

# Consistency issues in the best worst method: Measurements and thresholds<sup>☆</sup>

Fuqi Liang<sup>a,\*</sup>, Matteo Brunelli<sup>b</sup>, Jafar Rezaei<sup>a</sup>

<sup>a</sup>Faculty of Technology, Policy and Management, Delft University of Technology, Delft, the Netherlands

<sup>b</sup>Department of Industrial Engineering, University of Trento, Trento, Italy

## ARTICLE INFO

### Article history:

Received 1 July 2019

Accepted 19 December 2019

Available online 20 December 2019

### Keywords:

Best worst method

Consistency threshold

Consistency measurement

## ABSTRACT

The Best-Worst Method (BWM) uses ratios of the relative importance of criteria in pairs based on the assessment done by decision-makers. When a decision-maker provides the pairwise comparisons in BWM, checking the acceptable inconsistency, to ensure the rationality of the assessments, is an important step. Although both the original and the extended versions of BWM have proposed several consistency measurements, there are some deficiencies, including: (i) the lack of a mechanism to provide immediate feedback to the decision-maker regarding the consistency of the pairwise comparisons being provided, (ii) the inability to consider the ordinal consistency into account, and (iii) the lack of consistency thresholds to determine the reliability of the results. To deal with these problems, this study starts by proposing a cardinal consistency measurement to provide immediate feedback, called the input-based consistency measurement, after which an ordinal consistency measurement is proposed to check the coherence of the order of the results (weights) against the order of the pairwise comparisons provided by the decision-maker. Finally, a method is proposed to balance cardinal consistency ratio under ordinal-consistent and ordinal-inconsistent conditions, to determine the thresholds for the proposed and the original consistency ratios.

© 2019 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

The Best Worst Method (BWM), which is a Multi-Criteria Decision Making (MCDM) method that was recently developed by Rezaei [37], uses ratios of the relative importance of criteria in pairwise comparisons provided by a decision-maker (DM), based on two evaluation vectors: the Best criterion against the Other criteria, and the Other criteria against the Worst criterion. The weights of the criteria are obtained by solving a nonlinear [37] or a linear model [38]. Compared to one of the most popular pairwise comparison-based MCDM methods, Analytic Hierarchy Process (AHP), BWM requires fewer comparison data, while being able to generate more consistent comparisons, allowing it to produce more reliable results according to previous analyses [37]. Thanks to its simplicity and reliability, BWM has been widely applied to address a host of different problems [29,39,49]. For more detailed

information, readers are referred to a recent survey on the BWM [32].

BWM and other pairwise comparisons methods, like AHP and ANP (Analytical Network Process), are based on a DM's evaluations of the relative priorities of the decision-making elements as captured in a complete pairwise comparison matrix [41], incomplete pairwise comparison matrix [19] or vectors [37]. One of the advantages of using pairwise comparisons is that they allow us to estimate the inconsistency of a DM's preferences. Usually, the consistency level of the judgements is related to the rationality of the DM and his/her ability to discriminate between criteria/alternatives [21]. The DM's judgments have to meet the cardinal transitivity condition to be perfectly consistent; otherwise, the DM is not fully consistent, which may imply some irrationality in the relative weight estimates.

To check how inconsistent (deviating from the condition of full consistency) a full set of pairwise comparisons may be, Saaty [40], in their seminal work on the AHP, proposed a consistency measurement (Saaty index), but since then, many other consistency indices have been proposed [10]. Basically, the existing consistency measurements can be divided into two groups: the input-based measurements and output-based measurements [28]. The

<sup>☆</sup> This paper was processed by Associate Editor Dias.

\* Corresponding author.

E-mail addresses: [f.liang-2@tudelft.nl](mailto:f.liang-2@tudelft.nl) (F. Liang), [matteo.brunelli@unitn.it](mailto:matteo.brunelli@unitn.it) (M. Brunelli), [J.Rezaei@tudelft.nl](mailto:J.Rezaei@tudelft.nl) (J. Rezaei).

measurements in the former group are based on the input, i.e. preferences assigned to pairwise comparisons, e.g. Koczkodaj index [24], while the output-based consistency measurements are based on the weights or rankings. In this group, there are, for instance, Saaty's index [40] and the geometric consistency index proposed by Crawford and Williams [13].

The consistency measurements mentioned above were initially designed for complete pairwise comparison matrices and we cannot use them to measure the consistency degree of incomplete pairwise comparison matrices where some judgments are missing [19]. To adapt the consistency indices to incomplete pairwise comparison matrices, one of the most popular approaches is to complete the pairwise comparison matrices [16,47] and then measure their consistency in the traditional manner [19,28]. Instead of completing the matrix, a graph-theoretic approach can be used to generate all possible preferences by enumerating all spanning trees, after which the variance of these preferences can be used as a measure of inconsistency [6,31,44]. Replacing triads with cycles [28] is another way to estimate the inconsistency.

One might see BWM as a special case of incomplete pairwise comparison matrix. Although the method only uses a specific subset of  $2n-3$  comparisons gathered in two representative vectors, these preferences can be represented equivalently by an incomplete pairwise comparison matrix. One could argue that we could then complete the two vectors to create a full matrix and measure the inconsistency by using the approaches mentioned above. However, not only will that make the measurement more difficult (unrealistic), it will also destroy the simplification (non-redundancy) philosophy embedded in BWM. Therefore, to check the consistency by using this specific method, Rezaei [37] proposed a consistency measurement (sometimes referred to as inconsistency measurement) in the original version of BWM. Later, the extended BWM methods also provided corresponding consistency measurements similar to the original consistency measurement. For example, Mou et al. [35] extended BWM to include intuitionistic fuzzy multiplicative preference relations, and provided a new definition for the consistency algorithm to check consistency, while Guo and Zhao [18] proposed a consistency ratio (also referred to as inconsistency ratio) for fuzzy BWM, and Aboutorab et al. [1] explained a corresponding consistency ratio for the Z-numbers BWM.

However, the existing studies on BWM lack a metric/tool to provide the DM/analyst with immediate feedback regarding the consistency of the pairwise comparisons. The consistency ratios obtained by the existing consistency measurements of BWM are based on the outputs instead of directly on the inputs. A DM can only obtain the consistency ratio and check the consistency after the entire optimization process is completed, by using the existing consistency measurements. However, it has been shown that confronting the DM with the inconsistencies in his/her assessments after he/she has already gone through the entire elicitation process is ineffective [34]. In addition, the consistency ratios obtained by the original BWM, graph-theoretic approach [6,31,44] and the methods of replacing triads with cycles [28] are overall indicators that show the consistency of the pairwise comparison system as a whole, so they cannot help the DM locate their most inconsistent judgments. A proper consistency measurement should indeed assist the DM in identifying the most inconsistent comparisons [14] and achieve sufficiently consistent preferences [17,36]. Although some input-based consistency measurements for general incomplete pairwise comparison matrices, including the Koczkodaj index [24] and the Salo and Hämäläinen index [43], can be applied to BWM, some of their properties are not as desirable as we expected, as discussed in Section 3.

Moreover, the existing studies on consistency measurement in the BWM thus far fail to take ordinal consistency into consideration. Consistency in pairwise comparisons can be divided into two

categories: cardinal consistency and ordinal consistency [45]. The existing consistency ratios of BWM only measure cardinal consistency. However, even if the judgements have a high level of cardinal consistency, they can be still contradictory, according to the research of Kwiesielewicz and Van Uden [30]. The contradiction is caused by the violation of ordinal consistency, i.e. there is a discrepancy in the criteria importance rankings obtained from the two pairwise comparison vectors in BWM. If the preferences are ordinal-consistent, the final ranking will not change with the cardinal consistency ratio, only the intensity could vary; but if they are ordinal-inconsistent, a change in the cardinal consistency ratio could affect the final ranking [45]. Thus, in order to ensure a DM provides a stable judgement, it is important to check his/her ordinal consistency status, and indicate to what extent the ordinal consistency has been violated. There are several ordinal consistency measurements for the complete pairwise comparison matrices, like the ordinal coefficient proposed by Jensen and Hicks [22], the dissonance measurement proposed by Siraj et al. [45,46]. However, they cannot be applied to incomplete pairwise comparison matrices or the two vectors used in BWM.

Furthermore, there is no threshold for the consistency ratio of BWM in existing literature. Although BWM has been widely used and the consistency measurements help a DM check the reliability of his/her preferences, the absence of threshold associated with the existing consistency measurements makes it hard to provide a meaningful interpretation. Without a consistency threshold, the DM/analyst is left with the major problem of having to decide when his/her judgments should be revised and when it should be accepted, not to mention the consideration of the number of criteria and the scale of evaluation, making the situation even more complicated. The 10% rule of thumb of AHP has long been criticised [5,7,33], and even Saaty later suggested additional threshold values of 5% and 8% for 3 and 4 criteria, respectively [42]. Although some other methods have been proposed to determine consistency thresholds [2,4,33], most of them are applied in complete pairwise comparison matrices, which cannot be used directly for incomplete pairwise comparison matrices. Thus, designing a threshold determination algorithm for BWM can fill this gap.

As such, the contribution of this study is threefold: (i) Developing a mechanism designed to provide a DM with immediate feedback regarding his/her consistency status and making the elicitation process more effective. To this end, we propose an input-based consistency measurement, which is simple to use and has several desirable properties; (ii) Developing an ordinal consistency ratio that shows a DM's violation level involving ordinal consistency and complements the cardinal consistency measurement. With this ratio, a DM can revise his/her judgments to meet the ordinal consistency condition, which is a minimum requirement for a logical and rational DM; (iii) The most significant contribution of this study is to establish thresholds for the consistency ratios (the proposed consistency ratio and the original consistency ratio) used in BWM.

The remainder of the paper is structured as follows: In Section 2, the original BWM and its consistency measurement are introduced. An input-based consistency ratio is proposed as an alternative to replace the original output-based consistency ratio in Section 3. An ordinal consistency measurement is formulated in Section 4. The threshold tables are presented in Section 5, followed by the conclusion in Section 6.

## 2. The best worst method and consistency measurement

In this part, the basic steps of the original BWM are briefly introduced, and the original output-based consistency measurement is reviewed.

### 2.1. The basic steps of BWM

As a pairwise comparison method, BWM uses ratios of the relative importance of criteria in pairs estimated by a DM, from the two evaluation vectors,  $A_{BO}$  and  $A_{OW}$ . The weights of the criteria can be obtained by solving the linear or nonlinear program [38]. The basic steps of the original BWM can be summarized as below:

- Step 1 Have the set of evaluation criteria  $\{C_1, C_2, \dots, C_n\}$  determined by the DM.
- Step 2 Have the best (e.g. the most influential or important) and the worst (e.g. the least influential or important) criteria determined by the DM.
- Step 3 Determine the preferences of the best over all the other criteria using a number from  $\{1, 2, \dots, 9\}$ . The obtained Best-to-Others vector is:  $A_{BO} = (a_{B1}, a_{B2}, \dots, a_{Bn})$ , where  $a_{Bj}$  represents the preference of the best criterion  $C_B$  over criterion  $C_j$ ,  $j = 1, 2, \dots, n$ .
- Step 4 Determine the preferences of all the criteria over the worst criterion using a number from  $\{1, 2, \dots, 9\}$ . The obtained Others-to-Worst vector is:  $A_{OW} = (a_{1W}, a_{2W}, \dots, a_{nW})$ , where  $a_{jW}$  represents the preference of criterion  $C_j$  over the worst criterion  $C_W$ ,  $j = 1, 2, \dots, n$ .
- Step 5 Determine the weights  $(w_1^*, w_2^*, \dots, w_n^*)$  by solving the following model:

$$\begin{aligned} \min \max_j & \left\{ \left| \frac{w_B}{w_j} - a_{Bj} \right|, \left| \frac{w_j}{w_W} - a_{jW} \right| \right\}, \\ \text{s.t.} & \\ & \sum_{j=1}^n w_j = 1, w_j \geq 0, \text{ for all } j. \end{aligned} \quad (1)$$

Model (1) can be transformed into the following model:

$$\begin{aligned} \min & \xi \\ \text{s.t.} & \\ & \left| \frac{w_B}{w_j} - a_{Bj} \right| \leq \xi, \text{ for all } j, \\ & \left| \frac{w_j}{w_W} - a_{jW} \right| \leq \xi, \text{ for all } j, \\ & \sum_{j=1}^n w_j = 1, w_j \geq 0, \text{ for all } j. \end{aligned} \quad (2)$$

### 2.2. The original consistency measurement

In the remainder of this paper, when we talk about a *pairwise comparison system*, we will refer to the set of judgments contained in vectors  $A_{BO}$  and  $A_{OW}$ . Given this notion, we are able to provide the definition of cardinal consistency for the set of preferences contained in a pairwise comparison system.

**Definition 1** (Cardinal consistency). A pairwise comparison system is *cardinal-consistent* if

$$a_{Bj} \times a_{jW} = a_{BW}, \text{ for all } j, \quad (3)$$

where  $a_{BW}$  is the preference of the best criterion over the worst criterion.

However, it is common practice to allow a pairwise comparison system to deviate, to some extent, from the condition of cardinal-consistency. Thus, a consistency ratio is necessary to indicate how inconsistent a DM is. The consistency measurement proposed in the original BWM is based on  $\xi^*$ , which is the optimal objective value (the output) of the optimization model (2), so we call it

an *output-based consistency measurement* (we will use an output-based consistency measurement instead of using the original consistency measurement in the remainder of the paper). The ratio used to indicate the consistency level is called *Output-based Consistency Ratio*, noted as  $CR^O$  (we will use output-based consistency ratio or  $CR^O$  to represent original consistency ratio from now on), was defined as follows [37]:

**Definition 2** (Output-based Consistency Ratio). The *Output-based Consistency Ratio*  $CR^O$  is defined as

$$CR^O = \frac{\xi^*}{\xi_{\max}} \quad (4)$$

where  $\xi^*$  is the optimal objective value of model (2) and  $\xi_{\max}$  is the maximum possible  $\xi$ , which can be derived from [37]:

$$\xi^2 - (1 + 2a_{BW})\xi + (a_{BW}^2 - a_{BW}) = 0. \quad (5)$$

The range of  $CR^O$  is  $[0, 1]$ . The closer  $CR^O$  is to 0, the more consistent the judgments are. In particular,  $CR^O = 0$  means that the comparisons are cardinally consistent.

### 3. The proposed consistency measurement

The consistency ratio proposed in the original BWM can only be obtained after the entire elicitation process has finished, which means it cannot provide a DM with immediate feedback involving his/her consistency. To overcome this problem and to provide a DM with a clear and immediate idea of his/her consistency level, we propose an input-based consistency measurement for BWM that is easy to compute and has clear and simple algebraic meaning and interpretation. Furthermore, we will see that it has several desirable properties (in comparison to the existing indices) and a high correlation with the output-based consistency measurement.

In accordance with the original index, the new inconsistency index proposed in the following section only attains value 1 when, given  $a_{BW}$ , there exists a  $C_j$  such that  $a_{Bj} = a_{jW} = a_{BW}$ . This is possible because the index considers the maximum violation of local inconsistencies and the value 1 can actually be attained. None of the indices studied by Kułakowski and Talaga [28] has this property. Besides this similarity, we will also show the resemblance between the old and the new index using some numerical analyses.

#### 3.1. The input-based consistency ratio

In contrast to the Output-based Consistency Ratio ( $CR^O$ ), the ratio we propose in this paper can immediately indicate a DM's consistency level by using the input he/she provides, i.e. his/her preferences, instead of going through the entire optimization process, which is why it is called an *Input-based Consistency Ratio* ( $CR^I$ ):

**Definition 3** (Input-based Consistency Ratio). The *Input-based Consistency Ratio*  $CR^I$  is formulated as follows:

$$CR^I = \max_j CR_j^I \quad (6)$$

where

$$CR_j^I = \begin{cases} \frac{|a_{Bj} \times a_{jW} - a_{BW}|}{a_{BW} \times a_{BW} - a_{BW}} & a_{BW} > 1, \\ 0 & a_{BW} = 1 \end{cases} \quad (7)$$

$CR^I$  is the *global* input-based consistency ratio for all criteria,  $CR_j^I$  represents the *local* consistency level associated with criterion  $C_j$ .

Compared to the output-based consistency measurement, the input-based consistency measurement has several advantages:

**Table 1**  
Input-based consistency ratio of each criterion.

	Price	Quality	Comfort	Safety	Style
$a_{Bj}$	1	2	4	3	8
$a_{jW}$	8	4	4	2	1
$CR_j^I$	0	0	0.14	0.04	0

1. It can provide immediate feedback. The input-based consistency measurement is based on the input (preferences), which means it is not necessary to complete the entire elicitation process. The output-based consistency measurement on the other hand, is based on the output (weights), making it a difficult way to determine the consistency level. By using the simple calculation of the input-based consistency measurement, it is easy to provide a DM with immediate feedback.
2. It is easy to interpret: it is the maximum normalized discrepancy between the value of  $a_{BW}$  and its estimated value calculated as the indirect comparison  $a_{Bj} \times a_{jW}$ .
3. It can provide a DM with a clear guideline on the revision of the inconsistent judgement(s). The  $CR^0$  indicates the global consistency level, but it cannot show the DM which judgement should be revised. The local  $CR^I$ , however, displays the consistency levels associated to individual criteria; after identifying the maximum local  $CR^I$ , the most inconsistent judgement can be located, after which a DM can revise his/her judgements accordingly, instead of modifying them without a guideline.
4. It is model-independent. This  $CR^I$  can be applied independently to measure the consistency level in various form of BWM models, e.g. a non-linear or linear model, or a multiplicative model [11]. For example, the linear BWM model [38] does not have an effective consistency measurement, while the non-linear BWM model [37] has a different interpretation than the multiplicative BWM model [11]. By using the input-based consistency ratio, however, they are the same in all three models. Actually, the input-based consistency measurement does not depend on the optimization models.

**Example 1.** To illustrate the proposed consistency measurement, we adopt the car evaluation example from the original BWM [38], in which the best criterion is price and the worst criterion style. The pairwise comparisons vectors of  $A_{BO}$  and  $A_{OW}$  are presented in the second and third rows respectively. By using the input-based consistency measurement in Eq. (7), the  $CR_j^I$ s are represented in the last row of Table 1.

From Table 1, by using the maximum measurement (6), we can obtain the global  $CR^I$ , 0.14. One of the advantages of the input-based consistency measurement is that we can immediately locate the most inconsistent pairwise comparison from this table, which in this case is the preferences regarding the criterion *comfort*. If the  $CR^I$  is too high, the DM's preferences have to be modified.

### 3.2. Properties of the input-based consistency measurement

As indicated by Brunelli [8], it is important that formal properties of inconsistency indices be investigated to check their technical soundness and rule out possible unreasonable behaviours. The next proposition will show that  $CR^I$  satisfies a number of reasonable properties.

**Proposition 1.** *The proposed consistency measurement,  $CR^I = \max_j CR_j^I$  satisfies the following properties:*

1.  $CR^I = 0$  if and only if the preferences are cardinal-consistent.
2.  $CR^I$  is invariant with respect to a permutation of the indices of the criteria.

3.  $CR^I$  is normalized, i.e.  $0 \leq CR^I \leq 1$ .
4. If we consider a fully consistent pairwise comparison system, moving one of the preferences  $a_{Bj}$  or  $a_{jW}$  away from their original value in the range  $[1, a_{BW}]$  will result in an increase of the value of  $CR^I$ .
5. When  $a_{BW} > 1$ ,  $CR^I$  is a continuous function with respect to the values of  $a_{Bj}$ ,  $a_{jW}$ ,  $a_{BW}$  for all  $j$ .
6. If we remove a criterion which is neither the best nor the worst from the decision problem, then the value of  $CR^I$  cannot increase.

**Proof.** It is useful to consider the ordered set

$$S = \langle CR_j^I | j = 1, \dots, n \rangle = \left\langle \frac{|a_{Bj}a_{jW} - a_{BW}|}{a_{BW}a_{BW} - a_{BW}} | j = 1, \dots, n \right\rangle$$

so that we can consider  $CR^I$  to be a function of  $S$ , i.e.  $CR^I(S)$ , and, ultimately, of the preferences of the decision-maker.

1. If the preferences are consistent, then  $a_{Bj}a_{jW} = a_{BW}$ , for all  $j$ , from which we obtain  $S = \langle 0, \dots, 0 \rangle$  and  $CR^I = 0$ . In the other direction  $CR^I = 0$  only if  $S = \langle 0, \dots, 0 \rangle$ . If  $a_{BW} = 1$ ,  $CR_j^I = 0$ ,  $CR^I = 0$ , and  $a_{Bj} = a_{jW} = 1$ ,  $a_{Bj}a_{jW} = 1 \times 1 = 1 = a_{BW}$ , it is fully cardinal-consistent; If  $a_{BW} \neq 1$ , then the only case leading to  $S = \langle 0, \dots, 0 \rangle$  is when the numerators of the elements of  $S$  are all equal to zero, which is possible only if  $a_{Bj}a_{jW} = a_{BW}$ , for all  $j$ , which is the consistency condition.
2. A reordering of the criteria corresponds to an application of a permutation map  $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$  to the indices  $j$ . The new set

$$S_\sigma = \langle CR_{\sigma(j)}^I | j = 1, \dots, n \rangle$$

has the same elements of  $S$ , but in a different order. However, since the max function is symmetric,  $\max(S) = \max(S_\sigma)$ , for all permutations of the indices.

3. The normalization,  $CR^I \in [0, 1]$ , follows from the definition  $CR_j^I$  together with the facts that (1)  $|a_{Bj}a_{jW} - a_{BW}| \geq 0$ , (2)  $a_{BW} \geq 1$ , (3) when  $a_{BW} = 1$ ,  $CR^I = 0$ ; when  $a_{BW} > 1$ ,  $a_{BW}a_{BW} - a_{BW} > 0$  and (4)  $|a_{Bj}a_{jW} - a_{BW}| \leq a_{BW}a_{BW} - a_{BW}$ , because when the left-hand side  $a_{Bj}a_{jW} \geq a_{BW}$ ,  $a_{Bj}a_{jW} \leq a_{BW}a_{BW}$ ,  $a_{Bj}a_{jW} - a_{BW} \leq a_{BW}a_{BW} - a_{BW}$ , so the inequality holds; when the left-hand side  $a_{Bj}a_{jW} < a_{BW}$ , then  $a_{BW}$  should be larger than 2, and the right-hand side  $a_{BW}a_{BW} - a_{BW} \geq a_{BW}$ , therefore  $a_{BW} - a_{Bj}a_{jW} \leq a_{BW}a_{BW} - a_{BW}$ , the inequality holds also.
4. For each  $j \neq B, W$ , we want to study the reaction of  $CR^I(S)$  to changes in a single comparison in the range  $[1, a_{BW}]$ . In this case  $1 \leq a_{jW}$ ,  $a_{Bj} \leq a_{BW}$ , and we can consider  $a_{BW}$  a constant. Let us consider the effect of a variation of  $a_{Bj}$  in  $CR^I$  by taking its partial derivative

$$\frac{\partial CR^I}{\partial a_{Bj}} = \frac{a_{jW}}{(a_{BW}a_{BW} - a_{BW})} \frac{(a_{Bj}a_{jW} - a_{BW})}{|a_{Bj}a_{jW} - a_{BW}|}.$$

We can see that

$$\frac{\partial CR^I}{\partial a_{Bj}} \begin{cases} < 0, & |a_{Bj}a_{jW} - a_{BW}| < a_{BW} \\ > 0, & |a_{Bj}a_{jW} - a_{BW}| > a_{BW} \end{cases},$$

which shows that  $CR^I(a_{Bj})$  is a U-shaped function in  $[1, a_{BW}]$ , with minimum in the consistent case ( $a_{Bj}a_{jW} = a_{BW}$ ). The same conclusion follows if we consider  $a_{jW}$  instead of  $a_{Bj}$ .

5. Straightforward.  $CR^I$  is a continuous function for all  $a_{BW} > 1$ .
6. If we assume that the criterion which is eliminated, say  $C_i$ , is neither the best nor the worst, then  $a_{BW}$  remains unchanged and we can define a new set  $S_{-i}$  which disregards  $C_i$ :

$$S_{-i} = \langle CR_j^I | j \in \{1, \dots, n\} \setminus \{i\} \rangle.$$

Now, since  $S_{-i} \subset S$  we know that  $\max_j(S) \geq \max_j(S_{-i})$ .  $\square$



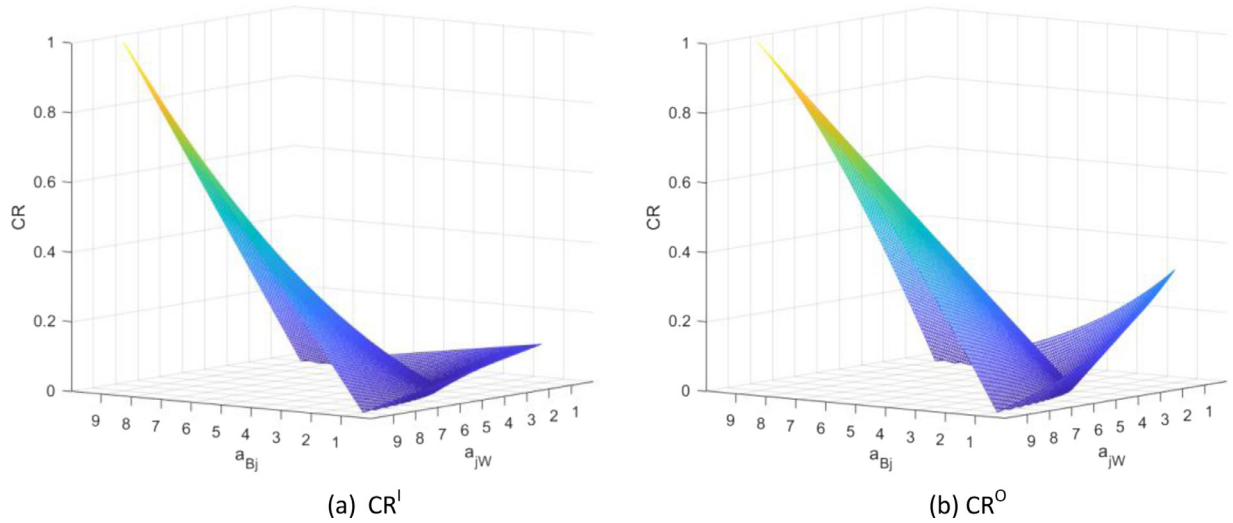


Fig. 1. The relationship between  $a_{Bj}$ ,  $a_{jw}$  and CR in the input-based (a) and the output-based (b) consistency measurements when the maximum scale is 9.

Note that these properties are adaptations of well-known properties already proposed and justified in the framework of pairwise comparison matrices. In particular, Properties 1, 2, 4 and 5 stem from those proposed by Brunelli and Fedrizzi [9], Property 3 from the normalization proposed by Koczkodaj et al. [26], and Property 6 from the contraction property proposed by Koczkodaj and Urban [25].

It is worth mentioning that Property 6 would not be satisfied by an approach based on the average of the local inconsistencies like Salo and Hämäläinen index [43].

### 3.3. Relationship between the input-based and output-based consistency ratio

In the input-based consistency measurement, when the number of criteria larger than 2, for two pairwise comparisons,  $a_{Bj}$  and  $a_{jw} \in \{1, 2, \dots, 9\}$ , the relationship between them and their corresponding  $CR^I$ 's is shown in Fig. 1(a). Likewise, we can calculate the relationship between  $a_{Bj}$ ,  $a_{jw} \in \{1, 2, \dots, 9\}$  and their  $CR^O$ 's for the output-based consistency measurement in BWM, which is shown in Fig. 1(b).

It is clear that these two relationship figures have similar shapes, which indicates they should have a high correlation.

To determine the agreement between these two indices, we analyse them from a statistical perspective by numerical simulations. Firstly, we randomly generated a set of 20,000 pairs of pairwise comparison vectors ( $A_{BO}$  and  $A_{OW}$ ) in a 9 criteria problem with 1–9 scales to represent the preferences provided by DMs in BWM. Then we computed the input-based consistency ratios and the output-based consistency ratios ( $CR^I$ ,  $CR^O$ ) for each pair of vectors in this 20,000 random pairs set. Each pair ( $CR^I$ ,  $CR^O$ ) is represented by a point in the scatter plot in Fig. 2.

As  $a_{Bj}$ ,  $a_{jw} \in \{1, 2, \dots, 9\}$  take values from a discrete scale, the possible  $CR^O$ 's and  $CR^I$ 's are limited. Thus, although we have obtained 20,000  $CR^O$ 's and  $CR^I$ 's, they distribute only in these limited possibilities, which is why there are much fewer than 20,000 dots in this scatter plot.

We compute the Pearson's correlation coefficient between  $CR^O$ 's and  $CR^I$ 's to check the linear correlation between them. The result of Pearson's correlation coefficient in this case is 0.9942, which means these  $CR^O$ 's and  $CR^I$ 's have a very high linear correlation. We also consider the Spearman index to measure the extent to which  $CR^O$ 's and  $CR^I$ 's are co-monotone. The result of the Spearman index

is 0.9963, which means these two variables are highly monotonically related.

When we calculate all the Pearson's and Spearman's correlation coefficients with respect to 3–9 criteria under maximal scale from 3 to 9, the minimum Pearson's and the minimum Spearman's correlation coefficients are 0.979 and 0.958, respectively. As such, based on these high correlation coefficients, the input-based consistency measurement and the output-based consistency measurement have a very good agreement, so they could be used interchangeably. Nevertheless, due to its advantages discussed in Section 3.1, there are valid reasons to prefer the input-based consistency measurement to the output-based consistency measurement.

## 4. Ordinal consistency measurement

In this section, an ordinal consistency ratio is proposed to determine the extent to which a DM violates the ordinal consistency. Some properties for this ratio are presented and the relationship between ordinal consistency and cardinal consistency is analysed.

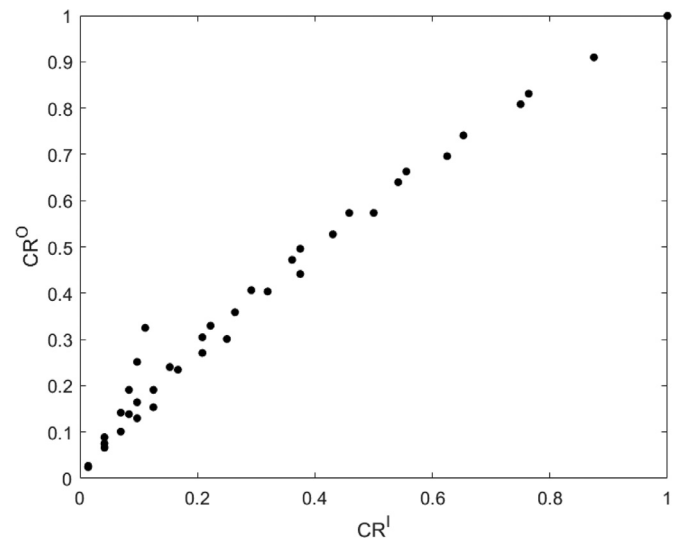


Fig. 2. Relation between  $CR^O$  and  $CR^I$  (9 criteria 9-scale).

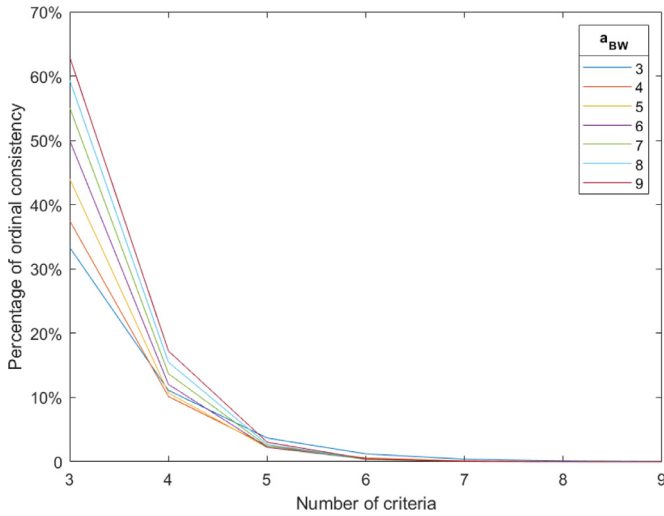


Fig. 3. The percentage of ordinal-consistent paired vectors.

#### 4.1. Ordinal consistency

Kwiesielewicz and Van Uden [30] have shown that, even if a pairwise comparison matrix passes the consistency test, it can still be contradictory. Therefore, in addition to calculating the cardinal consistency, it is also important to check whether the rankings of the criteria obtained from the two pairwise comparison vectors  $A_{BO}$  and  $A_{OW}$  are the same in BWM, in what we call *ordinal consistency condition*. The meaning of ordinal consistency in BWM is slightly different from that in early studies, which is mainly based on the circular triads [20,23,27]. We define the ordinal consistency in BWM as below:

**Definition 4** (Ordinal consistency). In the BWM, a pairwise comparison system is said to be *ordinal-consistent* if the order relations of the two paired comparison vectors ( $A_{BO}$  and  $A_{OW}$ ) are the same. That is, the following conditions should be satisfied:

$$(a_{Bi} - a_{Bj}) \times (a_{jW} - a_{iW}) > 0 \text{ or } (a_{Bi} = a_{Bj} = a_{iW}), \text{ for all } i \text{ and } j. \quad (8)$$

The ordinal consistency is the usual weak transitivity condition which should be the minimum requirement for a logical and rational DM [48]. Intuitively, one might consider ordinal consistency to be easily satisfied, but that is not true, especially when the number of criteria is large. To see how it develops, we randomly generated 100,000 paired vectors for each combination of criteria number from 3 to 9 to simulate the preferences for BWM. After categorizing, we can see the percentage of ordinal-consistent pairs is reduced dramatically as the number of criteria increases, as shown in Fig. 3. In reality, the situation is better than the randomly generated vectors, but after checking the data used in the original BWM, we found that only 24.4% of them are ordinal-consistent.

#### 4.2. Ordinal consistency ratio

Since the ordinal consistency has a vital impact on the ranking of the criteria, it is necessary to check whether the preferences violate the ordinal consistency, and, if so, to what extent. To do so, we need to define an index, which we call *Ordinal Consistency Ratio* (hereafter simply OR) in this study.

**Definition 5** (Ordinal Consistency Ratio). The *Ordinal Consistency Ratio* OR of a pairwise comparison system is defined as:

$$OR = \max_j OR_j \quad (9)$$

Table 2

Ordinal consistency ratio for each criterion.

	Price	Quality	Comfort	Safety	Style
$a_{Bj}$	1	2	4	3	8
$a_{jW}$	8	4	4	2	1
$OR_j$	0	0	0.3	0.2	0

where

$$OR_j = \frac{1}{n} \sum_{i=1}^n F((a_{Bi} - a_{Bj}) \times (a_{jW} - a_{iW})), \text{ for all } i \text{ and } j \quad (10)$$

where  $F(x)$  is a step function defined as:

$$F(x) = \begin{cases} 1 & \text{if } x < 0 \\ 0.5 & \text{if } x = 0 \text{ and } ((a_{Bi} - a_{Bj}) \neq 0 \text{ or } (a_{jW} - a_{iW}) \neq 0) \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

The rationale of  $OR_j$  formulation is that if criterion  $C_j$  overweighs criterion  $C_i$ , then the ordinal consistency should satisfy  $a_{Bi} > a_{Bj}$  and  $a_{jW} > a_{iW}$ , i.e.  $(a_{Bi} - a_{Bj}) \times (a_{jW} - a_{iW}) > 0$ . If only one of  $(a_{Bi} - a_{Bj})$  and  $(a_{jW} - a_{iW})$  is equal to 0, we say that, in this situation, it violates weak ordinal relation [12,15], but if both are equal to 0, it is ordinal-consistent.

$OR_j$  is called *local* ordinal consistency ratio, indicating the degree of consistency with respect to the  $j$ th criterion. With this ordinal consistency ratio ( $OR_j \in [0, 1]$ ), we can find out which criterion violates the relative order (and to what extent), and the higher the  $OR_j$  is, the more contradictory the preferences has regarding this criterion  $C_j$ .

OR is called *global* ordinal consistency ratio, which reflects the ordinal consistency of the pairwise comparison system provided by the DM.

**Example 2.** We use the car evaluation preferences example from the original BWM again (showed in the Example 1 in Section 3.1) to explain the ordinal consistency measurement. From the preference vector  $A_{BO}$ , we can easily get the ranking of the criteria: price > quality > safety > comfort > style. The ranking from the  $A_{WO}$  vector: price > quality ~ comfort > safety > style (">" means superior to, "~" means indifferent to). The orders of the criteria are different in these two vectors, thus the preferences of this DM violate the ordinal consistency. By using the ordinal consistency measurement from Eqs. (9)-(11), we can obtain the ordinal consistency ratios regarding each criterion in Table 2, which represent the ordinal violation level of each criterion. The global ordinal consistency ratios can be calculated from Eq. (9), which is 0.3 in this case.

Combining the cardinal and ordinal consistency ratios, a DM can check his/her rationality during the preference elicitation process. This immediate feedback helps the DM confronts his/her inconsistencies as soon as they arise, making this process more effective [34].

#### 4.3. Properties of the ordinal consistency ratio

The index OR (Eq. (9)) satisfies three basic properties. To enunciate the properties, we need to acknowledge that each vector  $A_{BO}$  and  $A_{OW}$  induces an order relation on the set of criteria. That is to say, for example,  $a_{Bi} > a_{Bj} \Rightarrow i < j$  and  $a_{iW} = a_{jW} \Rightarrow i \sim j$ .

1.  $OR(A_{BO}, A_{OW}) = 0$  if and only if the preferences in the vectors  $A_{BO}, A_{WO}$  induce the same order relation on the set of criteria.
2. OR is invariant with respect to permutations of criteria.

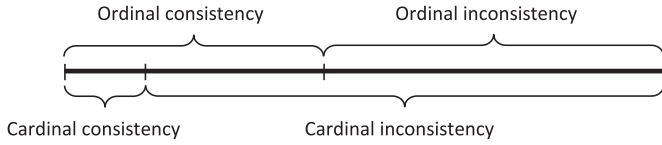


Fig. 4. The inclusion relation between ordinal and cardinal consistency.

- Given two vectors  $A_{BO}$  and  $A_{WO}$  representing the same order relation on the set of criteria, when we choose one preference (component of a vector) and we move it away from its original value in the range  $[1, a_{BW}]$ , this can only increase the value of OR or leave it unchanged.

Since these properties are similar to those in Proposition 1, the associated proof is omitted for the sake of brevity.

#### 4.4. The relationship between ordinal consistency and cardinal consistency

Analysing the data used in the original BWM [37,38], we can obtain the inclusion relation between cardinal and ordinal (in)consistency of the preferences obtained from different DMs, which is graphically presented in Fig. 4. For example, the pairwise comparison system with cardinal consistency is a subset of which, with ordinal consistency, the ordinal inconsistent system is a subset of cardinal inconsistency.

The inclusion relation between cardinal consistency and ordinal consistency shown in Fig. 4 is formalized in the Proposition 2 and Corollary 1.

**Proposition 2.** *If a pairwise comparison system is cardinal-consistent, it must be ordinal-consistent.*

**Proof.** Taking the cardinal consistency condition ( $a_{Bi} \times a_{iW} = a_{BW}$ ,  $a_{Bj} \times a_{jW} = a_{BW}$ , where  $a_{Bi}$ ,  $a_{iW}$ ,  $a_{Bj}$ ,  $a_{jW}$ ,  $a_{BW} \geq 1$ ), and ordinal consistency condition ( $(a_{Bi} - a_{Bj}) \times (a_{jW} - a_{iW}) > 0$  or  $(a_{Bi} = a_{Bj} \& a_{jW} = a_{iW})$ ), we shall show that, given a pairwise comparison system, cardinal consistency implies either (1).  $(a_{Bi} - a_{Bj}) \times (a_{jW} - a_{iW}) > 0$  or (2).  $a_{Bi} = a_{Bj} \& a_{jW} = a_{iW}$ .

- If  $a_{Bi} = a_{Bj}$ , then  $a_{jW} = \frac{a_{BW}}{a_{Bj}} = \frac{a_{BW}}{a_{Bi}} = a_{iW}$ , vice versa, the comparison is ordinal-consistent;
- If  $a_{Bi} \neq a_{Bj}$ , or  $a_{jW} \neq a_{iW}$ ,  $(a_{Bi} - a_{Bj}) \times (a_{jW} - a_{iW}) = a_{Bi} \times a_{jW} - a_{Bj} \times a_{jW} - a_{Bi} \times a_{iW} + a_{Bj} \times a_{iW}$

From the notion of cardinal consistency, we know that:

$$a_{Bi} \times a_{iW} = a_{BW}, a_{Bj} \times a_{jW} = a_{BW}, a_{jW} = \frac{a_{BW}}{a_{Bj}}, a_{iW} = \frac{a_{BW}}{a_{Bi}}$$

so,

$$a_{Bi} \times a_{jW} - a_{Bj} \times a_{jW} - a_{Bi} \times a_{iW} + a_{Bj} \times a_{iW}$$

$$= \frac{a_{Bi} \times a_{BW}}{a_{Bj}} + \frac{a_{Bj} \times a_{BW}}{a_{Bi}} - 2a_{BW}$$

$$= \frac{a_{BW}(a_{Bi}^2 + a_{Bj}^2)}{a_{Bi}a_{Bj}} - 2a_{BW}$$

$$= \frac{a_{BW}(a_{Bi}^2 + a_{Bj}^2 - 2a_{Bi}a_{Bj})}{a_{Bi}a_{Bj}}$$

$$= \frac{a_{BW}(a_{Bi} - a_{Bj})^2}{a_{Bi}a_{Bj}} > 0.$$

Therefore, the comparison is also ordinal-consistent.  $\square$

**Corollary 1.** *If a pairwise comparison system is ordinal-inconsistent, it must be cardinal-inconsistent.*

## 5. Thresholds for BWM

Even though we can easily identify the inconsistent judgement by using the consistency measurements proposed in this study, requiring the DM to achieve perfect cardinal and ordinal consistency is unrealistic. However, the question involving the degree to which inconsistency can be accepted has far been lacking in the study of BWM. As such, to bridge this gap, a threshold has to be defined. In the following section, based on the concept of ordinal and cardinal consistency measurement, a method to derive consistency thresholds is proposed.

### 5.1. A methodology for determining the thresholds

Inspired by Amenta et al. [3,4], we develop a method for determining the thresholds for BWM, which is based on the cardinal consistency measurement and the definition of ordinal consistency. The thresholds for BWM are established, not only for the input-based consistency measurement, but also for the output-based consistency measurement. However, we use the input-based consistency ratio ( $CR^I$ ) to illustrate this approach.

The basic idea is that, based on the concept of ordinal consistency, if a decision-maker is ordinal-consistent, the ranking of the final weights obtained from the two preference vectors ( $A_{BO}$  and  $A_{OW}$ ) will not change with  $CR^I$ , only the intensities may vary. In this sense, we can suggest that the preferences provided by the DM are reliable.

We use Monte-Carlo method to simulate the probability distribution of  $CR^I$ s. In this study, we analyse the entire problem space covering the weighting problems, with the number of criteria ranging from 3 to 9, and where the preferences can be assigned with the largest evaluation grade from 3 to 9, we call them 3-scale to 9-scale.<sup>1</sup> Consequently, in all, there are  $7 \times 7 = 49$  combinations to be analysed. For each combination, we randomly generated 10,000 pairs of ordinal-consistent vectors, each pair acting as the two vectors  $A_{BO}$  and  $A_{OW}$ . We categorized this group as an *acceptable group*, and calculated all the  $CR^I$ s of this group. Likewise, we randomly generated 10,000 pairs of ordinal-inconsistent vectors and calculated their  $CR^I$ s, which is categorized as an *unacceptable group*.

Theoretically, we can obtain all the possible  $CR^I$ s of the acceptable group in each situation, taking the *maximum* as a boundary (boundary 1), the  $CR^I$ s above this boundary are not acceptable, because they can only be ordinal-inconsistent. Although, practically, it is very difficult to traverse all the possibilities, we still assume that the *maximum*  $CR^I$  from 10,000 pair of vectors as the boundary 1, because the likelihood of having a higher value than this boundary is very low. For example, the maximum consistency value of 9-criterion and 9-scale ordinal-consistent pairwise comparison vectors is 0.7639, which means that, for any judgments whose  $CR^I$ s are bigger than this value in a 9-criterion and 9-scale size problem, they should be rejected.

However, that does not automatically mean that the  $CR^I$ s within that boundary are necessarily acceptable, because they could still be ordinal-inconsistent, and ordinal inconsistency is what we set out to reject. Based on this idea, the *minimum*  $CR^I$  could be used as a boundary (boundary 2), all of the  $CR^I$ s within this boundary are acceptable. For example, the minimum consistency value of 9-criterion and 9-scale ordinal-inconsistent paired vectors is 0.0694, if the  $CR^I$ s obtained are smaller than this boundary, they should be accepted.

Values of  $CR^I$  greater than boundary 1 are assumed to be totally unacceptable, while values below boundary 2 are assumed totally

<sup>1</sup> In each scale, we use discrete number from 1 to the largest grade which is actually the  $a_{BW}$ . For example, if we use 7-scale, the grades used in  $A_{BO}$  and  $A_{OW}$  are randomly selected from  $\{1, 2, \dots, 7\}$ .

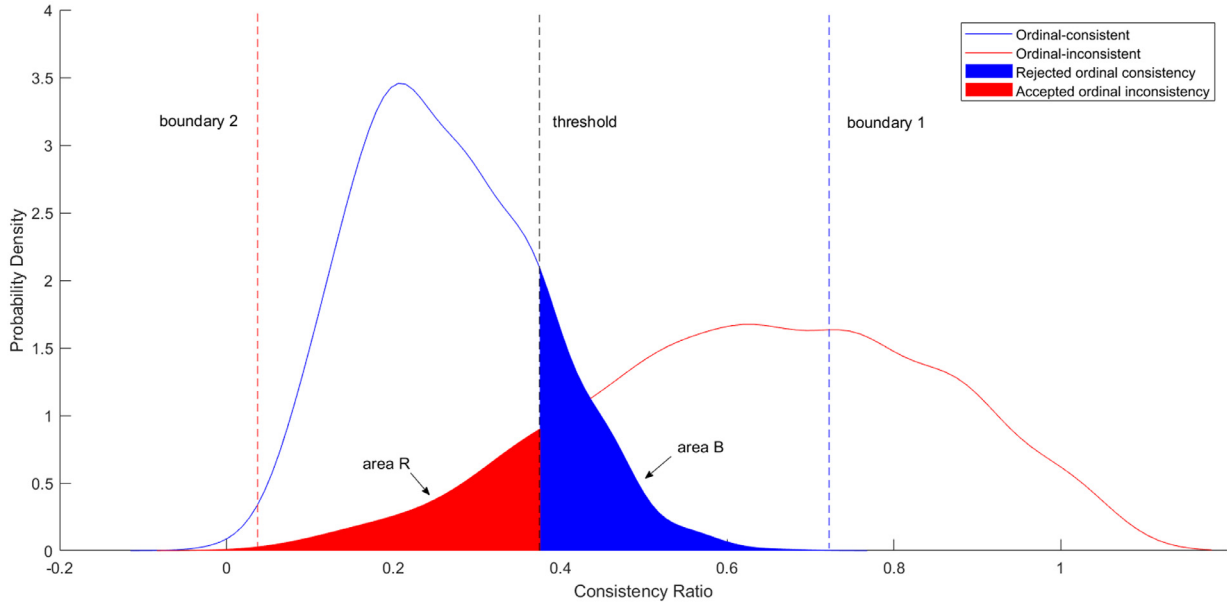


Fig. 5. The kernel distribution of  $CR^I$ s of the two groups (9-criteria 9-scale).

acceptable. Between boundary 1 and 2, we expect that there exists a threshold, making the proportion of ordinal inconsistency we accept as small as possible, and beyond the threshold, the proportion of ordinal consistency we reject should be as small as possible. In statistical terms, our goal is to minimize the sum of Type I error (false positive) and Type II error (false negative). This idea can be more clearly visualized in a kernel smoothing distribution in 9-criteria and 9-scale combination, as shown in Fig. 5.

From the idea explained above, the empirical cumulative distribution function can be used to achieve our purpose.

**Definition 6.** (Empirical cumulative distribution function). The empirical cumulative distribution function of  $CR^I$  can be defined as:

$$\hat{F}(\alpha) = \frac{1}{N} \sum_{i=1}^N I\{CR_i^I \leq \alpha\} \quad (12)$$

where  $I\{\cdot\}$  is the indicator function:

$$I\{CR_i^I \leq \alpha\} = \begin{cases} 1 & \text{if } CR_i^I \leq \alpha \\ 0 & \text{otherwise} \end{cases}, \quad (13)$$

and  $N$  is the pair number of pairwise comparisons,  $CR_i^I$  is the  $i$ th ( $i \in \{1, \dots, N\}$ ) input-based consistency ratio obtained from this  $N$  pairs of preferences,  $\alpha \in [0, 1]$  is the possible threshold.

We now distinguish the distribution function based on two groups: (1) for the *Acceptable* group, the cumulative distribution of  $CR^I$  in ordinal-consistent situation is denoted as  $\hat{F}^A(\alpha)$ ; (2) for the *Unacceptable* group, the cumulative distribution of  $CR^I$  in ordinal-inconsistent situation is denoted as  $\hat{F}^U(\alpha)$ .

The rejected part of the ordinal-consistent group is  $1 - \hat{F}^A(\alpha)$ , which can be seen in the blue area B in Fig. 5, and the accepted ordinal-inconsistent group is  $\hat{F}^U(\alpha)$ , which is the red area R. We can calculate the relative rejected proportion of the  $CR^I$ s in the acceptable group ( $P_{rejected}^A$ ) and the accepted proportion of the  $CR^I$ s in the unacceptable group ( $P_{accepted}^U$ ) using the following formulas:

$$P_{rejected}^A = \frac{1 - \hat{F}^A(\alpha)}{1 - \hat{F}^A(\alpha) + \hat{F}^U(\alpha)}, \quad (14)$$

$$P_{accepted}^U = \frac{\hat{F}^U(\alpha)}{1 - \hat{F}^A(\alpha) + \hat{F}^U(\alpha)}. \quad (15)$$

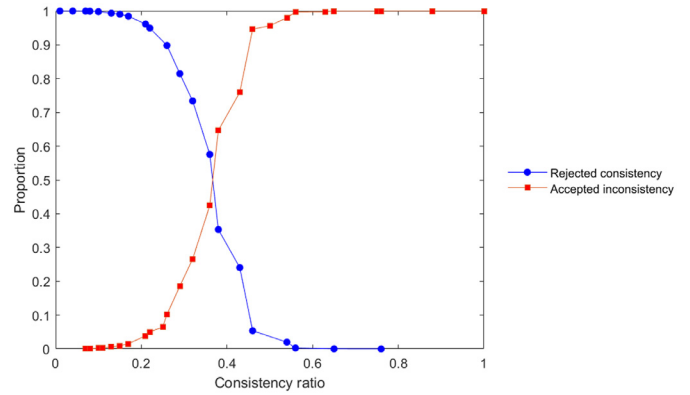


Fig. 6. The acceptance and rejection relative proportions of the two groups (9-criteria 9-scale).

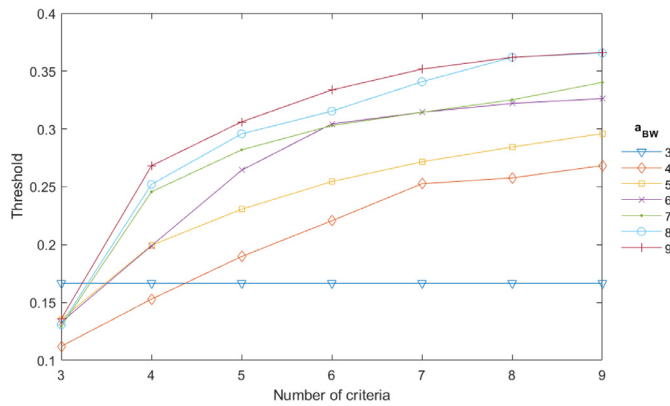
The relationship between these two proportions is shown in Fig. 6, which shows how the possibility of acceptance (red line with squares) and rejection (blue line with circles) distribute in the two groups according to the selected threshold from 0 to 1.

The goal is to obtain a threshold which makes the red and blue areas in Fig. 5 as small as possible, or makes the relative proportions of the two groups in Fig. 6 as close as possible. If there exists a  $CR^I$  obtained from the two groups which makes  $P_{rejected}^A = P_{accepted}^U$ , the two lines in Fig. 6 will intersect at that point, which means that the proportion of rejection in the acceptable group and the proportion of acceptance in the unacceptable group are the same. However, as the obtained  $CR^I$ s are discrete, there could be no  $CR^I$  at the intersection point, which means that we need to find out the intersecting coordinate of the two lines, using the corresponding  $CR^I$  as the threshold. The simulation algorithm for obtaining the threshold is illustrated in the Appendix.

## 5.2. Approximated thresholds for the input-based consistency ratio

Based on the algorithm presented above, we can finally establish the thresholds for BWM. In Table 3, we have obtained the consistency thresholds for combinations which range from 3–9 criteria





**Fig. 7.** Thresholds for different combinations using input-based consistency measurement.

**Table 3**  
Thresholds for different combinations using input-based consistency measurement.

Scales	Criteria						
	3	4	5	6	7	8	9
3	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667
4	0.1121	0.1529	0.1898	0.2206	0.2527	0.2577	0.2683
5	0.1354	0.1994	0.2306	0.2546	0.2716	0.2844	0.2960
6	0.1330	0.1990	0.2643	0.3044	0.3144	0.3221	0.3262
7	0.1294	0.2457	0.2819	0.3029	0.3144	0.3251	0.3403
8	0.1309	0.2521	0.2958	0.3154	0.3408	0.3620	0.3657
9	0.1359	0.2681	0.3062	0.3337	0.3517	0.3620	0.3662

with highest evaluation grades from 3 to 9 based on the input-based consistency measurement.

The thresholds in the combinations with 3-criteria and the combinations with 3-scale are relatively special. The thresholds in 3-scale problem remain unchanged even the number of criterion changes, because, no matter how many criteria there are, the maximum  $CR^I$  in the acceptable group and the minimum  $CR^I$  in the unacceptable group are equal to 0.1667. In most other cases, we can see that the thresholds have a tendency to increase along with the number of criteria and with the scale of the preferences, as shown in Fig. 7.<sup>2</sup>

### 5.3. Approximated thresholds for the output-based consistency ratio

By using the same algorithm in the Appendix, we can also determine the thresholds for the  $CR^O$  in different combinations, as shown in Table 4.<sup>3</sup>

Compared to the thresholds obtained from the input-based consistency measurement, the thresholds of the output-based consistency measurement are slightly higher.

Finally, by using the approximated consistency thresholds obtained above, we can check whether or not the consistency of the DM is acceptable. For instance, since the overall  $CR^I$  in the illustrative example in Section 3.1 is 0.14, which is less than the threshold of 0.2958 (in 5-criteria and 8-scale combination), as shown in Table 3, it is acceptable. If we use  $CR^O$ , which is 0.223, we can see that it is also below the threshold of 0.4029, as shown in Table 4.

<sup>2</sup> The combinations with 2-scale for the  $CR^I$  are not shown in Table 3 and Fig. 7, but it is worth mentioning that the threshold should be 0 in this case, because, when the preferences are ordinal-consistent,  $CR^I=0$ . Therefore, the DM should revise his or her preferences when  $CR^I > 0$ .

<sup>3</sup> The threshold for the  $CR^O$  in the combinations with 2-scale is 0, because when the preferences are ordinal-consistent,  $CR^O=0$ .

**Table 4**

Threshold for different combinations using output-based consistency measurement.

Scales	Criteria						
	3	4	5	6	7	8	9
3	0.2087	0.2087	0.2087	0.2087	0.2087	0.2087	0.2087
4	0.1581	0.2352	0.2738	0.2928	0.3102	0.3154	0.3273
5	0.2111	0.2848	0.3019	0.3309	0.3479	0.3611	0.3741
6	0.2164	0.2922	0.3565	0.3924	0.4061	0.4168	0.4225
7	0.2090	0.3313	0.3734	0.3931	0.4035	0.4108	0.4298
8	0.2267	0.3409	0.4029	0.4230	0.4379	0.4543	0.4599
9	0.2122	0.3653	0.4055	0.4225	0.4445	0.4587	0.4747

Thanks to these thresholds,  $CR^I$  and  $CR^O$  now have a meaningful interpretation, because we can now determine whether they are acceptable or not. The thresholds for  $CR^I$  can help a DM check his/her pairwise comparisons before solving the optimization program.

## 6. Conclusion

In this paper, we addressed the consistency issue in BWM. First, we argued that the output-based consistency measurement in BWM cannot provide immediate feedback to a DM, and only informs the DM about any inconsistencies in his/her assessments after the entire elicitation process has finished, which has been proven to be ineffective. In addition, existing consistency indices designed for the incomplete pairwise comparison matrices are not as desirable as we expected. To remedy that state of affairs, we propose an input-based consistency ratio, which has a number of desirable properties and a high correlation to the original ratio, to indicate the DM's consistency status during the preference elicitation process. This input-based consistency ratio is simple and is easy for a DM to identify his/her most inconsistent judgments. Then, to complement the cardinal consistency measurement, we proposed an ordinal consistency measurement to explicate the possible contradictions even in cases where the cardinal consistency of a DM's pairwise comparisons is considered to be good enough. This ratio not only shows how much a DM violates the ordinal consistency, but also provides a convenient way to identify and correct the conflicts involved. Finally, with the help of Monte-Carlo simulations, we determined the thresholds for the input-based and output-based consistency ratios in different scales with different numbers of criteria. The idea is to balance the ordinal consistency and inconsistency, making the portion of the cardinal consistency ratios that violate ordinal consistency to be accepted as small as possible and the portion of the cardinal consistency ratios that satisfy ordinal consistency to be rejected as small as possible. With these thresholds, a DM can decide whether or not to revise his/her earlier assessments. And because the input-based consistency measurement can indicate the consistency level regarding each criterion, it can be used in the preference revision process.

The method of determining the thresholds only considers whether the judgments are ordinal-consistent or not and has not taken the violation level into account. This will be examined in future studies. Similarly to the approach what was adopted in this paper, this method can also be applied to fuzzy consistency measurements to determine their corresponding thresholds.

## Acknowledgement

The authors gratefully acknowledge financial support for PhD study from China Scholarship Council (No. 201708440305).

## Appendix

The algorithm for obtaining the threshold for the  $CR^I$  is illustrated as follows and its graphical representation is shown in Fig. 8.

- Step 1 Generate pairwise comparison vectors. Suppose we have  $n$  criteria ( $n = 3, 4, \dots, 9$ ), two random vectors  $A_{BO} = (a_{B1}, \dots, a_{Bn})$  and  $A_{OW} = (a_{1W}, \dots, a_{nW})$  with the maximum scale  $m$  ( $m = 3, 4, \dots, 9$ ), are created to represent the pairwise comparisons vectors  $A_{BO}$  and  $A_{OW}$  in BWM. The elements in  $A_{BO}$  and  $A_{OW}$  are integers randomly selected from domain  $[1, m]$ .
- Step 2 Establish the ordinal-consistent group. After creating a pair of vectors  $a_B$  and  $a_W$ , it will be assigned to the ordinal-consistent group if it satisfies ordinal consistency condition (8), and  $i = i + 1$ .
- Step 3 Establish the ordinal-inconsistent group. If the paired vector generated in Step 1 does not satisfy the ordinal con-

sistency condition, it will be assigned to the ordinal-inconsistent group, and  $j = j + 1$ .

- Step 4 Continue to create the ordinal-consistent and ordinal-inconsistent groups through steps 1–3, until the size of both groups is 10,000.
- Step 5 Calculate the  $CR^I$  for all the paired vectors in these two groups by using Eqs. (6), (7).
- Step 6 Calculate the empirical cumulative distribution of  $CR^I$  for the two groups by using Eqs. (12), (13).
- Step 7 Calculate the relative rejected proportion of the  $CR^I$ s in the acceptable group ( $P_{rejected}^A$ ) and the accepted proportion of the  $CR^I$ s in the unacceptable group ( $P_{accepted}^U$ ) by using Eqs. (14), (15).
- Step 8 If there exists a  $CR_T^I$  making  $P_{rejected}^A = P_{accepted}^U$ , then this  $CR_T^I$  is the threshold. If not, go to next step.
- Step 9 Identify the cross point of the lines of  $P_{rejected}^A$  and  $P_{accepted}^U$ , the  $CR^I$  at this point is used as the threshold.

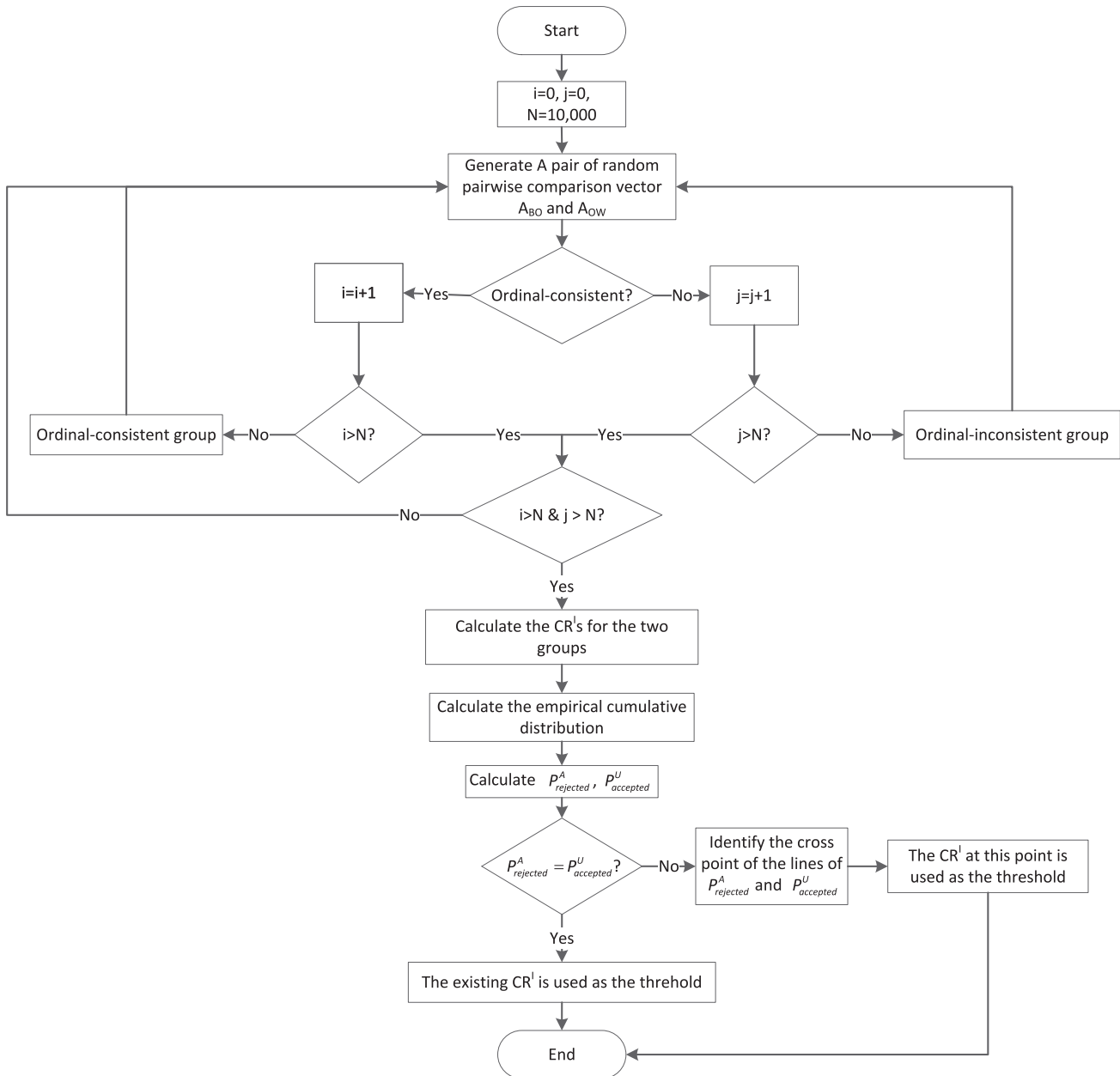


Fig. 8. Graphical representation of simulation algorithm.

## References

- [1] Aboutorab H, Saberi M, Asadabadi MR, Hussain O, Chang E. ZBWM: the Z-number extension of Best Worst Method and its application for supplier development. *Expert Syst Appl* 2018;107:115–25.
- [2] Aguarón J, Moreno-Jiménez JM. The geometric consistency index: approximated thresholds. *Eur J Oper Res* 2003;147(1):137–45.
- [3] Amenta P, Lucadamo A, Marcarelli G. Approximate thresholds for Salo-Hamalainen index. *IFAC-PapersOnLine* 2018;51(11):1655–9.
- [4] Amenta P, Lucadamo A, Marcarelli G. On the transitivity and consistency approximated thresholds of some consistency indices for pairwise comparison matrices. *Inf Sci* 2020;507:274–87.
- [5] Bozóki S, Rapcsák T. On Saaty's and Koczkodaj's inconsistencies of pairwise comparison matrices. *J Glob Optim* 2008;42(2):157–75.
- [6] Bozóki S, Tsyganok V. The (logarithmic) least squares optimality of the arithmetic (geometric) mean of weight vectors calculated from all spanning trees for incomplete additive (multiplicative) pairwise comparison matrices. *Int J Gener Syst* 2019;48(4):362–81.
- [7] Bozóki S, Fülöp J, Poesz A. On reducing inconsistency of pairwise comparison matrices below an acceptance threshold. *Cent Eur J Oper Res* 2015;23(4):849–66.
- [8] Brunelli M. A survey of inconsistency indices for pairwise comparisons. *Int J Gener Syst* 2018;47(8):751–71.
- [9] Brunelli M, Fedrizzi M. Axiomatic properties of inconsistency indices for pairwise comparisons. *J Oper Res Soc* 2015;66(1):1–15.
- [10] Brunelli M, Fedrizzi M. A general formulation for some inconsistency indices of pairwise comparisons. *Ann Oper Res* 2018;274(1–2):155–69.
- [11] Brunelli M, Rezaei J. A multiplicative best-worst method for multi-criteria decision making. *Oper Res Lett* 2019;47(1):12–15.
- [12] Cavallo B, D'Apuzzo L, Basile L. Weak consistency for ensuring priority vectors reliability. *J MultiCriteria Decis Anal* 2016;23(3–4):126–38.
- [13] Crawford G, Williams C. A note on the analysis of subjective judgment matrices. *J Math Psychol* 1985;29(4):387–405.
- [14] Ergu D, Kou G, Peng Y, Shi Y. A simple method to improve the consistency ratio of the pair-wise comparison matrix in ANP. *Eur J Oper Res* 2011;213(1):246–59.
- [15] Escobar MT, Aguarón J, Moreno-Jiménez JM. Some extensions of the precise consistency consensus matrix. *Decis Support Syst* 2015;74:67–77.
- [16] Fedrizzi M, Giove S. Incomplete pairwise comparison and consistency optimization. *Eur J Oper Res* 2007;183(1):303–13.
- [17] Fishburn P. Preference structures and their numerical representations. *Theor Comput Sci* 1999;217(2):359–83.
- [18] Guo S, Zhao H. Fuzzy best-worst multi-criteria decision-making method and its applications. *Knowl-Based Syst* 2017;121:23–31.
- [19] Harker PT. Incomplete pairwise comparisons in the analytic hierarchy process. *Math Model* 1987;9(11):837–48.
- [20] Iida Y. Ordinality consistency test about items and notation of a pairwise comparison matrix in AHP. In: *Proceedings of the international symposium on the analytic hierarchy process*; 2009.
- [21] Irwin FW. An analysis of the concepts of discrimination and preference. *Am J Psychol* 1958;71(1):152–63.
- [22] Jensen RE, Hicks TE. Ordinal data AHP analysis: a proposed coefficient of consistency and a nonparametric test. *Math Comput Model* 1993;17(4–5):135–50.
- [23] Kendall MG, Smith BB. On the method of paired comparisons. *Biometrika* 1940;31(3–4):324–45.
- [24] Koczkodaj WW. A new definition of consistency of pairwise comparisons. *Math Comput Model* 1993;18(7):79–84.
- [25] Koczkodaj WW, Urban R. Axiomatization of inconsistency indicators for pairwise comparisons. *Int J Approx Reason* 2018;94:18–29.
- [26] Koczkodaj WW, Magnot JP, Mazurek J, Peters JF, Rakhshani H, Soltys M, Strzałka D, Szybowski J, Tozzi A. On normalization of inconsistency indicators in pairwise comparisons. *Int J Approx Reason* 2017;86:73–9.
- [27] Kułakowski K. Inconsistency in the ordinal pairwise comparisons method with and without ties. *Eur J Oper Res* 2018;270(1):314–27.
- [28] Kułakowski K, & Talaga, D. (2019). Inconsistency indices for incomplete pairwise comparisons matrices. *arXiv preprint arXiv:1903.11873*.
- [29] Kumar A, Gupta H. Evaluating green performance of the airports using hybrid BWM and VIKOR methodology. *Tour Manag* 2020;76:103941.
- [30] Kwiesielewicz M, Van Uden E. Inconsistent and contradictory judgements in pairwise comparison method in the AHP. *Comput Oper Res* 2004;31(5):713–19.
- [31] Lundy M, Siraj S, Greco S. The mathematical equivalence of the "spanning tree" and row geometric mean preference vectors and its implications for preference analysis. *Eur J Oper Res* 2017;257(1):197–208.
- [32] Mi X, Tang M, Liao H, Shen W, Lev B. The state-of-the-art survey on integrations and applications of the best worst method in decision making: why, what, what for and what's next? *Omega* 2019;87:205–25.
- [33] Monsuur H. An intrinsic consistency threshold for reciprocal matrices. *Eur J Oper Res* 1997;96(2):387–91.
- [34] Monti S, Carenini G. Dealing with the expert inconsistency in probability elicitation. *IEEE Trans Knowl Data Eng* 2000;12(4):499–508.
- [35] Mou Q, Xu Z, Liao H. An intuitionistic fuzzy multiplicative best-worst method for multi-criteria group decision making. *Inf Sci* 2016;374:224–39.
- [36] Pereira V, Costa HG. Nonlinear programming applied to the reduction of inconsistency in the AHP method. *Ann Oper Res* 2015;229(1):635–55.
- [37] Rezaei J. Best-worst multi-criteria decision-making method. *Omega* 2015;53:49–57.
- [38] Rezaei J. Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega* 2016;64:126–30.
- [39] Rezaei J, Kothadiya O, Tavasszy L, Kroesen M. Quality assessment of airline baggage handling systems using SERVQUAL and BWM. *Tour Manag* 2018;66:85–93.
- [40] Saaty TL. A scaling method for priorities in hierarchical structures. *J Math Psychol* 1977;15(3):234–81.
- [41] Saaty TL. The analytic hierarchy process: planning, priority setting. *Resour Alloc* 1980;2.
- [42] Saaty TL. *Fundamentals of decision making and priority theory with the analytic hierarchy process*. RWS publications; 2000.
- [43] Salo AA, Hämäläinen RP. Preference programming through approximate ratio comparisons. *Eur J Oper Res* 1995;82(3):458–75.
- [44] Siraj S, Mikhailov L, Keane JA. Enumerating all spanning trees for pairwise comparisons. *Comput Oper Res* 2012;39(2):191–9.
- [45] Siraj S, Mikhailov L, Keane JA. Contribution of individual judgments toward inconsistency in pairwise comparisons. *Eur J Oper Res* 2015;242(2):557–67.
- [46] Siraj S, Mikhailov L, Keane J. A heuristic method to rectify intransitive judgments in pairwise comparison matrices. *Eur J Oper Res* 2012;216(2):420–8.
- [47] Ureña R, Chiclana F, Morente-Molina JA, Herrera-Viedma E. Managing incomplete preference relations in decision making: a review and future trends. *Inf Sci* 2015;302:14–32.
- [48] Xu Y, Gupta JND, Wang H. The ordinal consistency of an incomplete reciprocal preference relation. *Fuzzy Sets Syst* 2014;246:62–77.
- [49] Yadav G, Mangla SK, Luthra S, Jakhar S. Hybrid BWM-ELECTRE-based decision framework for effective offshore outsourcing adoption: a case study. *Int J Prod Res* 2018;56(18):6259–78.