The reliability analysis of rating systems in decision making

When scale meets multi-attribute additive value model

Zhao, Sihan; Dong, Yucheng; He, Ying

Published in:
Decision Support Systems

DOI:
10.1016/j.dss.2020.113384

Publication date:
2020

Document version:
Accepted manuscript

The reliability analysis of rating systems in decision making: When scale meets multi-attribute additive value model

Sihai Zhao[a], Yucheng Dong[a,*], Ying He[b]

[a]Business School, Sichuan University, Chengdu 610065, China

[b]Department of Business and Economics, University of Southern Denmark, Denmark

Abstract: A rating system (RS) comprises a rating metric defined by a discrete set of integers contained in an interval (e.g., $[0, N]$), and an aggregation rule. RSs are widely used in various fields to capture and summarize individuals' opinions on alternatives. In this paper we argue that the multi-attribute additive value model (MAVM) should be used as a benchmark to analyze the reliability of RSs, and present some tight bounds on the parameter N and overall rating scores, which guarantee the consistency between the RS $[0, N]$ and MAVM at the ranking and rating levels. Interestingly, the tight bounds at the rating level are twice as large as those at the ranking level. The results in this paper can provide new insights about the reliability analysis of RSs.

Keywords: Rating systems, multi-attribute additive value model, asymmetric information, reliability, consistency

1. Introduction

Rating systems (RSs) are widely used in business, management, education and many other fields as a method to capture and summarize individuals' opinions on alternatives [16, 17, 36, 38, 46, 47, 52]. Typically, the RS is an effective tool to deal with asymmetric information between

* Corresponding author.

Email addresses: zhaoshscu@126.com (S. Zhao), ycdong@scu.edu.cn (Y. Dong), yinghe@sam.sdu.dk (Y. He).

sellers and customers in the market [1, 8, 45]. Using RSs, products, candidates, or enterprises are often assessed to obtain rankings or ratings in different decision making situations. For example, a company may rate a candidate to decide whether to offer a job; an investor may consult the rating information of enterprises; and a bank may rate customers to determine their credit limits.

An RS usually comprises a rating metric defined by a set of integers contained in a finite interval, and an aggregation rule which combines individual rating information into an overall rating score [7]. For instance, a binary metric $[-1,1]$ refers to a 3-point scale containing three rating options: −1, 0, 1. The aggregation rule might be a simple weighted average to perform the aggregation process. Using the RSs information, decisions such as buying a product or selecting an excellent employee can be made.

The reliability analysis is one of the most challenging issues in the RS research because inconsistent decision outcomes may occur under different RSs. In the existing literature, three ways to analyze the reliability of RSs have been proposed:

(1) Misunderstanding of the meaning of the scale. Most research on rating scales empirically explores how people rate when given different rating scales [9, 13, 20, 21, 25, 35], and it is found that people often do not understand the meaning of the scale in an RS [2, 34, 39]. Preston and Colman [42] showed that the 2-point, 3-point and 4-point scales performed significantly worse on several indices of reliability, validity and discriminating power than scales with more response categories. Weijters et al. [50] pointed out that scale formats have strong effects on response distributions and mis-response to reversed items, and then formulated recommendations on the choice of a scale format. Schwarz et al. [44] found that using a scale $[-3,3]$ leads to different rating results from using a scale $[1,7]$, which implies that the actual numeric values used in an

RS will affect raters' expression of true opinions. If an RS cannot capture the details of people's true opinions, it may be difficult to make an effective decision, and thus the reliability of the RS is debatable. Similar studies can be found in the decision making models dealing with linguistic scales (e.g., [22, 28]).

(2) Strategic behaviors. Another explanation attributes inconsistent decision outcomes to raters' strategic behaviors. Cleveland and Murphy [10] and Murphy et al. [41] pointed out that raters may have different individual goals due to the fact that they may come from different domains or they all own particular interests. These goals will affect the effectiveness of the evaluation information they provide, which in turn affects the reliability of the final decision outcome. For example, a rater may use strategic behaviors to make his/her most preferred alternative(s) selected. In decision analysis, there exist some similar studies to examine the impacts of strategic behaviors on decision outcomes, and to provide some effective mechanisms for defending against strategic manipulations [11, 14, 15, 40, 49].

(3) Rounding process in RSs. Bargagliotti and Li [4] proposed a novel view that different decision outcomes may occur without human fault. Instead, the use of the rounding process in RSs plays an important role. They investigated the consistency between the scale metric $[1, N]$ and the binary metric $[-1,1]$ at the representation and aggregation levels. It is proved that any two opinions represented by the same rating in the scale metric are also represented by the same rating in the binary metric if and only if N is odd and $N-1$ is not divisible by 4. Besides, several sufficient conditions are provided to guarantee the same aggregation decision outcome when consulting both metrics.

The existing research has provided useful insights about the reliability analysis of RSs, but most of them empirically investigated the impact of people's psychological and behavioral factors on the reliability of RSs. Although Bargagliotti and Li [4] proposed an analytical view about the RSs' reliability, we find that there are some gaps that should be further filled:

(1) The ranking and rating consistency of multiple alternatives in RSs. Bargagliotti & Li [4] only investigated the consistency of a single alternative between the scale metric $[1, N]$ and the binary metric $[-1,1]$ at the representation and aggregation levels. It is common that we want to obtain the ranking or ratings of multiple alternatives in practical RS problems. For example, when someone is going to buy a smart phone online, what she may mostly care about is the ranking of the alternative smart phones on sale. Besides, a number of hotels are rated different stars (e.g., a five-star hotel). Therefore, we argue that it is of key importance to study the consistency of multiple alternatives in RSs at the ranking and rating levels to cope with real-world RS decision problems.

(2) The consistency between RSs and multi-attribute additive value model (MAVM) [31]. Essentially, RS is a method used to aggregate individuals' preferences. Such a problem has been extensively studied in economics in the fields of social choice and decision theory [3, 6, 18, 26, 32, 33, 37]. In decision theory, group utility is studied based on axiomatic approach, which provides a solid theoretical foundation for studying preferences aggregation. Axioms are observable conditions on the preferences, which make the utility theory testable on behavioral level. RS is proposed as a method to aggregate individuals' preferences to fulfill the same purpose as group utility theory; but it lacks such a theoretical foundation. Therefore, it is interesting to investigate how RS can be connected to the group utility theory. The MAVM is a classical model in multi-attribute utility theory [18, 19, 27, 31, 33, 48], which has been used to

axiomatize group utility [32]. It has the high interpretability of numerical scores that can be decomposed into individual values (e.g., [24, 30]). However, when evaluating products, candidates, or enterprises in business, we often use RSs (not MAVM). In essence, RS problems can be modeled by multiple attribute decision making because evaluating an alternative by multiple raters is equivalent to that by multiple attributes [29], and in the aggregation rules of RSs the additive model is widely employed [4, 7]. Therefore, RS can be considered as an approximate model of MAVM, and thus it is necessary to study the consistency between RS and MAVM. When the decision outcomes between RS and MAVM are consistent, we argue that the RS is reliable.

Similar to Bargagliotti & Li [4], it is assumed that the perceived state (PS) of a rater for the quality of an alternative exists, and the available metric requires the rater to choose a discrete rating to represent the PS for the alternative. In this paper, we model PS by a continuous measurement in the interval $[0,1]$, and the PSs are assumed to be linearly transformed (with rounding) into discrete integer ratings in the interval $[0, N]$, called the RS $[0, N]$. We argue that the PSs could be regarded as values of an alternative associated with multiple raters because the PSs and values are essentially the same thing and both of them measure how satisfied people are with an alternative [3, 12, 26, 31, 33]. Then, the MAVM is employed as the basis to analyze the reliability of RSs. We analytically present some tight bounds on the parameter N and overall rating scores to guarantee the consistency between the RS $[0, N]$ and MAVM at the ranking and rating levels. Interestingly, the tight bounds at the rating level are twice as large as those at the ranking level. Furthermore, we present the detailed simulation analysis to show more findings on the consistency of decision outcomes between RS and MAVM.

The results in this paper can provide useful insights for the reliability analysis of RSs. If PSs are available, MAVM can be employed to get reliable results. However, RS is widely used as an approximate model of MAVM in practical decision problems, which sometimes leads to unreliable results. For instance, two candidates (A and B) compete for a position, and three interviewers with same importance rate A and B on the scale $[0,5]$. The interviewers' PSs are 0.82, 0.91, 0.86 for A; and 0.86, 0.89, 0.88 for B. Using the RS $[0,5]$, the interviewers rate A as 4, 5, 4; and B as 4, 4, 4, respectively. Then the weighted average scores of A and B are 4.3 and 4, respectively, and thus A is more preferred. However, on the basis of MAVM, the aggregation results of PSs are 0.863 and 0.877 for A and B, respectively, and thus B is more preferred. Therefore, the RS leads to the inconsistent decision outcome with MAVM. In this paper, using MAVM as a benchmark, we provide some analytical conditions to show the key roles of the parameter N and overall rating scores in the reliability analysis of RSs.

The remainder of this paper is organized as follows. The next section introduces the framework to study the consistency between RS and MAVM. Then, we present the analytical consistency conditions between RS and MAVM at the ranking and rating levels in Section 3. Following this, in Section 4 simulation experiments are demonstrated to further explore the properties of RSs, and a hypothetical example of application is presented to illustrate the usability of the obtained results in practical RS problems. Finally, conclusion and future perspectives are included in Section 5.

2. Framework

An RS comprises a rating metric and an aggregation rule. The rating metric is a discrete set of integers contained in an interval. Let $int[e,f]$ be the set of integers between e and f which contains e and f. Let $t_{ij} \in int[e,f]$ be the rating of alternative i associated with rater j. For alternative i and m raters, $(int[e,f])^m$ denotes the space of all possible m-tuples $(t_{i1}, t_{i2}, \ldots, t_{im})$. The aggregation rule is some type of averaging function over $(t_{i1}, t_{i2}, \ldots, t_{im})$.

Definition 1 (Bargagliotti and Li [4]): An RS is the pair $(int[e,f], G)$ where G is an aggregation function from $(int[e,f])^m$ to the interval $[e,f]$:

$$G: (int[e,f])^m \to [e,f]$$

In this paper, we assume that the rating metric is the RS $[0,N]$, which will not change the essence of an RS, and we consider that the RS aggregation rule is a weighted average.

Similar to [4], it is assumed that the PS of a rater for an alternative exists, and the available metric requires the rater to choose a discrete rating to represent his/her PS for the alternative. In this paper, we model the PSs as continuous values in the interval $[0,1]$, and assume that there exists a linear transformation process (with rounding) from PSs to ratings. Let $v \in [0,1]$ be a PS, and the PS can be transformed into a rating, t, in a $[0,N]$ metric based on Eqs. (1) and (2):

$$v' = N \times v \tag{1}$$

$$t = rounding(v') \tag{2}$$

where v' is the value in the interval $[0,N]$ linearly transformed from v, and *rounding* is a rounding operation. When v' is located at halfway between two integer values, in this paper t will be the larger one. For instance, there is a PS that $v = 0.3$, and then its rating is $t = rounding(0.3 \times 5) = 2$ when the rating metric is $[0,5]$. This assumption is also made in the previous literature on RS. However, it should be noted that if another rounding point is adopted, all results of this paper are still valid.

The RS problem can be modeled by multiple attribute decision making because evaluating an alternative by multiple raters is equivalent to the group decision making problem where individuals' values are aggregated to a group value by multiple attribute value function [5, 6, 18, 29, 31]. Thus, in this paper we consider the following multiple attribute decision making problem:

Let $X = \{x_1, x_2, \dots, x_n\}$ be a set of alternatives; $A = \{a_1, a_2, \dots, a_m\}$ be a set of attributes, where an attribute is equivalent to a rater in the RS; and $w = \{w_1, w_2, \dots, w_m\}^T$ be a weight vector of attributes such that $w_i \geq 0$ and $\sum_{i=1}^{m} w_i = 1$. Let $V = [v_{ij}]_{n\times m}$, where v_{ij} denotes the PS of alternative x_i associated with attribute a_j. Let $T = [t_{ij}]_{n\times m}$, where $t_{ij} = rounding(N \times v_{ij})$, obtained via Eqs. (1) and (2), is the rating of alternative x_i associated with attribute a_j in the RS $[0, N]$.

The PS measures how satisfied people are with an alternative, essentially, v_{ij} also represents the value of alternative x_i over attribute a_j in the framework of MAVM [3, 12, 26, 31, 33]. Therefore, the MAVM is employed to weigh all attributes according to the PSs provided. Then the value of alternative x_i is determined by V_i:

$$V_i = \sum_{j=1}^{m} w_j v_{ij} \tag{3}$$

On the other hand, the evaluation information provided to the decision maker can be in the form of a collection of scale ratings capturing the quality of alternatives associated with raters (i.e., attributes). Then, the weighted average is used to derive the overall rating score of alternative x_i in the RS $[0, N]$ by T_i:

$$T_i = \sum_{j=1}^{m} w_j t_{ij} \tag{4}$$

Next, we introduce the concepts of the ranking and rating as follows:

(i) Ranking. Let $(V_1, V_2, \dots, V_n)$ be the values of alternatives using MAVM, from which we can obtain the ranking of alternatives: $Ranking_{MAVM}(x_i) = j$ if V_i is the jth largest element in $(V_1, V_2, \dots, V_n)$. For example, suppose that $(0.2,0.5,0.3)$ are the values of three alternatives using MAVM, then $Ranking_{MAVM}(x_2) = 1$. Similarly, let $(T_1, T_2, \dots, T_n)$ be the overall rating scores of alternatives using the RS $[0, N]$, from which the ranking of alternatives can be derived: $Ranking_{RS}(x_i) = j$ if T_i is the jth largest element in $(T_1, T_2, \dots, T_n)$.

Based on the ranking of alternatives, we can easily have the following results:

(1) $V_i > V_j \Leftrightarrow Ranking_{MAVM}(x_i) < Ranking_{MAVM}(x_j)$.

(2) $V_i = V_j \Leftrightarrow Ranking_{MAVM}(x_i) = Ranking_{MAVM}(x_j)$.

(3) $T_i > T_j \Leftrightarrow Ranking_{RS}(x_i) < Ranking_{RS}(x_j)$.

(4) $T_i = T_j \Leftrightarrow Ranking_{RS}(x_i) = Ranking_{RS}(x_j)$.

(ii) Rating. When considering MAVM, similar to Eqs. (1) and (2), we can transform $(V_1, V_2, \dots, V_n)$ into ratings in the RS $[0, N]$ by $Rating_{MAVM}(x_i) = rounding(N \times V_i)$. For example, let $(V_1, V_2, V_3) = (0.6,0.4,0.7)$ be the values of three alternatives using MAVM, and then $Rating_{MAVM}(x_3) = rounding(5 \times 0.7) = 4$ in the RS $[0,5]$. When considering the RS $[0, N]$, we can obtain the ratings of alternatives by $Rating_{RS}(x_i) = rounding(T_i)$.

Based on the ratings of alternatives, we can easily have the following results:

(1) $V_i > V_j \Rightarrow Rating_{MAVM}(x_i) \geq Rating_{MAVM}(x_j)$.

(2) $V_i = V_j \Rightarrow Rating_{MAVM}(x_i) = Rating_{MAVM}(x_j)$.

(3) $Rating_{MAVM}(x_i) > Rating_{MAVM}(x_j) \Rightarrow V_i > V_j$.

(4) $T_i > T_j \Rightarrow Rating_{RS}(x_i) \geq Rating_{RS}(x_j)$.

(5) $T_i = T_j \Rightarrow Rating_{RS}(x_i) = Rating_{RS}(x_j)$.

(6) $Rating_{RS}(x_i) > Rating_{RS}(x_j) \Rightarrow T_i > T_j$.

As mentioned above, in practical decision making problems, the inconsistency between RS and MAVM may occur. We use Example 1 to illustrate this issue.

Example 1: A decision maker is going to make a choice among two alternatives x_1 and x_2. The PSs of x_1 and x_2 over three attributes are $(v_{11}, v_{12}, v_{13}) = (0.50, 0.65, 0.45)$ and $(v_{21}, v_{22}, v_{23}) = (0.44, 0.74, 0.43)$, respectively. Let the RS be $[0,10]$, and let the weights of all attributes be equal. Then, the inconsistency of decision outcomes between RS and MAVM will arise:

(i) Considering the ranking level. $Ranking_{MAVM}(x_1) > Ranking_{MAVM}(x_2)$ because $V_1 = (0.50 + 0.65 + 0.45)/3 = 0.53$ and $V_2 = (0.44 + 0.74 + 0.43)/3 = 0.54$. In the RS $[0,10]$ the ratings of two alternatives over three attributes are $(t_{11}, t_{12}, t_{13}) = (5,7,5)$ and $(t_{21}, t_{22}, t_{23}) = (4,7,4)$, respectively. As a result, $Ranking_{RS}(x_1) < Ranking_{RS}(x_2)$ because $T_1 = (5 + 7 + 5)/3 = 5.7$ and $T_2 = (4 + 7 + 4)/3 = 5$.

(ii) Considering the rating level. $Rating_{MAVM}(x_1) = Rating_{MAVM}(x_2)$ because $Rating_{MAVM}(x_1) = rounding(10 \times 0.53) = 5$ and $Rating_{MAVM}(x_2) = rounding(10 \times 0.54) = 5$; while $Rating_{RS}(x_1) > Rating_{RS}(x_2)$ because $Rating_{RS}(x_1) = rounding(5.7) = 6$ and $Rating_{RS}(x_2) = rounding(5) = 5$.

The MAVM is a foundational and well-justified theory to evaluate alternatives, and thus we argue that MAVM should be used as a benchmark to study the reliability of RSs. If an RS is reliable, the decision outcomes between RS and MAVM should be consistent. Therefore, in order to analyze the reliability of RSs, we study the consistency issue between RS and MAVM at the ranking and rating levels. The consistency between RS and MAVM at the ranking and rating levels are formally defined in Definitions 2 and 3.

Definition 2: An RS is consistent with MAVM at the ranking level if the rankings of any two alternatives x_i and x_j conveyed by the RS and MAVM are the same, i.e., one of the following conditions holds,

(1) $\left(Ranking_{MAVM}(x_i) > Ranking_{MAVM}(x_j)\right) \& \left(Ranking_{RS}(x_i) > Ranking_{RS}(x_j)\right)$.

(2) $\left(Ranking_{MAVM}(x_i) = Ranking_{MAVM}(x_j)\right) \& \left(Ranking_{RS}(x_i) = Ranking_{RS}(x_j)\right)$.

(3) $\left(Ranking_{MAVM}(x_i) < Ranking_{MAVM}(x_j)\right) \& \left(Ranking_{RS}(x_i) < Ranking_{RS}(x_j)\right)$.

Definition 3: An RS is consistent with MAVM at the rating level if the ratings of any two alternatives x_i and x_j conveyed by the RS and MAVM are the same, i.e., one of the following conditions holds,

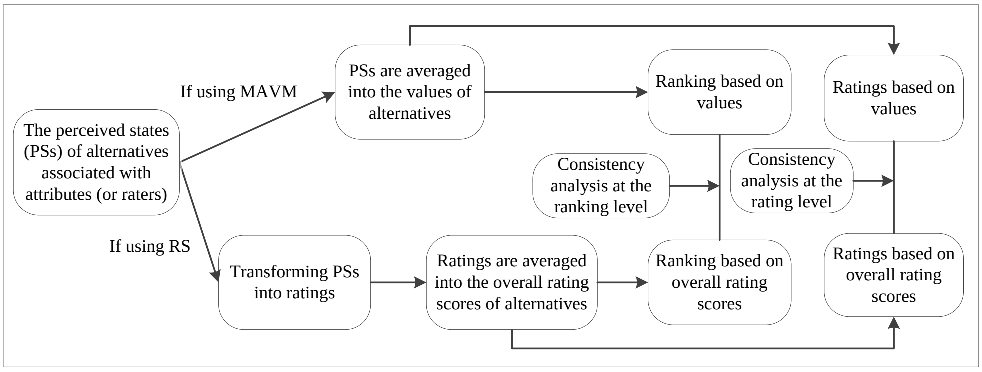
(1) $\left(Rating_{MAVM}(x_i) > Rating_{MAVM}(x_j)\right) \& \left(Rating_{RS}(x_i) > Rating_{RS}(x_j)\right)$.

(2) $\left(Rating_{MAVM}(x_i) = Rating_{MAVM}(x_j)\right) \& \left(Rating_{RS}(x_i) = Rating_{RS}(x_j)\right)$.

(3) $\left(Rating_{MAVM}(x_i) < Rating_{MAVM}(x_j)\right) \& \left(Rating_{RS}(x_i) < Rating_{RS}(x_j)\right)$.

Furthermore, the framework to analyze the consistency between RS and MAVM is presented in Fig. 1.

Fig. 1: Framework to analyze the consistency between RS and MAVM



The detailed procedures for the consistency analysis between RS and MAVM at the ranking and rating levels are presented below:

Step 1: Provide the PSs of alternatives associated with attributes (or raters).

Step 2: In the case of the MAVM, PSs are aggregated into the values of alternatives based on Eq. (3).

Step 3: In the case of the RS, PSs are transformed into ratings, and then ratings are aggregated into the overall rating scores of alternatives based on Eq. (4).

Step 4: Derive the ranking and ratings of alternatives based on the obtained values and overall rating scores, and then conduct the consistency analysis between RS and MAVM at the ranking and rating levels.

3. Consistency between RS and MAVM at the ranking and rating levels

In this section, we present the consistency conditions between RS and MAVM at the ranking and rating levels to study the reliability issue of RSs.

First, we present Lemma 1.

Lemma 1: Let $v_1, v_2 \in [0,1]$, $t_1 = rounding(N \times v_1)$ and $t_2 = rounding(N \times v_2)$, where $N \geq 1$ is an integer. Then, $v_2 - v_1 \in (\frac{t_2-t_1-1}{N}, \frac{t_2-t_1+1}{N})$.

The proof of Lemma 1 is provided in Appendix A.

Let x_r and x_s be any two alternatives in $X = \{x_1, x_2, \dots, x_n\}$, and they are evaluated on multiple attributes (or raters) $A = \{a_1, a_2, \dots, a_m\}$. Let v_{rj} and v_{sj} denote the PSs of alternatives x_r and x_s associated with attribute a_j, respectively. Let V_r and V_s be the values of alternatives x_r and x_s obtained by MAVM, i.e., $V_r = \sum_{j=1}^{m} w_j v_{rj}$ and $V_s = \sum_{j=1}^{m} w_j v_{sj}$, where $w = \{w_1, w_2, \dots, w_m\}^T$ is the weight vector of attributes. Let t_{rj} and t_{sj} denote the rating information of alternatives x_r and x_s associated with attribute a_j transformed from v_{rj} and v_{sj}. Let T_r and T_s be the overall rating scores of alternatives x_r and x_s obtained by RS, i.e., $T_r = \sum_{j=1}^{m} w_j t_{rj}$ and $T_s = \sum_{j=1}^{m} w_j t_{sj}$.

3.1 Tight bounds on N

Based on Definition 2 and Lemma 1, we present the consistency condition (a tight lower bound) about N between RS and MAVM at the ranking level (see Theorem 1). Notably, a tight lower bound means a maximum lower bound in mathematics.

Theorem 1: When using the RS $[0, N]$ to evaluate any two alternatives x_r and x_s, we have that

(1) When $V_r - V_s \neq 0$, there are two cases:

(i) for any $N \geq \frac{1}{|V_r - V_s|}$, the RS $[0, N]$ is always consistent with MAVM at the ranking level for alternatives x_r and x_s.

(ii) for any $N < \frac{1}{|V_r - V_s|}$, there exist some cases that the RS $[0, N]$ is inconsistent with MAVM at the ranking level for alternatives x_r and x_s.

(2) When $V_r - V_s = 0$, there exist some cases that the RS $[0, N]$ is inconsistent with MAVM at the ranking level for alternatives x_r and x_s for any $N \in \{1,2, \dots\}$.

The proof of Theorem 1 is provided in Appendix A.

According to Theorem 1, there exists a critical value of N, which is determined by $\frac{1}{|V_r - V_s|}$ if $V_r - V_s \neq 0$. When N is larger than the critical value, the RS $[0, N]$ is consistent with MAVM at the ranking level. Meanwhile, the critical value is tight, and for any $N < \frac{1}{|V_r - V_s|}$, there exist some cases that the RS $[0, N]$ is inconsistent with MAVM at the ranking level.

On the other hand, if $V_r - V_s = 0$, the critical value does not exist, and there exist some cases that the RS $[0, N]$ is inconsistent with MAVM at the ranking level for any $N \in \{1,2, \dots\}$.

Next, based on Definition 3 and Lemma 1, we present a tight lower bound on N, which guarantees the consistency between RS and MAVM at the rating level (see Theorem 2).

Theorem 2: When using the RS $[0, N]$ to evaluate any two alternatives x_r and x_s, we have that

(1) When $V_r - V_s \neq 0$, there are two cases:

(i) for any $N \geq \frac{2}{|V_r - V_s|}$, the RS $[0, N]$ is always consistent with MAVM at the rating level for alternatives x_r and x_s.

(ii) for any $N < \frac{2}{|V_r - V_s|}$, there exist some cases that the RS $[0, N]$ is inconsistent with MAVM at the rating level for alternatives x_r and x_s.

(2) When $V_r - V_s = 0$, there exist some cases that the RS $[0, N]$ is inconsistent with MAVM at the rating level for alternatives x_r and x_s for any $N \in \{1, 2, \dots\}$.

The proof of Theorem 2 is provided in Appendix A.

According to Theorem 2, if $V_r - V_s \neq 0$, the tight lower bound on N is $\frac{2}{|V_r - V_s|}$ at the rating level, and is twice as large as that at the ranking level.

Theorems 1 and 2 reveal that there exist two tights bounds on the parameter N determined by the value difference of alternatives (i.e., $|V_r - V_s|$): When N is larger than the bounds, there exists the consistency between RS and MAVM at the ranking and rating levels; Otherwise, inconsistent cases may occur.

3.2 Tight bounds on overall rating scores

Based on Definition 2 and Lemma 1, we present a tight lower bound on overall rating scores, which guarantees the consistency between RS and MAVM at the ranking level (see Theorem 3).

Theorem 3: When using the RS $[0, N]$ to evaluate any two alternatives x_r and x_s, we have that

(1) When $|T_r - T_s| \geq 1$, the RS $[0, N]$ is always consistent with MAVM at the ranking level

for alternatives x_r and x_s for any $N \in \{1,2,\ldots\}$.

(2) When $|T_r - T_s| < 1$, there exist some cases that the RS $[0, N]$ is inconsistent with MAVM at the ranking level for alternatives x_r and x_s for any $N \in \{1,2,\ldots\}$.

The proof of Theorem 3 is provided in Appendix A.

According to Theorem 3, there exists a critical value of 1. When $|T_r - T_s| \geq 1$, the RS $[0, N]$ is consistent with MAVM at the ranking level. Meanwhile, the critical value is tight, and when $|T_r - T_s| < 1$, there exist some cases that the RS $[0, N]$ is inconsistent with MAVM at the ranking level.

Now, based on Definition 3 and Lemma 1, we present the consistency condition about overall rating scores between RS and MAVM at the rating level (see Theorem 4).

Theorem 4: When using the RS $[0, N]$ to evaluate any two alternatives x_r and x_s, we have that

(1) When $|T_r - T_s| \geq 2$, the RS $[0, N]$ is always consistent with MAVM at the rating level for alternatives x_r and x_s for any $N \in \{1,2,\ldots\}$.

(2) When $|T_r - T_s| < 2$, there exist some cases that the RS $[0, N]$ is inconsistent with MAVM at the rating level for alternatives x_r and x_s for any $N \in \{1,2,\ldots\}$.

The proof of Theorem 4 is provided in Appendix A.

According to Theorem 4, the tight lower bound on overall rating scores is 2 at the rating level, and is twice as large as that at the ranking level.

Theorems 3 and 4 present two tights bounds determined by overall rating score difference of alternatives (i.e., $|T_r - T_s|$): There are consistent outcomes between RS and MAVM at the ranking level when $|T_r - T_s| \geq 1$, and at the rating level when $|T_r - T_s| \geq 2$.

Although the tight bounds of the parameter N and overall rating scores are both useful to present theoretical insights about the reliability of RSs, there are some key differences between them: The tight bounds on the parameter N are determined by the value difference of alternatives, and in practical RS problems we often don't know the values of alternatives, which limits its practical use. Instead, the tight bounds on overall rating scores depend directly on the rating information (not PSs), and thus provide a more practical tool to analyze the reliability of RSs. For example, an RS platform can directly judge whether the obtained RS results are reliable according to the data of T_r and T_s (see Section 4.2).

4. Simulation analysis and illustrative examples

In this section, we design simulation experiments to further explore the reliability of RSs. Moreover, we present a hypothetical example of application to illustrate the usability of the obtained results in practical RS problems.

4.1 Simulation analysis

The consistency conditions obtained analytically in Theorems 1-4 provide a basis to support the reliability analysis of RSs. But it is still unclear that:

(1) how the value difference of alternatives will influence the consistency between RS and MAVM when N is smaller than the critical values obtained in Theorems 1 and 2;

(2) how the overall rating score difference of alternatives will influence the consistency between RS and MAVM when the difference is smaller than the critical values obtained in Theorems 3 and 4.

Therefore, in this section, given a pair of alternatives (x_1 and x_2) and m attributes (raters), we use simulation experiments to further study the impact of the value difference and overall rating score difference of two alternatives on the consistency between RS and MAVM.

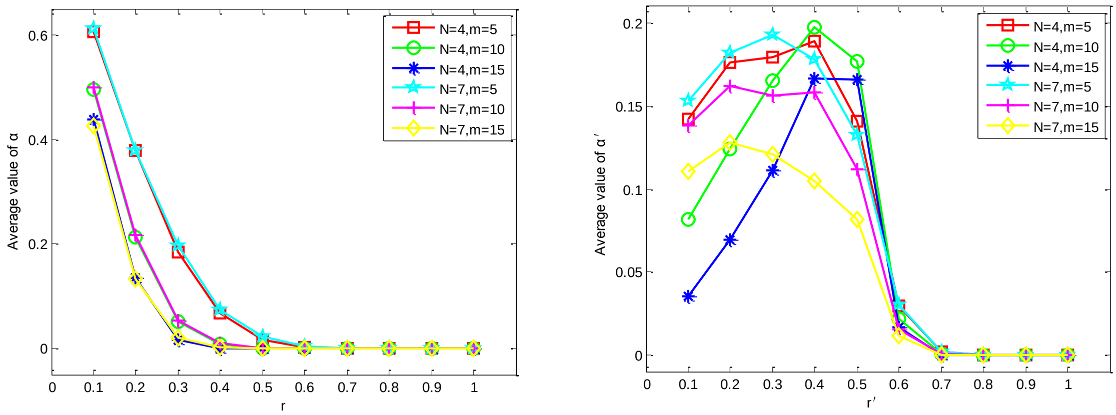
In Simulation experiment 1 we randomly generate V_1 and V_2 to satisfy $|V_1 - V_2| \in [\frac{r-0.1}{N}, \frac{r}{N})$ when setting $r \in \{0.1, 0.2, \ldots, 1\}$. Clearly, smaller r values show smaller differences between V_1 and V_2, and according to Theorem 1 the inconsistency between the RS $[0, N]$ and MAVM may occur at the ranking level for any $r \in \{0.1, 0.2, \ldots, 1\}$. Thus, in Simulation experiment 1 we study the impact of different $|V_1 - V_2|$ values on the inconsistency between RS and MAVM at the ranking level, by exploring the average inconsistency ratios under different r values.

Meanwhile, we present a revised version of Simulation experiment 1, called Simulation experiment 1', to randomly generate V_1 and V_2 to satisfy $|V_1 - V_2| \in [\frac{2(r\prime-0.1)}{N}, \frac{2r\prime}{N})$ when setting $r' \in \{0.1, 0.2, \ldots, 1\}$, and then to show the average inconsistency ratios under different r' values at the rating level.

The details of Simulation experiments 1 and 1' are provided in Appendix B.

We set different input parameters N, m, r and r', and run Simulation experiments 1 and 1' 10000 times to obtain the average values of α and α', which reflect the average inconsistency ratios of decision outcomes between RS $[0, N]$ and MAVM at the ranking and rating levels, respectively. The average values of α and α' are presented in Fig. 2.

Fig. 2: The average α and α' values in Simulation experiments 1 and 1'



(a) (b)

From Fig. 2, we find the following observations:

(i) Fig. 2 (a) and Fig. 2 (b) show that the inconsistency ratios (average α and α' values) will be close to 0 if $r > 0.5$ and $r' > 0.7$. The observation shows that there is still a quite high ratio of keeping consistency (close to 1) between RS and MAVM at the ranking and rating levels when the critical values obtained in Theorems 1 and 2 are reduced to certain degree. This finding indicates that a smaller N can keep the RS reliable in the vast majority of cases.

(ii) Fig. 2 (a) shows that increasing the values of r will result in decreasing the average α values under different N and m values. This means that a larger difference between V_1 and V_2 will lead to a smaller inconsistency ratio at the ranking level, and thus lead to higher reliability of RSs.

(iii) Fig. 2 (b) shows that a larger difference between V_1 and V_2 will lead to a smaller inconsistency ratio at the rating level, and thus lead to higher reliability of RSs when $r' \geq 0.4$. However, when $r' \leq 0.3$, Fig. 2 (b) shows lower average values of α' under different N and m values as the decrease of r'. This means that a very small difference between V_1 and V_2 will lead to a small inconsistency ratio at the rating level, and thus lead to high reliability of RSs. This observation is very different from the results at the ranking level, and can be explained as follows:

The $|V_1 - V_2|$ value is small when $r' \leq 0.3$, and thus the difference between T_1 and T_2 is small too. As a result, there is a high possibility that $Rating_{MAVM}(x_1) = Rating_{MAVM}(x_2) = Rating_{RS}(x_1) = Rating_{RS}(x_2)$, which leads to high reliability of RSs at the rating level.

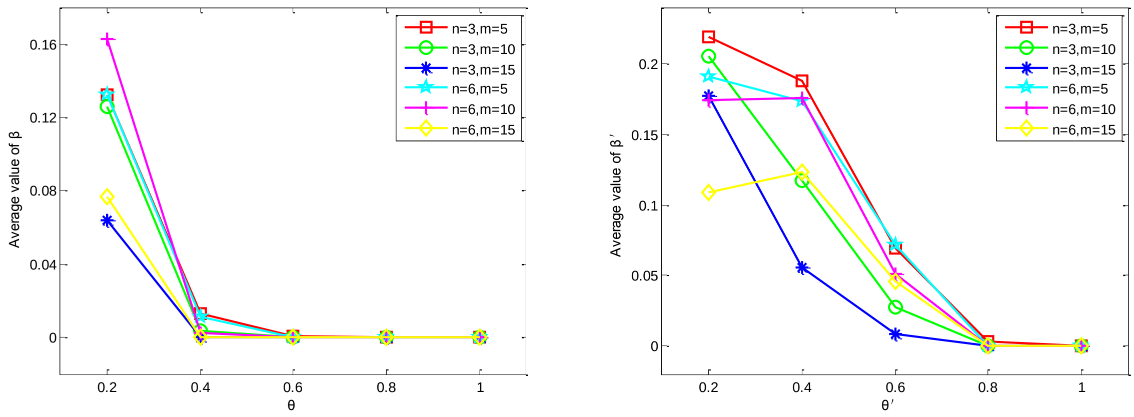
In Simulation experiment 2 we randomly generate T_1 and T_2 to satisfy $|T_1 - T_2| \in [\theta - 0.2, \theta)$ when setting $\theta \in \{0.2, 0.4, \dots, 1\}$. Clearly, smaller θ values show smaller differences between T_1 and T_2, and according to Theorem 3 the inconsistency between the RS $[0, N]$ and MAVM may occur at the ranking level for any $\theta \in \{0.2, 0.4, \dots, 1\}$. Thus, in Simulation experiment 2 we study the impact of different $|T_1 - T_2|$ values on the inconsistency between RS and MAVM at the ranking level, by exploring the average inconsistency ratios under different θ values.

Meanwhile, we present a revised version of Simulation experiment 2, called Simulation experiment 2', to randomly generate T_1 and T_2 to satisfy $|T_1 - T_2| \in 2[\theta' - 0.2, \theta')$ when setting $\theta' \in \{0.2, 0.4, \dots, 1\}$, and then to show the average inconsistency ratios under different θ' values at the rating level.

The details of Simulation experiments 2 and 2' are provided in Appendix B.

We set different input parameters N, m, θ and θ', and run Simulation experiments 2 and 2' 10000 times to obtain the average values of β and β', which reflect the average inconsistency ratios of decision outcomes between the RS $[0, N]$ and MAVM at the ranking and rating levels, respectively. The average values of β and β' are presented in Fig. 3.

Fig. 3: The average β and β' values in Simulation experiments 2 and 2′



(a) (b)

From Fig. 3, we have the following findings:

(i) Fig. 3 (a) and Fig. 3 (b) show that increasing the values of θ and θ' will result in decreasing the average β and β' values under different N and m values. This means that a larger difference between T_1 and T_2 will lead to a smaller inconsistency ratio at the ranking and rating levels, and thus lead to higher reliability of RSs.

(ii) Fig. 3 (a) shows that the inconsistency ratio will be close to 0 if $\theta \geq 0.4$. This observation shows that there is still a quite high ratio of keeping consistency (close to 1) between RS and MAVM at the ranking level when $|T_1 - T_2| \geq 0.4$. It indicates we can use a smaller critical value (e.g., 0.4) to guarantee the consistency between RS and MAVM at the ranking level. Similarly, Fig. 3 (b) shows that we can use a smaller critical value (e.g., 1.6) to guarantee the reliability of RSs at the rating level in the majority of cases.

As complementary to Theorems 1-4, the results in Simulation experiments 1, 1', 2 and 2' can further provide useful suggestions to improve the decision making reliability in RSs.

(1) Quite high reliability can still be kept at the ranking and rating levels when the obtained critical values in Theorems 1-4 are reduced to certain degree.

(2) The values and overall rating scores of alternatives play important roles in the reliability of

RSs. A large value difference or overall rating score difference of alternatives will lead to a low inconsistency ratio at the ranking level, and thus lead to higher reliability of RSs. When considering the rating level, a middle value difference of alternatives will result in a high inconsistency ratio, while a large overall rating score difference of alternatives will lead to a low inconsistency ratio.

4.2 Hypothetical application

In the following, we present a hypothetical example of application to illustrate the usability of the theoretical and simulation results. Suppose there are two cafes (x_1 and x_2) in a university campus. A team of marketing students conducts a consumer preference survey, and consumers (University staff and students) are invited to evaluate these two cafes from five attributes: Location (a_1), Product quality (a_2), Atmosphere (a_3), Waiting time (a_4) and Space available (a_5). In the example the weights of the attributes are equal. Here we only illustrate the usability of Theorems 1 and 3, and the cases of Theorems 2 and 4 are similar.

In Case A, we illustrate the results of Theorem 1 and Simulation experiment 1.

Case A: Consumers are asked to provide their PSs of the two cafes over the five attributes, and the PSs are presented in Table 1.

Table 1 PSs and the tight bounds of N at the ranking level

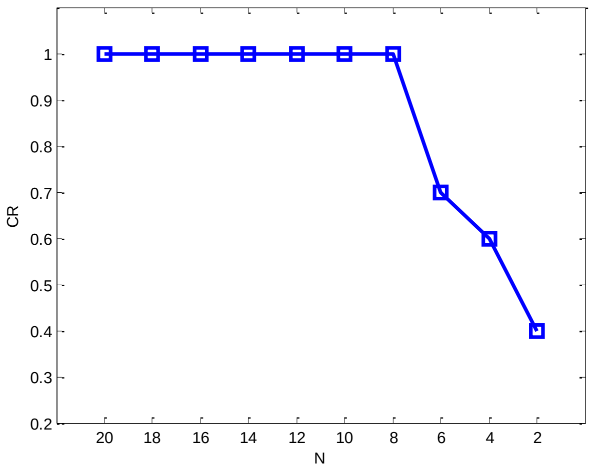
	x_1					x_2					Tight bounds at the ranking level (N)
	a_1	a_2	a_3	a_4	a_5	a_1	a_2	a_3	a_4	a_5	
Consumer 1	0.85	0.86	0.85	0.87	0.78	0.63	0.67	0.99	0.73	0.61	8.62
Consumer 2	0.70	0.86	0.84	0.73	0.90	0.76	0.95	0.79	0.64	0.64	20
Consumer 3	0.86	0.88	0.61	1.00	0.74	0.65	0.85	0.92	0.77	0.62	17.86
Consumer 4	0.99	0.87	0.77	0.79	0.77	0.92	0.94	0.88	0.93	0.95	11.63
Consumer 5	0.76	0.60	0.75	0.83	0.64	0.91	0.63	0.66	0.91	0.73	19.23
Consumer 6	0.85	0.93	0.70	0.91	0.92	0.82	0.75	0.84	0.82	0.72	13.89
Consumer 7	0.83	0.64	0.85	0.77	0.76	0.92	0.60	0.83	0.99	1.00	10.20
Consumer 8	0.98	0.82	0.87	0.92	0.86	0.63	0.93	0.88	0.71	0.71	8.47

Consumer 9	0.62	0.99	0.92	0.98	0.80	0.88	0.64	0.64	0.73	0.88	9.26
Consumer 10	0.87	0.92	0.86	0.92	0.88	0.68	0.65	0.71	0.73	0.76	5.43

It is assumed that the consumers will provide their rating information through the transformation of PSs (i.e., Eqs. (1) and (2)) based on the rationality principle. Then, based on Theorem 1 we can obtain the tight bounds of N (see Table 1) to check the consistency between MAVM and RS at the ranking level, which measures the reliability of the RSs used by consumers. In the example, when the consumers use the RS with $N \geq 20$, the 10 consumers' evaluation results about the ranking of the two cafes are all reliable in the sense of MAVM.

Further, let CR be the proportion of the consumers who have consistent evaluations between MAVM and RS at the ranking level. Clearly, $CR \in [0,1]$. The larger the value of CR, the higher proportion of consumers whose evaluation results are reliable. If $CR = 1$, all consumers' evaluation results are reliable. The CR values under different parameter N are present in Fig. 4.

Fig. 4. The CR values under different parameter N



According to Fig. 4, a high consistency proportion can still be kept when $N = 8$ even though most of the tight bounds on N are larger than 8 (See Table 1). This result indicates that there still exists high reliability in most cases when the tight bound on N at the ranking level is obviously reduced, which coincides with Simulation experiment 1.

Further, we use case B to illustrate the results of Theorem 3 and Simulation experiment 2.

Case B: In practical RS problems, we often don't know people's PSs, and Theorem 3 provides a useful tool to determine whether the consistency exists between MAVM and RS at the ranking level (i.e., the reliability of the RS can be guaranteed). Suppose ten consumers are asked to rate the two cafes using the RS $[0,6]$. Table 2 shows the rating information and the corresponding overall rating score differences.

Table 2 The ratings with the RS $[0,6]$ and the corresponding overall rating score differences

	x_1					x_2					$\|T_1 - T_2\|$	Consistency at the ranking level
	a_1	a_2	a_3	a_4	a_5	a_1	a_2	a_3	a_4	a_5		
Consumer 1	5	5	5	5	5	4	4	6	4	4	0.6	Yes
Consumer 2	4	5	5	4	5	5	6	5	4	4	0.2	No
Consumer 3	5	5	4	6	4	4	5	6	5	4	0	No
Consumer 4	6	5	5	5	5	6	6	5	6	6	0.6	Yes
Consumer 5	5	4	5	5	4	5	4	4	5	4	0.2	No
Consumer 6	5	6	4	5	6	5	5	5	5	4	0.4	Yes
Consumer 7	5	4	5	5	5	6	4	5	6	6	0.6	Yes
Consumer 8	6	5	5	6	5	4	6	5	4	4	0.8	Yes
Consumer 9	4	6	6	6	5	5	4	4	4	5	1	Yes
Consumer 10	5	6	5	6	5	4	4	4	4	5	1.2	Yes

Theorem 3 shows that $|T_1 - T_2| \geq 1$ can guarantee the consistency between MAVM and RS at the ranking level, and Simulation 2 indicates that there is still a high possibility to keep the consistency when $|T_1 - T_2| \geq 0.4$. The results in Table 2 coincide with Theorem 3 and Simulation 2. In summary, both the theoretical results and simulation results are validated in the hypothetical application.

5. Conclusion

RSs are widely used to obtain the ranking or ratings of alternatives in decision making, and one of the most challenging issues in RS problems is reliability because different RSs may lead to inconsistent decision outcomes. In this paper, we use MAVM as a benchmark to study the reliability of the RS $[0, N]$. Different from the existing research to investigate the impact of people's psychological and behavioral factors on the reliability of RSs, in this paper we analytically reveal the mechanism from the view of the rounding process to guarantee the reliability of RSs at the ranking and rating levels. The main contributions are concluded as follows:

(1) We find that there exist tight bounds on the parameter N, which are determined by the value difference of alternatives (see Theorems 1 and 2). When N is larger than the bounds, the consistency between RS and MAVM will hold at the ranking and rating levels; Otherwise, inconsistent cases may occur. Particularly, the tight bound on N at the rating level is twice as large as that at the ranking level.

(2) We present the tight bounds on overall rating scores (see Theorems 3 and 4): The consistency between RS and MAVM will hold when the overall rating score difference among alternatives is larger than 1 at the ranking level and larger than 2 at the rating level.

(3) We develop the detailed simulation analysis to show that the consistency between RS and MAVM at the ranking and rating levels can still be kept in most cases if the obtained tight bounds are reduced to certain degree. These simulation results further provide useful suggestions to improve the decision making reliability in RSs.

The tight bounds on the parameter N and overall rating scores both provide interesting insights about the reliability of RSs in theoretical aspects, but they behave differently in practical RS problems: We can't directly use the tight bounds of the parameter N to judge the reliability of the used RSs because we often don't know the values of alternatives (i.e., people's PSs) when using an RS. However, the tight bounds on overall rating scores depend directly on the rating information (not PSs), and thus provide more practical tools to analyze the reliability of RSs.

There are two limitations in this paper: (1) strategic behaviors are common in decision making activities [15, 23, 40, 43], but we assume that the decision makers honestly express their preferences using RSs; (2) the rating scores mean different things (the personalized individual semantics [51]) to different decision makers, but we don't consider this issue in RSs. Therefore, it will be interesting to study the reliability of RSs in a decision context with strategic behaviors and personalized individual semantics.

Acknowledgments

This work was supported by the grant (No. 71871149) from NSF of China, and the grants (Nos. sksyl201705 and 2018hhs-58 and LH2018004) from Sichuan University.

References

[1] Adomavicius, G., & Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. IEEE Intelligent Systems, 22(3), 48-55.
[2] Alwin, D. F., & Krosnick, J. A. (1985). The measurement of values in surveys: A comparison of ratings and rankings. Public Opinion Quarterly, 49(4), 535-552.
[3] Arrow, K. J. (1951). Social Choice and Individual Values, John Wiley & Sons, New York.
[4] Bargagliotti, A. E., & Li, L. (2013). Decision making using rating systems: When scale meets binary. Decision Sciences, 44(6), 1121-1137.
[5] Barron, F. H., & Barrett, B. E. (1996). Decision quality using ranked attribute weights. Management Science, 42(11), 1515-1523.
[6] Baucells, M., & Sarin, R. K. (2003). Group decisions with multiple criteria. Management Science, 49(8), 1105-1118.
[7] Bhattacharjee, R., & Goel, A. (2005). Avoiding ballot stuffing in ebay-like reputation systems. Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems, Philadelphia, PA.
[8] Chen, P. Y., Hong, Y., & Liu, Y. (2017). The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments. Management Science, 64(10), 4629-4647.
[9] Churchill Jr, G. A., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. Journal of Marketing Research, 21(4), 360-375.
[10] Cleveland, J. N., & Murphy, K. R. (1992). Analyzing performance appraisal as goal-directed

behavior. Research in Personnel and Human Resources Management, 10(2), 121-185.
[11] Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. Econometrica, 50, 1431-1451.
[12] Danielson, M., Ekenberg, L., & He, Y. (2014). Augmenting ordinal methods of attribute weight approximation. Decision Analysis, 11(1), 21-26.
[13] Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. International Journal of Market Research, 50(1), 61-104.
[14] Dong, Y., Liu, Y., Liang, H., Chiclana, F., & Herrera-Viedma, E. (2018). Strategic weight manipulation in multiple attribute decision making. Omega, 75, 154-164.
[15] Dong, Y., Zhang, H., & Herrera-Viedma, E. (2016). Integrating experts' weights generated dynamically into the consensus reaching process and its applications in managing non-cooperative behaviors. Decision Support Systems, 84, 1-15.
[16] Doumpos, M., & Figueira, J. R. (2019). A multicriteria outranking approach for modeling corporate credit ratings: An application of the Electre Tri-nC method. Omega, 82, 166-180.
[17] Doumpos, M., & Zopounidis, C. (2011). A multicriteria outranking modeling approach for credit rating. Decision Sciences, 42(3), 721-742.
[18] Dyer, J. S., & Sarin, R. K. (1979). Group preference aggregation rules based on strength of preference. Management Science, 25(9), 822-832.
[19] Farquhar, P. H., & Keller, L. R. (1989). Preference intensity measurement. Annals of Operations Research, 19(1), 205-217.
[20] Friedman, H. H., & Amoo, T. (1999). Rating the rating scales. Journal of Marketing Management, 9(3), 114–123.
[21] Friedman, H., & Friedman, E. (1997). A comparison of six overall evaluation rating scales. Journal of International Marketing and Marketing Research, 22(3), 129-138.
[22] García-Lapresta, J. L., & Pérez-Román, D. (2017). A consensus reaching process in the context of non-uniform ordered qualitative scales. Fuzzy Optimization and Decision Making, 16(4), 449-461.
[23] Gibbard, A. (1973). Manipulation of voting schemes: a general result. Econometrica, 41, 587-601.
[24] Greco, S., Kadzinski, M., Mousseau, V., & Slowinski, R. (2012). Robust ordinal regression for multiple criteria group decision: UTAGMS-GROUP and UTADISGMS-GROUP. Decision Support Systems, 52(3), 549-561.
[25] Green, P. E., & Rao, V. R. (1970). Rating scales and information recovery. How many scales and response categories to use? The Journal of Marketing, 34, 33-39.
[26] Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. Journal of Political Economy, 63(4), 309-321.
[27] Hauser, J. (1978). Consumer preference axioms: Behavioral postulates for describing and predicting stochastic choice. Management Science, 24(13), 1331-1341.
[28] Herrera, F., & Martínez, L. (2001). A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision-making. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 31(2), 227-234.
[29] Hochbaum, D. S., & Levin, A. (2006). Methodologies and algorithms for group-rankings decision. Management Science, 52(9), 1394-1408.
[30] Kadzinski, M., Ghaderi, M., Wąsikowski, J., & Agell, N. (2017). Expressiveness and robustness measures for the evaluation of an additive value function in multiple criteria

preference disaggregation methods: An experimental analysis. Computers & Operations Research, 87(87), 146-164.
[31] Keeney, R. L. & Raiffa, H. (1976). Decision with Multiple Objectives: Preferences and Value Tradeoffs, John Wiley & Son, New York.
[32] Keeney, R. L. (1976). A group preference axiomatization with cardinal utility. Management Science, 23(2), 140-145.
[33] Keeney, R. L., & Kirkwood, C. W. (1975). Group decision making using cardinal social welfare functions. Management Science, 22(4), 430-437.
[34] Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87(1), 72-107.
[35] Lehmann, D. R., & Hulbert, J. (1972). Are three-point scales always good enough? Journal of Marketing Research, 9(4), 444-446.
[36] Liu, J., Liao, X., Zhao, W., & Yang, N. (2016). A classification approach based on the outranking model for multiple criteria ABC analysis. Omega, 61, 19-34.
[37] Maccheroni, F., Marinacci, M., & Rustichini, A. (2012). Social decision theory: Choosing within and between groups. The Review of Economic Studies, 79(4), 1591-1636.
[38] Maldonado, S., Bravo, C., López, J., & Pérez, J. (2017). Integrated framework for profit-based feature selection and SVM classification in credit scoring. Decision Support Systems, 104, 113-121.
[39] McCarty, J. A., & Shrum, L. J. (2000). The measurement of personal values in survey research: A test of alternative rating procedures. Public Opinion Quarterly, 64(3), 271-298.
[40] Morgan, J., & Stocken, P. C. (2008). Information aggregation in polls. American Economic Review, 98(3), 864-896.
[41] Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. Journal of Applied Psychology, 89(1), 158-164.
[42] Preston, C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychologica, 104(1), 1-15.
[43] Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. Journal of Economic Theory, 10(2), 187-217.
[44] Schwarz, N., Knäuper, B., Hippler, H. J., Noelle-Neumann, E., & Clark, L. (1991). Rating scales numeric values may change the meaning of scale labels. Public Opinion Quarterly, 55(4), 570-582.
[45] Sun, X., Han, M., & Feng, J. (2019). Helpfulness of online reviews: Examining review informativeness and classification thresholds by search products and experience products. Decision Support Systems, 124, 113099.
[46] Tüselmann, H., Sinkovics, R. R., & Pishchulov, G. (2015). Towards a consolidation of worldwide journal rankings–a classification using random forests and aggregate rating via data envelopment analysis. Omega, 51, 11-23.
[47] Van Gestel, T., Baesens, B., Van Dijcke, P., Garcia, J., Suykens, J. A., & Vanthienen, J. (2006). A process model to develop an internal rating system: Sovereign credit ratings. Decision Support Systems, 42(2), 1131-1151.
[48] Von Winterfeldt, D. & Edwards, W. (1986). Decision Analysis and Behavioral Research, Cambridge: Cambridge University Press.
[49] Waung, M., & Highhouse, S. (1997). Fear of conflict and empathic buffering: Two explanations for the inflation of performance feedback. Organizational Behavior and Human

Decision Processes, 71(1), 37-54.
[50] Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. International Journal of Research in Marketing, 27(3), 236-247.
[51] Zhang, H., Dong, Y., Xiao, J., Chiclana, F., & Herrera-Viedma, E. (2020). Personalized individual semantics-based approach for linguistic failure modes and effects analysis with incomplete preference information. IISE Transactions, in press, doi: 10.1080/24725854.2020.1731774.
[52] Zhang, L., Mistry, K., Lim, C. P., & Neoh, S. C. (2018). Feature selection using firefly optimization for classification and regression models. Decision Support Systems, 106, 64-85.

Appendix A: Proofs

The proof of Lemma 1:

Let $v_1 \in [0,1]$ and $v_2 \in [0,1]$. Transform them into the interval $[0, N]$ via Eqs. (1) and (2), we have $t_1 = rounding(N \times v_1)$ and $t_2 = rounding(N \times v_2)$. Then, we have:

$$v_2 - v_1 \in \begin{cases} \left(\frac{t_2-t_1-1}{N}, \frac{t_2-t_1+1}{N}\right), & t_2 - t_1 \geq 2 \\ \left(0, \frac{2}{N}\right), & t_2 - t_1 = 1 \\ \left(-\frac{1}{N}, \frac{1}{N}\right), & t_2 - t_1 = 0 \\ \left(-\frac{2}{N}, 0\right), & t_2 - t_1 = -1 \\ \left(\frac{t_2-t_1-1}{N}, \frac{t_2-t_1+1}{N}\right), & t_2 - t_1 \leq -2. \end{cases} \tag{5}$$

Based on Eq. (5), we have $v_2 - v_1 \in \left(\frac{t_2-t_1-1}{N}, \frac{t_2-t_1+1}{N}\right)$.

This completes the proof of Lemma 1.

The proof of Theorem 1:

(1) We prove (i) based on reduction to absurdity as follows.

We assume $V_r > V_s$, i.e., $Ranking_{MAVM}(x_r) < Ranking_{MAVM}(x_s)$. Based on Lemma 1, we have:

$$\begin{aligned} V_r - V_s &= w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm}) \\ &< \frac{1}{N}\big(w_1(t_{r1} - t_{s1} + 1) + \cdots + w_i(t_{ri} - t_{si} + 1) + \cdots + w_m(t_{rm} - t_{sm} + 1)\big) \\ &= \frac{1}{N}\big(w_1(t_{r1} - t_{s1}) + \cdots + w_i(t_{ri} - t_{si}) + \cdots + w_m(t_{rm} - t_{sm}) \\ &\qquad + (w_1 + \cdots + w_i + \cdots + w_m)\big) \\ &= \frac{1}{N}\big(T_r - T_s + (w_1 + \cdots + w_i + \cdots + w_m)\big). \end{aligned} \tag{6}$$

If the inconsistency occurs at the ranking level, we have $Ranking_{RS}(x_r) \geq Ranking_{RS}(x_s)$, i.e., $T_r \leq T_s$ according to Definition 2. As a result, based on (6) we have

$$V_r - V_s < \frac{1}{N}(w_1 + \cdots + w_i + \cdots + w_m) = \frac{1}{N}. \tag{7}$$

Then, we have that $0 < V_r - V_s < \frac{1}{N}$, i.e., $N < \frac{1}{V_r - V_s}$, which contradicts $N \geq \frac{1}{|V_r - V_s|}$. Notably, we can prove (i) in the case of $V_r < V_s$ in the same way.

So, the RS $[0, N]$ is always consistent with MAVM at the ranking level for alternatives x_r and x_s if $N \geq \frac{1}{|V_r - V_s|}(V_r - V_s \neq 0)$.

This completes the proof of (i).

Next we prove (ii). We assume $V_r - V_s > 0$, i.e., $Ranking_{MAVM}(x_r) < Ranking_{MAVM}(x_s)$. Because $N < \frac{1}{|V_r - V_s|}$, we have that

$$N < \frac{1}{V_r - V_s} = \frac{1}{w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm})}. \tag{8}$$

Then, for any $N < \frac{1}{w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm})}$, there exist $v_{ri}, v_{si} (i = 1,2, \dots, m)$ satisfying Eq. (9).

$$\begin{cases} v_{ri} - v_{si} = \frac{1}{2N} \\ rounding(N \times v_{ri}) = rounding(N \times v_{si}). \end{cases} \tag{9}$$

Thus, we have $t_{ri} - t_{si} = rounding(N \times v_{ri}) - rounding(N \times v_{si}) = 0 \ (i = 1,2, \dots, m)$ and $T_r = T_s$, i.e., $Ranking_{RS}(x_r) = Ranking_{RS}(x_s)$. As a result, the inconsistency between the RS $[0, N]$ and MAVM occurs at the ranking level accoding to Definition 2. Notably, we can prove (ii) in the case of $V_r < V_s$ in the same way.

Therefore, for any $N < \frac{1}{|V_r - V_s|}(V_r - V_s \neq 0)$, there exist some inconsistency cases between the RS $[0, N]$ and MAVM at the ranking level for alternatives x_r and x_s.

This completes the proof of (ii).

(2) When $V_r - V_s = 0$, i.e., $Ranking_{MAVM}(x_r) = Ranking_{MAVM}(x_s)$, we have:

$$V_r - V_s = w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm}) = 0. \tag{10}$$

For any $N \in \{1,2, \dots\}$, we set $w_i = \frac{1}{m}(i = 1,2, \dots, m)$, and there exist $v_{ri}, v_{si} (i = 1,2, \dots, m)$ satisfying Eq. (11).

$$\begin{cases} (v_{ri} - v_{si}) \to 0, (t_{ri} - t_{si}) = -1, i = 1,2, \dots, m-1 \\ (v_{rm} - v_{sm}) \to 0, (t_{rm} - t_{sm}) = 0 \\ w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm}) = 0. \end{cases} \tag{11}$$

Then, in this case $T_r < T_s$, i.e., $Ranking_{RS}(x_r) > Ranking_{RS}(x_s)$. As a result, the inconsistency between the RS $[0, N]$ and MAVM occurs at the ranking level accoding to Definition 2.

So, for any $N \in \{1,2,\dots\}$, there exist some inconsistency cases between the RS $[0, N]$ and MAVM at the ranking level for alternatives x_r and x_s when $V_r - V_s = 0$.

This completes the proof of (2).

The proof of Theorem 2:

(1) We prove (i) based on reduction to absurdity as follows.

We assume $V_r > V_s$, i.e., $Rating_{MAVM}(x_r) \geq Rating_{MAVM}(x_s)$. Based on Lemma 1, we have Eq. (6), i.e., $V_r - V_s < \frac{1}{N}(T_r - T_s + w_1 + \cdots + w_i + \cdots + w_m)$.

If the inconsistency occurs at the rating level, we have

(a) $Rating_{MAVM}(x_r) > Rating_{MAVM}(x_s)$ and $Rating_{RS}(x_r) < Rating_{RS}(x_s)(T_r < T_s)$. As a result, based on Eq. (6) we have

$$V_r - V_s < \frac{1}{N}(w_1 + \cdots + w_i + \cdots + w_m) = \frac{1}{N}. \tag{12}$$

(b) $Rating_{MAVM}(x_r) > Rating_{MAVM}(x_s)$ and $Rating_{RS}(x_r) = Rating_{RS}(x_s)(|T_r - T_s| < 1)$. As a result, based on Eq. (6) we have

$$V_r - V_s < \frac{1}{N}(1 + w_1 + \cdots + w_i + \cdots + w_m) = \frac{2}{N}. \tag{13}$$

(c) $Rating_{MAVM}(x_r) = Rating_{MAVM}(x_s)(|V_r - V_s| < \frac{1}{N})$, then we have that

$$N < \frac{1}{|V_r - V_s|} = \frac{1}{|w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm})|}. \tag{14}$$

For any $N < \frac{1}{|w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm})|}$, there exist $v_{ri}, v_{si} (i = 1,2, \dots, m)$ satisfying Eq. (15).

$$\begin{cases} v_{ri} - v_{si} = \frac{1}{2N} \\ rounding(N \times v_{ri}) - rounding(N \times v_{si}) = 1. \end{cases} \tag{15}$$

Thus, we have $t_{ri} - t_{si} = rounding(N \times v_{ri}) - rounding(N \times v_{si}) = 1$ $(i = 1,2, \dots, m)$ and $T_r - T_s = 1$, i.e., $Rating_{RS}(x_r) > Rating_{RS}(x_s)$. As a result, the inconsistency between the RS $[0, N]$ and MAVM occurs at the rating level accoding to Definition 3.

Notably, we can prove (i) in the case of $V_r < V_s$ in the same way.

So, the RS $[0, N]$ is consistent with MAVM at the rating level for alternatives x_r and x_s if

$N \geq \frac{2}{|V_r - V_s|} (V_r - V_s \neq 0)$.

This completes the proof of (i).

Next, we prove (ii). We assume $V_r - V_s > 0$, i.e., $Rating_{MAVM}(x_r) \geq Rating_{MAVM}(x_s)$. Because $N < \frac{2}{|V_r - V_s|}$, we have that

$$N < \frac{2}{V_r - V_s} = \frac{2}{w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm})}. \tag{16}$$

Then, for any $N < \frac{2}{w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm})}$, we set $w_i = \frac{1}{m} (i = 1,2, \dots, m)$, and there exist $v_{ri}, v_{si} (i = 1,2, \dots, m)$ satisfying Eq. (17).

$$\begin{cases} (v_{ri} - v_{si}) \to 0, (t_{ri} - t_{si}) = -1, i = 1,2, \dots, m-1 \\ (v_{rm} - v_{sm}) \to \frac{1}{2N}, (t_{rm} - t_{sm}) = 0 \\ w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm}) > 0. \end{cases} \tag{17}$$

Then, in this case $T_r < T_s$, i.e., $Rating_{RS}(x_r) \leq Rating_{RS}(x_s)$. As a result, the inconsistency between the RS $[0, N]$ and MAVM occurs at the rating level accoding to Definition 3. Notably, we can prove (ii) in the case of $V_r < V_s$ in the same way.

Therefore, for any $N < \frac{2}{|V_r - V_s|} (V_r - V_s \neq 0)$, there exist some inconsistency cases between the RS $[0, N]$ and MAVM at the rating level for alternatives x_r and x_s.

This completes the proof of (ii).

(2) When $V_r - V_s = 0$, i.e., $Rating_{MAVM}(x_r) = Rating_{MAVM}(x_s)$, we have:

$$V_r - V_s = w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm}) = 0. \tag{18}$$

For any $N \in \{1,2, \dots\}$, we set $w_i = \frac{1}{m} (i = 1,2, \dots, m)$, and there exist $v_{ri}, v_{si} (i = 1,2, \dots, m)$ satisfying Eq. (19).

$$\begin{cases} (v_{ri} - v_{si}) \to 0, (t_{ri} - t_{si}) = -1, i = 1,2, \dots, m-1 \\ (v_{rm} - v_{sm}) \to 0, (t_{rm} - t_{sm}) = 0 \\ w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm}) = 0. \end{cases} \tag{19}$$

Then, in this case $T_r < T_s$, i.e., $Rating_{RS}(x_r) \leq Rating_{RS}(x_s)$. As a result, the inconsistency between the RS $[0, N]$ and MAVM occurs at the rating level accoding to Definition 3.

So, for any $N \in \{1,2, \dots\}$, there exist some inconsistency cases between the RS $[0, N]$ and MAVM at the rating level for alternatives x_r and x_s when $V_r - V_s = 0$.

This completes the proof of (2).

The proof of Theorem 3:

(1) According to Eq. (6), we have $V_r - V_s < \frac{T_r - T_s + 1}{N}$. Similarly, we can obtain that $V_r - V_s > \frac{T_r - T_s - 1}{N}$, i.e.,

$$\frac{T_r - T_s - 1}{N} < V_r - V_s < \frac{T_r - T_s + 1}{N}. \tag{20}$$

We first consider the case of $T_r - T_s \geq 1$. In this case, $Ranking_{RS}(x_r) < Ranking_{RS}(x_s)$. Based on Eq. (20), we have $V_r - V_s > 0$, i.e., $Ranking_{MAVM}(x_r) < Ranking_{MAVM}(x_s)$. Thus, for any $N \in \{1,2,\ldots\}$ the RS $[0, N]$ is consistent with MAVM at the ranking level for alternatives x_r and x_s. Notably, we can prove (1) in the case of $T_r - T_s \leq -1$ in the same way.

This completes the proof of (1).

(2) We assume $0 \leq T_r - T_s < 1$, i.e., $Ranking_{RS}(x_r) \leq Ranking_{RS}(x_s)$, then:

$$0 \leq T_r - T_s = w_1(t_{r1} - t_{s1}) + \cdots + w_i(t_{ri} - t_{si}) + \cdots + w_m(t_{rm} - t_{sm}) < 1. \tag{21}$$

For any $N \in \{1,2,\ldots\}$, we set $w_i = \frac{1}{m}(i = 1,2,\ldots,m)$, and there exist $v_{ri}, v_{si}(i = 1,2,\ldots,m)$ satisfying Eq. (22).

$$\begin{cases} (v_{ri} - v_{si}) \to 0, (t_{ri} - t_{si}) = 0, i = 1,2,\ldots,m-1 \\ (v_{rm} - v_{sm}) \to 0, (t_{rm} - t_{sm}) = 0 \; or \; 1 \\ w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm}) < 0 \\ 0 \leq w_1(t_{r1} - t_{s1}) + \cdots + w_i(t_{ri} - t_{si}) + \cdots + w_m(t_{rm} - t_{sm}) < 1. \end{cases} \tag{22}$$

Thus, in this case $V_r < V_s$, i.e., $Ranking_{MAVM}(x_r) > Ranking_{MAVM}(x_s)$. As a result, the inconsistency between the RS $[0, N]$ and MAVM occurs at the ranking level accoding to Definition 2. Notably, we can prove (2) in the case of $-1 < T_r - T_s \leq 0$ in the same way.

This completes the proof of (2).

The proof of Theorem 4:

(1) According to Eq. (20), we have

$$\frac{T_r - T_s - 1}{N} < V_r - V_s < \frac{T_r - T_s + 1}{N}. \tag{23}$$

We first consider the case of $T_r - T_s \geq 2$. In this case, $Rating_{RS}(x_r) > Rating_{RS}(x_s)$. Based on Eq. (23), we have $V_r - V_s > \frac{1}{N}$, i.e., $Rating_{MAVM}(x_r) > Rating_{MAVM}(x_s)$. Thus, for any $N \in \{1,2,\ldots\}$ the RS $[0, N]$ is consistent with MAVM at the rating level for alternatives x_r and x_s. Notably, we can prove (1) in the case of $T_r - T_s \leq -2$ in the same way.

This completes the proof of (1).

(2) We assume $0 \leq T_r - T_s < 2$, i.e., $Rating_{MAVM}(x_r) \geq Rating_{MAVM}(x_s)$. Then:

$$0 \leq T_r - T_s = w_1(t_{r1} - t_{s1}) + \cdots + w_i(t_{ri} - t_{si}) + \cdots + w_m(t_{rm} - t_{sm}) < 2. \tag{24}$$

For any $N \in \{1,2,\ldots\}$, we set $w_i = \frac{1}{m}(i = 1,2,\ldots,m)$, and there exist $v_{ri}, v_{si}(i = 1,2,\ldots,m)$ satisfying Eq. (25).

$$\begin{cases} (v_{ri} - v_{si}) \to -\frac{1}{N}, (t_{ri} - t_{si}) = 0, i = 1,2,\ldots,m-1 \\ (v_{rm} - v_{sm}) \to 0, (t_{rm} - t_{sm}) = 0 \; or \; 1 \\ w_1(v_{r1} - v_{s1}) + \cdots + w_i(v_{ri} - v_{si}) + \cdots + w_m(v_{rm} - v_{sm}) < 0 \\ 0 \leq w_1(t_{r1} - t_{s1}) + \cdots + w_i(t_{ri} - t_{si}) + \cdots + w_m(t_{rm} - t_{sm}) < 2. \end{cases} \tag{25}$$

Thus, in this case, $V_r < V_s$ i.e., $Rating_{MAVM}(x_r) \leq Rating_{MAVM}(x_s)$. As a result, the inconsistency between the RS $[0, N]$ and MAVM occurs at the rating level accoding to Definition 3. Notably, we can prove (2) in the case of $-2 < T_r - T_s \leq 0$ in the same way.

This completes the proof of (2).

Appendix B: Simulations

Simulation experiment 1

Input: N, m and r.

Output: α.

Step 1: Uniformly and randomly generate the PSs $v_{ij}(i = 1,2; j = 1,2,\ldots,m)$ for alternatives x_1 and x_2 over m attributes in the interval $[0,1]$. Let $V_1 = \frac{1}{m}\sum_{j=1}^{m} v_{1j}$ and $V_2 = \frac{1}{m}\sum_{j=1}^{m} v_{2j}$.

Step 2: If $|V_1 - V_2| \in [\frac{r-0.1}{N}, \frac{r}{N})$, go to next step; otherwise, go to Step 1.

Step 3: Transform v_{ij} into t_{ij} in the RS $[0, N]$ via Eqs. (1) and (2). Use Eq. (4) to obtain $T_1 = \frac{1}{m}\sum_{j=1}^{m} t_{1j}$ and $T_2 = \frac{1}{m}\sum_{j=1}^{m} t_{2j}$.

Step 4: Based on V_1, V_2, T_1 and T_2, the rankings of alternatives can be obtained by the RS $[0, N]$ and MAVM, respectively.

Step 5: If inconsistency occurs at the ranking level (Definition 2 is not satisfied), then $\alpha = 1$; otherwise $\alpha = 0$.

Simulation experiment 1'

Input: N, m and r'.

Output: α'.

Step 1: Same to Step 1 in **Simulation experiment 1.**

Step 2: If $|V_1 - V_2| \in [\frac{2(r'-0.1)}{N}, \frac{2r'}{N})$, go to next step; otherwise, go to Step 1.

Step 3: Same to Step 3 in **Simulation experiment 1.**

Step 4: Based on V_1, V_2, T_1 and T_2, the ratings of alternatives can be obtained by the RS $[0, N]$ and MAVM via Eqs (1) and (2), respectively.

Step 5: If inconsistency occurs at the rating level (Definition 3 is not satisfied), then $\alpha' = 1$; otherwise $\alpha' = 0$.

Simulation experiment 2

Input: N, m and θ.

Output: β.

Step 1: Uniformly and randomly generate the ratings $t_{ij}(i = 1,2; j = 1,2, \ldots, m)$ for alternatives x_1 and x_2 over m attributes in the interval $[0, N]$. Let $T_1 = \frac{1}{m}\sum_{j=1}^{m} t_{1j}$ and $T_2 = \frac{1}{m}\sum_{j=1}^{m} t_{2j}$.

Step 2: If $|T_1 - T_2| \in [\theta - 0.2, \theta)$, go to next step; otherwise, go to Step 1.

Step 3: Generate $v_{ij} \in \left[\frac{t_{ij}-0.5}{N}, \frac{t_{ij}+0.5}{N}\right)$ in the interval $[0,1]$. Use Eq. (3) to obtain $V_1 = \frac{1}{m}\sum_{j=1}^{m} v_{1j}$ and $V_2 = \frac{1}{m}\sum_{j=1}^{m} v_{2j}$.

Step 4: Based on V_1, V_2, T_1 and T_2, the rankings of alternatives can be obtained by the RS $[0, N]$ and MAVM, respectively.

Step 5: If inconsistency occurs at the ranking level (Definition 2 is not satisfied), then $\beta = 1$; otherwise $\beta = 0$.

Simulation experiment 2'

Input: N, m and θ.

Output: β'.

Step 1: Same to Step 1 in **Simulation experiment 2.**

Step 2: If $|T_1 - T_2| \in 2[\theta - 0.2, \theta)$, go to next step; otherwise, go to Step 1.

Step 3: Same to Step 3 in **Simulation experiment 2.**

Step 4: Based on V_1, V_2, T_1 and T_2, the ratings of alternatives can be obtained by the RS $[0, N]$ and MAVM via Eqs. (1) and (2), respectively.

Step 5: If inconsistency occurs at the rating level (Definition 3 is not satisfied), then $\beta' = 1$; otherwise $\beta' = 0$.