journal of Operational Research*, *252*(2), 587–600. https://doi.org/10.1016/j.ejor.2016.01.059

Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The critic method. *Computers & Operations Research*, *22*(7), 763–770. https://doi.org/10.1016/0305-0548(94)00059-H

Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. Wiley-Interscience.

Faramondi, L., Oliva, G., Setola, R., & Bozóki, S. (2023). Robustness to rank reversal in pairwise comparison matrices based on uncertainty bounds. *European Journal of Operational Research*, *304*(2), 676–688. https://doi.org/10.1016/j.ejor.2022.04.010

Garcia-Cascales, M. S., & Lamata, M. T. (2012). On rank reversal and TOPSIS method. *Mathematical and Computer Modelling*, *56*(5-6), 123–132. https://doi.org/10.1016/j.mcm.2011.12.022

Gomes, L. F. A. M., & Lima, M. M. P. P. (1991). TODIM: Basics and application to multi-criteria ranking of projects with environmental impacts. *Foundations of Computing and Decision Sciences*, (16), 113–127.

Huntington, E. (1938). A paradox in the scoring of competing teams. *Science*, *88*(2282), 287–288. https://doi.org/10.1126/science.88.2282.287

Johnson, C. R., Beine, W. B., & Wang, T. J. (1979). Right-left asymmetry in an eigenvector ranking procedure. *Journal of Mathematical Psychology*, *19*, 61–64. https://doi.org/10.1016/0022-2496(79)90005-1

Kondratev, A. Y., Ianovski, E., & Nesterov, A. S. (2023). How should we score athletes and candidates: Geometric scoring rules. *Operations Research*, *in press*. https://doi.org/10.1287/opre.2023.2473

Leoneti, A. B., & Gomes, L. F. A. M. (2021). A novel version of the TODIM method based on the exponential model of prospect theory: The ExpTODIM method. *European Journal of Operational Research*, *295*(3), 1042–1055. https://doi.org/10.1016/j.ejor.2021.03.055

Leskinen, P., & Kangas, J. (2005). Rank reversals in multi-criteria decision analysis with statistical modelling of ratio-scale pairwise comparisons. *Journal of the Operational Research Society*, *56*(7), 855–861. https://doi.org/10.1057/palgrave.jors.2601925

Leung, L. C., & Cao, D. (2001). On the efficacy of modeling multi-attribute decision problems using AHP and Sinarchy. *European Journal of Operational Research*, *132*(1), 39–49. https://doi.org/10.1016/S0377-2217(00)00111-9

Liao, H., & Wu, X. (2020). DNMA: A double normalization-based multiple aggregation method for multi-expert multi-criteria decision making. *Omega*, *94*, 102058. https://doi.org/10.1016/j.omega.2019.04.001

Liu, X., & Ma, Y. (2021). A method to analyze the rank reversal problem in the ELECTRE II method. *Omega*, *102*, 102317. https://doi.org/10.1016/j.omega.2020.102317

Millet, I., & Saaty, T. L. (2000). On the relativity of relative measures—accommodating both rank preservation and rank reversals in the AHP. *European Journal of Operational Research*, *121*(1), 205–212. https://doi.org/10.1016/S0377-2217(99)00040-5

Mirhedayatian, S. M., & Farzipoor Saen, R. (2011). A new approach for weight derivation using data envelopment analysis in the analytic hierarchy process. *Journal of the Operational Research Society*, *62*(8), 1585–1595. https://doi.org/10.1057/jors.2010.105

Mousavi-Nasab, S. H., & Sotoudeh-Anvari, A. (2018). A new multi-criteria decision making approach for sustainable material selection problem: A critical study on rank reversal problem. *Journal of Cleaner Production*, *182*, 466–484. https://doi.org/10.1016/j.jclepro.2018.02.062

Ramanathan, R. (2006). Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process. *Computers & Operations Research*, *33*(5), 1289–1307. https://doi.org/10.1016/j.cor.2004.09.020

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49–57. https://doi.org/10.1016/j.omega.2014.11.009

Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega*, *64*, 126–130. https://doi.org/10.1016/j.omega.2015.12.001

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*(3), 234–281. https://doi.org/10.1016/0022-2496(77)90033-5

Saaty, T. L. (1987). Rank generation, preservation, and reversal in the analytic hierarchy decision process. *Decision Sciences*, *18*(2), 157–177. https://doi.org/10.1111/j.1540-5915.1987.tb01514.x

Saaty, T. L. (2006). Rank from comparisons and from ratings in the analytic hierarchy/network processes. *European Journal of Operational Research*, *168*(2), 557–570. https://doi.org/10.1016/j.ejor.2004.04.032

Schenkerman, S. (1994). Avoiding rank reversal in AHP decision-support models. *European Journal of Operational Research*, *74*(3), 407–419. https://doi.org/10.1016/0377-2217(94)90220-8

Schoner, B., Wedley, W. C., & Choo, E. U. (1992). A rejoinder to Forman on AHP, with emphasis on the requirements of composite ratio scales. *Decision Sciences*, *23*(2), 509–517. https://doi.org/10.1111/j.1540-5915.1992.tb00403.x

Schotanus, F., van den Engh, G., Nijenhuis, Y., & Telgen, J. (2022). Supplier selection with rank reversal in public tenders. *Journal of Purchasing and Supply Management*, *28*(2), 100744. https://doi.org/10.1016/j.pursup.2021.100744

Soltanifar, M., & Shahghobadi, S. (2014). Survey on rank preservation and rank reversal in data envelopment analysis. *Knowledge-Based Systems*, *60*, 10–19. https://doi.org/10.1016/j.knosys.2013.12.021

Stam, A., & Silva, A. P. D. (1997). Stochastic judgments in the AHP: The measurement of rank reversal probabilities. *Decision Sciences*, *28*(3), 655–688. https://doi.org/10.1111/j.1540-5915.1997.tb01326.x

Szádoczki, Z., Bozóki, S., Juhász, P., Kadenko, S. V., & Tsyganok, V. (2023). Incomplete pairwise comparison matrices based on graphs with average degree approximately 3. *Annals of Operations Research*, *326*, 783–807. https://doi.org/10.1007/s10479-022-04819-9

Tomashevskii, I., & Tomashevskii, D. (2021). A non-heuristic multicriteria decision-making method with verifiable accuracy and reliability. *Journal of the Operational Research Society*, *72*(1), 78–92. https://doi.org/10.1080/01605682.2019.1650621

Tomashevskii, I. (2015). Eigenvector ranking method as a measuring tool: Formulas for errors. *European Journal of Operational Research*, *240*(3), 774–780. https://doi.org/10.1016/j.ejor.2014.07.050

Tomashevskii, I. (2018). Optimization methods to estimate alternatives in AHP: The classification with respect to the dependence of irrelevant alternatives. *Journal of the Operational Research Society*, *69*(7), 1114–1124. https://doi.org/10.1080/01605682.2017.1390533

Tu, J. C., & Wu, Z. B. (2023). Analytic hierarchy process rank reversals: Causes and solutions. *Annals of Operations Research*, *in press*. https://doi.org/10.1007/s10479-023-05278-6

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*(10), 1179–1189. https://doi.org/10.1287/mnsc.39.10.1179

Vargas, L. G. (1994). Reply to schenkerman's avoiding rank reversal in AHP decision support models. *European Journal of Operational Research*, *74*(3), 420–425. https://doi.org/10.1016/0377-2217(94)90221-6

Wang, X., & Triantaphyllou, E. (2008). Ranking irregularities when evaluating alternatives by using some ELECTRE methods. *Omega*, *36*(1), 45–63. https://doi.org/10.1016/j.omega.2005.12.003

Wang, Y. M., & Elhag, T. M. (2006). An approach to avoiding rank reversal in AHP. *Decision Support Systems*, *42*(3), 1474–1480. https://doi.org/10.1016/j.dss.2005.12.002

Wang, Y. M., Parkan, C., & Luo, Y. (2008). A linear programming method for generating the most favorable weights from a pairwise comparison matrix. *Computers & Operations Research*, *35*(12), 3918–3930. https://doi.org/10.1016/j.cor.2007.05.002

Wu, X., & Liao, H. (2019). A consensus-based probabilistic linguistic gained and lost dominance score method. *European Journal of Operational Research*, *272*(3), 1017–1027. https://doi.org/10.1016/j.ejor.2018.07.044

Zanakis, S. H., Solomon, A., Wishart, N., & Dublish, S. (1998). Multi-attribute decision making: A simulation comparison of select methods. *European Journal of Operational Research*, *107*(3), 507–529. https://doi.org/10.1016/S0377-2217(97)00147-1

Appendices

Appendix A

We recall the method to obtain the criterion weights from evaluations proposed by Diakoulaki et al. (1995). Let C_j, $j = 1, 2, \ldots, m$, and A_i, $i = 1, 2, \ldots, n$, represent m criteria and n alternatives.

Let w_j be the weight of criterion C_j. The evaluation of alternative A_i on criterion C_j is denoted by e_{ij}.

Step 1. Calculate the performance score of A_i on C_j as

$$s_{ij} = (e_{ij} - \min_{l=1,2,\ldots,n} e_{lj})/(\max_{l=1,2,\ldots,n} e_{lj} - \min_{l=1,2,\ldots,n} e_{lj})\,.$$

Step 2. Calculate the standard deviation σ_j of all performance scores on criterion C_j. Diakoulaki et al. (1995) consider that when σ_j is large, C_j contains more information. Hence, the larger the value of σ_j is, the more important C_j is.

Step 3. Calculate the linear correlation coefficient of the performance scores of criteria C_{j_1} and C_{j_2} as $r_{j_1}^{j_2}$. If $r_{j_1}^{j_2}$ is close to 1, then the performance scores on these criteria have a strong linear correlation, which means that if alternative A_1 has a better performance than alternative A_2 on C_{j_1}, then A_1 is highly likely to perform also better than A_2 on C_{j_2} (Diakoulaki et al., 1995). In this situation, we just need one of C_{j_1} and C_{j_2}. Conversely, if C_j has low linear correlation coefficients with other criteria, then C_j is important because the performance scores on C_j are different from those on other criteria. Therefore, the larger the value of $\sum_{l_1=1,2,\ldots,n} 1 - r_j^{l_1}$ is, the more important C_j is.

Step 4. Calculate the weight of C_j. Referring to the above discussion, if both $\sum_{l_1=1,2,\ldots,n} 1 - r_j^{l_1}$ and σ_j are large, then C_j is important. Diakoulaki et al. (1995) therefore considered their product. After normalization, the weight of C_j is obtained as

$$w_j = \frac{\sigma_j \times \sum_{l_1=1,2,\ldots,m} (1 - r_j^{l_1})}{\sum_{l_2=1,2,\ldots,m} \sigma_{l_2} \times \sum_{l_1=1,2,\ldots,m} (1 - r_{l_2}^{l_1})}\,.$$

Appendix B

We recall the method to obtain consistent and inconsistent data in CIS.

Let A_i, $i = 1, 2, \ldots, n$, represent n alternatives. The ratio scale is denoted by S. The pairwise comparison of A_{i_1} to A_{i_2} is denoted by $p_{i_1 i_2}$. To obtain inconsistent pairwise comparison data, we randomly pick up a number from S for every $p_{i_1 i_2}$, $i_1, i_2 = 1, 2, \ldots, n$. It is worth noting that this method can generate pairwise comparison data with high inconsistency levels. To generate inconsistent data with moderate inconsistency levels, we refer to Szádoczki et al. (2023) and Csató (2024).

To obtain consistent pairwise comparison data, we first randomly select one alternative, *e.g.*, A_1. Then, we randomly pick up a number from S for every p_{1i_2}, $i_2 = 1, 2, \ldots, n$. The other pairwise comparisons can be calculated from p_{1i_2} (referring to Saaty, 1977). For example, $p_{34} = p_{31} \times p_{14} = 1/p_{13} \times p_{14}$.

BWM is particular. To obtain consistent pairwise comparison data, we randomly select an alternative from n alternatives to represent the best alternative A_B. Then, we randomly pick a number larger than 1 from S for every p_{Bi}, $i = 1, 2, \ldots, n$. Other pairwise comparisons can be calculated from p_{Bi}. We find the worst alternative A_W whose p_{Wi}, $i = 1, 2, \ldots, n$, are all not larger than 1. When there is more than one such alternative, we randomly select one. To obtain inconsistent pairwise comparison data, we randomly select two alternatives from n alternatives to represent the best and worst alternatives. We then randomly pick a number not smaller than 1 (resp. a number not larger than 1) from S for every p_{Bi} (resp. p_{Wi}), $i = 1, 2, \ldots, n$. When we delete the first or last alternative in the simulation, we delete $n+1$ pairwise comparisons, *i.e.*, p_{Bi}, $i = 1, 2, \ldots, n$, and p_{WB}. When we delete the other alternatives, *e.g.*, A_2, which are not the best or the worst alternative, we delete two pairwise comparisons, *e.g.*, p_{B2} and p_{W2}. Here, it is worth noting that when we delete the best (or the worst) alternative in BWM, the remaining pairwise comparisons will be consistent.

It should be noted that p_{ii} is always 1.