Best-Worst Method: Inconsistency, Uncertainty, Consensus, and Range Sensitivity

Liang, F.

DOI
10.4233/uuid:0fd5201c-09cd-46e1-899c-47c0d9ad1f6b

Publication date
2021

Document Version
Final published version

Citation (APA)
Liang, F. (2021). *Best-Worst Method: Inconsistency, Uncertainty, Consensus, and Range Sensitivity*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:0fd5201c-09cd-46e1-899c-47c0d9ad1f6b

Best-Worst Method

INCONSISTENCY, UNCERTAINTY, CONSENSUS AND RANGE SENSITIVITY

Fuqi Liang

Best-Worst Method: Inconsistency, Uncertainty, Consensus, and Range Sensitivity

Fuqi Liang

Delft University of Technology

Best-Worst Method: Inconsistency, Uncertainty, Consensus, and Range Sensitivity

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
woensdag, 8 december, 2021 om 10:00 uur

door

Fuqi LIANG
Master of Management Science,
Shenzhen University, China
geboren te Maoming, Guangdong, China

Dit proefschrift is goedgekeurd door de:
promotor: Dr. J. Rezaei
promotor: Dr. M. Brunelli

Samenstelling van de promotiecommissie:

Rector Magnificus	voorzitter
Dr. J. Rezaei	Technische Universiteit Delft, promotor
Dr. M. Brunelli	University of Trento, promotor

Onafhankelijke leden:

Prof.dr. B. Lev	Drexel University
Prof.dr. S. Greco	University of Catania
Prof.dr. N. Agell Jané	ESADE
Prof.mr.dr. J. A. de Bruijn	Technische Universiteit Delft
Prof.dr.ir. C. G. Chorus	Technische Universiteit Delft

This dissertation is funded by China Scholarship Council under grant 201708440305 and Delft University of Technology.

Keywords: Best-Worst Method, inconsistency, uncertainty, consensus, range sensitivity

Cover Design: Fuqi Liang

ISBN: 978-94-6366-484-4

An electronic version of this dissertation is available at:

http://repository.tudelft.nl/.

Printed in China

Dedicated to my family

献给我的家人

Preface

I am Lucky.

As I am going to complete my Ph.D. journey, I would like to take this opportunity to tell you exactly how lucky I am, and express my sincere gratitude to those who made this journey joyful and fruitful.

First and foremost, I am lucky to have Dr. Jafar Rezaei as my promoter. Without you, my Ph.D. life would have been much harder. You gave me maximum flexibility to plan my research, without much pressure. You continuously provided encouragement and were always willing and enthusiastic to help me in any way you could throughout my research project. It was such a pleasure to have you with me on my journey. We talked about life, the future, education, passion and friends. You always put yourself into other people's shoes, showing your understanding and patience. I hope I can have more time working with you, since there's so much I can learn from you. I am also lucky to have Dr. Matteo Brunelli as my supervisor. Your carefulness and detailed feedback helped safeguard the quality of our papers. You asked me very often what we could do next and what you could help me. I am thoroughly impressed by your enthusiasm when it comes to conducting research.

I am lucky to have a group of amazing colleagues. Prof. Caspar Chorus, your charming personality made me want to listen to all your speeches, which are full of humor and wisdom. Prof. Lori Tavasszy, it was really pleasant to be in your company, and I really appreciate your kindness and calmness. Prof. Bert van Wee, you are thoughtful and knowledgeable, but humble and careful as well, providing an excellent example of what a researcher should be. Shahrzad, my beautiful and intelligent office mate, you are so great, I was so lucky to sit next to you, chat with you and laugh with you. It was such a pity that I haven't spent more time hanging out with you. Huang Bing, you are a passionate and sensitive girl, I think if I had been mature enough, I would have cherished your kindness more back then. Kartika, Masoud, Kailan, thank you for providing your valuable, constructive comments and helping me to improve my papers. I want

to thank my MSc students Kevin, Kyle, and the BSc student group, Karel, Martijn, Mick, Olivier, for their supervision, giving me many new insights. Wenhua, Yan Hong, Tanzhe, Patrick, Baiba, Nicholas, Teodóra, Betty, Ellen, Li Na, Wang Ni, Longjian, Hanxin, Renxing, Wenting, Haiyan, Luo Jisiwei, Javanshir, Majid, Haiyan, Longxiao, it was you who made our faculty a happy place, and I have learned so much from you.

I am lucky to have made a group of great friends. In particular, I would like to express my deep gratitude to Shahnoor, the sweetest girl ever. I believe it was fate that arranged for you to live next door, and know that my life would be less joyful without you. I will always remember our romantic dinners, our guitar nights, our happiness and sadness. You have never wavered in your support for me, I know you will be there if I come to you any time in the future. I also want to thank Lin Bin, I was so lucky to be your neighbor, it was such fun cooking and playing ping pong with you, which was particularly important for me in making it through these dark corona days. Haopeng, Chang Jin, Feng Yan, I was so lucky to know you when I was here, knowing nobody, you helped me get through the difficult times early on. Deng Zhe, Zeyu, Yunlong, Zhou Lin, Shen Qi, Boxin, Josiah, Anita, Nianlei, Tingkai, Yimu, Bejond, Indra, Shi yan, Zhongkai, Yan Bo, thank you all for letting me have a happy badminton time in Delft University of Technology. There are still so many people I need to thank, I can’t mention you all here, but I will thank all those who brought me laughter, gave me suggestions and helped me grow, it was my privilege to meet you.

I was lucky to be given the scholarship by the Chinese Government, I am really grateful and hope to do my part to fulfill the great Chinese dream. I was lucky to have the opportunity to stay in this beautiful country for four years, to study at TU Delft, to be part of a great research team, to make a group of amazing friends. The place and the people nourished me, helped me to be a better person. If I have brought some laughter and joy to the people around me, that would be my great honor.

As I am about to embark on a new journey, I hope I could still be lucky, and have the courage to follow my heart. To those who I love, I wish you all happiness.

Lucky for you.

Fuqi Liang

July 2021

Content

1 Introduction

Our decisions shape our lives. Making good decisions is not a simple task. Decision analysis can help people and organizations to improve the quality of their judgments. Methods like multi-criteria decision-making methods are widely used nowadays, and they are especially helpful when confronting circumstances that are characterized by multi-dimensionality. This thesis focuses on one of the latest proposed multi-criteria decision-making method, Best-Worst Method, examines critical issues related to this method and proposes corresponding solutions, in order to make the method more comprehensive and useful in practice.

1.1 Decision-making analysis

We make decisions every day, consciously or subconsciously, and our choices impact our lives in trivial and important ways. Simple decisions, like when to get up, what to eat for lunch, require little analysis, but for some complex problems, for instance what career to pursue or what kind of house to buy, which supplier should the company chooses and in which project should the government invest, we need to build a mental model of the problems so as to organize our thoughts and choose wisely.

A decision-making process usually involves several steps: identifying the problems, eliciting the preferences, evaluating the alternatives and determining the best options (Simon, 1960; Keeney, 1982; Keeney and Raiffa, 1993; Kleindorfer et al., 1993). Formal decision-making analysis can generally speaking be divided into three types (Bell et al., 1988; Kleindorfer et al., 1993; Tzeng and Huang, 2011): 1) Descriptive analysis, which concerns with the ways that decision makers (DMs) actually solve the problems, 2) Prescriptive analysis, which considers the methods that DMs ought to use to improve their decisions and 3) Normative analysis, which focuses on the methods DMs should follow to ideally address the problems.

People have probably been analyzing decisions since they started making them. Although we cannot know when and where it all started, but we can see many examples in history. For instance, in the Eighteenth century, Benjamin Franklin designed a decision-making approach in a letter to Joseph, *Priestly, on September 1772.* He made a list with two columns and assigned weights to the entries to reach a balance (tradeoff):

My way is to divide half a sheet of paper by a line into two columns; writing over the one Pro and over the other Con. Then during three or four days' consideration, I put down under the different heads short hints of the different motives, that at different time occur to me, for or against the measure. When I have thus got them altogether in one view, I endeavor to estimate their respective weights; and where I find two, one on each side, that seem equal, I strike them both out. If I judge some two reasons con equal to some three reasons pro, I strike out five; and thus proceeding, I find where the balance lies; and if after a day or two of further consideration, nothing new that is of importance occurs on either side, I come to a determination accordingly.

Although Benjamin Franklin's procedure may not be as rigorous as the ones we use today, people have constantly tried to improve their decision-making skills. As a result, the body of literature dedicated to making decisions is growing all the time, and the increasingly complex contexts nowadays demand more mathematically sound decision analysis methodologies to make more sensible and logical decisions.

The origins of modern decision analysis may trace back to Von Neumann and Morgenstern (1953), who are credited for developing the first axiomatic foundation of expected utility theory. Fishburn is another prolific contributor to utility theory, with his well-known books *Decision and Value Theory* (Fishburn, 1964) and *Utility Theory for Decision Making* (Fishburn, 1970), both of which advanced the utility theory and helped pave the way for Multi-Attribute Utility Theory (MAUT). Keeney and Raiffa (1976) summarized prior research in their book *Decision Analysis with Multiple Conflicting Objectives*, laying a solid foundation for the coming extensive decision analysis researches.

Against the mainstream of economics in 1950s, Simon (1955) argued that decision-making does not obey the postulates of the 'rational man', people do not solve problems by maximizing utility, instead, they are 'satisfiers', looking for solutions that meet the setting aspiration levels. In other words, people are rationally limited or bounded. Based on this argument, Simon developed a behavioral theory, based on which Tversky and Kahneman (1974) proposed their famous Prospect Theory (Kahneman and Tversky, 1979) and the Cumulative Prospect Theory (Tversky and Kahneman, 1992), which later spurred much research in behavioral decision analysis.

Thanks to the above-mentioned and other important works, decision analysis has become increasingly systematic and comprehensive. In recent decades, many directions have evolved in decision analysis, with multi-criteria decision-making being one of the most popular branches.

1.2 Multi-criteria decision-making

One of the critical requirements of the decision analysis methods that help us to make better decisions is the ability to deal with multi-dimensional situations. Multi-Criteria Decision-Making (MCDM) is one of the most well-known branches of decision-making and it has the capacity to handle problems with a multitude of, often conflicting, objectives (Greco et al., 2016). It can be divided into two categories: continuous MCDM, also known as Multiple Objective Decision-Making (MODM), and discrete MCDM, also known as Multiple Attitude Decision-Making (MADM) (Tzeng and Huang, 2011). In this thesis, we focus on discrete MCDM, which means MADM, but we use the abbreviation MCDM in the remainder of this thesis.

The MCDM methods are used to select, rank or sort a finite number of alternatives based on a number of criteria, where each criterion level approximates the level of achievement of one of the objectives. This field has been studied extensively and many methods have been proposed (Greco et al., 2016), for example: Analytic Hierarchy Process (AHP) (Saaty, 1977), ELimination and Choice Expressing the REality (ELECTRE methods) (Roy, 1968; Figueira et al., 2013; Figueira et al., 2016), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) (Gupta and Barua, 2018), Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE) (Behzadian et al., 2010), to name but a few.

According to the interpretation of Keeney and Raiffa (1976), MCDM is "*not descriptive, because most people do not attempt to think systematically about hard choices under uncertainty. It is also not normative, since it is not an idealized theory designed for the superrational being with an all-powering intellect. It is, instead, a prescriptive approach designed for normally intelligent people who want to think hard and systematically about some important real problems.*"

In most cases, there are two phases in an MCDM process[1]: first, the alternatives are evaluated with respect to the various criteria, after which the evaluations are aggregated into an overall score (aggregated value) for each alternative.

Many methods have been developed for such an aggregation, including the weighted average or simple additive weighting method, the ordered weighted averaging method and Quasi-arithmetic means (Grabisch et al., 2011). The right choice of aggregating model depends on the situation and the preference of the DM. However, among these models, the weighted sum model is the earliest and probably the most widely used one (Triantaphyllou, 2000), and it is also the one used in this thesis.

If there are m alternatives and n criteria, the aggregated value for each alternative is evaluated by the following function (Fishburn, 1967):

$$V_i = \sum_{j=1}^{n} a_{ij} w_j \,, \text{ for } i = 1,2, \cdots, m, \tag{1.1}$$

where V_i is the aggregated value score of alternative i, n is the number of criteria, a_{ij} is the normalized value of the alternative i with respect to criterion j, and w_j is the weight of criterion j.

After identifying the set of criteria and their corresponding evaluations (which are not included in this study), the weights (or scaling constants) of the criteria are the most important factor left to aggregate the performance of the alternatives into single values.

Various methods have been proposed to elicit the weights of criteria (Zardari et al., 2015; Dias et al., 2018). The existing weighting methods can be classified into two groups, according to whether Decision-Makers (DMs) provide their preferences or not: the former group considers the judgements from the DMs to determine the weights of criteria, while the latter group does not involve the subjective judgments of DMs. In the former group, the popular methods are AHP (Saaty, 1977), Simple Multi-Attribute Rating Technique (SMART) (Edwards, 1977), Direct Rating method (Bottomley and Doyle, 2001), SWING (Von Winterfeldt and Edwards, 1986), the Tradeoff method (Keeney and Raiffa, 1976) and the Best-Worst Method (BWM) (Rezaei, 2015; 2016). The latter group includes the Entropy Method (Deng et al., 2000), the

[1] Some other well-known MCDM methods are not applying this process, for example, the outranking methods, like ELECTRE and the PROMETHEE method, the procedure is different: first of all, the alternatives are pairwise compared to determine an outranking relation, then it is exploited to obtain a recommendation (see e.g. (Roy, 1991)).

Regression Method (Johnson, 2000), the Standard Deviation Method (Diakoulaki et al., 1995), and others (Zardari et al., 2015). Generally speaking, the former category is much more popular, as real decision-making problems are naturally based on the judgments provided by decision-makers.

1.3 Best Worst Method

The Best-Worst Method (BWM), a recently developed method proposed by Rezaei in 2015 (2015), uses pairwise comparisons to determine the weights of criteria (and alternatives). Thanks to its simplicity, flexibility and general applicability, the BWM has been applied in various areas since its introduction, including supply chains (Badri Ahmadi et al., 2017; Gupta and Barua, 2017), energy (Gupta et al., 2017), technology (Ren, 2018), tourism (Rezaei et al., 2018; Kumar et al., 2020) and manufacturing (Moktadir et al., 2018; Mi et al., 2019).

BWM has been studied from a theoretical perspective as well. For example, a concentration ratio has been proposed by Rezaei (2020) to check the concentration of the optimal intervals obtained from the nonlinear BWM. A linear BWM to generate a unique solution and an interval weight analysis to deal with inconsistent comparisons with more than three criteria has also been proposed, because in some cases, the original BWM can result in multi-optimality (Rezaei, 2016). The original BWM has been combined with other techniques, like TOPSIS (Gupta and Barua, 2018), VIKOR (VIseKriterijumska Optimizacija I Kompromisno Resenje) (Kumar et al., 2020), ELECTRE methods (Roy, 1968; Figueira et al., 2013; Figueira et al., 2016) and PROMETHEE (Behzadian et al., 2010). For a more exhaustive review, see the review study by Mi et al. (2019), and the bibliographical report[2].

To better understand the existing issues surrounding the use of the BWM in the next part, it is necessary to present the basic procedure of the method, which can be summarized as follows:

Step 1. The set of criteria $\{C_1, C_2, \cdots, C_n\}$ is determined by the DM.

Step 2. The best (e.g. the most influential or the most important) and worst (the least influential or the least important) criteria are determined by the DM. The two criteria are shown by C_B and C_W, respectively. The two reference points are used in Step 3 and 4 for conducting pairwise comparisons.

Step 3. The preference of the best over all the other criteria is determined by the DM using a number from $\{1, 2, \ldots, 9\}$, where 1 means 'equally important' and 9 means 'extremely more important' and the other numbers represent the preference of the DM from equally more important to extremely more important. The obtained Best-to-Others vector is: $A^{BO} = (a_{B1}, a_{B2}, \cdots, a_{Bn})$, where a_{Bj} represents the preference of the best criterion C_B over criterion C_j, $j = 1,2, \cdots, n$.

Step 4. The preferences of all the criteria over the worst criterion are determined by the DM using a number from $\{1, 2, \ldots, 9\}$. The obtained Others-to-Worst vector is: $A^{OW} = (a_{1W}, a_{2W}, \cdots, a_{nW})$, where a_{jW} represents the preference of criterion $C_j, j = 1, 2, \cdots, n,$ over the worst criterion C_W.

Step 5. The optimal weights $(w_1^*, w_2^*, \cdots, w_n^*)$ are found by solving the following model:

2 From https://bestworstmethod.com

$$\begin{aligned} \text{minimize} \quad & \max_j \left\{ \left| \frac{w_B}{w_j} - a_{Bj} \right|, \left| \frac{w_j}{w_W} - a_{jW} \right| \right\}, \\ \text{subject to} \quad & w_1 + \cdots + w_n = 1, \\ & w_j > 0,\ j = 1, \ldots, n. \end{aligned} \tag{1.2}$$

1.4 Statement of problems

Although BWM has been widely applied and studied in some areas, like the combination of fuzzy theory and group decision-making, other important problems that occur very frequently in practice have not yet been examined, including: 1. the inconsistency in the preferences provided by DMs, 2. the uncertain information embedded in the DMs' judgement, 3. the consensus reaching problem in group decision-making, 4. the range sensitivity in MCDM problems that lack consideration in BWM.

Inconsistency: One of the advantages of BWM, and other pairwise comparisons methods, like AHP and ANP (Analytical Network Process), is that they allow the analysts to assess the inconsistency of a DM's preferences, which usually has to do with the rationality of the DM and their ability to discriminate between criteria/alternatives (Irwin, 1958). Traditionally, we say that a DM is perfectly consistent when the cardinal transitivity condition of their preferences is met. To check the inconsistency degree (the deviation from the condition of full consistency) of the provided preferences, Saaty (1977), in his seminal work on the AHP, proposed a (in)consistency measurement (the Saaty index). After which, many other (in)consistency indices have been proposed for a complete pairwise comparison matrix (Brunelli and Fedrizzi, 2019). In the original version of BWM, an (in)consistency measurement is proposed (Rezaei, 2015). Similarly, the extended BWM methods also provided corresponding (in)consistency measurements in the same logic with the original (in)consistency measurement (Mou et al., 2016; Guo and Zhao, 2017; Aboutorab et al., 2018).

However, the existing (in)consistency measurements for BWM can only provide the (in)consistency feedback after the entire optimization process has been completed, and the results are dependent on the optimization models, since different models (linear (Rezaei, 2016), multiplicative (Brunelli and Rezaei, 2019), and Bayesian BWM (Mohammadi and Rezaei, 2020)) can be chosen by DMs. Moreover, consistency in pairwise comparisons can be divided into cardinal consistency and ordinal consistency (Siraj et al., 2015). The existing studies on (in)consistency measurement in the BWM thus far can only measure cardinal consistency and they fail to take ordinal consistency into consideration. Furthermore, although the existing (in)consistency measurements can help a DM to check the reliability of their preferences, the absence of a threshold associated with them makes it hard to provide a meaningful interpretation and the DM/analyst is left with the major problem of having to decide when their judgments should be revised and when they should be accepted.

Uncertainty: Handling uncertain information is one of the critical issues in MCDM, including BWM. There are many different meanings of uncertainty. Among the many interpretations of uncertainty, we found that the definition proposed by Zimmermann (2000) is well-suited to the motivation of this study:

"Uncertainty implies that in a certain situation a person does not dispose about information which quantitatively and qualitatively is appropriate to describe, prescribe or predict deterministically and numerically a system, its behavior or other characteristics."

Typically, uncertainty can be divided into three categories: fuzziness (or vagueness), which results from the imprecise boundaries of fuzzy sets; discord (or strife), which expresses conflicts among the various sets of alternatives; and non-specificity (or imprecision), which is connected to the sizes (cardinalities) of relevant sets of alternatives (Klir and Wierman, 1999). For example, a fuzzy set represents fuzziness, a probability distribution represents only discord, while a classical set simply represents non-specificity (Jousselme et al., 2006).

Although many researchers have tried to manage uncertainty by extending BWM, most only handle fuzziness (Mou et al., 2016; Guo and Zhao, 2017; Aboutorab et al., 2018; Nie et al., 2018; Pamučar et al., 2018; Hafezalkotob et al., 2020), and most of which, incidentally, merely complicate the simple situations, preferences with discord and non-specificity have not been treated properly in the existing extensions. A belief structure defined in the Dempster-Shafer theory (D-S theory) framework (Shafer, 1976) can be a good way to handle judgement both with discord and non-specificity (Jousselme et al., 2006).

Consensus: In many cases, a single DM or expert can easily analyze the situation and make the decision on their own. However, in some complex circumstances, it is difficult for a single DM or expert to take all the relevant aspects of a decision-making problem into account. Therefore, it is helpful to have a group of experts/DMs with different knowledge and backgrounds work together to arrive at a more comprehensive and rational solution. However, the diversified professional fields of the experts/DMs may lead to conflict and make it difficult to reach a consensus within the group.

Although several studies on BWM have incorporated group decision-making (Jia and Wang, 2016; Mou et al., 2016; You et al., 2016; Hafezalkotob and Hafezalkotob, 2017; Mou et al., 2017; Safarzadeh et al., 2018; Hajek and Froelich, 2019; Mohammadi and Rezaei, 2019; Hafezalkotob et al., 2020), most use the traditional way to assign weights to the experts/DMs, rather than using a consensus model to reach agreement based on the preferences of experts/DMs. In addition, the importance of the reliability of the preferences provided by the experts/DMs is underestimated and rarely considered in the existing group BWM approaches. Moreover, it is likely that different experts/DMs have different sets of criteria when assessing the alternatives in real life, which cannot be handled appropriately in the traditional way, and it has not been examined in existing group BWM studies.

After obtaining the overall value of each alternative for each DM, we need to calculate the aggregated value of each alternative for all the DMs. When using nonlinear BWM, the results of the weights are sometimes intervals (Rezaei, 2016), and so will the aggregated values of each DM if we use the additive value function (Keeney and Raiffa, 1976). Traditionally, the technique used in literature for aggregating intervals is to average the interval centers (Yaniv, 1997). However, that approach does not take the ranges of the intervals into account (Lyon et al., 2015) and overlooks the overlapping areas of the intervals.

Range sensitivity: In decision analysis, the weights of the criteria should depend on the ranges of the criteria (the outcome intervals of the criteria), i.e. with others unchanged, the greater the range of a criterion, the greater its weight should be, which is referred to as the *range-sensitivity principle* (Fischer, 1995). Without incorporating the ranges of the criteria, some weighting methods may produce some distortion or biases in the elicitation of weights (Fischer et al., 1987; Montibeller and Von Winterfeldt, 2015). According to previous studies (von Nitzsch and Weber, 1993; Fischer, 1995), even if the range of the criterion is mentioned, DMs often do not adjust their judgements on the weights properly, which means that methods that do not consider the ranges, like simple ranking or direct rating methods, should only be used with great care (von Nitzsch and Weber, 1993; Fischer, 1995).

Although BWM encourages DMs to consider the range of criteria in advance, in practice, this is not done systematically. In this sense, methods like SWING (Von Winterfeldt and Edwards,

1986) and Tradeoff method (Keeney and Raiffa, 1976) which require DMs to provide their preference based on the range of criteria could handle this problem better than BWM could. How to take the range of criterion into consideration in BWM in an explicit and systematic way to avoid distortion or biases, is an important issue that requires further investigation.

1.5 Research objectives and research questions

The primary aim of this study is to address the issues encountered in BWM. As discussed in the problem statement, there are basically four main problems that have to be examined, namely, inconsistency, uncertainty, consensus and range sensitivity issues. As such, the principal objective of this thesis is as follows:

Managing inconsistency, uncertainty, consensus, and range sensitivity in the Best Worst Method.

This overarching objective can be achieved by four separate studies, as explained below.

1.5.1 Study 1: Checking the (in)consistency of preferences in BWM

Checking the (in)consistency of the preferences provided by DMs is important to ensure the reliability of their judgments. Since the existing (in)consistency measurements are model-dependent, the resulting (in)consistency ratios vary for different extended BWM models. It is desirable to have a universal (in)consistency measurement that does not depend on the model being used. Besides, since minor inconsistencies are often unavoidable in pairwise comparisons, it is necessary to know when the preferences are unacceptable. In other words, we need to establish thresholds for the (in)consistency ratios so that we can check whether or not the preferences are reliable, and that requires a methodology to construct the (in)consistency thresholds.

This study answers the following research questions:

Q1.1 How to develop a BWM model-independent (in)consistency measurement?

Q1.2 How to take the ordinal (in)consistency into account?

Q1.3 How to determine the (in)consistency thresholds?

1.5.2 Study 2: Managing uncertain information in BWM

Due to imprecision in assessment, unfamiliarity with the problem at hand or a lack of data, there is very often uncertainty when DMs provide their preferences. Since the original BWM can only handle certain preferences, extending it to incorporate a DM's hesitation is necessary in real-life situations. How to capture the uncertain judgments with discord and non-specificity is especially important and has thus far received insufficient attention, since most uncertainty studies in BWM focus on fuzziness. After considering the uncertain information, a new BWM model is needed to incorporate this new form of information to elicit weights. Besides, excessively uncertain preferences usually lead to unreliable results, so measuring uncertainty is a good way to monitor the reliability of a DM's judgments.

This study answers the following research questions:

Q2.1 How to capture DM's ambiguity information?

Q2.2 How to elicit weights based on the uncertain preferences?

1.5.3 Study 3: Reaching a consensus in group BWM

There are two issues in this group consensus problem. Firstly, in existing group BWM studies, stakeholders can only consider a fixed set of criteria, but in reality, different stakeholders often have different sets of criteria regarding the same decision problem. How to incorporate these different criteria sets and compare the results of different stakeholders could be a problem for the group BWM in literature. Secondly, because the weights obtained from the nonlinear BWM are usually intervals, traditional approaches to aggregation usually take the average of the intervals, whereas the overlaps of the intervals are preferred in reaching a consensus. How to take these interval features into account and reach the best consensus among multiple stakeholders could be a valuable study for group BWM.

This study answers the following research questions:

Q3.1 How to accommodate different criteria sets in group BWM?

Q3.2 How to reach consensus with interval weights?

1.5.4 Study 4: Accounting range sensitivity of criteria in BWM

In this study, we examine the importance of considering the ranges of criteria in MCDM methods, and the consequences of neglecting them. By reviewing existing literature, we try to determine how the other MCDM methods deal with criteria ranges. Since BWM does not systematically take the ranges of criteria into consideration, we need to find a way to incorporate them to avoid possible range insensitivity bias.

This study answers the following research question:

Q4.1 How to account ranges (sensitivity) of criteria in BWM?

1.6 The structure of the thesis

The thesis is a collection of four separate articles that address the four main topics mentioned above. After introducing the research questions in this chapter, the readers have a basic understanding of the four critical issues in the practice and theory of BWM. Since checking inconsistency is a necessity for BWM procedures, the inconsistency issue elaborated in Chapter 2 connects to all the other studies. The study of uncertain information management in BWM is presented in Chapter 3, and the method proposed in this study is applied to an evaluation of large infrastructure projects in Indonesia. The consensus study is illustrated in Chapter 4, along with an application in an inland terminal selection project in Germany. The range sensitivity problem in BWM is discussed in Chapter 5, a new method is proposed and a port performance evaluation case study in the Netherlands is used to examine the feasibility of the proposed method. The conclusion and discussions of this thesis are presented in Chapter 6, including reflections, limitations and recommendations for further research. The framework of the entire thesis is shown in Figure 1-1.

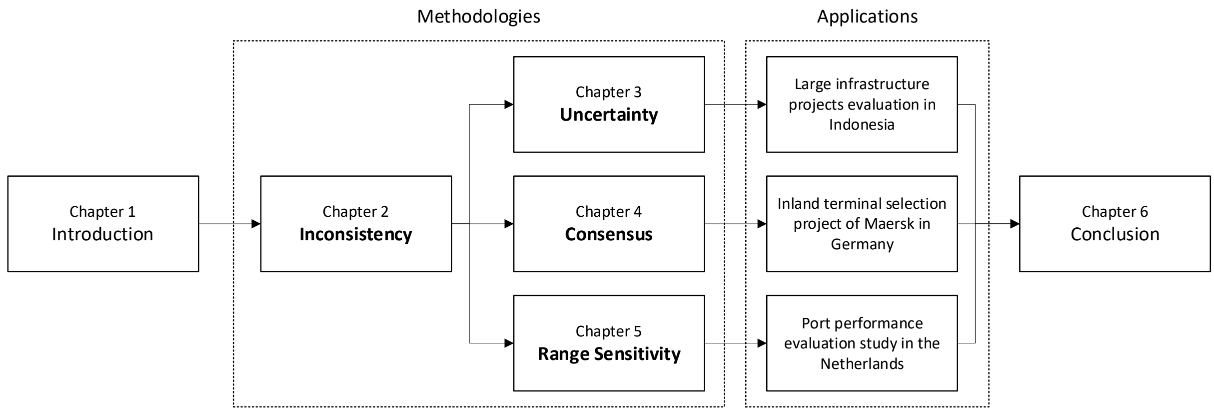


Figure 1-1. The framework of the thesis

References

Aboutorab, H., Saberi, M., Asadabadi, M. R., Hussain, O. & Chang, E. (2018). ZBWM: The Z-number extension of Best Worst Method and its application for supplier development. *Expert Systems with Applications*, *107*, 115-125.

Badri Ahmadi, H., Kusi-Sarpong, S. & Rezaei, J. (2017). Assessing the social sustainability of supply chains using Best Worst Method. *Resources, Conservation and Recycling*, *126*, 99-106.

Behzadian, M., Kazemzadeh, R. B., Albadvi, A. & Aghdasi, M. (2010). PROMETHEE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, *200*(1), 198-215.

Bell, D. E., Raiffa, H. & Tversky, A. (1988). Decision making: Descriptive, normative, and prescriptive interactions, cambridge university Press.

Bottomley, P. A. & Doyle, J. R. (2001). A comparison of three weight elicitation methods: Good, better, and best. *Omega*, *29*(6), 553-560.

Brunelli, M. & Fedrizzi, M. (2019). A general formulation for some inconsistency indices of pairwise comparisons. *Annals of Operations Research*, *274*(1-2), 155-169.

Brunelli, M. & Rezaei, J. (2019). A multiplicative best-worst method for multi-criteria decision making. *Operations Research Letters*, *47*(1), 12-15.

Deng, H., Yeh, C. & Willis, R. J. (2000). Inter-company comparison using modified TOPSIS with objective weights. *Computers & Operations Research*, *27*(10), 963-973.

Diakoulaki, D., Mavrotas, G. & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The critic method. *Computers & Operations Research*, *22*(7), 763-770.

Dias, L. C., Morton, A. & Quigley, J. (2018). Elicitation. Switzerland, Springer International Publishing.

Edwards, W. (1977). How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, *7*(5), 326-340.

Triantaphyllou, E. (2000). Multi-criteria decision making methods. Multi-criteria decision making methods: A comparative study, Springer: 5-21.

Figueira, J. R., Greco, S., Roy, B. & Słowiński, R. (2013). An overview of ELECTRE methods and their recent extensions. *Journal of Multi-Criteria Decision Analysis*, *20*(1-2), 61-85.

Figueira, J. R., Mousseau, V. & Roy, B. (2016). ELECTRE methods. Multiple criteria decision analysis, Springer: 155-185.

Fischer, G. W. (1995). Range sensitivity of attribute weights in multiattribute value models. *Organizational Behavior and Human Decision Processes, 62*(3), 252-266.

Fischer, G. W., Damodaran, N., Laskey, K. B. & Lincoln, D. (1987). Preferences for proxy attributes. *Management Science, 33*(2), 198-214.

Fishburn, P. C. (1964). Decision and Value Theory. New York, Wiley.

Fishburn, P. C. (1967). Methods of estimating additive utilities. *Management science*, *13*(7), 435-453.

Fishburn, P. C. (1970). Utility theory for decision making. New York, John Wiley and Sons.

Grabisch, M., Marichal, J., Mesiar, R. & Pap, E. (2011). Aggregation functions: Means. *Information sciences*, *181*(1), 1-22.

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

Guo, S. & Zhao, H. (2017). Fuzzy best-worst multi-criteria decision-making method and its applications. *Knowledge-Based Systems*, *121*, 23-31.

Gupta, H. & Barua, M. K. (2017). Supplier selection among SMEs on the basis of their green innovation ability using BWM and fuzzy TOPSIS. *Journal of Cleaner Production*, *152*, 242-258.

Gupta, H. & Barua, M. K. (2018). A framework to overcome barriers to green innovation in SMEs using BWM and Fuzzy TOPSIS. *Science of The Total Environment*, *633*, 122-139.

Gupta, P., Anand, S. & Gupta, H. (2017). Developing a roadmap to overcome barriers to energy efficiency in buildings using best worst method. *Sustainable Cities and Society*, *31,* 244-259.

Hafezalkotob, A. & Hafezalkotob, A. (2017). A novel approach for combination of individual and group decisions based on fuzzy best-worst method. *Applied Soft Computing*, *59*, 316-325.

Hafezalkotob, A., Hafezalkotob, A., Liao, H. & Herrera, F. (2020). Interval MULTIMOORA Method Integrating Interval Borda Rule and Interval Best–Worst-Method-Based Weighting Model: Case Study on Hybrid Vehicle Engine Selection. *IEEE Transactions on Cybernetics*, *50*(3), 1157-1169.

Hajek, P. & Froelich, W. (2019). Integrating TOPSIS with interval-valued intuitionistic fuzzy cognitive maps for effective group decision making. *Information Sciences*, *485*, 394-412.

Irwin, F. W. (1958). An analysis of the concepts of discrimination and preference. *The American Journal of Psychology*, *71*(1), 152-163.

Jia, F. & Wang, X. Y. (2016). BWM-TOPSIS multi-criteria group decision-making method based on rough number. *Kongzhi yu Juece/Control and Decision*, *31*(10), 1915-1920.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*(1), 1-19.

Jousselme, A., Liu, C., Grenier, D. & Bossé, É. (2006). Measuring ambiguity in the evidence theory. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 36*(5), 890-903.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263-292.

Keeney, R. L. (1982). Decision analysis: an overview. *Operations research*, *30*(5), 803-838.

Keeney, R. L. & Raiffa, H. (1976). Decision analysis with multiple conflicting objectives. New York, John Wiley & Sons.

Keeney, R. L. & Raiffa, H. (1993). Decisions with multiple objectives: preferences and value trade-offs, Cambridge university press.

Kleindorfer, P. R., Kunreuther, H. G. & Schoemaker, P. J. (1993). Decision sciences: An integrative perspective, Cambridge University Press.

Klir, G. & Wierman, M. (1999). Uncertainty-based information: Elements of generalized information theory, Physica-Verlag.

Kumar, A., Alora, A. & Gupta, H. (2020). Evaluating green performance of the airports using hybrid BWM and VIKOR methodology. *Tourism Management*, *76*, 103941.

Lyon, A., Wintle, B. C. & Burgman, M. (2015). Collective wisdom: Methods of confidence interval aggregation. *Journal of Business Research*, *68*(8), 1759-1767.

Mi, X., Tang, M., Liao, H., Shen, W. & Lev, B. (2019). 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*, *87*, 205-225.

Mohammadi, M. & Rezaei, J. (2019). Bayesian best-worst method: A probabilistic group decision making model. *Omega*, 102075.

Mohammadi, M. & Rezaei, J. (2020). Bayesian best-worst method: A probabilistic group decision making model. *Omega*, *96*, 102075.

Moktadir, M. A., Ali, S. M., Kusi-Sarpong, S. & Shaikh, M. A. A. (2018). Assessing challenges for implementing Industry 4.0: Implications for process safety and environmental protection. *Process Safety and Environmental Protection*, *117*, 730-741.

Montibeller, G. & Von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Analysis*, *35*(7), 1230-1251.

Morgenstern, O. & Von Neumann, J. (1953). Theory of games and economic behavior, Princeton university press.

Mou, Q., Xu, Z. & Liao, H. (2016). An intuitionistic fuzzy multiplicative best-worst method for multi-criteria group decision making. *Information Sciences*, *374*, 224-239.

Mou, Q., Xu, Z. & Liao, H. (2017). A graph based group decision making approach with intuitionistic fuzzy preference relations. *Computers & Industrial Engineering*, *110*, 138-150.

Nie, R., Tian, Z., Wang, X., Wang, J. & Wang, T. (2018). Risk evaluation by FMEA of supercritical water gasification system using multi-granular linguistic distribution assessment. *Knowledge-Based Systems*, *162*, 185-201.

Pamučar, D., Petrović, I. & Ćirović, G. (2018). Modification of the Best-Worst and MABAC methods: A novel approach based on interval-valued fuzzy-rough numbers. *Expert Systems with Applications*, *91*, 89-106.

Ren, J. (2018). Technology selection for ballast water treatment by multi-stakeholders: A multi-attribute decision analysis approach based on the combined weights and extension theory. *Chemosphere*, *191*, 747-760.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49-57.

Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega*, *64*, 126-130.

Rezaei, J. (2020). A concentration ratio for nonlinear best worst method. *International Journal of Information Technology & Decision Making*, *19*(3), 891-907.

Rezaei, J., Kothadiya, O., Tavasszy, L. & Kroesen, M. (2018). Quality assessment of airline baggage handling systems using SERVQUAL and BWM. *Tourism Management*, *66*, 85-93.

Roy, B. (1968). Classement et choix en présence de points de vue multiples. *Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle*, *2*(1), 57-75.

Roy, B. (1991). The outranking approach and the foundations of electre methods. *Theory and Decision*, *31*(1), 49-73.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*(3), 234-281.

Safarzadeh, S., Khansefid, S. & Rasti-Barzoki, M. (2018). A group multi-criteria decision-making based on best-worst method. *Computers & Industrial Engineering*, *126*, 111-121.

Shafer, G. (1976). A mathematical theory of evidence, Princeton university press.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99-118.

Simon, H. A. (1960). The New Science of Management Decision, Prentice-Hall.

Siraj, S., Mikhailov, L. & Keane, J. A. (2015). Contribution of individual judgments toward inconsistency in pairwise comparisons. *European Journal of Operational Research*, *242*(2), 557-567.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297-323.

Tzeng, G. & Huang, J. (2011). Multiple attribute decision making: methods and applications, Chapman and Hall/CRC.

von Nitzsch, R. & Weber, M. (1993). The effect of attribute ranges on weights in multiattribute utility measurements. *Management Science*, *39*(8), 937-943.

Von Winterfeldt, D. & Edwards, W. (1986). Decision analysis and behavioral research. Cambridge, Cambridge University Press.

Yaniv, I. (1997). Weighting and trimming: Heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes*, *69*(3), 237-249.

You, X., Chen, T. & Yang, Q. (2016). Approach to multi-criteria group decision-making problems based on the Best-Worst-Method and ELECTRE method. *Symmetry*, *8*(9), 95.

Zardari, N. H., Ahmed, K., Shirazi, S. M. & Yusop, Z. B. (2015). Weighting methods and their effects on multi-criteria decision making model outcomes in water resources management. New York, Springer International Publishing.

Zimmermann, H. J. (2000). An application-oriented view of modeling uncertainty. *European Journal of Operational Research*, *122*(2), 190-198.

2 Consistency issues in the Best Worst Method: Measurements and thresholds

Liang, F., Brunelli, M. & Rezaei, J. (2020). Consistency issues in the best worst method: Measurements and thresholds. Omega, 96, 102175.

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.

2.1 Introduction

The Best Worst Method (BWM), which is a Multi-Criteria Decision Making (MCDM) method that was recently developed by Rezaei (2015), 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 (Rezaei, 2015) or a linear model (Rezaei, 2016). 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 (Rezaei, 2015). Thanks to its simplicity and reliability, BWM has been widely applied to address a host of different problems (Rezaei et al., 2018; Yadav et al., 2018; Kumar et al., 2020). For more detailed information, readers are referred to a recent survey on the BWM (Mi et al., 2019).

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 (Saaty, 1980), incomplete pairwise comparison matrix (Harker, 1987) or vectors (Rezaei, 2015). 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 (Irwin, 1958). 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 (1977), in their seminal work on the AHP, proposed a consistency measurement (Saaty index), but since then, many other consistency indices have been proposed (Brunelli and Fedrizzi, 2019). Basically, the existing consistency measurements can be divided into two groups: the input-based measurements and output-based measurements (Kułakowski and Talaga, 2019). The measurements in the former group are based on the input, i.e. preferences assigned to pairwise comparisons, e.g. Koczkodaj index (Koczkodaj, 1993), while the output-based consistency measurements are based on the weights or rankings. In this group, there are, for instance, Saaty's index (1977) and the geometric consistency index proposed by Crawford and Williams (1985).

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 (Harker, 1987). To adapt the consistency indices to incomplete pairwise comparison matrices, one of the most popular approaches is to complete the pairwise comparison matrices (Fedrizzi and Giove, 2007; Ureña et al., 2015) and then measure their consistency in the traditional manner (Harker, 1987; Kułakowski and Talaga, 2019). 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 (Siraj et al., 2012; Lundy et al., 2017; Bozóki and Tsyganok, 2019). Replacing triads with cycles (Kułakowski and Talaga, 2019) is another way to estimate the inconsistency.

BWM can be seen 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. We could 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

(2015) 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 and Xu et al. (2016) 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 (2017) proposed a consistency ratio (also referred to as inconsistency ratio) for fuzzy BWM, and Aboutorab and Saberi et al. (2018) 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 (Monti and Carenini, 2000). In addition, the consistency ratios obtained by the original BWM, graph-theoretic approach (Siraj et al., 2012; Lundy et al., 2017; Bozóki and Tsyganok, 2019) and the methods of replacing triads with cycles (Kułakowski and Talaga, 2019) 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 (Ergu et al., 2011) and achieve sufficiently consistent preferences (Fishburn, 1999; Pereira and Costa, 2015). Although some input-based consistency measurements for general incomplete pairwise comparison matrices, including the Koczkodaj index (Koczkodaj, 1993) and the Salo and Hämäläinen index (1995), can be applied to BWM, some of their properties are not as desirable as we expected, as discussed in Section 2.3.

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 (Siraj et al., 2015). 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 (2004). 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 (Siraj et al., 2015). 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 (1993), the dissonance measurement proposed by Siraj et al. (2012; 2015). 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 (Monsuur, 1997; Bozóki and Rapcsák, 2008; Bozóki et al., 2015), and even Saaty later suggested additional threshold values of 5% and 8% for 3 and 4 criteria, respectively (Saaty, 1994). Although some other methods have been proposed to determine consistency thresholds (Monsuur, 1997; Aguarón and Moreno-Jiménez, 2003; Amenta et al., 2020), 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 ratios and the original consistency ratio) used in BWM.

The remainder of the paper is structured as follows: In Section 2.2, 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 2.3. An ordinal consistency measurement is formulated in Section 2.4. The threshold tables are presented in Section 2.5, followed by the conclusion in Section 2.6.

2.2 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.

2.2.1 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 (Rezaei, 2016). The basic steps of original BWM can be summarized as below:

Step 1. Have the set of evaluation criteria $\{C_1, C_2, \cdots, C_n\}$ 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 $\{1,2,\ldots,9\}$. The obtained Best-to-Others vector is: $A_{BO} = (a_{B1}, a_{B2}, \cdots, a_{Bn})$, where a_{Bj} represents the preference of the best criterion C_B over criterion $C_j, j = 1,2,\cdots,n$.

Step 4. Determine the preferences of all the criteria over the worst criterion using a number from $\{1,2,\ldots,9\}$. The obtained Others-to-Worst vector is: $A_{OW} = (a_{1W}, a_{2W}, \cdots, a_{nW})$, where a_{jW} represents the preference of criterion C_j over the worst criterion C_W, $j = 1,2,\cdots,n$.

Step 5. Determine the weights $(w_1^*, w_2^*, \cdots, w_n^*)$ by solving the following model:

$$\begin{aligned}&\min \ \max_{j}\left\{\left|\frac{w_B}{w_j}-a_{Bj}\right|,\left|\frac{w_j}{w_W}-a_{jW}\right|\right\}\\&s.\ t.\\&\sum_{j=1}^{n} w_j = 1\\&w_j \geq 0 \ \text{, for all } j.\end{aligned} \tag{2.1}$$

Model (2.1) can be transformed into the following model:

$$\begin{aligned}&\min \xi\\&s.\ t.\\&\left|\frac{w_B}{w_j}-a_{Bj}\right| \leq \xi \ \text{, for all } j\\&\left|\frac{w_j}{w_W}-a_{jW}\right| \leq \xi\text{, for all } j\\&\sum_{j=1}^{n} w_j = 1,\\&w_j \geq 0 \ \text{, for all } j.\end{aligned} \tag{2.2}$$

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

$$a_{Bj} \times a_{jW} = a_{BW}\text{, for all } j \tag{2.3}$$

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 ξ^*, which is the optimal objective value (the output) of the optimization model (2.2), 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 (Rezaei, 2015):

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

$$CR^O = \frac{\xi^*}{\xi_{max}} \tag{2.4}$$

where ξ^* is the optimal objective value of model (2.2) and ξ_{max} is the maximum possible ξ, which can be derived from (Rezaei, 2015):

$$\xi^2 - (1 + 2a_{BW})\xi + (a_{BW}^2 - a_{BW}) = 0 \tag{2.5}$$

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.

2.3 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 (2019) has this property. Besides this similarity, we will also show the resemblance between the old and the new index using some numerical analyses.

2.3.1 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:

$$CR^I = \max_j CR_j^I \tag{2.6}$$

where

$$CR_j^I = \begin{cases} \frac{|a_{Bj} \times a_{jW} - a_{BW}|}{a_{BW} \times a_{BW} - a_{BW}} & a_{BW} > 1 \\ 0 & a_{BW} = 1 \end{cases} \tag{2.7}$$

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:

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.

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 $a_{Bj} \times a_{jW}$.

It can provide a DM with a clear guideline on the revision of the inconsistent judgement(s). The CR^O 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.

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 (Brunelli and Rezaei, 2019). For example, the linear BWM model (Rezaei, 2016) does not have an effective consistency measurement, while the non-linear BWM model (Rezaei, 2015) has a different interpretation than the multiplicative BWM model (Brunelli and Rezaei, 2019). 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 (Rezaei, 2016), 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 Equation (2.7), the $CR_j^I s$ are represented in the fourth row of Table 2-1.

Table 2-1. 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

From Table 2-1, by using the maximum measurement (2.6), 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 .

2.3.2 Properties of the input-based consistency measurement

As indicated by Brunelli (2018), 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 $\sigma: \{1, \dots, n\} \to \{1, \dots, n\}$ 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 $a_{Bj}, a_{jW}, a_{BW} \geq 1$, and $a_{Bj}, a_{jW} \leq a_{BW}$, so $a_{Bj}a_{jW} \leq a_{BW}a_{BW}$, $a_{Bj}a_{jW} - a_{BW} \leq a_{BW}a_{BW} - a_{BW}$, when $a_{Bj}a_{jW} - a_{BW} \geq 0$, the inequality $|a_{Bj}a_{jW} - a_{BW}| \leq a_{BW}a_{BW} - a_{BW}$ holds; when $a_{Bj}a_{jW} - a_{BW} < 0$, because $a_{Bj}, a_{jW} \geq 1$, then $a_{Bj}a_{jW} \geq 1$, now $a_{Bj}a_{jW} < a_{BW}$, and a_{BW} should be an integer, then $a_{BW} \geq 2$, so $a_{BW}a_{BW} - a_{BW} \geq a_{BW}$, and because $a_{Bj}a_{jW} < a_{BW}$, then $a_{BW} > a_{Bj}a_{jW} - a_{BW}$, therefore $a_{BW}a_{BW} - a_{BW} \geq a_{BW} > a_{Bj}a_{jW} - a_{BW}$, 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

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

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

$$S_{-i} = \langle CR_j | j \in \{1, \dots, n\} \backslash \{i\} \rangle$$

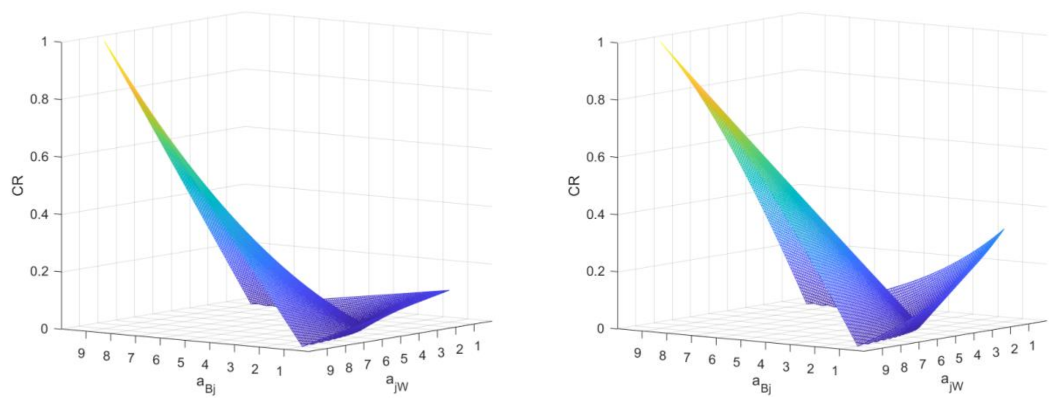
Now, since $S_{-i} \subset S$ we know that $\max_j(S) \geq \max_j(S_{-i})$. □

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 (2015), Property 3 from the normalization proposed by Koczkodaj et al. (2017), and Property 6 from the contraction property proposed by Koczkodaj and Urban (2018).

Note that Property 6 would not be satisfied by an approach based on the average of the local inconsistencies like Salo and Hämäläinen index (1995).

2.3.3 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 Figure 2-1 (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 Figure 2-2 (b).



(a) CR^I (b) CR^O

Figure 2-1. 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

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 Figure 2-2.

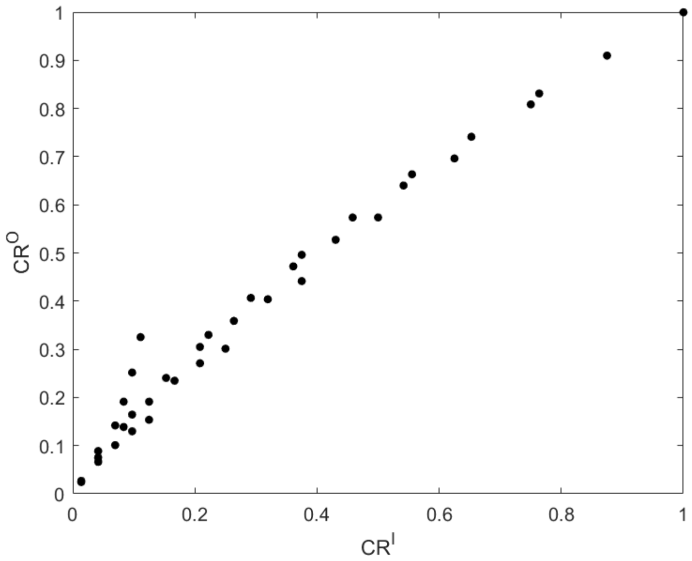


Figure 2-2. Relation between CR^O and CR^I (9 criteria 9-scale)

As a_{Bj}, $a_{jW} \in \{1,2,\ldots,9\}$ 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 2.3.1, there are valid reasons to prefer the input-based consistency measurement to the output-based consistency measurement.

2.4 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.

2.4.1 Ordinal consistency

Kwiesielewicz and Van Uden (2004) 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 (Kendall and Smith, 1940; Iida, 2009; Kułakowski, 2018). 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:

$$\left(a_{Bi} - a_{Bj}\right) \times \left(a_{jW} - a_{iW}\right) > 0 \text{ or } \left(a_{Bi} = a_{Bj} \,\&\, a_{jW} = a_{iW}\right) \text{ for all } i \text{ and } j \tag{2.8}$$

The ordinal consistency is the usual weak transitivity condition which should be the minimum requirement for a logical and rational DM (Xu et al., 2014). 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 Figure 2-3. 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 (Rezaei, 2015).

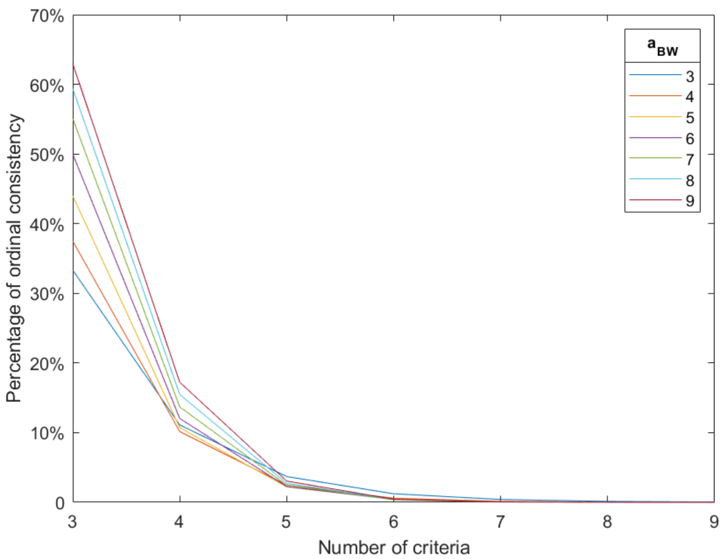


Figure 2-3. The percentage of ordinal-consistent paired vectors

2.4.2 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:

$$OR = \max_j OR_j \tag{2.9}$$

where

$$OR_j = \frac{1}{n}\sum_{i=1}^{n} F\left(\left(a_{Bi} - a_{Bj}\right) \times \left(a_{jW} - a_{iW}\right)\right) \text{ for all } i \text{ and } j \tag{2.10}$$

where $F(x, y)$ is a step function, where $x = a_{Bi} - a_{Bj}$, $y = a_{jW} - a_{iW}$, it is defined as:

$$F(x,y) = \begin{cases} 1 & \text{if } x \times y < 0 \\ 0.5 & \text{if } x \times y = 0 \text{ and } (x \neq 0 \text{ or } y \neq 0) \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

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. $\left(a_{Bi} - a_{Bj}\right) \times \left(a_{jW} - a_{iW}\right) > 0$. If only one of $\left(a_{Bi} - a_{Bj}\right)$ and $\left(a_{jW} - a_{iW}\right)$ is equal to 0, we say that, in this situation, it violates weak ordinal relation (Escobar et al., 2015; Cavallo et al., 2016), 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 jth 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 2.3.1) to explain the ordinal consistency measurement. From the preference vector A_{BO}, we can easily get the ranking of the criteria: price $\succ$ quality $\succ$ safety $\succ$ comfort $\succ$ style. The ranking from the A_{WO} vector: price $\succ$ quality $\sim$ comfort $\succ$ safety $\succ$ style ("$\succ$" means superior to, "$\sim$" 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 Equations (2.9)-(2.11), we can obtain the ordinal consistency ratios regarding each criterion in Table 2-2, which represent the ordinal violation level of each criterion. The global ordinal consistency ratios can be calculated from Equation (2.9), which is 0.3 in this case.

Table 2-2. 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

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 (Monti and Carenini, 2000).

2.4.3 Properties of the ordinal consistency ratio

The index OR (Equation (2.9)) 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, $a_{Bi} > a_{Bj} \Rightarrow i \prec j$ 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.
3. 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.

2.4.4 The relationship between ordinal consistency and cardinal consistency

Analysing the data used in the original BWM (Rezaei, 2015; 2016), we can obtain the inclusion relation between cardinal and ordinal (in)consistency of the preferences obtained from different DMs, which is graphically presented in Figure 2-4. 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.

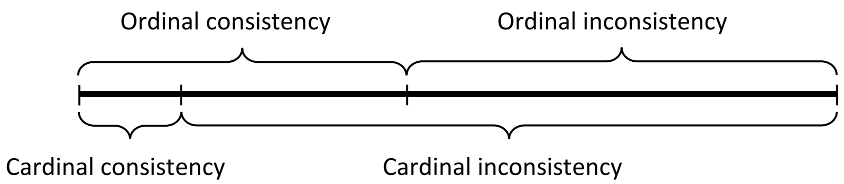


Figure 2-4. The inclusion relation between ordinal and cardinal consistency

The inclusion relation between cardinal consistency and ordinal consistency shown in Figure 2-4 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 ($\left(a_{Bi} - a_{Bj}\right) \times \left(a_{jW} - a_{iW}\right) > 0$ or $\left(a_{Bi} = a_{Bj} \,\&\, a_{jW} = a_{iW}\right)$), we shall show that, given a pairwise comparison system, cardinal consistency implies either (1). $\left(a_{Bi} - a_{Bj}\right) \times \left(a_{jW} - a_{iW}\right) > 0$ or (2). $a_{Bi} = a_{Bj} \,\&\, a_{jW} = a_{iW}$.

(1) $a_{Bi} = a_{Bj}$, if and only if $a_{jW} = \frac{a_{BW}}{a_{Bj}} = \frac{a_{BW}}{a_{Bi}} = a_{iW}$, then the comparison is ordinal-consistent;

(2) If $a_{Bi} \neq a_{Bj}$, or $a_{jW} \neq a_{iW}$, $\left(a_{Bi} - a_{Bj}\right) \times \left(a_{jW} - a_{iW}\right) = a_{Bi} \times a_{jW} - a_{Bj} \times a_{jW} - a_{Bi} \times a_{iW} + a_{Bj} \times a_{iW}$

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,

$$
\begin{aligned}
&a_{Bi} \times a_{jW} - a_{Bj} \times a_{jW} - a_{Bi} \times a_{iW} + a_{Bj} \times a_{iW} \\
&= \frac{a_{Bi} \times a_{BW}}{a_{Bj}} + \frac{a_{Bj} \times a_{BW}}{a_{Bi}} - 2a_{BW} \\
&= \frac{a_{BW}\left(a_{Bi}^2 + a_{Bj}^2\right)}{a_{Bi}a_{Bj}} - 2a_{BW} \\
&= \frac{a_{BW}\left(a_{Bi}^2 + a_{Bj}^2 - 2a_{Bj}a_{Bi}\right)}{a_{Bi}a_{Bj}} \\
&= \frac{a_{BW}\left(a_{Bi} - a_{Bj}\right)^2}{a_{Bi}a_{Bj}} > 0
\end{aligned}
$$

Therefore, the comparison is also ordinal-consistent. □

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

2.5 Thresholds for BWM

Even though we can easily identify the inconsistent judgment 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.

2.5.1 A methodology for determining the thresholds

Inspired by Amenta et al. (2018; 2020), 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[3]. Consequently, in all, there are 7×7 = 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 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 Figure 2-5.

[3] 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 $\{1,2,\dots,7\}$.

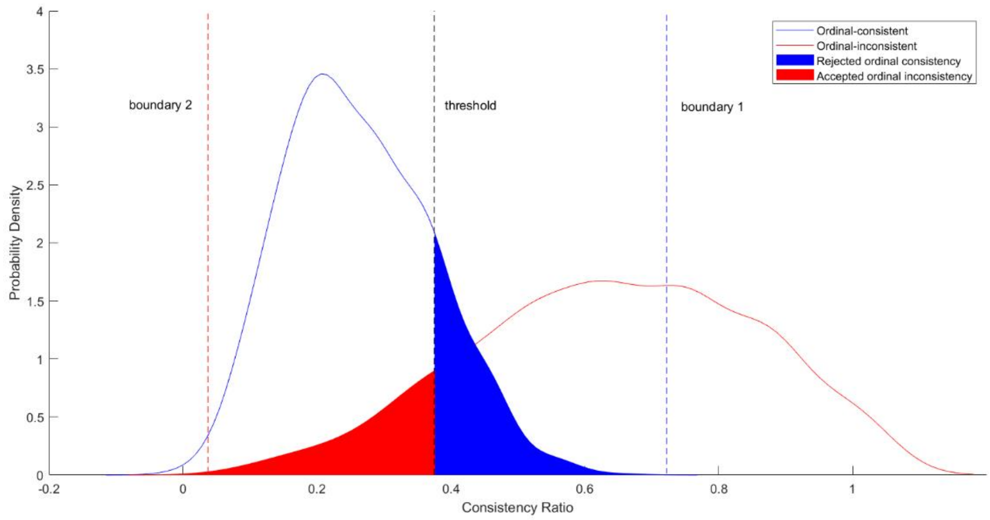


Figure 2-5. The kernel distribution of $CR^I s$ of the two groups (9-criteria 9-scale)

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:

$$\hat{F}(\alpha) = \frac{1}{N}\sum_{i=1}^{N} I\{CR_i^I \leq \alpha\} \tag{2.12}$$

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

$$I\{CR_i^I \leq \alpha\} = \begin{cases} 1 & \text{if } \ CR_i^I \leq \alpha \\ 0 & \text{othrwise} \end{cases} \tag{2.13}$$

where N is the pair number of pairwise comparisons, CR_i^I is the ith ($i \in \{1, \cdots, N\}$) 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 in Figure 2-5, and the accepted ordinal-inconsistent group is $\hat{F}^U(\alpha)$, which is the red area. We can calculate the relative rejected proportion of the $CR^I s$ in the acceptable group ($P^A_{rejected}$) and the accepted proportion of the $CR^I s$ in the unacceptable group ($P^U_{accepted}$) using the following formulas:

$$P^A_{rejected} = \frac{1-\hat{F}^A(\alpha)}{1-\hat{F}^A(\alpha)+\hat{F}^U(\alpha)} \tag{2.14}$$

$$P^U_{accepted} = \frac{\hat{F}^U(\alpha)}{1-\hat{F}^A(\alpha)+\hat{F}^U(\alpha)} \tag{2.15}$$

The relationship between these two proportions is shown in Figure 2-6, which shows how the possibility of acceptance (red line) and rejection (blue line) distribute in the two groups according to the selected threshold from 0 to 1.

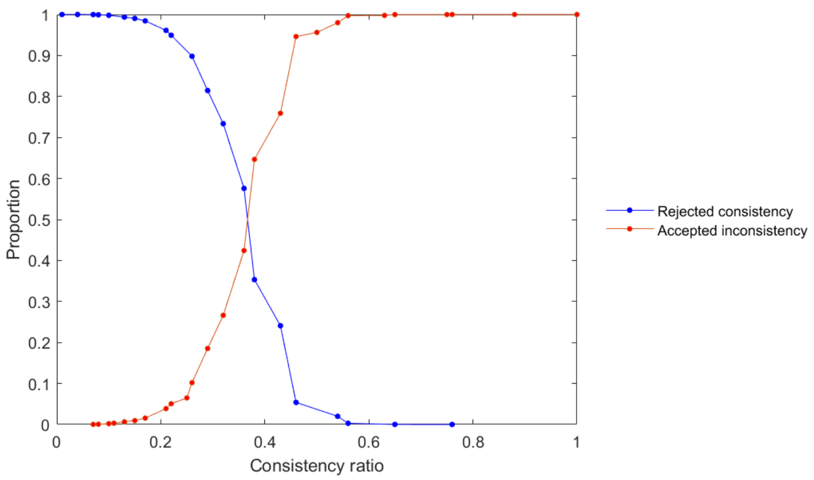


Figure 2-6. The acceptance and rejection relative proportion of the two groups (9-criteria 9-scale)

The goal is to obtain a threshold which makes the red and blue areas in Figure 2-5 as small as possible, or makes the relative proportions of the two groups in Figure 2-6 as close as possible. If there exists a CR^I obtained from the two groups which makes $P^A_{rejected} = P^U_{accepted}$, the two lines in Figure 2-6 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^Is 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.

2.5.2 Approximated thresholds for the input-based consistency ratio

Based on the algorithm presented above, we can finally establish the thresholds for BWM. In Table 2-3, we have obtained the consistency thresholds for combinations which range from 3-9 criteria with highest evaluation grades from 3 to 9 based on the input-based consistency measurement.

Table 2-3. Thresholds for different combinations using input-based consistency measurement

	Criteria						
Scales	3	4	5	6	7	8	9
3	0.17	0.17	0.17	0.17	0.17	0.17	0.17
4	0.11	0.15	0.19	0.22	0.25	0.26	0.27
5	0.14	0.20	0.23	0.25	0.27	0.28	0.30
6	0.13	0.20	0.26	0.30	0.31	0.32	0.33
7	0.13	0.25	0.28	0.30	0.31	0.33	0.34
8	0.13	0.25	0.30	0.32	0.34	0.36	0.37
9	0.14	0.27	0.31	0.33	0.35	0.36	0.37

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 Figure 2-7[4].

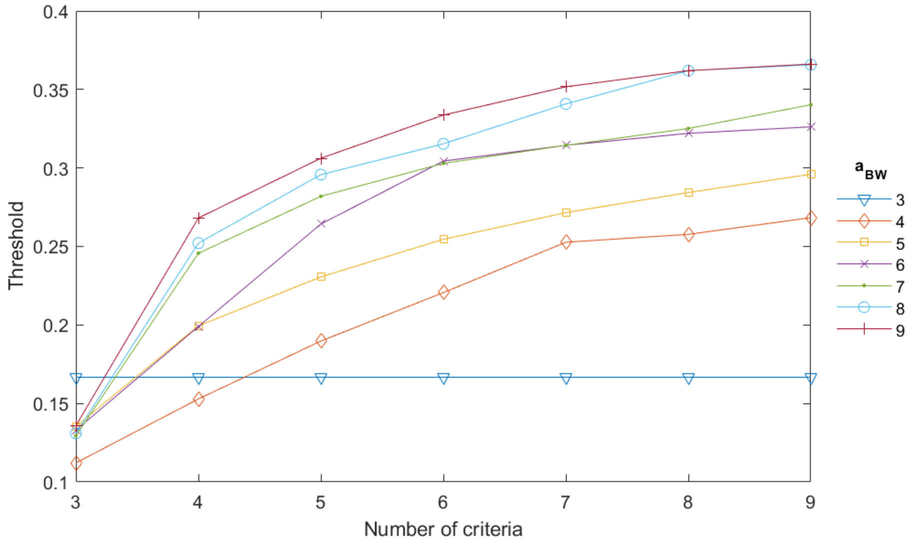


Figure 2-7. Thresholds for different combinations using input-based consistency measurement

2.5.3 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 2-4[5].

Table 2-4. Threshold for different combinations using output-based consistency measurement

	Criteria						
Scales	3	4	5	6	7	8	9
3	0.21	0.21	0.21	0.21	0.21	0.21	0.21
4	0.16	0.24	0.27	0.29	0.31	0.32	0.33
5	0.21	0.28	0.30	0.33	0.35	0.36	0.37
6	0.22	0.29	0.36	0.39	0.41	0.42	0.42
7	0.21	0.33	0.37	0.39	0.40	0.41	0.43
8	0.23	0.34	0.40	0.42	0.44	0.45	0.46
9	0.21	0.37	0.41	0.42	0.44	0.46	0.47

Compared to the thresholds obtained from the input-based consistency measurement, the thresholds of the input-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 2.3.1 is 0.14, which is less than the threshold of 0.2958 (in 5-criteria and 8-scales combination), as shown in Table 2-3, 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 2-4.

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

[5] The threshold for the CR^O in the combinations with 2-scale is 0, because when the preferences are ordinal-consistent, $CR^O = 0$.

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.

2.6 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.

Appendix

The algorithm for obtaining the threshold for the CR^I is illustrated as follows and its graphical representation is shown in Figure 2-8.

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 (2.8), and $i = i + 1$.

Step 3: Establish the ordinal-inconsistent group. If the paired vector generated in Step 1 does not satisfy the ordinal consistency 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 Equations (2.6)-(2.7).

Step 6: Calculate the empirical cumulative distribution of CR^I for the two groups by using Equations (2.12)-(2.13).

Step 7: Calculate the relative rejected proportion of the $CR^I s$ in the acceptable group ($P^A_{rejected}$) and the accepted proportion of the $CR^I s$ in the unacceptable group ($P^U_{accepted}$) by using Equations (2.14)-(2.15).

Step 8: If there exists a CR^I_T making $P^A_{rejected} = P^U_{accepted}$, then this CR^I_T is the threshold. If not, go to next step.

Step 9: Identify the cross point of the lines of $P^A_{rejected}$ and $P^U_{accepted}$, the CR^I at this point is used as the threshold.

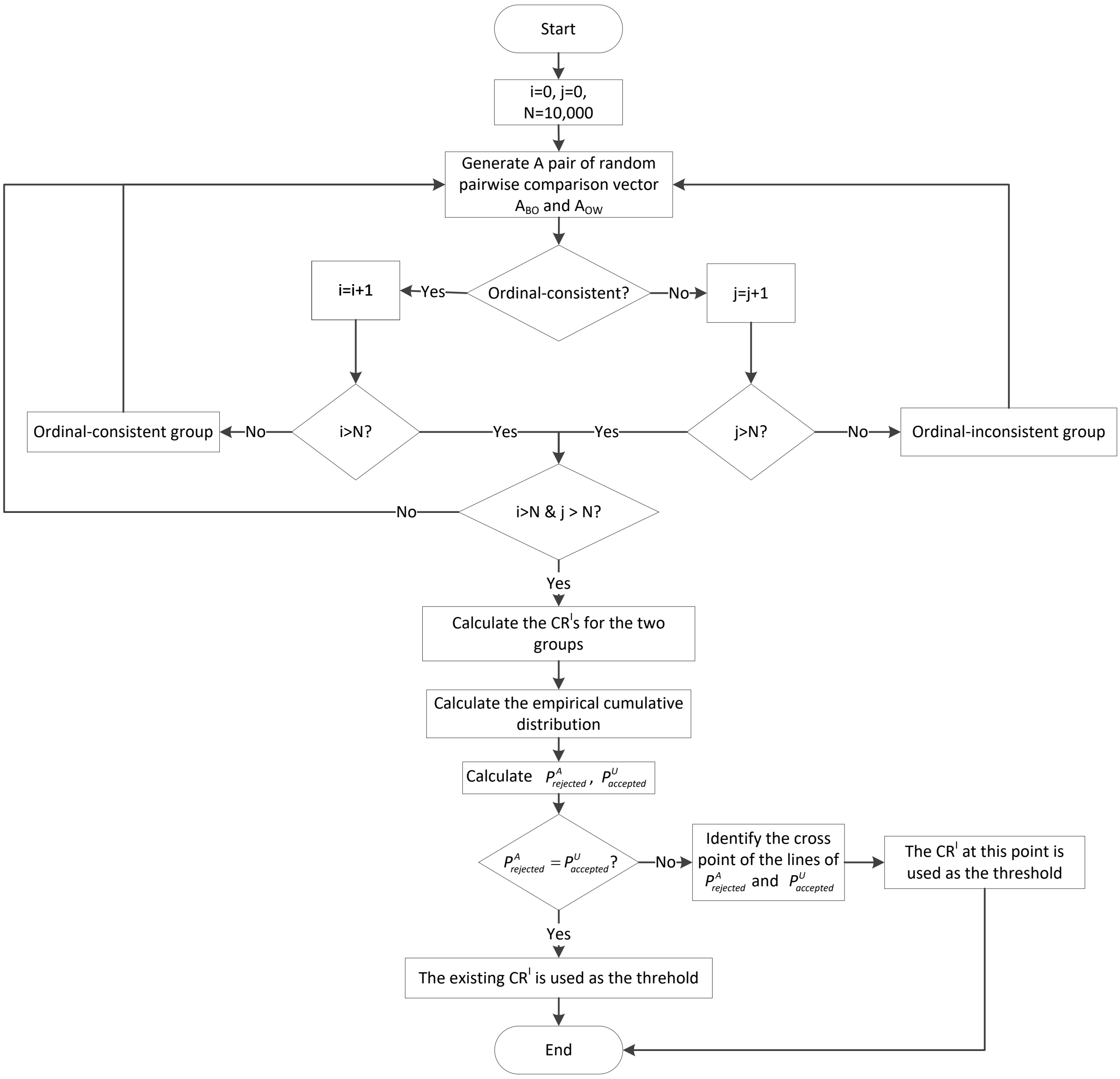


Figure 2-8. Graphical representation of simulation algorithm

Acknowledgment

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

References

Aboutorab, H., Saberi, M., Asadabadi, M. R., Hussain, O. & Chang, E. (2018). ZBWM: The Z-number extension of Best Worst Method and its application for supplier development. *Expert Systems with Applications*, *107*, 115-125.

Aguarón, J. & Moreno-Jiménez, J. M. (2003). The geometric consistency index: Approximated thresholds. *European Journal of Operational Research*, *147*(1), 137-145.

Amenta, P., Lucadamo, A. & Marcarelli, G. (2018). Approximate thresholds for Salo-Hamalainen index. *IFAC-PapersOnLine*, *51*(11), 1655-1659.

Amenta, P., Lucadamo, A. & Marcarelli, G. (2020). On the transitivity and consistency approximated thresholds of some consistency indices for pairwise comparison matrices. *Information Sciences*, *507*, 274-287.

Bozóki, S. & Rapcsák, T. (2008). On Saaty's and Koczkodaj's inconsistencies of pairwise comparison matrices. *Journal of Global Optimization*, *42*(2), 157-175.

Bozóki, S. & Tsyganok, V. (2019). 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. *International Journal of General Systems*, *48*(4), 362-381.

Bozóki, S., Fülöp, J. & Poesz, A. (2015). On reducing inconsistency of pairwise comparison matrices below an acceptance threshold. *Central European Journal of Operations Research*, *23*(4), 849-866.

Brunelli, M. (2018). A survey of inconsistency indices for pairwise comparisons. *International Journal of General Systems*, *47*(8), 751-771.

Brunelli, M. & Fedrizzi, M. (2015). Axiomatic properties of inconsistency indices for pairwise comparisons. *Journal of the Operational Research Society*, *66*(1), 1-15.

Brunelli, M. & Fedrizzi, M. (2019). A general formulation for some inconsistency indices of pairwise comparisons. *Annals of Operations Research*, *274*(1-2), 155-169.

Brunelli, M. & Rezaei, J. (2019). A multiplicative best-worst method for multi-criteria decision making. *Operations Research Letters*, *47*(1), 12-15.

Cavallo, B., D'Apuzzo, L. & Basile, L. (2016). Weak consistency for ensuring priority vectors reliability. *Journal of Multi-Criteria Decision Analysis*, *23*(3-4), 126-138.

Crawford, G. & Williams, C. (1985). A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, *29*(4), 387-405.

Ergu, D., Kou, G., Peng, Y. & Shi, Y. (2011). A simple method to improve the consistency ratio of the pair-wise comparison matrix in ANP. *European Journal of Operational Research*, *213*(1), 246-259.

Escobar, M. T., Aguarón, J. & Moreno-Jiménez, J. M. (2015). Some extensions of the precise consistency consensus matrix. *Decision Support Systems*, *74*, 67-77.

Fedrizzi, M. & Giove, S. (2007). Incomplete pairwise comparison and consistency optimization. *European Journal of Operational Research*, *183*(1), 303-313.

Fishburn, P. (1999). Preference structures and their numerical representations. *Theoretical Computer Science*, *217*(2), 359-383.

Guo, S. & Zhao, H. (2017). Fuzzy best-worst multi-criteria decision-making method and its applications. *Knowledge-Based Systems*, *121*, 23-31.

Harker, P. T. (1987). Incomplete pairwise comparisons in the analytic hierarchy process. *Mathematical Modelling*, *9*(11), 837-848.

Iida, Y. (2009). Ordinality consistency test about items and notation of a pairwise comparison matrix in AHP. Proceedings of the international symposium on the analytic hierarchy process.

Irwin, F. W. (1958). An analysis of the concepts of discrimination and preference. *The American Journal of Psychology*, *71*(1), 152-163.

Jensen, R. E. & Hicks, T. E. (1993). Ordinal data AHP analysis: A proposed coefficient of consistency and a nonparametric test. *Mathematical and Computer modelling*, *17*(4-5), 135-150.

Kendall, M. G. & Smith, B. B. (1940). On the method of paired comparisons. *Biometrika*, *31*(3-4), 324-345.

Koczkodaj, W. W. (1993). A new definition of consistency of pairwise comparisons. *Mathematical and Computer Modelling*, *18*(7), 79-84.

Koczkodaj, W. W. & Urban, R. (2018). Axiomatization of inconsistency indicators for pairwise comparisons. *International Journal of Approximate Reasoning*, *94*, 18-29.

Koczkodaj, W. W., Magnot, J. P., Mazurek, J., Peters, J. F., Rakhshani, H., Soltys, M., Strzałka, D., Szybowski, J. & Tozzi, A. (2017). On normalization of inconsistency indicators in pairwise comparisons. *International Journal of Approximate Reasoning*, *86*, 73-79.

Kułakowski, K. (2018). Inconsistency in the ordinal pairwise comparisons method with and without ties. *European Journal of Operational Research*, *270*(1), 314-327.

Kułakowski, K. & Talaga, D. (2019). Inconsistency indices for incomplete pairwise comparisons matrices. *arXiv preprint arXiv:1903.11873*.

Kumar, A., Alora, A. & Gupta, H. (2020). Evaluating green performance of the airports using hybrid BWM and VIKOR methodology. *Tourism Management*, *76*, 103941.

Kwiesielewicz, M. & Van Uden, E. (2004). Inconsistent and contradictory judgements in pairwise comparison method in the AHP. *Computers & Operations Research*, *31*(5), 713-719.

Lundy, M., Siraj, S. & Greco, S. (2017). The mathematical equivalence of the "spanning tree" and row geometric mean preference vectors and its implications for preference analysis. *European Journal of Operational Research*, *257*(1), 197-208.

Mi, X., Tang, M., Liao, H., Shen, W. & Lev, B. (2019). 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*, *87*, 205-225.

Monsuur, H. (1997). An intrinsic consistency threshold for reciprocal matrices. *European Journal of Operational Research*, *96*(2), 387-391.

Monti, S. & Carenini, G. (2000). Dealing with the expert inconsistency in probability elicitation. *IEEE Transactions on Knowledge and Data Engineering*, *12*(4), 499-508.

Mou, Q., Xu, Z. & Liao, H. (2016). An intuitionistic fuzzy multiplicative best-worst method for multi-criteria group decision making. *Information Sciences*, *374*, 224-239.

Pereira, V. & Costa, H. G. (2015). Nonlinear programming applied to the reduction of inconsistency in the AHP method. *Annals of Operations Research*, *229*(1), 635-655.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49-57.

Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega*, *64*, 126-130.

Rezaei, J., Kothadiya, O., Tavasszy, L. & Kroesen, M. (2018). Quality assessment of airline baggage handling systems using SERVQUAL and BWM. *Tourism Management*, *66*, 85-93.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*(3), 234-281.

Saaty, T. L. (1980). The analytic hierarchy process: planning, priority setting. *Resource Allocation*, 2.

Saaty, T. L. (1994). Fundamentals of decision making and priority theory with the AHP. RWS publications.

Salo, A. A. & Hämäläinen, R. P. (1995). Preference programming through approximate ratio comparisons. *European Journal of Operational Research*, *82*(3), 458-475.

Siraj, S., Mikhailov, L. & Keane, J. A. (2012). Enumerating all spanning trees for pairwise comparisons. *Computers & Operations Research*, *39*(2), 191-199.

Siraj, S., Mikhailov, L. & Keane, J. A. (2015). Contribution of individual judgments toward inconsistency in pairwise comparisons. *European Journal of Operational Research*, *242*(2), 557-567.

Siraj, S., Mikhailov, L. & Keane, J. (2012). A heuristic method to rectify intransitive judgments in pairwise comparison matrices. *European Journal of Operational Research*, *216*(2), 420-428.

Ureña, R., Chiclana, F., Morente-Molinera, J. A. & Herrera-Viedma, E. (2015). Managing incomplete preference relations in decision making: A review and future trends. *Information Sciences*, *302*, 14-32.

Xu, Y., Gupta, J. N. D. & Wang, H. (2014). The ordinal consistency of an incomplete reciprocal preference relation. *Fuzzy Sets and Systems*, *246*, 62-77.

Yadav, G., Mangla, S. K., Luthra, S. & Jakhar, S. (2018). Hybrid BWM-ELECTRE-based decision framework for effective offshore outsourcing adoption: a case study. *International Journal of Production Research*, *56*(18), 6259-6278.

3 Belief-based Best Worst Method

Liang, F., Brunelli, M., Septian, K. & Rezaei, J. (2021). Belief-Based Best Worst Method. International Journal of Information Technology & Decision Making, 20(01), 287-320.

Abstract

The Best-Worst Method (BWM) is a Multi-Criteria Decision Making (MCDM) method that has recently been introduced. The original BWM assumes that decision-makers are always certain about their judgments even if, in reality, decision-makers often express uncertain preferences. To deal with uncertainty, we introduce a belief structure in the BWM, a concept involving the preference degree adopted via Dempster-Shafer theory. A new approach is proposed to allow BWM to cope with this kind of information, where the level of belief in preferences being expressed is taken into account. In addition, an inconsistency measurement and an uncertainty measurement are proposed for the belief-based BWM, providing the foundation for a reliability degree of the decision-makers, after which the belief-based BWM is extended to include a group of decision-makers. Based on their reliability degrees and the weights of the criteria obtained from the various individuals, the overall criteria weights can be aggregated accordingly. Finally, a case study on the assessment of the infrastructure project criteria system in Indonesia is provided to demonstrate the applicability and feasibility of the proposed method.

3.1 Introduction

Multi-Criteria Decision-Making (MCDM) is an important area of operations research. It refers to finding an optimal result or ranking from a finite number of alternatives that are characterized in terms of multiple, usually conflicting, criteria (Zeleny, 1982). There is a large and growing body of literature that has so far investigated MCDM methods (Greco et al., 2016). One of the latest MCDM methods is the Best-Worst Method (BWM), proposed by Rezaei (2015), which uses pairwise comparisons to determine the weights of criteria. Thanks to its simplicity, flexibility and general applicability, since its inception, the BWM has been applied in a number of areas, including quality assessment (Rezaei et al., 2018), supply chain management (Badri Ahmadi et al., 2017; Gupta and Barua, 2017), energy (Gupta et al., 2017), technology selection (Ren, 2018), cloud service selection (Nawaz et al., 2018), web service selection (Serrai et al., 2017), and hybrid vehicle engine selection (Hafezalkotob et al., 2019). In addition to its practical applications, many researchers have extended the BWM from a theoretical perspective as well. For example, since the original BWM can in some cases result in multi-optimality, Rezaei (2016) proposed an interval weight analysis to deal with inconsistent comparisons with more than three criteria, as well as providing a linear BWM to generate a unique solution. Some researchers tried to combine subjective weights and objective weights together on the basis of BWM (Nie et al., 2018; Ren, 2018). For a more exhaustive review, see the review study by Mi et al. (Mi et al., 2019), and the bibliographical report[6].

One of the critical issues in the BWM is the way it deals with uncertainty. Typically, there are three types of uncertainty, according to the summary of Klir and Wierman (1999): *fuzziness* (or vagueness), which results from the imprecise boundaries of fuzzy sets; *discord* (or strife), which expresses conflicts among the various sets of alternatives; and *non-specificity* (or imprecision), which is connected to sizes (cardinalities) of relevant sets of alternatives. For example, a fuzzy set represents fuzziness, while a probability distribution represents only discord, and a classical set simply represents non-specificity (Jousselme et al., 2006). Although researchers have extended BWM to deal with uncertainty, most of them can only handle fuzziness (Mou et al., 2016; Guo and Zhao, 2017; Aboutorab et al., 2018; Nie et al., 2018; Pamučar et al., 2018; Hafezalkotob et al., 2019). A decision-maker (DM) who wants to provide his preferences with discord and non-specificity cannot be handled properly in BWM. However, a belief structure defined in the Dempster-Shafer theory (D-S theory) framework (Shafer, 1976) can handle both discord and non-specificity (Jousselme et al., 2006). Therefore, incorporating the belief system into the BWM will complement existing literature and make it possible to include these two types of uncertainty.

In D-S theory, subjective probabilities are replaced by “degrees of belief” within a belief structure, which can be used to express the extent to which a decision-maker (DM) believes a specific proposition to be true (Yager and Alajlan, 2015). Consider, for example, the comparison of the criteria price and quality in a sample involving cars, where a customer may state that he is 50% sure that price is slightly more important than quality, 20% sure that price is far more important than quality, and 30% sure that price is extremely more important than quality. These ‘belief degrees’ can be assigned to any subsets, making it possible to handle uncertainty and ignorance in a belief matrix. Such uncertainty and ignorance could be caused by imprecision in assessment, unfamiliarity with the problem at hand, a lack of data or the absence of certain stakeholders in a group decision (Durbach and Stewart, 2012). Moreover, by using distribution assessment, the belief structure in question can capture precise data and as well as different types of uncertainties, such as probabilities and ambiguity in subjective judgments. As such, when modelling uncertainty by belief structure, D-S theory is more flexible and versatile than the traditional Bayes theory, where probabilities can only be assigned to individual hypotheses, instead of providing an explicit mechanism for dealing with ignorance (Xu et al., 2006). Belief structure was introduced to MCDM by Yang and Singh (1994) in an Evidential Reasoning approach, since then, there have been a plethora of studies into the belief structure (Beynon et al., 2000; Xu et al., 2006; Yang et al., 2006; Ng and Chuah, 2014; Zhou et

[6] From https://bestworstmethod.com/papers-and-slides/

al., 2018) and its extensions (Yang and Xu, 2002, Guo et al., 2009). A recent study has tried to extend BWM to the belief structure (Fei et al., 2020), however, since it uses pignistic probability function and weighted sum method to obtain an intermediate value, which is then used as input of the BWM, essentially speaking, it makes no change to the original BWM.

Next to uncertainty, complexity is another important issue that is considered in MCDM (including the BWM). In real-world decisions, it is difficult for a single DM to take all the relevant aspects of a decision-making problem into account. As a result, a group of DMs from different areas provides the advantages of synergy and information-sharing compared to the decisions that are made by a single individual. Thus, many of the decision–making processes that occur in the real world involve group settings designed to make the decision-making process more comprehensive and rational. Multi-criteria group decision making (MCGDM), in which multiple DMs provide their evaluations regarding all the criteria of a decision-making problem, has been one of the most important and promising parts in modern decision-making theory (Li et al., 2016; Zhang et al., 2019).

To date, several extensions of the BWM to group decision-making have been proposed (Jia and Wang, 2016; Mou et al., 2016; You et al., 2016; Hafezalkotob and Hafezalkotob, 2017; Mou et al., 2017; Safarzadeh et al., 2018; Hafezalkotob et al., 2019; Hajek and Froelich, 2019; Mohammadi and Rezaei, 2019). However, in the existing group BWM approaches, the impact of the reliability of DMs is underestimated and rarely considered. The reliability of DMs in group BWM can be defined as their ability to provide a *certain* and *consistent* evaluation using pairwise comparisons. In the existing MCGDM research, the experts or DMs are usually assumed to be both rational and reliable. However, according to Simon (Simon, 1955; 1956), our rationality is bounded due to our limited computational ability, selective memory and perception. As such, the judgements expressed by DMs in the BWM may be inconsistent and include some degree of uncertainty and imprecision (Hafezalkotob and Hafezalkotob, 2017). Also, because the DMs reliability has a significant impact on the rationality and validity of the results (Fu et al., 2015), neglecting it could lead to system accidents (Wang et al., 2011). In other words, being able to measure that reliability effectively and apply it within the group aggregation process is significantly important to the group BWM.

The objective of this study is to incorporate information regarding the belief structure into BWM and enable the method to handle the opinions of a group of experts. Specifically, the belief structure preference is applied to pairwise comparisons and the original BWM is extended to handle that type of information. In order to solve the multiple optimal solutions problem of the nonlinear model, two models are used to obtain the boundary of the weights. Moreover, in order to check the reliability of DMs when they apply belief structure during the elicitation process, a reliability degree is defined based on the inconsistency and uncertainty levels of the DMs in question, and a group belief BWM framework is proposed. With the reliability degrees of DMs, the final weights of criteria can be determined by integrating the criteria and the weights obtained from each individual.

The remainder of this study is organized as follows. In Section 3.2, the original BWM is reviewed and the concept of belief structure introduced. In Section 3.3, new BWM models are proposed to deal with belief structure preferences. In Section 3.4, a reliability measurement is proposed based on the inconsistency measurement (or consistency measurement) and the uncertainty measurement (or certainty measurement). The proposed method is then extended to include group decision-making problems, in Section 3.5. In Section 3.6, an application to the evaluation of the infrastructure project criteria system in Indonesia is provided to demonstrate the applicability and feasibility of the proposed method. Finally, some concluding remarks are presented in Section 3.7.

3.2 Preliminaries

In this section, we review the original BWM and discuss the basic terminology and definitions of the belief structure in D-S theory. The overall uncertainty measurement of belief structure, designed to measure random and non-specific uncertainty, is also introduced.

3.2.1 The original BWM

As a weighting method based on pairwise comparisons, the BWM uses ratios of the relative importance of criteria in pairs, as estimated by a DM, from two evaluation vectors, the best criterion in relation to the other criteria, and the other criteria in relation to the worst criterion, whereby the weights of the criteria can be obtained by solving an optimization problem (Rezaei, 2015). The basic steps of original BWM can be summarized as follows:

Step 1. Determine the set of evaluation criteria $\{C_1, C_2, \cdots, C_n\}$.

Step 2. Determine the best (e.g. the most influential or the most important) and the worst (e.g. the least influential or the least important) criteria.

Step 3. Determine the preferences of the best criterion over all the other criteria, using a number between 1 to 9. The obtained Best-to-Others (BO) vector is: $A_{BO} = (a_{B1}, a_{B2}, \cdots, a_{Bn})$, where $.a_{Bj}.$ represents the preference of the best criterion C_B over other criterion $C_j, j = 1,2, \cdots, n$.

Step 4. Determine the preferences of all the criteria over the worst criterion. The obtained Others-to-Worst (OW) vector is: $A_{OW} = (a_{1W}, a_{2W}, \cdots, a_{nW})$, where a_{jW} represents the preference of other criterion C_j over the worst criterion C_W, $j = 1,2, \cdots, n$.

Step 5. Determine the weights $(w_1^*, w_2^*, \cdots, w_n^*)$ by solving the following model:

$$\begin{aligned}
&\min \ \max_j \left\{ \left| \frac{w_B}{w_j} - a_{Bj} \right|, \left| \frac{w_j}{w_W} - a_{jW} \right| \right\} \\
&s.\ t. \\
&\textstyle\sum_{j=1}^{n} w_j = 1 \\
&w_j \geq 0 \text{ , for all } j.
\end{aligned} \tag{3.1}$$

Model (3.1) can be transferred into the following model:

$$\begin{aligned}
&\min \xi \\
&s.\ t. \\
&\left| \frac{w_B}{w_j} - a_{Bj} \right| \leq \xi. \text{, for all } j \\
&\left| \frac{w_j}{w_W} - a_{jW} \right| \leq \xi, \text{ for all } j \\
&\sum_{j=1}^{n} w_j = 1 \\
&w_j \geq 0, \text{ for all } j.
\end{aligned} \tag{3.2}$$

When the preferences are not fully consistent, the nonlinear model (3.2) usually generates multiple optimal solutions. Rezaei (2016) proposed two models to derive interval weights, which include all the possible solutions, as well as a linear alternative designed to obtain a unique solution.

3.2.2 Belief structure

The basic concepts of belief structure is introduced in this part. The pignistic probability function and uncertainty measurement for belief structure are also discussed here, to be used at a later point.

Basic terminology

Suppose the DM is using a finite set of assessment grades $\Omega = \{h_1, h_2, \cdots, h_K\}$ to express his preferences, which is commonly called *frame of discernment* in the D-S theory. These grades are mutually exclusive and collectively exhaustive for all of the evaluations. The power set of Ω, which is the set of all the subsets of Ω, can be presented as:

$2^{\Omega} = \{H_l\} = \{H_1, H_2, \ldots, H_{2^K}\} = \{\emptyset, \{h_1\}, \ldots, \{h_K\}, \{h_1, h_2\}, \ldots, \{h_1, h_K\}, \ldots, \{h_1, \ldots, h_{K-1}\}, \Omega\}$, $l = 1, \cdots, 2^K$

Definition 7 (Shafer, 1976): A *basic probability assignment* to all subsets H_l of 2^{Ω} is a function $m: 2^{\Omega} \rightarrow [0,1]$, which satisfies:

$m(\emptyset) = 0$ and $\sum_{H_l \in 2^{\Omega}} m(H_l) = 1$.

The value $m(H_l)$ is assigned only to the set H_l and not to a smaller subset. Any subset H_l with $m(H_l) > 0$ is called a *focal element*. The set of all the focal elements is denoted with F. The pair $\langle F, m \rangle$ is called the body of evidence.

Based on the degree of belief, some other measures of confidence can be defined.

A *belief measure* is a function $Bel: 2^{\Omega} \rightarrow [0,1]$, which represents our confidence that the concerned element belongs to H or any of its subsets B and is defined by:

$$Bel(H_l) = \sum_{B \subseteq H_l} m(B). \tag{3.3}$$

A *plausibility measure* is a function $Pls: 2^{\Omega} \rightarrow [0,1]$, defined by:

$$Pls(H_l) = \sum_{B \cap H_l \neq \emptyset} m(B). \tag{3.4}$$

$Pls(H_l)$ represents the extent to which we fail to disbelieve H_l. Thus, $Bel(H_l)$ and $Pls(H_l)$ can be interpreted as the lower and upper bound of probability to which H_l is supported (Yager, 1987).

Definition 8 (Smets and Kennes, 1994): For a $m(H_l)$ on 2^{Ω}, its associated *pignistic probability function* $\beta_m: \Omega \rightarrow [0,1]$ is defined as:

$$\beta(h_k) = \sum_{H_l: h_k \in H_l} \frac{m(H_l)}{|H_l|}, \tag{3.5}$$

where $|H_l|$ is the cardinality of H_l.

The principle underlying the pignistic probability function is called the generalized insufficient reason principle, because the insufficient reason principle is used at the level of each focal element of the belief function. This pignistic probability can be interpreted as the degree of belief in each element of the frame of discernment Ω.

Uncertainty measurement for belief structure

A noteworthy uncertainty measure called the *Aggregated Uncertainty (AU) measure*, which was proposed by Harmanec and Klir (1994) to quantify the total uncertainty of a belief function, is adopted in this paper to measure the uncertainty of the given preferences, because it can measure both discord and non-specificity, and it satisfies all the basic requirements for a meaningful measure of aggregate uncertainty in evidence theory .

Definition 9 (Harmanec and Klir, 1994): Let Ω be a finite frame of discernment, and Bel be a belief measure on Ω. The *Aggregated Uncertainty* AU associated with Bel is measured by:

$$AU(Bel) = \max_{p_x \text{ consistent with } Bel} [-\textstyle\sum_{x\in\Omega} p_x \log_2 p_x], \tag{3.6}$$

where the maximum is taken over all distributions $\{p_x\}_{x\in\Omega}$ that are consistent with Bel, and $\{p_x\}_{x\in\Omega}$ should satisfy the following constraints:

$$s.t. \begin{cases} p_x \in [0,1], \ \forall x \in \Omega \\ \sum_{x\in\Omega} p_x = 1 \\ Bel(A) \le \sum_{x\in A} p_x \le Pls(A), \ \forall A \subseteq \Omega \end{cases}.$$

As can be seen from the definition, AU is the maximum (upper) Shannon entropy of all probability distribution under the constraints according to the given basic belief assignments. This measure can capture both non-specificity and discord, and it is a well-justified method to measure uncertainty within the D-S theory. It has been proven that AU satisfies a number of reasonable properties for uncertainty measures in evidential theory (Abellán and Masegosa, 2008).

3.3 The BWM with belief structure

In this section, a belief-based BWM is proposed to deal with uncertain information by using belief structures. Because the proposed method may generate multiple optimal solutions, we introduce a method to obtain the interval weights that can comprise all the possible solutions.

3.3.1 Belief structure for pairwise comparison

Suppose a finite set of assessments $\Omega = \{h_1, h_2, \cdots, h_K\}$ is used by a DM to provide his pairwise comparison preferences, these assessments are assumed to be mutually exclusive.

In BWM, a set of 1-9 grades is usually defined to determine the preference of one criterion over another, to show their relative importance, serving as the frame of discernment:

$\Omega^{Importance} = \{h_1, \ h_2, \ h_3, \ h_4, \ h_5, \ h_6, \ h_7, \ h_8, \ h_9\}$.

Each element in this frame of discernment refers to a verbal judgment and a scale, as shown in Table 3-1.

Table 3-1. The linguistic terms and scales for the importance of pairwise comparisons

Grades	Verbal description	Scales
h_1	Equally important	1
h_2	Equally to slightly more important	2
h_3	Slightly more important	3
h_4	Slightly to strongly more important	4
h_5	Strongly more important	5
h_6	Strongly to very strongly more important	6
h_7	Very strongly more important	7
h_8	Very strongly to extremely more important	8
h_9	Extremely more important	9

After determining the frame of discernment, the DM can compare criteria C_i to C_j with subset H_l $(l = 1, \cdots, 2^K)$ from 2^{Ω} to evaluate his preference and assign $m_{l,ij}$ (instead of calling this the *basic probability assignment*, we call it *basic belief assignment* in BWM*)* to express his basic belief degree with regard to $H_l \subseteq \Omega$.

By using the pignistic probability function in Definition 8, the belief degree $\beta_{k,ij}$ (pignistic probability) associated to each grade $h_{k,ij}$ when comparing criteria C_i to C_j under the frame of discernment Ω can be obtained:

$$\beta_{k,ij} = \sum_{H_l:h_{k,ij} \in H_l} \frac{m(H_l)}{|H_l|}, \tag{3.7}$$

Then the pair of assessment of each grade $h_{k,ij}$ $(h_{k,ij} \in H_l)$ and the *belief degree* $\beta_{k,ij}$ ($\langle h_{k,ij}, \beta_{k,ij} \rangle$) form the *body of assessment* (similar to the body of evidence in D-S theory), which can be profiled by a belief structure (denoted as S_{ij}):

$$S_{ij} = \{(h_{k,ij}, \beta_{k,ij}),\ k = 1, \cdots, K\}. \tag{3.8}$$

Example 1. When a DM wants to buy a car and compares the relative importance of the criterion price over criterion style, suppose he decides to take $\Omega = \{h_1, h_2, h_3\}$={Equally important ($h_1$), equally to slightly more important (h_2), slightly more important (h_3)} as the frame of discernment, then he constructs his belief evaluations as:

m: $m\{\emptyset\} = 0$, $m\{h_1\} = 0$, $m\{h_2\} = 0$, $m\{h_3\} = 0.6$, $m\{h_1, h_2\} = 0$, $m\{h_1, h_3\} = 0$, $m\{h_2, h_3\} = 0.1$, $m\{h_1, h_2, h_3\} = 0.3$,

which means he is 60% sure that the price is *slightly more important than* style (grade h_3), 10% sure on grades h_2 and h_3, which leaves 30% belief for the remaining set, which represents his degree of ignorance.

According to Equation (3.7), the belief degree (β_k) to each grade h_k can be computed as:

$$\beta_1 = m\{h_1\} + \frac{m\{h_1,h_2\}}{2} + \frac{m\{h_1,h_3\}}{2} + \frac{m\{h_1,h_2,h_3\}}{3} = 0 + 0 + 0 + 0.1 = 0.1,$$

$$\beta_2 = m\{h_2\} + \frac{m\{h_1,h_2\}}{2} + \frac{m\{h_2,h_3\}}{2} + \frac{m\{h_1,h_2,h_3\}}{3} = 0 + 0 + 0.05 + 0.1 = 0.15,$$

$$\beta_3 = m\{h_3\} + \frac{m\{h_1,h_3\}}{2} + \frac{m\{h_2,h_3\}}{2} + \frac{m\{h_1,h_2,h_3\}}{3} = 0.6 + 0 + 0.05 + 0.1 = 0.75.$$

Then the belief structure of comparing criterion price over criterion style can be constructed as:

$S_{price,style} = \{(h_1, \beta_1), (h_2, \beta_2), (h_3, \beta_3)\} = \{(h_1, 0.1), (h_2, 0.15), (h_3, 0.75)\}$.

3.3.2 The procedure of BWM with belief structure

To incorporate the belief structure into the BWM, the model's procedure can be provided as follows:

Step 1. DM determines the set of evaluation criteria and the frame of discernment.

To evaluate a MCDM problem, the DM should identify the corresponding set of criteria to evaluate the performance of the alternatives involved. Here, we suppose there are n criteria $C = \{C_1, C_2, \cdots, C_n\}$. A set of grades are identified by DMs to evaluate the pairwise comparisons, assuming that the frame of discernment consists of K grades: $\Omega = \{h_1, h_2, \cdots, h_K\}$.

Step 2. DM selects the best (e.g. the most influential or the most important) and the worst (e.g. the least influential or the least important) criteria.

In this step, the DM is asked to identify the best and worst criteria, based on the criteria set. The best criterion is represented as C_B, the worst criterion as C_W.

Step 3. DM assigns the preference of the best criterion over all the other criteria, with basic belief assignments.

The DM needs to provide his preferences in comparing the best criterion C_B to the other criteria C_j under the set of identified assessment grade Ω. The entire subset H_l of 2^Ω will be complemented with the basic belief assignment $m_{l,Bj} \in [0,1]$. The subsets with $m_{l,Bj} > 0$ make up the body of assessment.

Step 4. DM assigns the preference of all the other criteria over the worst criterion, with basic belief assignments.

The DMs assigns basic belief scores ($m_{l,jW}$) to the entire subset H_l of 2^Ω when comparing the other criteria C_j to the worst criterion C_W. The body of assessment is made up by the subsets with $m_{l,jW} > 0$.

Step 5. Construct belief structures according to the pignistic probability function.

Determine the belief degree $\beta_{k,Bj}$ to each grade $h_{k,Bj}$ by using the pignistic probability function Equation (3.7), after which the belief structure involved in comparing the best criterion to the others can be constructed as:

$S_{Bj} = \{(h_{k,Bj}, \beta_{k,Bj}),\ k = 1, \cdots, K\}$.

The resulting Best-to-Others (BO) vector is: $S_B = (S_{B1}, S_{B2}, \cdots, S_{Bn})$, where S_{Bj} represents the preference of the best criterion C_B over the other criterion $C_j, j = 1,2, \cdots, n$.

Similarly, the belief structure of comparing the others to the worst criterion can be constructed as:

$S_{jW} = \{(h_{k,jW}, \beta_{k,jW}),\ k = 1, \cdots, K\}$.

The resulting Others-to-Worst (OW) vector is: $S_W = (S_{1W}, S_{2W}, \cdots, S_{nW})$, where S_{jW} represents the preference of other criterion C_j over the worst criterion C_W, $j = 1,2, \cdots, n$.

Step 6. Determine the weights $(w_1^*, w_2^*, \cdots, w_n^*)$.

To determine the optimal weights with respect to a belief structure, we need to make each pair of $\frac{w_B}{w_j}$ and $\frac{w_j}{w_W}$ as close as possible to the grade $h_{k,Bj}^*$ ($h_{k,jW}^*$) with the maximum belief degree $\beta_{k,Bj}^*$ ($\beta_{k,jW}^*$) in the corresponding belief structure S_{Bj} (S_{jW}). The underlying idea is that the grade with the higher belief score should be valued more, and the grade with the lower belief score should be valued less. To operate this idea for all j, the maximum difference between $\frac{w_B}{w_j}$ and $h^*_{k,Bj}$ ($\frac{w_j}{w_W}$ and $h^*_{k,jW}$) for all j should be minimized, which means that the constrained optimization problem to determine the optimal weights is constructed as follows:

$$\begin{aligned}
&\min \ \max\left\{\left|\frac{w_B}{w_j} - h_{k,Bj}\right|\beta_{k,Bj}, \left|\frac{w_j}{w_W} - h_{k,jW}\right|\beta_{k,jW}\right\} \\
&s.\ t. \\
&\sum_{j=1}^{n} w_j = 1 \\
&w_j \geq 0, \text{for all } j.
\end{aligned} \tag{3.9}$$

Model (3.9) can be transferred into the following model:

$$\begin{aligned}
&\min \ \xi \\
&s.\ t. \\
&\left|\frac{w_B}{w_j} - h_{k,Bj}\right|\beta_{k,Bj} \leq \xi, \text{for all } j \text{ and } k \\
&\left|\frac{w_j}{w_W} - h_{k,jW}\right|\beta_{k,jW} \leq \xi, \text{for all } j \text{ and } k \\
&\sum_{j=1}^{n} w_j = 1 \\
&w_j \geq 0, \text{for all } j.
\end{aligned} \tag{3.10}$$

Solving problem (3.10), the optimal weights $(w_1^*, w_2^*, \dots, w_n^*)$ are obtained. The optimal value ξ^* obtained from this program indicates that the closer it is to 0, the more consistent the DM is.

Example 2. We use the same case that was studied by Rezaei (2016) and suppose that the frame of discernment is $\Omega = \{h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8\}$ (Step 1). The second criterion, Price (C_2), is identified as the best criterion, and the fifth criterion Style (C_5) is identified as the worst criterion (Step 2). Next, the DM provides his basic belief assignments (only values for focal elements are listed) with regard to the best criterion compared to the others, and the other criteria compared to the worst, as seen in Table 3-2 and Table 3-3 (Steps 3 and 4):

Table 3-2. Assessments of the Best criterion to the others

Best to Others	Best criterion: C_2
Quality (C_1)	$m_{21}\{h_2, h_3\} = 1$
Price (C_2)	$m_{22}\{h_1\} = 1$
Comfort (C_3)	$m_{23}\{h_3, h_4, h_5, h_6, h_7\} = 0.3$, $m_{23}\{h_4, h_5, h_6\} = 0.5$, $m_{23}\{h_2, h_3, h_4\} = 0.1$, $m_{23}\{\Omega\} = 0.1$
Safety (C_4)	$m_{24}\{h_2\} = 0.6$, $m_{24}\{h_2, h_3\} = 0.4$
Style (C_5)	$m_{25}\{h_8\} = 0.8$, $m_{25}\{h_6, h_7, h_8\} = 0.2$

Table 3-3. Assessments of the other criteria to the Worst

Others to Worst	Worst criterion: C_5
Quality (C_1)	$m_{15}\{h_4\} = 1$
Price (C_2)	$m_{25}\{h_7, h_8\} = 0.6,\ m_{25}\{h_8\} = 0.2,\ m_{25}\{\Omega\} = 0.2$
Comfort (C_3)	$m_{35}\{h_2, h_3, h_4\} = 0.7,\ m_{35}\{h_3\} = 0.3$
Safety (C_4)	$m_{45}\{h_4\} = 0.8,\ m_{45}\{h_3, h_4, h_5\} = 0.2$
Style (C_5)	$m_{55}\{h_1\} = 1$

After applying the pignistic probability function (3.7), the basic belief assignments can be transformed into belief structures (Step 5):

$S_{21} = \{(h_2, 0.5), (h_3, 0.5)\ \}$

$S_{22} = \{(h_1, 1)\ \}$

$$S_{23} = \begin{Bmatrix} (h_1, 0.013), (h_2, 0.046), (h_3, 0.106), (h_4, 0.276), \\ (h_5, 0.239), (h_6, 0.239), (h_7, 0.073), (h_8, 0.013) \end{Bmatrix}$$

$S_{24} = \{(h_2, 0.8), (h_3, 0.2)\ \}$

$S_{25} = \{(h_6, 0.067), (h_7, 0.067), (h_8, 0.867)\ \}$

$S_{15} = \{(h_4, 1)\ \}$

$$S_{25} = \begin{Bmatrix} (h_1, 0.025), (h_2, 0.025), (h_3, 0.025), (h_4, 0.025), \\ (h_5, 0.025), (h_6, 0.025), (h_7, 0.325), (h_8, 0.525) \end{Bmatrix}$$

$S_{35} = \{(h_2, 0.233), (h_3, 0.533), (h_4, 0.233)\ \}$

$S_{45} = \{(h_3, 0.067), (h_4, 0.867), (h_5, 0.067)\ \}$

$S_{55} = \{(h_1, 1)\ \}$

Figure 3-1 visualizes the distribution of the belief degrees involving each individual grade. For example, the belief structure S_{21} has 0.5 belief degree on grade 2 and grade 3 respectively.

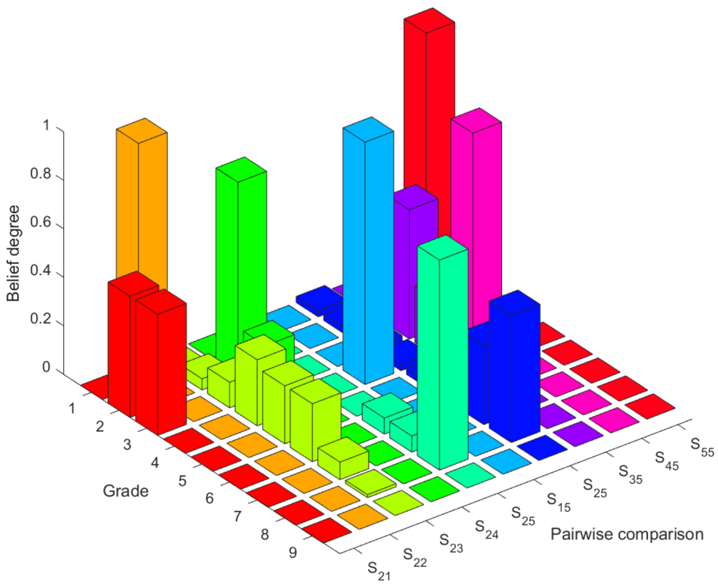


Figure 3-1. The distribution of belief degrees in Example 2

Solving the optimization problem, we can obtain one set of the optimal weights and ξ^* as follows (Step 6):

$w_1^* = 0.196,\ w_2^* = 0.453,\ w_3^* = 0.113,\ w_4^* = 0.185,\ w_5^* = 0.054$, and $\xi^* = 0.475$.

The multiple optimal solutions issue is addressed in Section 3.3.3.

From the results, we can give another interpretation to the belief-based BWM. For instance, in the assessment C_2 over C_1, $m_{21}\{h_2, h_3\} = 1$, $S_{21} = \{(h_2, 0.5), (h_3, 0.5)\}$, the DM hesitates between h_2 and h_3, and the result $a_{21}^* = \frac{w_2^*}{w_1^*} = 2.5$ can capture this hesitation, since it lies in the middle. Also, for assessment C_4 over C_5, the basic belief assignment is $m_{45}\{h_4\} = 0.8$, $m_{45}\{h_3, h_4, h_5\} = 0.2$, so we expect the result can focus more on h_4 because the DM has expressed greater belief and certainty, instead of h_3 and h_5. The result of C_4 over C_5 is $a_{45}^* = \frac{w_4^*}{w_5^*} = 4$, which shows that it weighs more the strongest belief h_4.

The algorithm and analysis present the features of the belief-based BWM. The method not only allows a DM to provide his basic belief assignments in a more flexible way, it also balances the hesitation of the DM, taking all the preferences and beliefs into account and trying to come closer to the preferences with stronger beliefs and move further away from preferences associated with weaker beliefs.

If each belief structure provided by a DM is 100% sure on one single grade, that would mean the DM has no uncertainty at all, and this belief structure-based BWM in essence becomes the original BWM.

3.3.3 Models to derive interval weights

The nonlinear BWM can have multiple optimal solutions when the pairwise comparisons are not fully consistent. In order to handle that problem, we propose a method to obtain the minimum and maximum weights of each criterion. Two models are proposed to calculate the lower and upper bounds of the weights of criterion C_j based on the ξ^*, that is the optimal solution of models (3.9) and (3.10).

$$
\begin{aligned}
&\min \ w_j \\
&s.\ t. \\
&\left|\frac{w_B}{w_j} - h_{k,Bj}\right| \beta_{k,Bj} \le \xi^*, \text{for all } j \text{ and } k \\
&\left|\frac{w_j}{w_W} - h_{k,jW}\right| \beta_{k,jW} \le \xi^*, \text{for all } j \text{ and } k \\
&\sum_{j=1}^{n} w_j = 1 \\
&w_j \ge 0, \text{for all } j,
\end{aligned}
\tag{3.11}
$$

$$\begin{aligned} &\max \ w_j \\ &s.\ t. \\ &\left|\frac{w_B}{w_j} - h_{k,Bj}\right| \beta_{k,Bj} \leq \xi^*, \text{for all } j \text{ and } k \\ &\left|\frac{w_j}{w_W} - h_{k,jW}\right| \beta_{k,jW} \leq \xi^*, \text{for all } j \text{ and } k \\ &\sum_{j=1}^{n} w_j = 1 \\ &w_j \geq 0, \text{for all } j. \end{aligned} \tag{3.12}$$

After solving these two models for all criteria, the optimal value of the objective function of (3.11) is taken as the minimum w_j^{*-} and, similarly, the optimal value of (3.12) is the maximum w_j^{*+}. Together, they identify intervals $[w_j^{*-}, w_j^{*+}]$. For the operations of interval weights and the method of ranking the criteria, the reader might refer to (Rezaei, 2016).

Example 3. From Example 2, we obtain $\xi^* = 0.4753$, which indicates that the system of pairwise comparisons is not fully consistent, and the nonlinear belief-based BWM model can generate multiple optimal solutions. To solve that problem, we use the interval weights to contain all the possible solutions. The optimal interval weights of belief-based BWM obtained thanks to the optimization problems (3.11) and (3.12), are:

$w_1^* = [0.178,0.216]$, $w_2^* = [0.418,0.456]$, $w_3^* = [0.104,0.114]$, $w_4^* = [0.18,0.2312]$, $w_5^* = [0.049，0.054]$.

The mean of the all the optimal intervals can be used to indicate the middle position of these interval weights, the result being: $w_1^*(\text{mean}) = 0.195$, $w_2^*(\text{mean}) = 0.439$, $w_3^*(\text{mean}) = 0.11$, $w_4^*(\text{mean}) = 0.204$, $w_5^*(\text{mean}) = 0.052$. The interval weights and their means are shown in Figure 3-2.

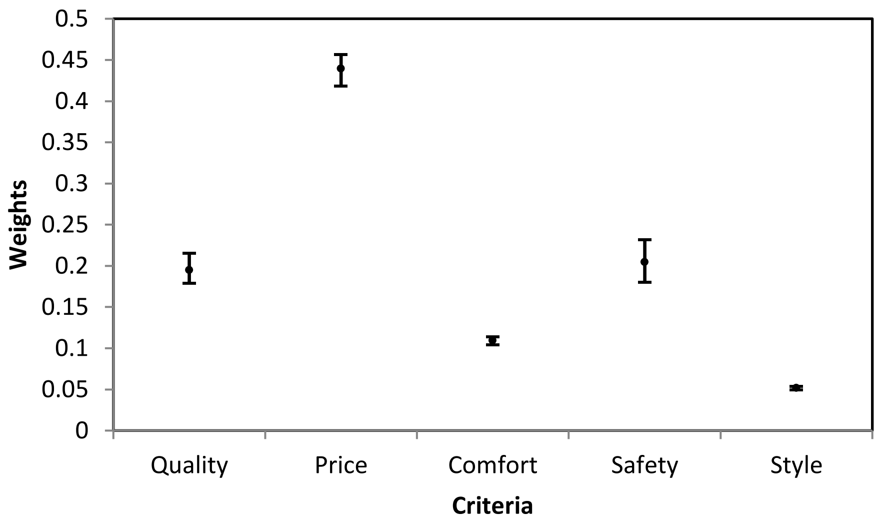


Figure 3-2. The interval weights of belief BWM

3.4 The reliability measurement

After determining the weights, it is very important to check the reliability of the results. It has been a long debate on the measurement of the reliability or expertise of an expert/DM, especially when there is no external standard to verify (Shanteau et al., 2003). Traditionally, the reliability of an expert is measured by the consensus with the other experts (Fu et al., 2015; Du et al., 2018). However, according to psychological investigations and empirical studies (Shanteau et al., 2003; Weiss and Shanteau, 2004), the agreement with other experts is neither necessary nor sufficient for expertise, rather, intra-individual consistency is a necessity.

Besides, the uncertainty degree of an expert is also highly related to his/her reliability. For example, if an expert provides a preference profile like {(1,0.5),(9,0.5)}, or like the highly nonspecific belief distributions {({1,2,3,4,5},0.5),({5,6,7,8,9},0.5)}, the expert faces randomness and the non-specificity problems (Pal et al., 1992; Klir and Wierman, 1999). Both cases could yield unreliable results, because the expert essentially has not provided sufficient information for a decision.

Therefore, in this section, we discuss a method designed to measure the reliability degree of an expert's judgments incorporate his inconsistency and uncertainty levels. To that end, an inconsistency measurement and an uncertainty measurement are proposed based on belief structure-based BWM.

3.4.1 The inconsistency measurement for belief BWM

The original BWM uses pairwise comparisons of criteria based on DMs' evaluations of the relative priorities of decision-making elements. As such, the pairwise comparisons are said to be perfectly (cardinal-) consistent if they satisfy the transitivity condition $a_{Bj} \times a_{jW} = a_{BW}$; otherwise, the DM is not fully consistent, which may imply some irrationality in the relative weight estimates (Kou et al., 2014).

In belief-based BWM, to handle the information of belief structures, the utility-based approach (Yang, 2001) can be adopted to compute the value of belief structures. The expected utility of a belief structure S_{ij} is noted as u_{ij}, and can be computed as follows:

$$u_{ij} = \sum_{k=1}^{K} u(h_k)\beta_{k,ij}, \tag{3.13}$$

where $u(h_k) = k$. Then the value of a_{ij} in the original BWM can be replaced by the expected utility u_{ij}, thus the transitivity condition is transformed into:

$$u_{Bj} \times u_{jW} = u_{BW}. \tag{3.14}$$

According to the definition of belief structure, suppose the DM identifies a set of evaluation grades $\Omega = \{h_1, h_2, \cdots, h_K\}$ which is applied to pairwise comparisons, then $S_{BW} = (h_K, 1)$ is the maximum belief structure that can generate the highest possible value to u_{BW}. If $u_{Bj} \times u_{jW} \neq u_{BW}$, the inconsistency will occur, whether $u_{Bj} \times u_{jW}$ is higher or lower than u_{BW}. When u_{Bj} and u_{jW} have the highest value, which is equal to u_{BW}, that will result in the largest inequality. According to $\left(\frac{w_B}{w_j}\right) \times \left(\frac{w_j}{w_W}\right) = \frac{w_B}{w_W}$, the following equation can be obtained:

$$(u_{Bj} - \xi) \times (u_{jW} - \xi) = u_{BW} + \xi. \tag{3.15}$$

For the maximum inconsistency of belief structure, $u_{Bj} = u_{jW} = u_{BW}$, Equation (3.16) can be written as :

$$(u_{BW} - \xi) \times (u_{BW} - \xi) = u_{BW} + \xi, \tag{3.16}$$

and formulated as:

$$\xi^2 - (1 + 2u_{BW})\xi + (u_{BW}^2 - u_{BW}) = 0. \tag{3.17}$$

Because $u_{BW} = u(K) = K,\ K \in \{1,2,3,\cdots\}$, Equation (3.17) becomes:

$$\xi^2 - (1 + 2K)\xi + (K^2 - K) = 0. \tag{3.18}$$

After solving Equation (3.18) for different K, the maximum possible ξ can be obtained and used as the inconsistency index for belief-based BWM. The result of the inconsistency index is shown in Table 3-4. The inconsistency index obtained for belief-based BWM is the same as the original BWM, because the u_{BW} in the belief BWM is the same as a_{BW} in the original BWM.

Table 3-4. Inconsistency index table

K	1	2	3	4	5	6	7	8	9
Inconsistency index	0	0.44	1	1.63	2.3	3	3.73	4.47	5.23

We can now use the ξ^* obtained from the belief-based BWM models (3.10) to calculate the Inconsistency Ratio (IR)[7]:

$$IR = \frac{\xi^*}{Inconsistency\ Index}. \tag{3.19}$$

$IR \in [0,1]$, and the closer IR is to 0, the more consistent the judgments are. When $IR = 0$, the judgments of a DM are said to be *fully consistent*.

3.4.2 The uncertainty measurement for belief BWM

We stated earlier that the advantage of the belief-based BWM is the way it deals with uncertain preferences. However, it is important to quantify this very same uncertainty as it can be related with the reliability and the stability of the final results. The final goal of such an analysis would be to identify excessively uncertain preferences.

The measure of uncertainty for the belief-based BWM can be formulated as:

$$AU(Bel) = \max_{p_{k,ij}\ consistent\ with\ Bel} \left[-\sum_{k \in \Omega} p_{k,ij} \, log_2 \, p_{k,ij} \right]$$
$$s.t. \begin{cases} p_{k,ij} \in [0,1],\ \forall k \in \Omega \\ \sum_{k \in \Omega} p_{k,ij} = 1 \\ Bel(H_l) \le \sum_{k \in H_l} p_{k,ij} \le Pls(H_l),\ \forall H_l \subseteq \Omega \end{cases}. \tag{3.20}$$

[7] In case of K=1, the preferences are always fully consistent, hence the IR is zero.

The uncertainty measure algorithm for belief-based BWM

To compute the AU function, an algorithm was proposed by Harmanec et al.(1996), which, in spite of being proved to be correct by Klir and Wierman (1999), is too complex in some cases, and it is why Liu et al. (2007) proposed using another algorithm to reduce the computational complexity, which, unfortunately, was flawed, and it was subsequently corrected by Huynh and Nakamori (2010) with an improved algorithm. This uncertainty measure for belief structure-based BWM uses Huynh and Nakamori's algorithm (2010), which is presented in its adapted form in Table 3-5.

Table 3-5. The algorithm of uncertainty measurement

Input: The set of focal elements F of belief function Bel and their corresponding basic belief assignments.

Output: $AU(Bel)$, $\{p_k\}_{k\in\Omega}$ such that $AU(Bel) = -\sum_{k\in\Omega} p_{k,Bj}\, log_2\, p_{k,Bj}$ and $AU(Bel) = -\sum_{k\in\Omega} p_{k,jW}\, log_2\, p_{k,jW}$.

1. Initialize $AU(Bel) = 0$.
2. Compute the belief measures for all elements of $U(F)$, which is the union of the focal elements from F.
3. Find a set $H_l \in U(F)$, $(l = 1, \cdots, 2^K)$ such that $Bel(H_l)/|H_l|$ is maximal. If there is more than one such set H_l, the one with the largest cardinality should be selected.
4. For $k \in H_l$, put $p_{k,Bj} = Bel(H_l)/|H_l|$ and $p_{k,jW} = Bel(H_l)/|H_l|$; calculate $AU(Bel) := AU(Bel) - Bel(H_l) \times log_2 p_{k,Bj}$ and $AU(Bel) := AU(Bel) - Bel(H_l) \times log_2 p_{k,jW}$.
5. Set $F^{'} = \{H_f \backslash H_l | H_f \in F\} \backslash \{\emptyset\}$.
   1) If $F^{'} = \emptyset$, stop.
   2) Otherwise, for each $S \in F^{'}$, put

      $m(S) = \sum_{H_f \in F,\ H_f \backslash H_l = S} m(H)$

      and set $F = F^{'}$.
6. If $|F| > 1$, return to step 2.
7. If $|F| = 1$ and $F = \{S\}$, put $p_{k,Bj} = m(S)/|S|$ (or $p_{k,jW} = m(S)/|S|$) and $AU(Bel) := AU(Bel) - m(S) \times log_2 p_{k,Bj}$ and$AU(Bel) := AU(Bel) - m(S) \times log_2 p_{k,jW}$.

Global uncertainty

The AU measure is used to quantify the total uncertainty of a given belief structure. To measure the global uncertainty of a DM, we need to take all the basic belief assignments into consideration. The DM's global uncertainty can be calculated as the average uncertainty of the given preferences:

$$\overline{AU} = \frac{1}{2n-3}\left(\max_{p_{k,Bj}\ consistent\ with\ Bel}\left[-\sum_{k\in\Omega} p_{k,Bj}\, log_2\, p_{k,Bj}\right] + \max_{p_{k,jW}\ consistent\ with\ Bel}\left[-\sum_{k\in\Omega} p_{k,jW}\, log_2\, p_{k,jW}\right]\right) \tag{3.21}$$

To compare the uncertainty degrees of different frames of discernment with different grades, we need to normalize the uncertainty degrees in the interval [0,1]. As the maximum value of $\overline{AU}$ is $log_2 K$, where K is the cardinality of discernment frame, the normalization of $\overline{AU}$ can be formulated as follows:

$$\widetilde{AU} = \frac{\overline{AU}}{log_2 K}. \tag{3.22}$$

The range of $\widetilde{AU}$ is $[0,1]$, the closer $\widetilde{AU}$ is to 0, the more certain the judgments are. When $\widetilde{AU} = 0$, the judgments of a DM are said to be *fully certain*.

3.4.3 The reliability degree

The original BWM considers the reliability of a DM's assessments only through his inconsistency level, regardless of whether they use certain numbers or uncertain terms. However, highly uncertain judgments are unstable and lead to unreliable results. Therefore, in addition to looking at the inconsistency level, the uncertainty level also has to be taken into account to determine the reliability of a DM's judgments. In light of these considerations, we define the following reliability index.

Definition 10: The pairwise comparisons of a DM are said to be *fully reliable* if they are fully consistent and completely certain. The Reliability Degree (RD) of a DM's judgments can be formulated as:

$$RD = 1 - \frac{\sqrt{(IR)^2 + (\widetilde{AU})^2}}{\sqrt{2}}. \tag{3.23}$$

The RD ranges from 0 to 1, and when it is closer to 1, we say that the pairwise comparisons provided by this DM are more reliable, because they are more consistent and more certain. As illustrated in Figure 3-3. When $RD = 1$, the DM is considered to be fully reliable.

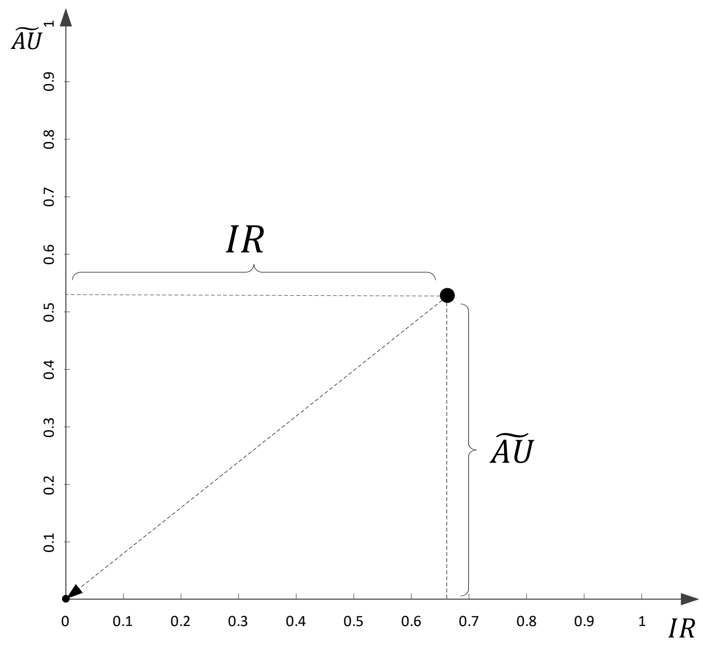


Figure 3-3. The illustration of reliability

Unlike other formulations, e.g. $\frac{(IR+\widetilde{AU})}{2}$, our proposed formula for RD has a clear geometric interpretation: it is the *distance* from the point $(IR,\ \widetilde{AU})$ to (1, 1). In case of considering expertise of DM as part of the reliability degree, we can use a generalized form: $Generalized\ reliability\ degree =\ \alpha RD + (1-\alpha) expertise\ degree$, where $expertise\ degree$ and $\alpha\ \in [0,1]$.

3.5 Group BWM with belief structure

Due to the complexity of MCDM problems, it is common for several experts from different fields to form a group to assess the problems together. In addition, if the problems involve more than one stakeholder or multiple decision-makers, a group decision-making method is needed to aggregate the individual preferences.

The existing aggregation methods for group-based BWM rarely take the reliability level of the DMs' judgments into account. As discussed in Section 3.4, we assume that the inconsistency level and uncertainty level contribute equally to a DM's reliability level. We propose an aggregation method for the group- and belief structure-based BWM, which uses the reliability degrees to determine suitable weights for the DMs.

We can extend the belief-based BWM proposed in Section 3.3 to multi-criteria group decision-making problems. We assume that the DMs express their preferences honestly, which means that the preferences reflect their inconsistency and uncertainty levels. The procedure is illustrated below and the flowchart of the steps involved is shown in Figure 3-4.

Step 1. The group of DMs $D = \{D_1, D_2, \cdots D_G\}$ negotiate and determine the set of evaluation criteria $\{C_1, C_2, \cdots, C_n\}$ and the frame of discernment, which contains K grades: $\Omega = \{h_1, h_2, \cdots, h_K\}$.

Step 2. Each DM D_g determines his best and worst criteria (C_B^g and C_W^g, respectively).

Step 3. Each DM D_g assesses the best criterion over all the other criteria with basic belief assignments .

Step 4. Each DM D_g assesses all the other criteria over the worst criterion with basic belief assignments .

Step 5. Construct belief structures according to the pignistic probability function in Equation (3.7).

Step 6. The non-linear belief-based program (3.10) and the two decomposed models (3.11) and (3.12) in Section 3.3 are used to find the optimal criteria weights $w_j^g = \{w_1^g, w_2^g, \cdots, w_n^g\}$ for each DM D_g.

Step 7. From the preferences that have been provided, each DM D_g can obtain his IR^g and uncertainty degree $\widetilde{AU}^g$ by using the consistency measurement and uncertainty measurement discussed in Section 3.4.

Step 8. The weight of each DM is assumed to be a function of his reliability degree(s) RD^g obtained by Equation (3.23). Under this assumption, we suggest deriving the weight of each DM D_g (λ^g) by means of:

$$\lambda^g = \frac{RD^g}{\sum_{g=1}^{G} RD^g} \tag{3.24}$$

Step 9. Aggregate all the criteria weights from each DM D_g into an overall weight $\widetilde{w}_j$, which can be calculated by

$$\widetilde{w}_j = \sum_{g=1}^{G} \lambda^g w_j^g. \tag{3.25}$$

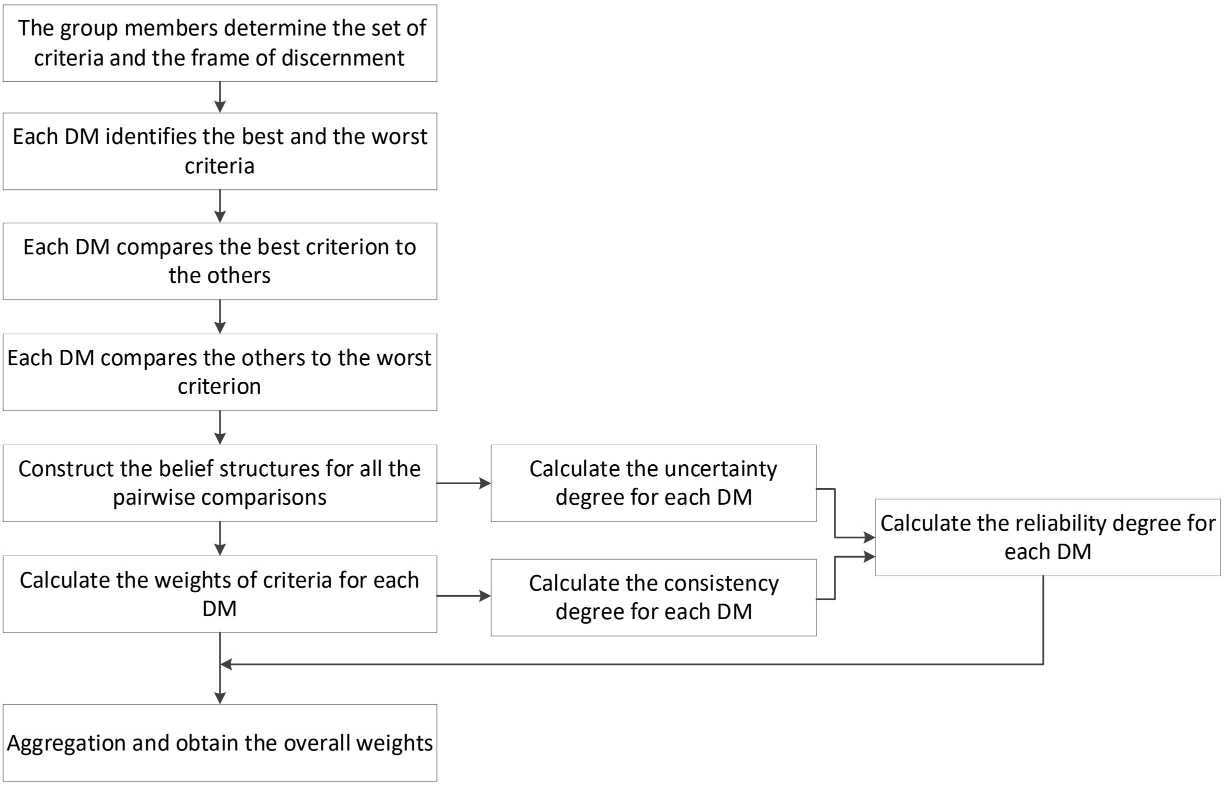


Figure 3-4. The procedure of group BWM with belief structure

3.6 Case Study

As one of the new emerging markets, Indonesia is striving to boost its economic development by making efforts to accelerate strategic projects which can be realized within a short period of time. Each of these projects and programs has its own objectives and responsibility, but due to the lack of coordination between various stakeholders in government and private sectors, there is potential to cause delay to the implementation[8]. Therefore, to deal with this problem, the Committee for Acceleration of Priority Infrastructure Delivery (KPPIP, shortly in Indonesian) was established. The mission of KPPIP is to screen and select the National Strategic Projects, and carry out monitoring activities for National Strategic Projects, as well as conduct high level debottlenecking strategies for Priority Projects (KPPIP, 2017).

Before providing coordination in debottlenecking efforts for the 247 National Strategic Projects and programs, due to limited resources, KPPIP should shortlist 37 projects as priority projects in line with the criteria established by KPPIP (Figure 3-5). KPPIP will then monitor the shortlisted projects and ensure that they comply with quality standards and regulations. This case study will focus only on determining the importance level of the established criteria, not considering the monitoring and implementation part.

[8] The basic information of this case study is from KPPIP's website: https://kppip.go.id/en/

Figure 3-5. List of 37 KPPIP priority projects (KPPIP, 2017)

To support the decision making process, KPPIP is equipped with a Project Management Office (PMO), which comprises of professional experts in their respective fields. These experts are responsible for providing recommendations to the implementation team in selecting priority projects. To evaluate various infrastructure projects, four sectors are formed in KPPIP, i.e. Energy and Electricity sector (EE), Road and Bridge sector (RB), Transportation sector(TT), Water and Sanitation sector(WS). The organizational structure of KPPIP can be seen in Figure 3-6.

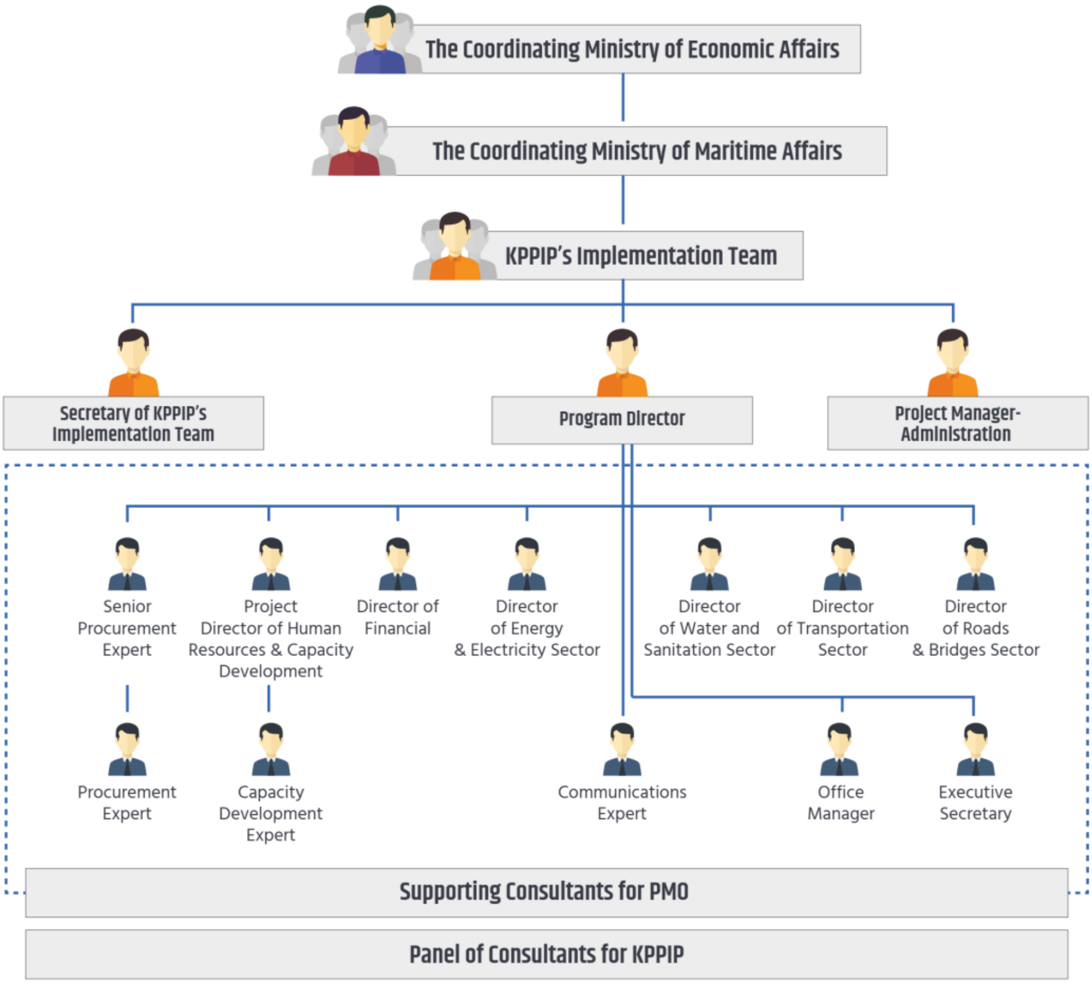


Figure 3-6. The organizational structure of KPPIP (KPPIP, 2017)

The National Strategic Projects are complex to evaluate. After discussion with the experts, 20 criteria were identified by KPPIP and classified into four categories as shown in Table 3-6.Almost all of these criteria were assigned equal weights initially (Executive Direction 0.08, Issuance of project permits and Number of authorities involved are 0.12, the others are all 0.04), which is unreasonable according to an interview we conducted with the leader of KPPIP. In addition to the arbitrariness, the assignment of weights to the criteria did not consider the variety of the four different sectors and the reliability degree of the experts in these sectors. Therefore, the weights of the criteria were suggested to be reevaluated by KPPIP with a more structured/analytic methodology.

In this study, we invited 4 experts from the four sectors (one from each sector - EE, RB, TT, WS) in KPPIP to reevaluate the importance of the given criteria by using the proposed method. They are asked to follow Step 1 to Step 4 of the belief based BWM in Section 3.5 and provided their assessments. Table 3-7 presents the pairwise comparison assessment for the main categories from the four sectors in KPPIP. Table 3-8 to Table 3-11 show the assessment for all the criteria in each category from the four KPPIP sectors.

Table 3-6. The criteria identified by KPPIP

Category	Criteria	Definition
Project Preparation (PP)	PP1: Outline Business Case comprehensiveness	Preliminary thoughts, which contains the information, such as outcomes, benefits, and potential risks associated with the proposal.
	PP2: Economic benefits	It considers the benefit to the economy, environmental, and also social.
	PP3: Technical planning complexity	It considers environments that can bring the project into a complex development, such as land-use plan, environmental dispute, and relocation.
	PP4: Project Development Fund support	A programmatic approach to the funding of the cost for early tasks to encourage contracting agencies to use best practice.
	PP5: Infrastructure readiness/requirement surrounding the project	The government intends to accelerate the development in the country; it is implemented by integrating infrastructures that can carry out or perform more economic activities in the society surrounding the project.
Funding (F)	F1: Acquisition of interest from the investor(s)	Investors are one of the primary sources of funding for the project that is required to develop multiple projects
	F2: Determination of funding scheme	It shows the scheme of the funding which considers the strategic issue on the availability of the investors' interest.
	F3: Funding resources synchronization	It is a body provided by the central government when it is needed to assist the team in organizing the funding of the project.
	F4: Public Sector Organization structuration	Public Sector Organization is an entity that is formed to manage the policy and operating requirements that enable a government to achieve its goals of public governance.
	F5: Granting of credit risk	Credit risk is the possibility of a loss resulting from a debtor's failure to meet the obligations. It measures the availability of the assurance of the projects.
	F6: Granting of business feasibility support	The business feasibility support refers to the availability of elements, which support the continuity of the project development.
Coordination (C)	C1: Stakeholder Buy-in	Process of involving all the related stakeholders to reach consensus.
	C2: Land acquisition coordination	Most infrastructure projects mostly involve many areas to be cleared for the project and needs complicated coordination.
	C3: Spatial plan synchronization	Most infrastructure projects involve many areas, and sometimes, it has a different land-use plan that can create some dispute.
	C4: Number of authorities involved	It refers to the complexity that happens due to the administrative and coordination time needed in the project.
	C5: Implementation of procurement between government and business entity	Some projects involve coordination between private parties or other stakeholders that do not have or little experience in the field or can be due to an innovative project.
	C6: Synchronization with other National Strategic Projects	This criterion intends to synchronize between two or more National Strategic Projects that relate to each other.
Policy (P)	P1: Executive Direction	The president's vision of the country to distribute the wealth to the society, and have national-range impacts.
	P2: Publishing of supporting policies	The government try to accelerate the development by publishing some sectoral/general policies.
	P3: Issuance of project permits	A project permit is a critical milestone of the project, which lets the team continue/start/operate an activity in the project.

In this case, the experts were suggested to evaluate the criteria by assigning each of the basic belief assignments to only one grade (for the sake of simplicity), and the unassigned degree represents the ignorance. For example, the assessment {(2; 0,3), (3; 0,7)} and {(5; 0,5), (Ω; 0,5)} in the bottom right in Table 3-7, can be interpreted as: the Water and Sanitation sector compared the best category, which is Project Preparation (PP), to Policy (P) with 30% confidence that Project Preparation is equally to slightly more important (grade h_2) than Policy, and 70% confidence that Project Preparation is slightly more important (grade h_3) than Policy; This sector compared Policy to the worst category, which is Coordination (C), with 50% confidence that Policy is strongly more important (grade h_5) than Coordination, and the remaining 50% allocated to ignorance.

Table 3-7. The KPPIP main category assessment from the four sectors.

EE			RB			TT			WS		
Category	Best to Others	Others to Worst	Category	Best to Others	Others to Worst	Category	Best to Others	Others to Worst	Category	Best to Others	Others to Worst
PPB	{(1; 1)}	{(7; 1)}	PPB	{(1; 1)}	{(7; 0,7), (Ω; 0,3)}	PPB	{(1; 1)}	{(2; 0,2), (3; 0,8)}	PPB	{(1; 1)}	{(5; 0,8), (Ω; 0,2)}
F	{(1; 1)}	{(7; 1)}	F	{(3; 0,7), (Ω; 0,3)}	{(3; 0,8), (Ω; 0,2)}	F	{(1; 0,2), (2; 0,8)}	{(2; 0,8), (3; 0,2)}	F	{(2; 1)}	{(4; 0,7), (Ω; 0,3)}
CW	{(7; 1)}	{(1; 1)}	C	{(5; 0,8), (Ω; 0,2)}	{(1; 0,6), (Ω; 0,4)}	C	{(1; 0,1), (2; 0,9)}	{(2; 0,9), (3; 0,1)}	CW	{(5; 0,8), (Ω; 0,2)}	{(1; 1)}
P	{(1; 1)}	{(7; 1)}	PW	{(7; 0,7), (Ω; 0,3)}	{(1; 1)}	PW	{(2; 0,2), (3; 0,8)}	{(1; 1)}	P	{(2; 0,3), (3; 0,7)}	{(5; 0,5), (Ω; 0,5)}

B = Best category, W = Worst category

Table 3-8. Assessment for all the criteria in Project Preparation category from four sectors.

EE			RB			TT			WS		
Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst
PP1	{(1; 1)}	{(9; 1)}	PP1	{(2; 0,7), (Ω; 0,3)}	{(5; 0,8), (Ω; 0,2)}	PP1^{B}	{(1; 1)}	{(3; 0,1), (4; 0,9)}	PP1	{(2; 0,7), (Ω; 0,3)}	{(4; 0,8), (Ω; 0,2)}
PP2	{(1; 0,2), (2; 0,8)}	{(9; 1)}	PP2	{(3; 0,8), (Ω; 0,2)}	{(6; 0,8), (Ω; 0,2)}	PP2	{(1; 0,1), (2; 0,9)}	{(3; 0,8), (4; 0,2)}	PP2^{W}	{(6; 0,7), (Ω; 0,3)}	{(1; 1)}
PP3	{(1; 0,2), (2; 0,8)}	{(9; 1)}	PP3^{B}	{(1; 1)}	{(7; 0,7), (Ω; 0,3)}	PP3	{(2; 0,2), (3; 0,8)}	{(2; 0,8), (3; 0,2)}	PP3	{(2; 0,8), (Ω; 0,2)}	{(3; 0,8), (Ω; 0,2)}
PP4^{B}	{(1; 1)}	{(9; 1)}	PP4	{(2; 0,7), (Ω; 0,3)}	{(6; 0,7), (Ω; 0,3)}	PP4	{(3; 0,2), (4; 0,8)}	{(2; 0,9), (3; 0,1)}	PP4^{B}	{(1; 1)}	{(6; 0,7), (Ω; 0,3)}
PP5^{W}	{(9; 1)}	{(1; 1)}	PP5^{W}	{(7; 0,7), (Ω; 0,3)}	{(1; 1)}	PP5^{W}	{(4; 0,1), (5; 0,9)}	{(1; 1)}	PP5	{(4; 0,9), (Ω; 0,1)}	{(5; 0,9), (Ω; 0,1)}

B = Best criterion, W = Worst criterion

Table 3-9. Assessment for all the criteria in Funding category from four sectors.

EE			RB			TT			WS		
Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst
F1	{(9; 1)}	{(1; 1)}	F1^{W}	{(6; 0,8), (Ω; 0,2)}	{(1; 1)}	F1	{(4; 0,9), (5; 0,1)}	{(3; 0,1), (4; 0,9)}	F1	{(1; 1)}	{(5; 1)}
F2^{B}	{(1; 1)}	{(9; 1)}	F2^{B}	{(1; 1)}	{(6; 0,8), (Ω; 0,2)}	F2^{B}	{(1; 1)}	{(5; 0,9), (6; 0,1)}	F2	{(1; 1)}	{(5; 0,7), (Ω; 0,3)}
F3^{W}	{(9; 1)}	{(1; 1)}	F3	{(2; 0,6), (Ω; 0,4)}	{(3; 0,6), (Ω; 0,4)}	F3	{(3; 0,9), (4; 0,1)}	{(2; 0,1), (3; 0,9)}	F3^{B}	{(1; 1)}	{(6; 0,8), (Ω; 0,2)}
F4	{(9; 1)}	{(1; 1)}	F4	{(5; 0,7), (Ω; 0,3)}	{(2; 0,6), (Ω; 0,4)}	F4	{(4; 0,9), (5; 0,1)}	{(2; 0,1), (3; 0,9)}	F4	{(2; 0,7), (Ω; 0,3)}	{(5; 0,8), (Ω; 0,2)}
F5	{(9; 1)}	{(1; 1)}	F5	{(3; 0,6), (Ω; 0,4)}	{(3; 0,6), (Ω; 0,4)}	F5	{(4; 0,9), (5; 0,1)}	{(2; 0,1), (3; 0,9)}	F5^{W}	{(6; 0,8), (Ω; 0,2)}	{(1; 1)}
F6	{(9; 1)}	{(1; 1)}	F6	{(5; 0,7), (Ω; 0,3)}	{(5; 0,7), (Ω; 0,3)}	F6^{W}	{(5; 0,9), (6; 0,1)}	{(1; 1)}	F6	{(3; 0,8), (Ω; 0,2)}	{(2; 0,8), (Ω; 0,2)}

B = Best criterion, W = Worst criterion

The nonlinear belief-based BWM is used in this case to determine the weights of the main categories for each sector, the interval weights for the main categories obtained from model (3.11) and (3.12) are shown in Table 3-12 (Step 5 and 6). In this table, the IRs and uncertainty degrees of each sector are obtained following the procedure described in Section 3.4 (Step 7), and the respective reliability degrees (weights) are derived via Equation (3.24) (Step 8). The last row of Table 3-12 contains the aggregated weights for each main category (Step 9).

Table 3-10. Assessment for all the criteria in Coordination category from four sectors.

EE			RB			TT			WS		
Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst
$C1^B$	{(1; 1)}	{(7; 1)}	$C1^B$	{(1; 1)}	{(7; 0,9), (Ω; 0,1)}	$C1^B$	{(1; 1)}	{(4; 0,7), (5; 0,3)}	$C1^B$	{(1; 1)}	{(6; 0,4), (Ω; 0,6)}
C2	{(1; 1)}	{(7; 1)}	C2	{(7; 0,9), (Ω; 0,1)}	{(3; 0,7), (Ω; 0,3)}	C2	{(2; 0,8), (3; 0,2)}	{(2; 0,2), (3; 0,8)}	C2	{(4; 0,6), (Ω; 0,4)}	{(4; 0,9), (Ω; 0,1)}
C3	{(1; 1)}	{(7; 1)}	C3	{(2; 0,8), (Ω; 0,2)}	{(5; 0,8), (Ω; 0,2)}	C3	{(2; 0,9), (3; 0,1)}	{(2; 0,1), (3; 0,9)}	$C3^W$	{(6; 0,4), (Ω; 0,6)}	{(1; 1)}
$C4^W$	{(7; 1)}	{(1; 1)}	C4	{(2; 0,8), (Ω; 0,2)}	{(6; 0,8), (Ω; 0,2)}	$C4^W$	{(4; 0,7), (5; 0,3)}	{(1; 1)}	C4	{(1; 0,8), (Ω; 0,2)}	{(4; 0,8), (Ω; 0,2)}
C5	{(1; 1)}	{(7; 1)}	C5	{(5; 0,7), (Ω; 0,3)}	{(2; 0,7), (Ω; 0,3)}	C5	{(3; 0,8), (4; 0,1), (Ω; 0,1)}	{(3; 0,1), (4; 0,8), (Ω; 0,1)}	C5	{(4; 0,7), (Ω; 0,3)}	{(5; 0,7), (Ω; 0,3)}
C6	{(1; 1)}	{(7; 1)}	$C6^W$	{(7; 0,9), (Ω; 0,1)}	{(1; 1)}	C6	{(2; 0,9), (3; 0,1)}	{(2; 0,1), (3; 0,9)}	C6	{(3; 0,8), (Ω; 0,2)}	{(3; 0,8), (Ω; 0,2)}

B = Best criterion, W = Worst criterion

Table 3-11. Assessment for all the criteria in Policy category from four sectors.

EE			RB			TT			WS		
Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst	Criteria	Best to Others	Others to Worst
$P1^W$	{(8; 1)}	{(1; 1)}	P1	{(1; 0,8), (Ω; 0,2)}	{(2; 0,8), (Ω; 0,2)}	$P1^B$	{(1; 1)}	{(4; 0,2), (5; 0,8)}	$P1^W$	{(2; 1)}	{(1; 1)}
$P2^B$	{(1; 1)}	{(8; 1)}	$P2^W$	{(2; 0,8), (Ω; 0,2)}	{(1; 1)}	P2	{(3; 0,2), (4; 0,8)}	{(3; 0,8), (4; 0,2)}	P2	{(1; 0,9), (Ω; 0,1)}	{(2; 0,9), (Ω; 0,1)}
P3	{(1; 1)}	{(8; 1)}	$P3^B$	{(1; 1)}	{(2; 0,8), (Ω; 0,2)}	$P3^W$	{(4; 0,2), (5; 0,8)}	{(1; 1)}	$P3^B$	{(1; 1)}	{(2; 1)}

B = Best criterion, W = Worst criterion

Table 3-12. Weights of the main categories.

Sector	Category						
	PP	F	C	P	IR	AU	λ
EE	0.32	0.32	0.05	0.32	0	0	0.32
RB	[0.47, 0.61]	[0.17, 0.33]	[0.08, 0.16]	[0.09, 0.11]	0.21	0.53	0.19
TT	[0.4, 0.41]	[0.24, 0.24]	[0.22, 0.23]	[0.13, 0.13]	0.04	0.20	0.27
WS	[0.39, 0.45]	[0.21, 0.33]	[0.06, 0.07]	[0.22, 0.26]	0.17	0.41	0.22
Aggregated	[0.39, 0.43]	[0.24, 0.3]	[0.10, 0.12]	[0.20, 0.21]	-	-	-

Although the IRs and uncertainty degrees are relatively high, we did not ask the experts to revise their preferences in this study, because without a threshold for the belief-based BWM (which could be developed in the future), there is no way to determine whether or not the experts are sufficiently consistent and certain. As such, we accept all the experts' judgment, but with different weights for the experts based on their reliability degrees.

The IRs and AUs in Table 3-12 are pictured in Figure 3-7, which shows how far the experts in the four sectors are from the perfect reliability status: the closer the coordinate is to the origin, the greater the reliability. As we can see, EE is the most reliable, so that his assessments carries higher weight (0.3189) than the others.

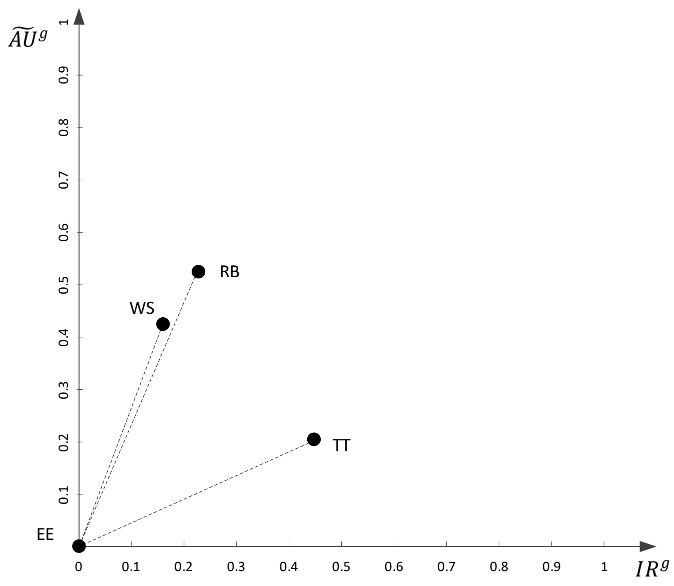


Figure 3-7. Reliability of the four sectors

Similarly, we can calculate the local weights for each criterion from Table 3-8 to Table 3-11 for each sector. By combining the aggregated weights of the main categories with the local weights of the criteria (see (2016) for the interval operations), we can obtain the global weights for each criterion for each sector. Then, we follow the Steps 5 to 9 again to obtain the overall weight for each criterion. The results can be seen in Figure 3-8 and Table 3-13.

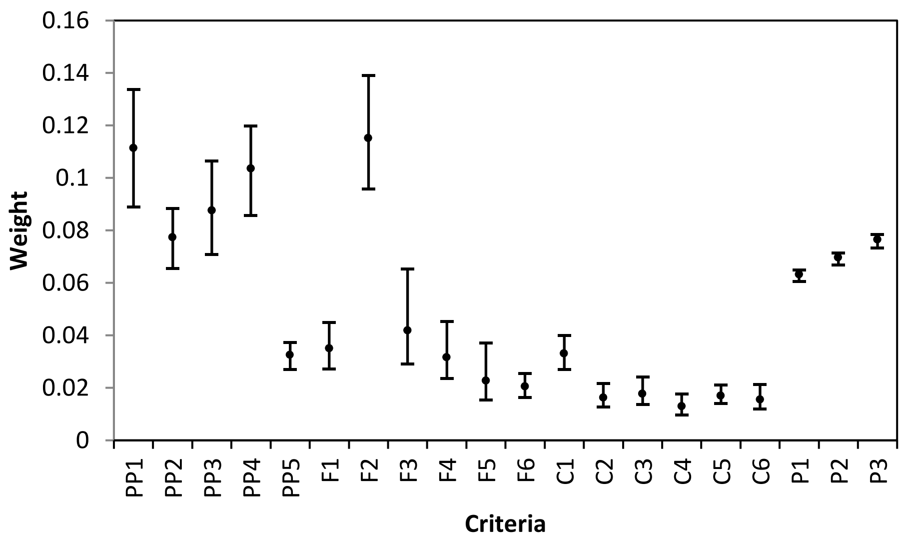


Figure 3-8. The overall weights of the KPPIP criteria

Table 3-13. Overall weights of each criterion

Category	Aggregated weights of categories	Criteria	EE	RB	TT	WS	Overall weights of criteria
PP	[0.39, 0.43]	PP1	[0.23, 0.26]	[0.15, 0.25]	[0.37, 0.43]	[0.11, 0.29]	[0.09, 0.13]
		PP2	[0.22, 0.23]	[0.17, 0.21]	[0.2, 0.29]	[0.04, 0.06]	[0.07, 0.09]
		PP3	[0.23, 0.23]	[0.23, 0.3]	[0.37, 0.11]	[0.18, 0.1]	[0.07, 0.1]
		PP4	[0.26, 0.27]	[0.18, 0.29]	[0.11, 0.13]	[0.35, 0.48]	[0.09, 0.12]
		PP5	[0.03, 0.03]	[0.04, 0.04]	[0.08, 0.09]	[0.15, 0.2]	[0.03, 0.04]
F	[0.24, 0.3]	F1	0.07	[0.04, 0.06]	[0.13, 0.17]	[0.19, 0.29]	[0.03, 0.04]
		F2	0.64	[0.31, 0.49]	[0.34, 0.44]	[0.17, 0.21]	[0.1, 0.14]
		F3	0.07	[0.08, 0.28]	[0.09, 0.24]	[0.24, 0.34]	[0.03, 0.07]
		F4	0.07	[0.04, 0.16]	[0.08, 0.16]	[0.18, 0.24]	[0.02, 0.05]
		F5	0.07	[0.07, 0.27]	[0.08, 0.16]	[0.02, 0.04]	[0.02, 0.04]
		F6	0.07	[0.11, 0.18]	[0.05, 0.07]	[0.04, 0.05]	[0.02, 0.03]
C	[0.1, 0.12]	C1	0.19	[0.31, 0.41]	[0.25, 0.36]	[0.32, 0.4]	[0.03, 0.04]
		C2	0.19	[0.05, 0.07]	[0.1, 0.24]	[0.11, 0.17]	[0.01, 0.02]
		C3	0.19	[0.15, 0.28]	[0.1, 0.24]	[0.04, 0.05]	[0.01, 0.02]
		C4	0.03	[0.19, 0.32]	[0.05, 0.067]	[0.14, 0.23]	[0.01, 0.02]
		C5	0.19	[0.05, 0.11]	[0.13, 0.19]	[0.13, 0.16]	[0.01, 0.02]
		C6	0.19	[0.03, 0.05]	[0.1, 0.24]	[0.08, 0.19]	[0.01, 0.02]
P	[0.2, 0.21]	P1	0.06	[0.36, 0.36]	[0.67, 0.67]	[0.2, 0.2]	[0.06, 0.06]
		P2	0.48	[0.2, 0.2]	[0.22, 0.22]	[0.38, 0.38]	[0.07, 0.07]
		P3	0.47	[0.44, 0.44]	[0.11, 0.11]	[0.42, 0.42]	[0.07, 0.08]

From the results we can see that the overall weights are rather different from the original weights used by KPPIP. Determination of funding scheme (F2) is considered to be one of the most important criteria with maximum weight of 0.14. The importance of the criteria in Coordination category are relatively low.

We also checked the weights obtained by the proposed belief-based BWM with the leader of KPPIP, and he confirmed that our findings are much more reasonable than the original ones used by KPPIP.

3.7 Conclusions

This aim of this study was to develop an extended BWM model to deal with belief structure-related information. Compared to the original BWM, the superiority of the proposed belief-based BWM method has to do with the fact that it can capture different types of uncertainties, including probabilities and vagueness in subjective judgments, and, as discussed in the introduction, that it is more flexible than the fuzzy BWM.

In the belief-based BWM, we first ask the DM to indicate his preferences in pairwise comparisons, with basic belief assignments, which are then transformed into the belief degrees (pignistic probabilities) associated with each grade. These degrees are then used to construct an optimization problem, to obtain the weights of the criteria. Since the nonlinear belief-based BWM was able to generate multiple solutions in cases where DMs are inconsistent, two models are developed to derive the interval weights of criteria.

In the real-world contexts, it is likely that a group-based decision-making process is preferred over individual decisions, because of the complexity of the problems. The decision-making processes that take place in group settings tend to make the decisions more comprehensive and reasonable. However, the uncertainty contained in the estimations provided by the different DMs in such a group, and the inconsistency involved in the pairwise comparisons, can produce unreliable and unstable results, making it necessary to measure the uncertainty and inconsistency degree of the preferences being expressed, since these two degrees can reflect the reliability of a DM. To date, few studies have including the reliability of the judgments made by a DM. To remedy that state of affairs, this study proposes a method to measure the reliability degree of a DM, based on his inconsistency and uncertainty levels. Based on the degree of inconsistency and uncertainty obtained from the preferences of the DMs, we can measure the relative reliability of DMs, which can then be used to assign weights to different DMs, based on which we can aggregate the weights of criteria from the belief BWM, and obtain the final weights of the criteria involved.

It is worth noticing that, instead of weighing the preferences of the decision makers according to how much mutually supportive they are, we propose an approach to weigh the experts based on the quality of their preferences at an individual level. Although there is not a "gold standard" to aggregate preferences, our approach is supported by some empirical and psychological studies, e.g. (Shanteau et al., 2002; Shanteau et al., 2003; Weiss and Shanteau, 2004), which consider a number of factors contributing to the expertise of a decision maker. Among these factors, there are the experience, which is reflected in the precision of the judgments, and their internal coherence, i.e. the consistency. In our proposal both these factors are taken into account.

The ideas underlying the belief-based BWM have been illustrated by numerical examples after each proposed model, and a real-world case study of infrastructure project criteria system assessment in Indonesia is demonstrating the applicability and feasibility of the models.

One of the aims of future research will be to increase our understanding of the inconsistency and uncertainty measures, so that we can determine the thresholds for acceptable levels uncertainty and inconsistency. In addition, it is important to take a closer look at the links between inconsistency and uncertainty, and to examine how they affect one another. In that regard, it might be also interesting to check the relation between the *concentration* of the weights provided by the non-linear BWM (Rezaei, 2020) and their inconsistency and uncertainty. Furthermore, it is also possible to extend the belief-based BWM to linear model, but because it does not fit the framework of the group decision making, so we leave it to another separate study. And finally, combining other MCDM methods (e.g. TOPSIS, VIKOR, ELECTRE methods) with belief-based BWM may provide another possible direction of further research.

Acknowledgment

An initial version of this paper was presented at the 25th International Conference on Multiple Criteria Decision Making, June 2019, Istanbul, and we thank the participants for their constructive comments. The authors gratefully acknowledge financial support for the PhD project from the China Scholarship Council (No. 201708440305).

References

Abellán, J. & Masegosa, A. (2008). Requirements for total uncertainty measures in Dempster–Shafer theory of evidence. *International Journal of General Systems*, *37*(6), 733-747.

Aboutorab, H., Saberi, M., Asadabadi, M. R., Hussain, O. & Chang, E. (2018). ZBWM: The Z-number extension of Best Worst Method and its application for supplier development. *Expert Systems with Applications*, *107*, 115-125.

Badri Ahmadi, H., Kusi-Sarpong, S. & Rezaei, J. (2017). Assessing the social sustainability of supply chains using Best Worst Method. *Resources, Conservation and Recycling*, *126*, 99-106.

Beynon, M., Curry, B. & Morgan, P. (2000). The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modelling. *Omega*, *28*(1), 37-50.

Du, Y., Yang, N., Zhou, W. & Li, C. (2018). A Reliability-Based Consensus Model for Multiattribute Group Decision-Making with Analytically Evidential Reasoning Approach. *Mathematical Problems in Engineering*, *2018*, 1-14.

Durbach, I. N. & Stewart, T. J. (2012). Modeling uncertainty in multi-criteria decision analysis. *European Journal of Operational Research*, *223*(1), 1-14.

Fei, L., Lu, J. & Feng, Y. (2020). An extended best-worst multi-criteria decision-making method by belief functions and its applications in hospital service evaluation. *Computers & Industrial Engineering*, *142*, 106355.

Fu, C., Yang, J. & Yang, S. (2015). A group evidential reasoning approach based on expert reliability. *European Journal of Operational Research*, *246*(3), 886-893.

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

Guo, M., Yang, J. B., Chin, K. S., Wang, H. W. & Liu, X. B. (2009). Evidential reasoning approach for multiattribute decision analysis under both fuzzy and Interval uncertainty. *IEEE Transactions on Fuzzy Systems*, *17*(3), 683-697.

Guo, S. & Zhao, H. (2017). Fuzzy best-worst multi-criteria decision-making method and its applications. *Knowledge-Based Systems*, *121*, 23-31.

Gupta, H. & Barua, M. K. (2017). Supplier selection among SMEs on the basis of their green innovation ability using BWM and fuzzy TOPSIS. *Journal of Cleaner Production*, *152*, 242-258.

Gupta, P., Anand, S. & Gupta, H. (2017). Developing a roadmap to overcome barriers to energy efficiency in buildings using best worst method. *Sustainable Cities and Society*, *31*, 244-259.

Hafezalkotob, A. & Hafezalkotob, A. (2017). A novel approach for combination of individual and group decisions based on fuzzy best-worst method. *Applied Soft Computing*, *59*, 316-325.

Hafezalkotob, A., Hafezalkotob, A., Liao, H. & Herrera, F. (2020). Interval MULTIMOORA Method Integrating Interval Borda Rule and Interval Best–Worst-Method-Based Weighting Model: Case Study on Hybrid Vehicle Engine Selection. *IEEE Transactions on Cybernetics, 50*(3), 1157-1169.

Hajek, P. & Froelich, W. (2019). Integrating TOPSIS with interval-valued intuitionistic fuzzy cognitive maps for effective group decision making. *Information Sciences*, *485*, 394-412.

Harmanec, D. & Klir, G. J. (1994). Measuring total uncertainty in Dempster-Shafer theory: A novel approach. *International journal of general system*, *22*(4), 405-419.

Harmanec, D., Resconi, G., Klir, G. J. & Pan, Y. (1996). On the computation of uncertainty measure in Dempster-Shafer theory. *International Journal Of General System*, *25*(2), 153-163.

Huynh, V. & Nakamori, Y. (2010). Notes on "Reducing algorithm complexity for computing an aggregate uncertainty measure". *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *40*(1), 205-209.

Jia, F. & Wang, X. Y. (2016). BWM-TOPSIS multi-criteria group decision-making method based on rough number. *Kongzhi yu Juece/Control and Decision*, *31*(10), 1915-1920.

Jousselme, A., Liu, C., Grenier, D. & Bossé, É. (2006). Measuring ambiguity in the evidence theory. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *36*(5), 890-903.

Klir, G. & Wierman, M. (1999). Uncertainty-based information: Elements of generalized information theory, Physica-Verlag.

Kou, G., Ergu, D. & Shang, J. (2014). Enhancing data consistency in decision matrix: Adapting Hadamard model to mitigate judgment contradiction. *European Journal of Operational Research*, *236*(1), 261-271.

KPPIP (2017). KPPIP Report: Period of January-June 2017, KPPIP.

Li, G., Kou, G. & Peng, Y. (2016). A group decision making model for integrating heterogeneous information. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48*(6), 982-992.

Liu, C., Grenier, D., Jousselme, A. & Bosse, E. (2007). Reducing algorithm complexity for computing an aggregate uncertainty measure. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *37*(5), 669-679.

Mi, X., Tang, M., Liao, H., Shen, W. & Lev, B. (2019). 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*, *87*, 205-225.

Mohammadi, M. & Rezaei, J. (2020). Bayesian best-worst method: A probabilistic group decision making model. *Omega*, *96*, 102075.

Mou, Q., Xu, Z. & Liao, H. (2016). An intuitionistic fuzzy multiplicative best-worst method for multi-criteria group decision making. *Information Sciences*, *374*, 224-239.

Mou, Q., Xu, Z. & Liao, H. (2017). A graph based group decision making approach with intuitionistic fuzzy preference relations. *Computers & Industrial Engineering*, *110*, 138-150.

Nawaz, F., Asadabadi, M. R., Janjua, N. K., Hussain, O. K., Chang, E. & Saberi, M. (2018). An MCDM method for cloud service selection using a Markov chain and the best-worst method. *Knowledge-Based Systems*, *159*, 120-131.

Ng, C. Y. & Chuah, K. B. (2014). Evaluation of design alternatives' environmental performance using AHP and ER approaches. *IEEE Systems Journal*, *8*(4), 1185-1192.

Nie, R., Tian, Z., Wang, X., Wang, J. & Wang, T. (2018). Risk evaluation by FMEA of supercritical water gasification system using multi-granular linguistic distribution assessment. *Knowledge-Based Systems*, *162*, 185-201.

Pal, N. R., Bezdek, J. C. & Hemasinha, R. (1992). Uncertainty measures for evidential reasoning I: A review. *International Journal of Approximate Reasoning*, *7*(3), 165-183.

Pamučar, D., Petrović, I. & Ćirović, G. (2018). Modification of the Best-Worst and MABAC methods: A novel approach based on interval-valued fuzzy-rough numbers. *Expert Systems with Applications*, *91*, 89-106.

Ren, J. (2018). Technology selection for ballast water treatment by multi-stakeholders: A multi-attribute decision analysis approach based on the combined weights and extension theory. *Chemosphere*, *191*, 747-760.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49-57.

Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega*, *64*, 126-130.

Rezaei, J. (2020). A Concentration Ratio for Nonlinear Best Worst Method. *International Journal of Information Technology & Decision Making*, 1-17.

Rezaei, J., Kothadiya, O., Tavasszy, L. & Kroesen, M. (2018). Quality assessment of airline baggage handling systems using SERVQUAL and BWM. *Tourism Management*, *66*, 85-93.

Safarzadeh, S., Khansefid, S. & Rasti-Barzoki, M. (2018). A group multi-criteria decision-making based on best-worst method. *Computers & Industrial Engineering*, *126*, 111-121.

Serrai, W., Abdelli, A., Mokdad, L. & Hammal, Y. (2017). Towards an efficient and a more accurate web service selection using MCDM methods. *Journal of Computational Science*, *22*, 253-267.

Shafer, G. (1976). A mathematical theory of evidence, Princeton university press.

Shanteau, J., Weiss, D. J., Thomas, R. P. & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, *136*(2), 253-263.

Shanteau, J., Weiss, D. J., Thomas, R. P., Pounds, J. & Hall, B. (2003). How can you tell if someone is an expert? Empirical assessment of expertise. *Emerging perspectives on judgment and decision research*, 620-641.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, *69*(1), 99-118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129.

Smets, P. & Kennes, R. (1994). The transferable belief model. *Artificial Intelligence*, *66*(2), 191-234.

Wang, A., Luo, Y., Tu, G. & Liu, P. (2011). Quantitative Evaluation of Human-Reliability Based on Fuzzy-Clonal Selection. *IEEE Transactions on Reliability*, *60*(3), 517-527.

Weiss, D. J. & Shanteau, J. (2004). The vice of consensus and the virtue of consistency. *Psychological investigations of competent decision making*, 226-240.

Xu, D., Yang, J. & Wang, Y. (2006). The evidential reasoning approach for multi-attribute decision analysis under interval uncertainty. *European Journal of Operational Research*, *174*(3), 1914-1943.

Yager, R. R. (1987). On the dempster-shafer framework and new combination rules. *Information Sciences*, *41*(2), 93-137.

Yager, R. R. & Alajlan, N. (2015). Evaluating belief structure satisfaction to uncertain target values. *IEEE Transactions on Cybernetics*, *46*(4), 869-877.

Yang, J. B., Wang, Y. M., Xu, D. L. & Chin, K. S. (2006). The evidential reasoning approach for MADA under both probabilistic and fuzzy uncertainties. *European Journal of Operational Research*, *171*(1), 309-343.

Yang, J. (2001). Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties. *European Journal of Operational Research*, *131*(1), 31-61.

Yang, J. & Singh, M. G. (1994). An evidential reasoning approach for multiple-attribute decision making with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics*, *24*(1), 1-18.

Yang, J. & Xu, D. (2002). On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *32*(3), 289-304.

You, X., Chen, T. & Yang, Q. (2016). Approach to multi-criteria group decision-making problems based on the Best-Worst-Method and ELECTRE method. *Symmetry*, *8*(9), 95.

Zeleny, M. (1982). Multiple Criteria Decision Making. New York, McGraw Hill.

Zhang, H., Kou, G. & Peng, Y. (2019). Soft consensus cost models for group decision making and economic interpretations. *European Journal of Operational Research*, *277*(3), 964-980.

Zhou, M., Liu, X., Chen, Y. & Yang, J. (2018). Evidential reasoning rule for MADM with both weights and reliabilities in group decision making. *Knowledge-Based Systems*, *143*, 142-161.

4 Inland terminal location selection using the multi-stakeholder Best-Worst Method

Liang, F., Verhoeven, K., Brunelli, M. & Rezaei, J. (2021). Inland terminal location selection using the multi-stakeholder best-worst method. International Journal of Logistics Research and Applications, 1-23.

Abstract

The aim of this study is to develop an inland terminal location selection methodology. This methodology is viewed from the perspective of the shipping line designing the inland transport chain while also taking the objectives of multiple other stakeholders into account. To that end, we develop a consensus model for a group Best-Worst Method (BWM) in order to aggregate the evaluations of the various stakeholders. The proposed method is applied to a real-life case study involving the Maersk shipping line, in which nine experts representing three different types of stakeholders assess six possible locations. After the evaluation, the market volume potential is identified as one of the most important criteria. Furthermore, a sensitivity analysis indicates that a varying influx of the container volume has no impact on the most desirable location.

4.1 Introduction

Over the last three decades, the shipping industry has evolved from a highly segmented sector into a more integrated sector (Franc and Van der Horst, 2010). Traditionally shipping lines were merely involved with maritime transport between seaports across the globe, but they are increasingly trying to integrate their ocean transport setups with connecting inland transport services to provide door-to-door business propositions to their customers (Frémont, 2009; Franc and Van der Horst, 2010). This is known as vertical integration, which allows shipping lines to improve the coordination of container flows and inland repositioning tactics (Song and Dong, 2011; Van den Berg and De Langen, 2015; Wan et al., 2016). Vertical integration helps increase the (cost) efficiency of the provided hinterland operations, which in turn attracts customers and increases the shipping lines' market share in, and control over, the hinterland.

With regard to the cost efficiency of the inland transport chain, intermodal transport has significant economic advantages due to the possibility of putting multiple containers on larger vehicles and reducing the costs per transported container (Simina et al., 2012). A major component of intermodal transport is the inland terminal, where containers are transshipped between trucks and intermodal vehicles, or vice versa (Teye et al., 2017; Teye et al., 2018). Because of the necessity of transshipment operations for intermodal transport, inland terminals have a considerable impact on the cost efficiency of the broader inland transport chain (Rodrigue and Notteboom, 2009). Therefore, active engagement with inland terminals by shipping lines enables them to effectively use these facilities in their inland transport configurations (Van den Berg and De Langen, 2015). Because a number of competing stakeholders also use these facilities (e.g., freight forwarders, logistics service providers and competing shipping lines), setting up an inland terminal dedicated to the needs of the shipping line and its (potential) customers makes it possible to create an inland transport chain (Franc and Van der Horst, 2010; Tan et al., 2018).

The location of an inland terminal is a crucial factor due to its ability to contribute to the effectiveness of the inland transport chain, mainly because that location determines the distances for container movements in the main haulage and pre-/end-haulage legs (Pekin, 2010; Tsao and Thanh, 2019). As such, selecting the inland terminal location is an essential task for the shipping line when designing the inland transport chain in which the terminal has to contribute to its efficiency and effectiveness (Van Nguyen et al., 2020). Although numerous studies have examined location selection (Bontekoning et al., 2004; Alumur and Kara, 2008; Teye et al., 2017; Teye et al., 2018), so far few have approached the issue from the perspective of shipping lines, which is exactly what this study is intended to remedy.

Moreover, because the operations of the inland terminal business are not conducted by the shipping line alone but also by other companies in the inland transport chain, multiple stakeholders play a role in selecting the best location (Franc and Van der Horst, 2010; Rodrigue et al., 2010; Wilmsmeier et al., 2011; Monios and Wilmsmeier, 2012; De Langen et al., 2013), which is why the problem definition with regard to the shipping line is approached by taking the different objectives of those stakeholders into account. In addition to the objectives of the shipping line itself are the objectives of terminal operators and the transport companies that use the terminal for their operations. Accordingly, to select a desirable location we need to find a compromise between the different objectives. The stakeholders involved are very likely to apply different sets of criteria when evaluating the location problem, next to the transportation cost (Limbourg and Jourquin, 2009). To date, few studies have attempted to aggregate the preferences of a heterogeneous group with different sets of criteria. To remedy that state of affairs, we propose a consensus framework based on the Best-Worst Method (BWM) (Rezaei, 2015).

This study offers a methodological contribution along with a real-case application. That is, we formalize the inland terminal location selection problem from the scope of a shipping line and develop a group consensus decision-making method using different sets of criteria. First, this study adds to the existing literature regarding inland terminals from the perspective of the inland transport chain by approaching the specific inland terminal location selection problem from the perspective of the shipping line. Various researchers have studied the components, activities, and dynamics of container port hinterlands (Lee and Yang, 2018). Multiple studies include analyses involving container transport markets (e.g., (De Langen et al., 2013; Rodrigue and Notteboom, 2013)) and the optimisation of hinterland transport efficiency (e.g., (Caris et al., 2012; Notteboom and Rodrigue, 2017)). However, few studies adopt the point of view of the shipping line as a leading stakeholder in that context, mostly because this is a relatively new development, both professionally and academically, especially with regard to studies involving inland terminals and inland terminal location selection, where shipping lines were originally not the main stakeholders involved. Where the studies of Franc and Van der Horst (2010) and Van den Berg and De Langen (2015) touched upon the relationship between inland service integration and inland terminals, this study adds to contemporary literature and follows up on their notions by focusing on the development and operation of inland terminals from the perspective of the shipping line as a key stakeholder, with the main purpose of improving the shipping line's inland transport chain services offered to its customers. In addition, this research contributes to the current literature on vertical integration in hinterland container transport markets by obtaining insights into the differently valued criteria involved in inland terminal location decisions from the perspective of the shipping lines themselves and of other companies involved in the inland transport chains.

Second, this study adds to the existing literature on Group Multi-Criteria Decision-Making (GMCDM) problems by considering the varying preferences for individually relevant criteria stemming from the distinct objective(s) of each involved stakeholder. In most primary studies on GMCDM, the views of the multiple stakeholders involved in the research are taken into account by having them evaluate all the criteria that are considered to be relevant within a fixed set of criteria (e.g., (Kayikci, 2010; Regmi and Hanaoka, 2013; Roso et al., 2015)). Based on the evaluations of the stakeholders, criteria weights are then calculated and used for a further assessment. The basic requirement of this approach is that the calculated weights and the resulting values need to stem from the same fixed set of criteria so they can be compared to one another. However, it has been argued that decision-making criteria (which are originally stored in one fixed set) are not necessarily relevant to the particular objectives of all the involved stakeholders, which implies that criteria that are *irrelevant* to certain stakeholders are subjected to their assessment, while not actually being the right criteria to be used to reflect their preferences (Macharis et al., 2012). In fact, in real-life situations different stakeholders are likely to use different sets of criteria. The aim of this study is to develop a method to compare the weights of the criteria of different stakeholders in a meaningful way.

In the remainder of this study, the theoretical framework for the transportation system under investigation is defined in Section 4.2, in addition to a review of the relevant literature regarding inland terminal evaluation factors. Next, the methodology used in this study is presented in Section 4.3, while Section 4.4 discusses the application of the location selection methodology to a case study involving the Maersk shipping line. Based on the results of the case study, the conclusions and discussion are presented in Section 4.5, including practical implications and recommendations for further research.

4.2 Theoretical background

This section starts with a review of the four-layer framework: the relevant stakeholders are identified from the literature, and the factors considered in inland terminal location selection studies are reviewed.

4.2.1 Stakeholders in the inland terminal location selection system

The research system of the inland terminal location is viewed as a component of the broader container port hinterland, which is a structure consisting of multiple layers. By using an adapted four-layer framework[9] from the study by Notteboom and Rodrigue (2017), which is applicable to the container port hinterland, it is possible to identify the logistical layer, the transport layer, the infrastructural layer and the locational layer (see Figure 4-1), all of which are considered to be important for assessing the container port hinterland due to *demand pull effects* from a higher layer in relation to the layer below and the *valorisation effects* from a lower layer towards the layer above. From each layer, components, activities, and related stakeholders relevant to the evaluation and selection of the inland terminal location can be extracted. Regarding the stakeholders, a distinction can be drawn between *key stakeholders* and *contextual stakeholders*. Key stakeholders are directly involved in the main activities within the respective layers, while contextual stakeholders (including the government) are associated with but not actively involved in those activities (at least within the scope of this research). The latter are therefore not included as stakeholders in the remainder of the study.

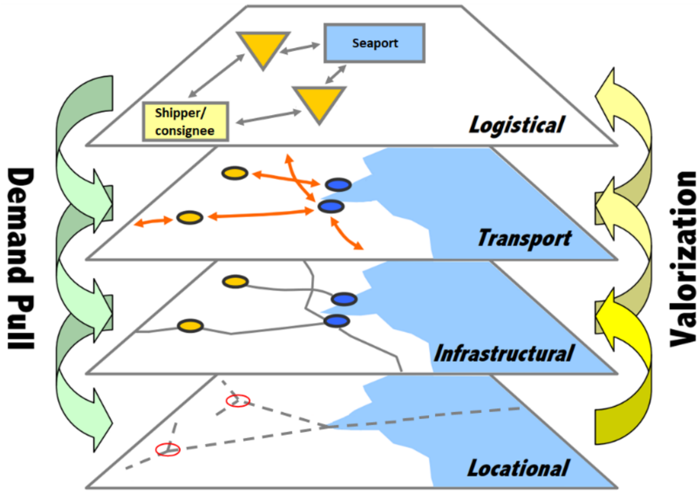


Figure 4-1. Four-layer framework expanding the structure of the container port hinterland (adapted from Notteboom and Rodrigue (2017))

The **logistical layer** contains the organization of the supply and transport chains, which were originally designed mostly by freight forwarders or other third parties for shippers/consignees in so-called third-party logistics (Douma, 2008; De Langen et al., 2013; Van den Berg and De Langen, 2015). However, in recent times shipping lines have increasingly tended to expand their scope and take greater control of the design of transport chains (which is the basis of this research), with the aim of increasing the *Carrier Haulage* setups (compared to *Merchant Haulage* setups) in the hinterlands, which in turn increases their scope from port-to-port to door-to-door transport (Frémont, 2009; Franc and Van der Horst, 2010). A way to encourage this development is through an active engagement of inland terminals in the hinterland transport network, which would facilitate the transport and transshipment operations contributing to the effectiveness of the inland transport chains (Franc and Van der Horst, 2010). Hence, with regard to the shipping line's design objectives for its inland transport chains in the logistical layer, the

[9] In the original four-layer framework, the focus is on the seaport as a node between ocean and hinterland transport. In the adapted framework, this focal point has shifted from the seaport node (which is less relevant in this study) towards the inland terminal, as a node between the main haulage transport and the pre-/end-haulage transport.

shipping line is considered to be a key stakeholder, while freight forwarders and shippers/consignees are viewed as contextual stakeholders (as competitors and customers, respectively).

The **transport layer** contains the transport and transshipment operations that realize the designed transport chain services described above. The main haulage, in the form of either a rail or barge transport, is performed via intermodal transport operators, while pre-/end-haulage activities are conducted by truck transport operators (De Langen et al., 2013). Since these transport operators actively use inland terminals in their operations, they are seen as key stakeholders in regard to selecting the best inland terminal location. Transshipment operations are performed at inland terminals by inland terminal operators, which are regarded as key stakeholders in the transport layer. In addition to basic logistics services, value-added services, which involve extra services aimed at improving the (cost) efficiency of the broader transport chain, can also be provided (Rodrigue et al., 2010).

The **infrastructural layer** contains the transport and transshipment infrastructure used to facilitate the transport and transshipment operations described above. Transport infrastructure (e.g., roads, railways, and inland waterways) is usually developed and owned by government stakeholders based on maintaining and increasing public wellbeing in a larger sense than merely in relation to inland terminals (De Langen et al., 2013). As such, the decisions made regarding the development of transport infrastructure are considered to fall outside of the scope of this research. However, the availability of transport infrastructure is taken into account because it does affect the selection of an inland terminal location. Thus, in this research the government is not considered to be a key stakeholder with regard to infrastructure. Node infrastructures, such as inland terminals, are usually owned by private or public-private entities (Bergqvist and Monios, 2014). With regard to inland terminals, the key stakeholders are the organizations that own and operate the inland terminals, implying their presence in the transport as well as the infrastructural layer (Bergqvist et al., 2015). Whereas inland terminal operators are key stakeholders, public (government) stakeholders are not involved in inland terminal ownership because of their relatively nonexecutive roles. With regard to the infrastructural function of the inland terminal within the inland transport chain, the load centre that facilitates integrated transport and transshipment solutions close to the locations of shippers/consignees is the most applicable to the shipping line, which aims at setting up the facility as a component of the inland services it provides.

The **locational layer** contains the geographical locations of the infrastructural components discussed above within the economic space of the container port hinterland. These infrastructure locations are relative to the container volume (the number of containers imported and/or exported from a certain area). Generating/attracting points in this economic space define the distances between these locations and the actual infrastructure and consequently the relative effectiveness of the infrastructure. Accordingly, the location of infrastructure can contribute to, as well as be dependent on, the economic space (Rodrigue et al., 2010). The selection of infrastructure locations is shaped by these relationships as well (Rodrigue and Notteboom, 2009). The infrastructure operator is a key stakeholder in selecting the location of the infrastructure and is involved in the actual provision and operation of the infrastructure. Since this study is aimed at selecting a location for an inland terminal infrastructure to be used specifically within the inland transport chain designed by the shipping line, the latter is also a key stakeholder in the location selection process.

An overview of the contents of the container port hinterland layer applicable to this study, including the key activities and the stakeholders making the (final) decisions with regard to these activities, are shown in Figure 4-1.

Table 4-1. The corresponding key activities and stakeholders from the container port hinterland layers

Layer	Key activity	Key stakeholder
Logistical	Organize inland transport chains	Shipping line
Transport	Transport containers	Intermodal transport operator, truck transport operator
	Transship containers	Inland terminal operator
Infrastructural	Provide transshipment infrastructure	Inland terminal operator
Locational	Select infrastructure location	Shipping line, inland terminal operator

The information from the container port hinterland structure review is used as the system input in the remainder of this study. Based on the outcomes of that review, the inland terminal location selection process discussed in this study can be configured as follows:

- The stakeholders making the actual decisions with regard to the selection of the location of the inland terminal are the **shipping line** and the inland **terminal operator**.
  - The shipping line evaluates an inland terminal location and decides to select it based on the objective of *incorporating the terminal into the designed inland transport chain*.
  - The terminal operator evaluates an inland terminal location and decides to select it based on the objective of *ensuring the profitability of transshipment operations at the site*.
- Although the transport operators, i.e., the **terminal users**, are not actually involved in the decision-making process regarding the location of the terminal, they are *affected* by the decision because the eventual location of the terminal determines the locations in the broader network transport, between which operations have to be conducted. As such, the fit of the selected location within the transport operation scheme of the terminal user affects the (cost) efficiency of the transport operations and, ultimately, of the entire inland transport chain. Accordingly, the evaluation of the inland terminal location by the transport operators is based on the objective of *using the inland terminal to optimize their transport operations*. This evaluation is ultimately important in terms of selecting a location that is beneficial to these transport operations and ultimately to the designed inland transport chain.

Figure 4-2 shows the graphical configuration of this study. Note that there is a clear distinction between the stakeholders evaluating and ultimately selecting the inland terminal location and the stakeholders that only *evaluate* the location. The dotted line indicates the importance of the evaluation by the terminal user on top of the evaluations by the stakeholders that ultimately select the location.

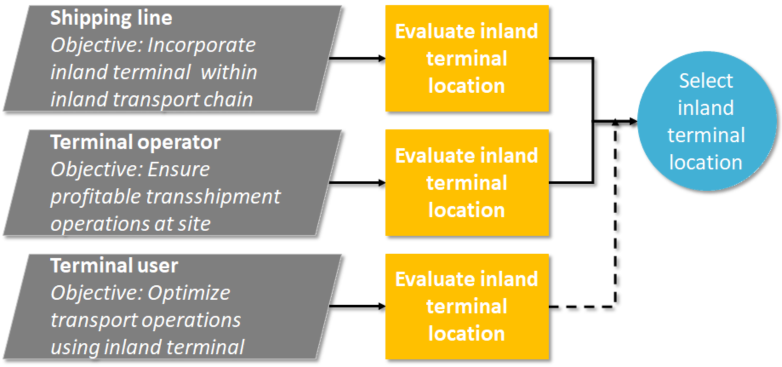


Figure 4-2. Inland terminal location selection configuration

4.2.2 Factors (or criteria) considered in inland terminal location selection studies

The selection of a suitable location for an inland terminal involves multiple factors related to the different interrelated features of the container port hinterland. Both quantitative and qualitative factors are often used in location studies (Notteboom, 2011). In MCDM studies these factors are referred to as criteria, and their evaluations by stakeholders ultimately determine the outcomes of decision-making models. This section describes the most commonly used criteria with regard to the location-related decision-making problem for inland terminals. The review was carried out using Scopus[10] as a primary academic database and Google Scholar[11] as an additional bibliographic search engine. As such, literature specifically concerned with the location selection of inland terminals was considered. Since the terminology used to denote inland terminal facilities often varies, attention was also paid to location selection studies involving *dry ports* and *freight villages*. Although these terms and definitions may vary, what the referenced location selection studies have in common is the fact that the studies in question all focus on inland intermodal transshipment of *containerized cargo,* which ensures that the reviewed factors are appropriate for the aim of this study.

The literature reviewed in this study consists of MCDM studies specifically focused on inland terminal location selection. Not all the criteria are included in the remainder of this study, first of all for practical reasons: it is not desirable to have an overly long list of criteria because that implies (time-)intensive data gathering and criteria-weighting processes. In addition, as the (combinations of) stakeholders, (geographical) scopes and methods each study uses vary, so too do the criteria. As such, not all the factors mentioned in the reviewed studies are considered eligible for this particular research. To include only the most suitable factors, *the observed criteria are prioritized*, and the factors that are not relevant to the scope of our study are filtered out. These include:

- Factors that are not generically applicable because they are location-specific (e.g., criteria aimed at particular local legislation).
- Factors indicating *existing* properties/performances of a facility, not applicable because this study is specifically aimed at identifying a location for a *new* inland terminal.
- Factors considered with conditions that are preliminarily considered when selecting alternative locations, and thus are not relevant to assess (e.g., connection to infrastructure network).
- Factors concerned with certain terminal functionalities not applicable to the considered *load centre* terminal type.

The prioritized criteria are further examined in the remainder of this section and structured according to the container port hinterland layer structure. The factors are assigned to the logistical layer, the transport layer, the infrastructural layer, and the locational layer.

Logistical layer factors

Logistical factors used for the selection of an inland terminal location are mostly connected to local market characteristics and related indicators, which affect the decisions involved in the organization of transport chains (at those locations). One of the most frequently observed factors in this regard is *market volume potential*, which relates to the entities in a certain area generating and/or attracting freight volumes and is often expressed as *demand* (Regmi and

---

[10] https://www.scopus.com/
[11] https://scholar.google.com/

Hanaoka, 2013; Nguyen and Notteboom, 2016), usually as the freight volume (e.g., TEU (twenty-foot equivalent units)) moved to/from an area in a certain time unit (e.g., (Roso et al., 2015; Rožić et al., 2016)). Other frequently proposed economic factors related to organizing inland transport chains are the *labour market,* as a resource for conducting inland facility operations (e.g., (Long and Grasman, 2012; Karaşan and Kahraman, 2019)), as well as the more general *socioeconomic development* of an area, often indicated by such indicators as the area's per capita GRP (Gross Regional Product) (e.g., (Kayikci, 2010; Li et al., 2011). Multiple factors are used to indicate a local investment climate. At a market level, *transport and logistics competition* is used to indicate the number of potential competitors offering inland facility services (Long and Grasman, 2012; Karaşan and Kahraman, 2019). At an administrative level, *government policy-related* factors are proposed to indicate the local/regional/national regulatory and/or political stances on the development of inland facilities at certain locations (e.g., (Ka, 2011; Roso et al., 2015)). These factors indicate the broad perspectives being used in most inland facility location studies, ranging from factors that focus explicitly on transport and/or logistics (e.g., *market volume potential*) to factors describing more general market indicators (e.g., *socioeconomic development*). In this regard, the number of criteria related to the logistical layer is relatively high because multiple characteristics from several types of market(-related) components/developments may affect the decisions being made regarding the organization of the transport chain. An overview of all prioritized logistical layer factors included in this study is presented in Table 2-1.

Table 4-2. Logistical layer criteria observed in the literature

Factor	Explanation	Observations in the literature
Market volume potential	Amount of container volumes predicted to be generated in and/or attracted to the area	Ka (2011); Kayikci (2010); Nguyen and Notteboom (2016); Roso et al. (2015); Rožić et al. (2016); Wei et al. (2010); Li et al. (2011); Komchornrit (2017); Regmi and Hanaoka (2013)
Local labour market	Local supply of sufficiently skilled labour for inland terminal-related activities	Karaşan and Kahraman (2019); Rožić et al. (2016); Long and Grasman (2012); Nguyen and Notteboom (2016)
Regional economic development	Local/regional socio-economic characteristics indicating the development of population and economy	Ka (2011); Kayikci (2010); Roso et al. (2015)
Regional transport/logistics competition	Number of companies involved in inland terminal(-related) activities in the area	Karaşan and Kahraman (2019); Long and Grasman (2012); Regmi and Hanaoka (2013)
Government policy	Local/regional political, administrational, and regulatory circumstances with regards to the inland terminal(-related) developments/activities	Ka (2011); Karaşan and Kahraman (2019); Nguyen and Notteboom (2016); Long and Grasman (2012); Li et al. (2011); Roso et al. (2015)

Transport layer factors

Factors used for the selection of inland facility locations in connection to the transport layer are directly related to the operational activities at and around the inland facility. These can be *conditions* under which these operations (have to) take place but also, for instance, the effects of such operations. Operational costs are often used, including *costs for transport* (e.g., (Ka, 2011; Regmi and Hanaoka, 2013)) and *costs for operating the inland facility* (Ka, 2011; Nguyen and Notteboom, 2016). Furthermore, local traffic characteristics that affect transport operations are also mentioned, mostly in terms of *traffic congestion* indicators (e.g., (Wei et al., 2010; Li et al., 2011)), which can sometimes be translated directly into *delivery times* (Karaşan and Kahraman, 2019). In addition, the environmental effects of the operations are also included. Whereas it is sometimes indicated whether those effects take place on a local scale (e.g., (Nguyen and Notteboom, 2016; Özceylan et al., 2016)) or on a global scale (Kayikci, 2010), this is not always the case[12]. Other transport- and transshipment-related factors in this context are *noise pollution*, which can be viewed as a local effect (e.g., (Roso et al., 2015)), and *energy consumption*, which can be approximated by the indirect increase or reduction of emissions as a result of the energy used in the transport and transshipment operations at or near a site (Kayikci, 2010). An overview of all the transport layer-related factors included in this study is presented in Table 4-3.

Table 4-3. Transport layer criteria observed in the literature

Factor	Explanation	Observations in the literature
Total inland transport costs	Overall costs for inland transport, including (if applicable) trucking costs, rail/barge costs and inland terminal handling costs.	Nguyen and Notteboom (2016); Kayikci (2010); Ka (2011); Regmi and Hanaoka (2013); Wei et al. (2010)
Traffic congestion	Local congested infrastructure causing delays in transport flows.	Nguyen and Notteboom (2016); Karaşan and Kahraman (2019); Long and Grasman (2012); Kayikci (2010); Wei et al. (2010)
Environmental effects	Effects of inland terminal(-related) operations on the environment, e.g., release of hazardous materials or emissions in surroundings.	Nguyen and Notteboom (2016); Kayikci (2010); Regmi and Hanaoka (2013); Özceylan et al. (2016)
Inland terminal operational costs	Costs for operating inland terminal and related activities (e.g., handling).	Nguyen and Notteboom (2016); Ka (2011)

Infrastructural layer factors

The factors that are used for the selection of an inland facility location selection and that are related to the infrastructural layer are first connected to the local infrastructure and its characteristics. In this sense, *local transport infrastructure* metrics (e.g., (Rožić et al., 2016; Komchornrit, 2017)) are used to indicate the properties of the infrastructure in relation to (potential) inland facilities in the area. Criteria concerning the *development/construction* of the infrastructure are also frequently mentioned, which are associated with the infrastructural layer, since that also involves the *provision* of the infrastructures. The involved factors refer to the investments needed to set up an inland facility, which are usually subdivided into *costs for land* (e.g., (Yıldırım and Önder, 2014; Özceylan et al., 2016)), *costs for construction* (e.g., (Regmi and Hanaoka, 2013; Karaşan and Kahraman, 2019)) and *other types of investment costs* (Ka, 2011; Nguyen and Notteboom, 2016), which are grouped under the overarching inland terminal CAPEX (CAPital EXpenditure) in the criteria selection survey we sent to the stakeholders in the transport chain. In addition to monetary factors, the resource availability factor of *expansion*

[12] As existing literature is often unclear on whether the environmental effects are local or global, this criterion is subdivided into Local environmental effects and Global environmental effects in the criteria selection survey that was sent to the transport chain experts.

possibilities indicates the ability to develop more inland facility infrastructure if necessary/desirable (e.g., (Roso et al., 2015; Özceylan et al., 2016)). In this regard, the *spatial development* criterion is also included to indicate potentially unfavourable land-use types close to the (potential) inland facility (e.g., (Kayikci, 2010; Komchornrit, 2017)). An overview of all the infrastructural factors included in this study is presented in Table 4-4.

Table 4-4. Infrastructure layer criteria observed in the literature

Factor	Explanation	Observations in the literature
Transport infrastructure network in area	Characteristics of a transport infrastructure network (e.g., lengths, density) in the area	Kayikci (2010); Ka (2011); Regmi and Hanaoka (2013); Komchornrit (2017); Karaşan and Kahraman (2019); Roso et al. (2015); Rožić et al. (2016); Li et al. (2011)
Land purchase costs	Costs of purchasing land for the inland terminal	Nguyen and Notteboom (2016); Regmi and Hanaoka (2013); Yıldırım and Önder (2014); Özceylan et al. (2016)
Construction costs	Costs of building the inland terminal	Nguyen and Notteboom (2016); Regmi and Hanaoka (2013); Karaşan and Kahraman (2019)
Other investment costs	Other costs with regards to setting up the inland terminal (e.g., for equipment)	Nguyen and Notteboom (2016); Ka (2011)
Land use near location *	Land-use at sites near the inland terminal location	Nguyen and Notteboom (2016); Kayikci (2010); Komchornrit (2017)
Expansion possibilities	Available land that could potentially be used to physically expand the inland terminal	Nguyen and Notteboom (2016); Karaşan and Kahraman (2019); Yıldırım and Önder (2014); Özceylan et al. (2016); Roso et al. (2015)

*Although often called *Spatial development* in the literature, we use *Land use near the location* in the criteria selection survey we sent to transport chain experts to more clearly indicate the factor representation.

Locational layer factors

Factors used for the selection of an inland facility location and related to locational aspects only play a role with regard to *proximity* measures, such as the distance between a given inland facility location and various other objects in the economic space of the container port hinterland represented by the locational layer. The most frequently mentioned factor in the literature, which is also the most applicable to this study, is *market proximity*, i.e., the distance between the inland facility and locations in the area at/to which a certain amount of container volumes are generated/attracted (Long and Grasman, 2012; Yıldırım and Önder, 2014; Roso et al., 2015; Nguyen and Notteboom, 2016; Özceylan et al., 2016; Karaşan and Kahraman, 2019). The entities at these locations make up the total market volume potential in a given area, as previously described in Logistical layer factors of Section 4.2.2. As such, the *market proximity* is the only locational factor included in this study.

4.3 A consensus-building model for BWM group decision-making

Because of the multiple actors and multiple criteria involved in the decision-making process regarding the selection of an inland terminal location, we developed a group BWM that consists of 6 phases:

Phase 1: Define the location selection problem and determine the stakeholders

The first phase of the research methodology involves defining the inland terminal location selection problem of the particular case study by determining the scope of the study and the potential alternatives within that scope, and then selecting the experts from the stakeholders identified by the shipping line. The total number of experts is indicated as K.

Phase 2: Determination of criteria

The basis of the criteria selection procedure is formed by the criteria we identified in the literature review or proposed by experts. Through a *criteria selection survey*, a list containing the observed criteria is sent to and assessed by the experts, in accordance with Phase 1. They are then asked to indicate which criteria they consider to be important and relevant to the selection problem. Based on these indications, a list of relevant decision-making criteria is assembled for each stakeholder. It is worth mentioning that, unlike conventional methods, each expert could provide a different list of criteria according to their own backgrounds.

Phase 3: Criteria operationalization and data gathering

Next, we need to determine the measurement of each criterion and gather the corresponding data. Determining the measuring units makes it possible to know the types of data that have to be gathered to sufficiently represent the criteria, to define the criteria in comparable metrics and to determine whether the criteria are of the benefit or cost type. The evaluations are collected in a decision-making matrix $R = \{r_{ij}\}_{m\times n}$, where r_{ij} represents the evaluation on the ith alternative with respect to the jth criterion, with m alternatives and n criteria.

Phase 4: Determination of criteria weights using BWM

The reason we use the BWM to calculate the criteria weights is that (i) it is a structural pairwise comparison method that requires fewer pairwise comparisons than the AHP (Analytic Hierarchy Process) method, (ii) by considering two opposite references (best and worst), it helps mitigate an anchoring bias, and (iii) it can generate more consistent and reliable results (Rezaei, 2015; 2020). The BWM has been widely used in different areas, including location selection problems (Pamučar et al., 2017; Stević et al., 2018; Kheybari et al., 2019). For more information, see a recent review of BWM (Mi et al., 2019).

To calculate the weight w_j^k of criterion j for expert k ($k = 1,2,\cdots,K$), we first need to ask the experts to indicate which is the most influential or important (best) and the least influential or important (worst) criterion.

Then, the experts are asked to determine their preferences for their most important criteria over all the other criteria, using a number from $\{1,2,\ldots,9\}$, where, for example, 1 represents "equally important", while 9 represents "extremely important than". The obtained Best-to-Others vector for expert k is $A_{Bj}^k = \left(a_{B1}^k, a_{B2}^k, \cdots, a_{Bn}^k\right)$, where a_{Bj}^k represents the preference of the best criterion B over criterion j given by expert k. Suppose expert k has determined n criteria.

Next, the experts are asked to determine their preferences for all their selected criteria over their worst criterion using a number from $\{1,2,\ldots,9\}$. The obtained Others-to-Worst vector for expert k is $A_{OW}^k = \left(a_{1W}^k, a_{2W}^k, \cdots, a_{nW}^k\right)$, where a_{jW}^k represents the preference of criterion j over the worst criterion W given by expert k. Note that the best and worst criteria may be different for each expert.

The next phase involves calculating the optimal weights w_j^{k*}. The optimal weights for the criteria are determined by setting the conditions where, for each pair of w_B^k/w_j^k and w_j^k/w_W^k, $w_B^k/w_j^k = a_{Bj}^k$ and $w_j^k/w_W^k = a_{jW}^k$. To find a good approximation for all j, a solution in which the maximum absolute differences $\left|\frac{w_B^k}{w_j^k} - a_{Bj}^k\right|$ and $\left|\frac{w_j^k}{w_W^k} - a_{jW}^k\right|$ for all j are minimized is formulated in the following model:

$$\begin{aligned}
&\min \max_j \left\{ \left|\frac{w_B^k}{w_j^k} - a_{Bj}^k\right|, \left|\frac{w_j^k}{w_W^k} - a_{jW}^k\right| \right\} s.t. \\
&s.t. \\
&\sum_{j=1}^{n} w_j^k = 1 \\
&w_j > 0, \forall j.
\end{aligned} \tag{4.1}$$

This model can generate multiple optimal solutions. By using the two models proposed by Rezaei (2016), we can include these solutions in the form of interval weights $\bar{w}_j^k = [w_j^{k,min}, w_j^{k,max}]$, where $w_j^{k,min}$ and $w_j^{k,max}$ represent the minimum and maximum weights for criterion j for expert k.

To assess the reliability of the comparisons provided by expert k, we consider the consistency ratio (CR^k) proposed by Liang et al. (2020):

$$CR^k = \max_j \frac{\left|a_{Bj}^k \times a_{jW}^k - a_{BW}^k\right|}{a_{BW}^k \times a_{BW}^k - a_{BW}^k}, \tag{4.2}$$

When $a_{BW}^k = 1$, $CR^k = 0$.

After we obtain the consistency ratios, we need to check whether the judgements are consistent enough and acceptable according to these CRs, which means that thresholds are needed. We use the consistency thresholds (Table 2-3) from the study by Liang et al. (2020). This threshold table consists of combinations of scales (a_{BW}^k) from 3 to 9 and number of criteria (n) from 3 to 9. The CRs obtained in the manner indicated above are compared to the thresholds: if the CRs are smaller than the thresholds, the judgements are acceptable, and vice versa.

Table 4-5. Thresholds for different combinations[13]

Scales	Criteria						
	3	4	5	6	7	8	9
3	0.17	0.17	0.17	0.17	0.17	0.17	0.17
4	0.11	0.15	0.19	0.22	0.25	0.26	0.27
5	0.14	0.20	0.23	0.25	0.27	0.28	0.30
6	0.13	0.20	0.26	0.30	0.31	0.32	0.33
7	0.13	0.25	0.28	0.30	0.31	0.33	0.34
8	0.13	0.25	0.30	0.32	0.34	0.36	0.37
9	0.14	0.27	0.31	0.33	0.35	0.36	0.37

[13] The thresholds for the combinations with 2-scale should be 0.

Phase 5: Alternatives′ value determination using the proposed consensus model

To compare these alternatives, we use an additive value function (Keeney and Raiffa, 1976) to determine the overall value V_i^k of each alternative i for expert k based on the weights of criteria w_j^k and the normalized evaluations p_{ij}:

$$V_i^k=\sum_{j=1}^{n} w_j^k p_{ij}. \tag{4.3}$$

where

$$p_{ij}=\begin{cases}\frac{r_{ij}-r_j^{min}}{r_j^{max}-r_j^{min}}, \text{ for benefit criteria}\\ \frac{r_j^{max}-r_{ij}}{r_j^{max}-r_j^{min}}, \text{ for cost criteria}\end{cases}. \tag{4.4}$$

It is, of course, possible to use other normalization formulas instead of (4.4). The weights obtained from the BWM may be intervals, which means we have to consider the interval calculation of the values of alternatives ($\bar{V}_i^k$). Because we assumed that the value function has an additive form, it makes sense to proceed in parallel with the approach commonly used in Multi-Attribute Value Theory (MAVT) and take the weighted average of the contributions of each criterion, just by using interval arithmetic and the following formula:

$$\bar{V}_i^k=\sum_{j=1}^{n} \bar{w}_j^k p_{ij}=\left[\sum_{j=1}^{n} p_{ij} w_j^{k,min}, \sum_{j=1}^{n} p_{ij} w_j^{k,max}\right]. \tag{4.5}$$

However, it is well known that the obtained interval would be unrealistically wide. In fact, if we require the weight vector to be nonnegative and include components summing up to one, it is easy to show that the lower and upper bounds of equation (4.5) cannot be reached. For instance, if all the weights were at their lowest level, $w_j^{k,min}$, they would not add up to 1 (in which case there is no full consistency), which means that, together, they would not represent a weight vector. To solve this problem, we need to use constrained interval arithmetic (Lodwick, 2007) and use the following approach:

$$\bar{V}_i^k=\left[V_i^{k,min}, V_i^{k,max}\right], \tag{4.6}$$

where

$$V_i^{k,min}=min\left\{\sum_{j=1}^{n} p_{ij} w_j \middle| w_j \in \left[w_j^{k,min}, w_j^{k,max}\right]\ j=1,\dots,n, \sum_{j=1}^{n} w_j=1\right\},$$

$$V_i^{k,max}=max\left\{\sum_{j=1}^{n} p_{ij} w_j \middle| w_j \in \left[w_j^{k,min}, w_j^{k,max}\right]\ j=1,\dots,n, \sum_{j=1}^{n} w_j=1\right\}.$$

This generates stakeholder-specific *values*, where the lower and upper bounds result from stakeholder k's criteria evaluations: $\left[V_i^{k,min}, V_i^{k,max}\right]$. With the value of each stakeholder k for each alternative i, each alternative's set of values is used to determine the overall value score for the alternative.

After obtaining the value for each expert, we then need to calculate the aggregated value for all the experts. A traditional technique for aggregating intervals in the literature is to take the average of the interval centres (Yaniv, 1997). However, that approach does not take the ranges of the intervals into account (Lyon et al., 2015) and overlooks the overlapping areas of the intervals.

Thus, in this section a group consensus model is proposed to solve this problem, the rationale of which is to eliminate outliers and place the aggregated value in the overlapping areas to the greatest extent possible because they represent the consensual opinions of the experts. To that end, we need to calculate the minimum of the sum of the differences between the aggregated values (which we first need to determine), as well as each stakeholder's extremes of the interval $\bar{V}_i^k \in \left[V_i^{k,min}, V_i^{k,max}\right]$, for which we use the following approach:

$$\bar{V}_i^{agg} = \left\{x^* | x^* = arg\,\min_x \sum_{k=1}^{K} \left|x - V_i^{k,max}\right| + \left|x - V_i^{k,min}\right|\right\}, \tag{4.7}$$

where $\bar{V}_i^{agg}$ itself could possibly be an interval. It can be observed that if we rank from the smallest to the greatest all the $V_i^{k,min}$ and the $V_i^{k,max}$ values for all the experts and we rename them in a unique ordered set $\{y_1, \dots, y_{2K}\}$, we then obtain $V_i^{agg} = [y_K, y_{K+1}]$. Figure 4-3 clarifies the approach with two simple examples with three experts each.

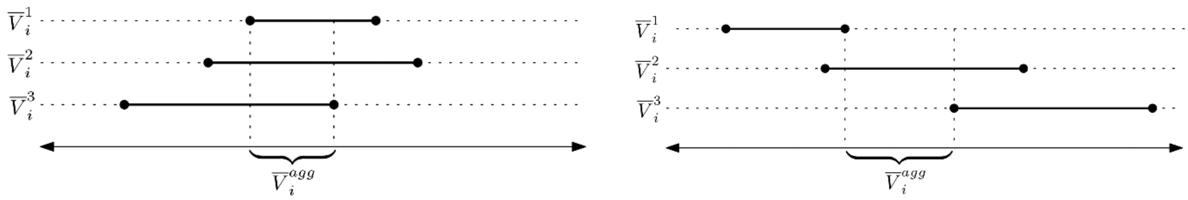


Figure 4-3: Two examples of computing $\bar{V}_i^{agg}$.

Our decision to use this formulation is based on the fact that it allows us to eliminate outliers, similar to what the median does for a set or real numbers[14].

Phase 6: Location selection

The resulting values can now be ranked[15] to select the most desirable alternative.

4.4 Application to the location selection problem

The method described above is applied to the inland terminal location selection case study of the Maersk shipping line. The results of the case study are discussed in this section. The framework of the group BWM methodology proposed in Section 4.3 applied to this study is presented in Figure 4-4.

[14] It can actually be shown that when the intervals collapse into real numbers our approach identifies the median of these real numbers. Following the conventions of probability theory, we interpret the median of an even number of observations as the interval that has the two middlemost values as extremes.

[15] For the ranking method of intervals, we refer readers to Rezaei (2016).

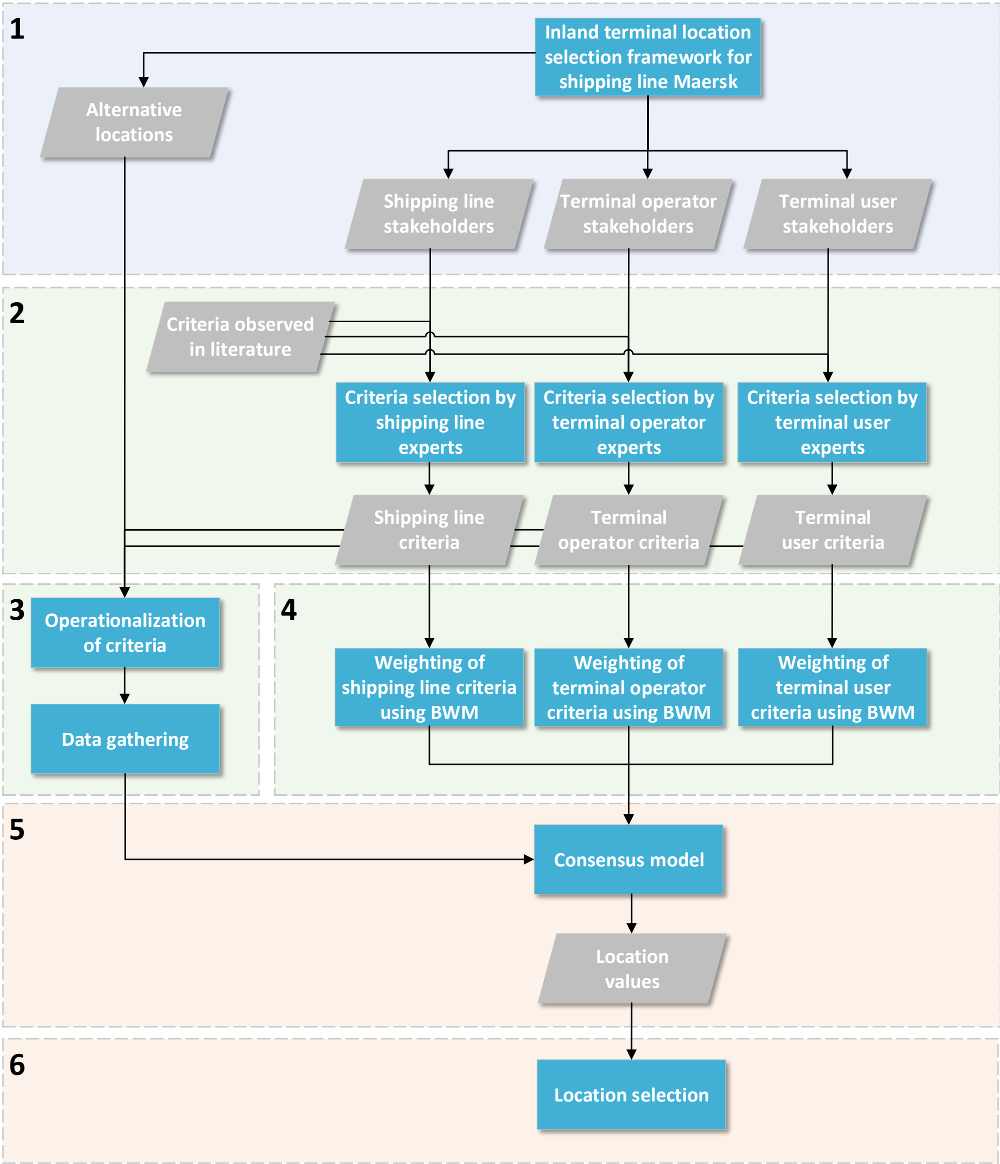


Figure 4-4. The framework of multi-stakeholder BWM inland terminal location selection

4.4.1 Results of Phase 1: Location selection problem definition

The inland terminal location selection problem examined in this study involves a specific geographical region (for reasons of confidentiality, Maersk requested not to reveal the name of the locations), which, broadly speaking, encompasses the urban and catchment areas of the cities we shall refer to as C, D and E. In total, six alternative locations were submitted for evaluation (two per city), referred to generically as C1, C2, D1, D2, E1 and E2. All alternatives meet the conditions of being at least 50 hectares in size and being located next to road and rail infrastructure, as indicated in advance by Maersk.

Next, relevant stakeholders are included by the Maersk shipping line. In addition to including the shipping line and terminal operator as decision-making stakeholders, two types of terminal user stakeholders were also included for the evaluations: truck transport operators, as representatives of pre-/end-haulage transport operators, and rail transport operators, as representatives of intermodal transport operators. Only rail operators were approached due to the geographical scope of this study, since the region in question does not have an extensive inland waterway network and/or a correspondingly extensive (focus on the) barge transport market in connection to intermodal hinterland transport[16]. A selection of relevant experts of Maersk and the relevant vendors[17] of the company is also included to represent the stakeholder expert groups. These experts were nominated by the managers of Maersk as representatives of the different areas. In total, 12 shipping line experts, eight terminal operator experts, three rail transport operator experts and two truck transport operator experts were contacted.

Despite the fact that the resulting stakeholder configuration consists of the actual decision-making stakeholders (i.e., the shipping line and terminal operator stakeholders) and the evaluating stakeholders (i.e., terminal users), as discussed in Section 4.2, it was decided to equally include all the stakeholders and their preferences in the initial consensus model setup[18]. As such, no specific weights were assigned to specific stakeholders in order to reflect their assumed importance within the actual decision-making process.

4.4.2 Results of Phase 2: Decision criteria

The stakeholder-specific decision-making criteria are based on the way the experts assessed the criteria we identified in the literature review. Through a survey, the criteria resulting from the literature review were presented to the stakeholders, who were asked to indicate the ones they found most important with regard to evaluating the location of an inland terminal. The criteria included for each stakeholder were based on the number of times they were mentioned by the experts. Because the shipping line provided more experts than the other stakeholders, a criterion had to be mentioned at least twice by experts from the shipping line to be included, as opposed to at least once for the other stakeholders. In addition, the experts were asked to suggest additional important criteria they felt were missing from the list we provided. The resulting criteria for each stakeholder are listed in Table 4-6.

[16] This means that barge operators are excluded from the surveys and inland waterway networks and features are excluded from the analyses.

[17] The vendors are the suppliers of the shipping line, i.e., the terminal and transport operators performing the inland services offered by Maersk.

[18] The decision-maker at Maersk agreed to this arrangement.

Table 4-6. Criteria to be considered in a study based on the criteria selection survey results

Criterion	Stakeholder			
	Shipping line	Terminal operator	Rail transport operator	Truck transport operator
Anchor customer proximity*		✓		
Enabling modality shift*		✓		
Expansion possibilities	✓		✓	✓
Government policy	✓			
Inland terminal CAPEX	✓	✓	✓	
Intermodal market profitability*	✓			✓
Land use near location				✓
Local depot capacity*	✓			
Market proximity	✓		✓	✓
Market volume potential	✓	✓	✓	✓
Regional economic development		✓		
Regional transport/logistics competition	✓	✓		✓
Terminal market profitability*				✓
Total inland transport costs	✓	✓		✓
Transport infrastructure network in area	✓	✓	✓	✓

** Criteria additionally added based on the experts' inputs*[19].

4.4.3 Results of Phase 3: Operationalized criteria and data

Phase 3 involves the operationalization of the criteria into measurable and comparable metrics, which leads to particular data-gathering methods and, ultimately, to a collection of quantitative data corresponding to each criterion for each alternative. The resulting data are listed in Table 4-7.

[19] *Anchor customer proximity*: the anchor customer volume (in FEU) within an area. *Enabling modality shift*: the potential volume to be shifted from the modality in an area. *Intermodal market profitability*: the margins gained from providing/practising intermodal transport services in a certain area. *Local depot capacity*: the total container volume that could possibly be stored in the broader area in which an alternative location is situated. *Terminal market profitability*: the margins gained from providing inland terminal operations in a certain area.

Table 4-7. Evaluation of alternatives based on the criteria

Criteria	Alternatives					
	C1	C2	D1	D2	E1	E2
Anchor customer proximity (FEU) [20]	17431	15931	1337	1337	0	0
Enabling modality shift (FEU)	23215.21	21950.21	1337	1337	2277.56	2946.06
Expansion possibilities (m^2)	831919	0	1039121	174705	102817	0
Government policy (index)[21]	1	1	1	1	0	0
Inland terminal CAPEX (million €)	5.85175	5.77675	5.67675	5.75175	5.65175	5.62675
Intermodal market profitability (€/TEU)	87.78	93.10	127.01	151.97	147.07	144.85
Land use near location (index)[22]	1	1	1	0	0	0
Local depot capacity (TEU)	5650	5650	2950	2950	1000	1000
Market proximity (FEU/km)	3357.62	3670.63	875.27	887.38	740.62	744.45
Market volume potential (FEU)	31509.73	29993.73	4751.15	4870.10	5947.68	7427.05
Regional economic development (€)	31600	31600	30300	30300	28300	28300
Regional transport/logistics competition (number)[23]	108	108	130	130	134	134
Terminal market profitability (€)[24]	-24.06	-24.06	-27.67	-27.67	-15	-15
Total inland transport costs (€/TEU)	432.22	426.90	555.49	530.53	495.43	497.65
Transport infrastructure network in area (km/100 km^2)	78.18	61.86	33.94	42.55	112.45	111.15

4.4.4 Results of Phase 4: Criteria weights

In this phase, the criteria discussed above are further evaluated to assign weight factors to each stakeholder via the BWM. To that end, a preference statement survey was sent to the expert group, in which they indicate the Best and Worst criterion, the relative preferences of the remaining criteria compared to the Best criterion and the relative preferences of the Worst criterion compared to the remaining criteria. Based on the survey, which was ultimately completed by nine respondents (three experts from the shipping line, three from the terminal operator, two from the rail transport operator and one from the truck transport operator), the criteria weight intervals for each expert were calculated by means of solving the nonlinear

[20] Forty-foot Equivalent Unit (FEU)

[21] A government policy index is proposed to indicate whether the local government within the region is willing to support inland terminal development, where -1 is negative, 0 is neutral, and 1 is positive.

[22] We use 1 and 0 to represent whether the land-use near the location has a positive or negative effect on the (operations of) the potential inland terminal.

[23] The regional number of companies offering transport and logistics services on one hand implies a certain level of competition, while on the other hand also indicates the potential for cooperation. Since the line between these two is not directly clear, the *competition* factor is simplified to the *number of companies offering transport, logistics and terminal services in the area.*

[24] Terminal market profitability involves the margins gained from providing inland terminal operations in a certain area. These margins thus depend on the local market, which is quantitatively assessed through the rates applied by the locally existing and operating terminal service providing companies. As the rates applied by the shipping line are not yet known (because it currently does not operate any terminal it owns, no rates have been developed), only the *costs* for terminal handlings are used as in indication for the possible margins gained per area (the higher the costs, the lower the margins).

BWM model with the respective preference statement inputs, as shown in Table 4-8 to Table 4-11.

Table 4-8. Criteria weights for shipping lines

Criterion	Expert					
	Shipping line 1		Shipping line 2		Shipping line 3	
	Lower	Upper	Lower	Upper	Lower	Upper
Expansion possibilities	0.005	0.01	0.004	0.004	0.08	0.08
Government policy	0.03	0.03	0.02	0.02	0.06	0.06
Inland terminal CAPEX	0.03	0.03	0.17	0.17	0.06	0.06
Intermodal market profitability	0.18	0.18	0.40	0.40	0.19	0.19
Local depot capacity	0.02	0.02	0.02	0.02	0.03	0.03
Market proximity	0.24	0.27	0.03	0.03	0.27	0.27
Market volume potential	0.19	0.24	0.17	0.20	0.16	0.16
Regional transport/logistics competition	0.07	0.08	0.07	0.10	0.10	0.10
Total inland transport costs	0.16	0.16	0.05	0.05	0.02	0.02
Transport infrastructure network in area	0.04	0.04	0.04	0.04	0.02	0.02

Table 4-9. Criteria weights for terminal operators

Criterion	Expert					
	Terminal operator 1		Terminal operator 2		Terminal operator 3	
	Lower	Upper	Lower	Upper	Lower	Upper
Anchor customer proximity	0.09	0.17	0.14	0.16	0.27	0.28
Enabling modality shift	0.08	0.10	0.03	0.04	0.03	0.05
Inland terminal CAPEX	0.03	0.04	0.10	0.11	0.08	0.09
Market volume potential	0.12	0.21	0.28	0.31	0.13	0.13
Regional economic development	0.23	0.30	0.14	0.16	0.02	0.03
Regional transport/logistics competition	0.05	0.10	0.14	0.16	0.13	0.13
Total inland transport costs	0.09	0.15	0.04	0.06	0.27	0.28
Transport infrastructure network in area	0.11	0.16	0.06	0.08	0.03	0.04

Table 4-10. Criteria weights for rail transport operators

Criterion	Expert			
	Rail transport operator 1		Rail transport operator 2	
	Lower	Upper	Lower	Upper
Expansion possibilities	0.04	0.04	0.11	0.11
Inland terminal CAPEX	0.17	0.19	0.11	0.11
Market proximity	0.09	0.13	0.05	0.05
Market volume potential	0.38	0.43	0.36	0.36
Transport infrastructure network in area	0.23	0.31	0.36	0.36

Table 4-11. Criteria weights for truck transport operators

Criterion	Truck transport operator expert	
	Lower	Upper
Expansion possibilities	0.03	0.05
Intermodal market profitability	0.13	0.22
Land use near location	0.02	0.03
Market proximity	0.15	0.25
Market volume potential	0.24	0.30
Regional transport/logistics competition	0.05	0.07
Terminal market profitability	0.07	0.09
Total inland transport costs	0.10	0.13
Transport infrastructure network in area	0.03	0.06

In addition to calculating the weight intervals, the comparison consistencies are checked using the consistency threshold values. Using equation (4.2) we can calculate the consistency ratios and then compare them to Table 2-3. We found that, except for Rail transport operator 1 (whose CR is 0.2857, against a consistency threshold of 0.2844) [25], all the CRs of the other experts are below the corresponding thresholds.

4.4.5 Results of Phase 5: Values of the alternatives

First, the original evaluation data in Table 4-7 need to be normalized using equation (4.4), which results in Table 4-12, where we use "+" to present the benefit criteria and "-" to present the cost criteria.

Table 4-12. Normalized data scores for all criteria sets

Criteria	Alternatives					
	C1	C2	D1	D2	E1	E2
Anchor customer proximity (+)	1	0.91	0.08	0.08	0	0
Enabling modality shift (+)	1	0.94	0	0	0.04	0.07
Expansion possibilities (+)	0.80	0	1	0.17	0.10	0
Government policy (+)	1	1	1	1	0	0
Inland terminal CAPEX (-)	0	0.33	0.78	0.44	0.89	1
Intermodal market profitability (+)	0	0.08	0.61	1	0.92	0.89
Land use near location (+)	1	1	1	0	0	0
Local depot capacity (-)[26]	0	0	0.58	0.58	1	1
Market proximity (+)	0.89	1	0.05	0.05	0	0
Market volume potential (+)	1	0.94	0	0	0.04	0.10
Regional economic development (+)	1	1	0.61	0.61	0	0
Regional transport/logistics competition (-)	1	1	0.15	0.15	0	0
Terminal market profitability (+)	0.29	0.29	0	0	1	1
Total inland transport costs (-)	0.96	1	0	0.19	0.47	0.45
Transport infrastructure network in area (+)	0.56	0.36	0	0.11	1	0.98

[25] Because the CR of Rail transport operator 1 lies slightly above the threshold, and because the relevant comparisons are ordinally consistent (for the checking method, refer to Liang et al. (2020)), we consider this situation to be acceptable and the expert was not asked to revise the preferences.

[26] Local depot capacity is a cost criterion because it measures the amount of depot capacity already present in the area. In this sense, the greater the capacity already present, the more competition there is in the area and the lower the demand for a new depot capacity.

The normalized data, in combination with the criteria weight intervals as determined in Phase 4, are used to calculate the value assigned to the alternatives. The separate values based on each stakeholder's preferences and the corresponding data are added to generate a set of values for each alternative using equation (4.5). The resulting values for the various individual experts are shown in Figure 4-5. The error bars represent the interval values of the alternatives, while the columns represent the middle values of the intervals.

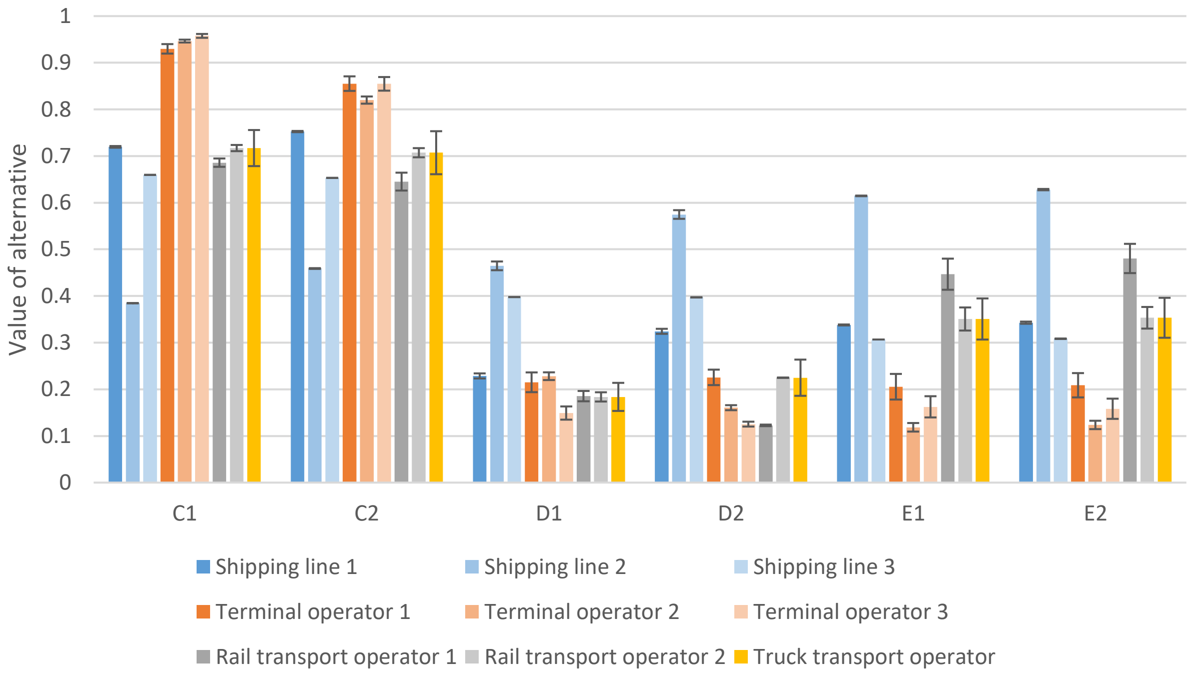


Figure 4-5. Values for each alternative of each expert

The individual values are then aggregated using the group consensus model (4.7). The aggregate values for each alternative are shown in Figure 4-6, where the average values are presented numerically.

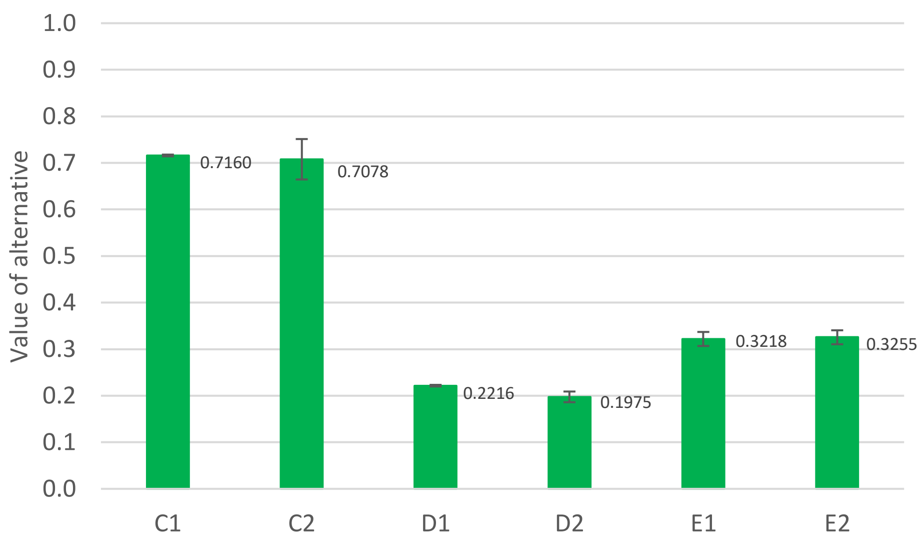


Figure 4-6. The aggregated value of each alternative

4.4.6 Results of Phase 6: Location selection

The final values of C1 and C2 (see Figure 4-6) are very close to one another. In fact, C1 and C2 are two sites in one district.

4.4.7 Sensitivity analysis

A sensitivity analysis was conducted to test how the input affects the final results, which can be assessed by adapting the corresponding model parameters and comparing the results to those of the model in Phase 5.

Because the container volume is considered to be relatively important by most stakeholders, it accounts for a substantial share of the value for the alternatives, especially for the locations in area C, where the potential container volume is significantly higher than those in the other areas, which increases the relative value of C compared to the others. However, due to particular economic events, container volumes generated in/attracted to a certain area may have more fluctuations in shorter time periods, for example, when a (new) shipper/consignee opens a facility in a particular region (e.g., a factory or distribution centre), which is sufficiently large to cause a relatively high influx of annual container volumes. As such, *the market volume potential* also turns out to be a critical factor in the MCDM study.

We use several hypothetical (albeit fairly realistic) scenarios to assess a sudden growth in container volume areas D and E, which are selected because they currently have the lowest container volumes. This allows us to measure the effect of such an increase in volume on their performance as potential locations compared to option C, which already has higher volumes. As the hypothetical market volume potential influx is unlikely to occur at both locations at the same time, we decided to assess them individually by increasing their *market volume potential* incrementally[27].

The incremental development of the aggregate values[28] for the alternatives as a result of an increase in container volume areas D and E are shown in Figure 4-7 and Figure 4-8, respectively.

[27] An incremental market volume potential influx is considered; the minimal increase begins with 10,000 FEU extra potential p.a. (which is approximately the size of the largest shippers/consignees in the current study region in eastern Germany), up to a maximum increase of 30,000 FEU extra potential p.a. (which is approximately the size of the largest shippers/consignees in all of Germany), with intermediate steps of 5,000 FEU p.a.

[28] For a clear presentation, we only use the average values instead of using intervals, but this has no impact on the final implications.

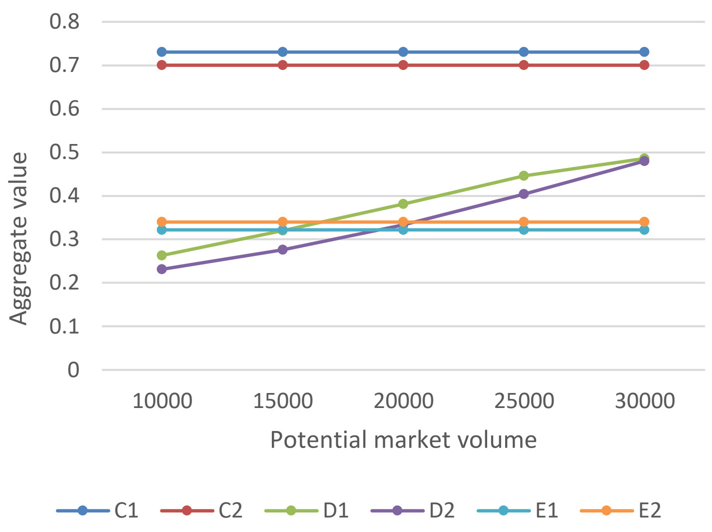


Figure 4-7. Aggregate values of alternatives with respect to a varying container volume influx in area D.

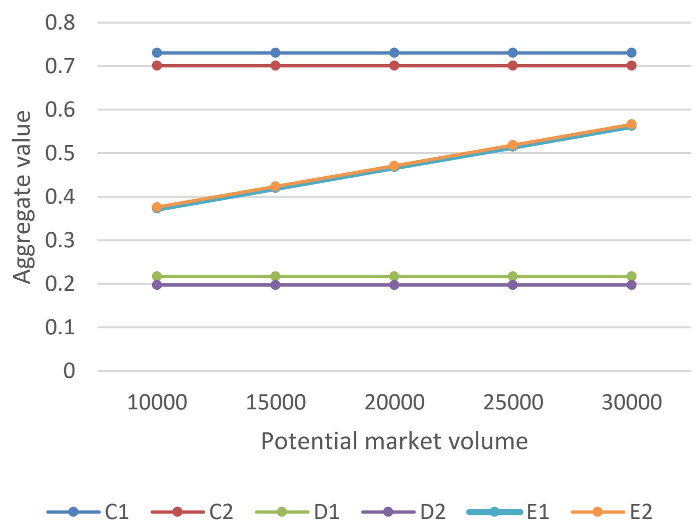


Figure 4-8. Aggregate values of alternatives with respect to a varying container volume influx in area E.

As can be seen, their values as alternative locations increase along with the container volume, and when the market volume potential of area D surpasses 20,000 FEU p.a. (per annum), the value of area D exceeds that of area E (see Figure 4-7). However, the growth in the market volume in areas D and E does not change the fact that area C is the best alternative, which means that, between 10,000 and 30,000, an increase in the potential market volume in an alternative location has no impact on the final selection.

This conclusion can be supported by noting that the weight of the criterion Potential market volume is, at the most, 0.43, as specified by Rail transport operator 1, which is less than the value difference between the two best alternatives and the remaining four, which means that, even if we increase the Potential market volume to the maximum level for alternatives D1, D2, E1 and E2, that does not make them preferable to C1 and C2.

4.5 Conclusions and Discussion

Summary: The aim of this study was to identify a desirable location for a shipping line to set up an inland terminal that can be incorporated into the inland transport services the company provides to its customers, which meant we had to take the interests of a number of different stakeholders into account, including the shipping line, terminal operator, and terminal user. To that end, a group BWM consensus model was proposed to determine the criteria weights for each stakeholder and aggregate the interval values of the relevant alternatives.

Results and discussions: In addition to the existing criteria, several new criteria (anchor customer proximity, enabling modality shift intermodal market profitability, local depot capacity, and terminal market profitability) were added to this specific project to shed light on the criteria set. Based on these criteria, different stakeholders apply different subsets of criteria to assess their particular areas. While the shipping line assigns the greatest value to *intermodal market profitability* and *market proximity*, terminal operators and users indicate that they place greater value on container volume-related (e.g., *market volume potential*) factors. Overall, these container-related characteristics are rated highly by all the stakeholders involved, including the shipping line. As a result, the two locations in city C were the best alternatives, in large part thanks to the potential container volume in area C. Sensitivity analyses indicated that the influence of a growth in the market volume at the other locations has a limited impact on the final selection.

Furthermore, it can be concluded that most of the weights the different stakeholders placed on specifically relevant criteria ultimately lead to comparable results in terms of preferability, which shows that although the business models and objectives of the different stakeholders may vary, an increasingly (cost) efficient inland transport chain ultimately benefits everyone involved. Based on the insights from the relevant stakeholders and combined with the data used as input for the MCDM model, location C1 emerged as the optimal choice, especially in terms of local market conditions and expansion possibilities.

Limitation and future study: First, in a real-life situation there may be more stakeholders involved in selecting inland terminals than those included in this study, including the government and local shippers. These stakeholders were not included in this study because of the nonexecutive role they play in the container transport domain. However, since the government (can) also take(s) part in the development of terminal infrastructure in areas under its jurisdiction, it would be interesting to include its preferences with regard to the relevant criteria. Second, as far as shippers are concerned, the decision to not include them in the research stems from the Maersk's objective to first choose a location near some potential partner shippers and conduct more in-depth research including other shippers at a later date. Third, further research should look at the inland terminals themselves and on the terminals within the container port hinterland in general, rather than focusing exclusively on the specific framework applied in this manuscript.

In addition, it may be worthwhile to take a closer look at the shipping line's vertical integration into the container port hinterland in a broader sense, which is an increasingly interesting development. In the past, before shipping lines began providing door-to-door services, the inland transport services were managed entirely by third parties such as freight forwarders, making them *customers* of the shipping lines. However, because of the developments described in this research, shipping lines are increasingly becoming *competitors* of former customers, thus it would be interesting to explore the changing dynamics between shipping lines and their customer competitors, especially since the implications of these changing dynamics could also affect the suitability of inland terminals located in certain inland transport chains.

Acknowledgment

The authors especially appreciate the help from Matej Vybiral from Maersk, Bart Wiegmans from Delft University of Technology, and all the other experts from Maersk, APM Terminals and Maersk's vendors. Additionally, the authors gratefully acknowledge the financial support for PhD study from the China Scholarship Council (No. 201708440305).

References

Alumur, S. & Kara, B. Y. (2008). Network hub location problems: The state of the art. *European Journal of Operational Research*, *190*(1), 1-21.

Bergqvist, R. & Monios, J. (2014). The role of contracts in achieving effective governance of intermodal terminals. *World Review of Intermodal Transportation Research*, *5*(1), 18-38.

Bergqvist, R., Macharis, C., Meers, D. & Woxenius, J. (2015). Making hinterland transport more sustainable a multi actor multi criteria analysis. *Research in Transportation Business & Management*, *14*, 80-89.

Bontekoning, Y. M., Macharis, C. & Trip, J. J. (2004). Is a new applied transportation research field emerging?--A review of intermodal rail-truck freight transport literature. *Transportation Research Part A: Policy and Practice*, *38*(1), 1-34.

Caris, A., Macharis, C. & Janssens, G. K. (2012). Corridor network design in hinterland transportation systems. *Flexible Services and Manufacturing Journal*, *24*(3), 294-319.

De Langen, P. W., Fransoo, J. C. & van Rooy, B. (2013). Business models and network design in hinterland transport. Handbook of Global Logistics, Springer: 367-389.

Douma, A. (2008). Aligning the operations of barges and terminals through distributed planning., University of Twente, Enschede, Netherlands.

Franc, P. & Van der Horst, M. (2010). Understanding hinterland service integration by shipping lines and terminal operators: a theoretical and empirical analysis. *Journal of Transport Geography*, *18*(4), 557-566.

Frémont, A. (2009). Shipping lines and logistics. *Transport Reviews*, *29*(4), 537-554.

Ka, B. (2011). Application of fuzzy AHP and ELECTRE to China dry port location selection. *The Asian Journal of Shipping and Logistics*, *27*(2), 331-353.

Karaşan, A. & Kahraman, C. (2019). A novel intuitionistic fuzzy DEMATEL-ANP-TOPSIS integrated methodology for freight village location selection. *Journal of Intelligent & Fuzzy Systems*, *36*(2), 1335-1352.

Kayikci, Y. (2010). A conceptual model for intermodal freight logistics centre location decisions. *Procedia - Social and Behavioral Sciences*, *2*(3), 6297-6311.

Keeney, R. L. & Raiffa, H. (1976). Decision analysis with multiple conflicting objectives. New York, John Wiley & Sons.

Kheybari, S., Kazemi, M. & Rezaei, J. (2019). Bioethanol facility location selection using best-worst method. *Applied Energy*, *242*, 612-623.

Komchornrit, K. (2017). The selection of dry port location by a hybrid CFA-MACBETH-PROMETHEE method: A case study of Southern Thailand. *The Asian Journal of Shipping and Logistics*, *33*(3), 141-153.

Lee, P. T. & Yang, Z. (2018). Multi-Criteria Decision Making in Maritime Studies and Logistics, Springer.

Li, F., Shi, X. & Hu, H. (2011). Location selection of dry port based on AP clustering-the case of southwest China. *Journal of System and Management Sciences*, 1(5), 79-88.

Liang, F., Brunelli, M. & Rezaei, J. (2020). Consistency issues in the best worst method: Measurements and thresholds. *Omega*, *96*, 102175.

Limbourg, S. & Jourquin, B. (2009). Optimal rail-road container terminal locations on the European network. *Transportation Research Part E: Logistics and Transportation Review*, *45*(4), 551-563.

Lodwick, W. A. (2007). Interval and fuzzy analysis: A unified approach. Advances in Imaging and Electron Physics. P. Hawkes, Elsevier. 148: 75-192.

Long, S. & Grasman, S. E. (2012). A strategic decision model for evaluating inland freight hub locations. *Research in Transportation Business & Management*, *5*, 92-98.

Lyon, A., Wintle, B. C. & Burgman, M. (2015). Collective wisdom: Methods of confidence interval aggregation. *Journal of Business Research*, *68*(8), 1759-1767.

Macharis, C., Turcksin, L. & Lebeau, K. (2012). Multi actor multi criteria analysis (MAMCA) as a tool to support sustainable decisions: State of use. *Decision Support Systems*, *54*(1), 610-620.

Mi, X., Tang, M., Liao, H., Shen, W. & Lev, B. (2019). 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*, *87*, 205-225.

Monios, J. & Wilmsmeier, G. (2012). Giving a direction to port regionalisation. *Transportation Research Part A: Policy and Practice*, *46*(10), 1551-1561.

Nguyen, L. C. & Notteboom, T. (2016). A multi-criteria approach to dry port location in developing economies with application to Vietnam. *The Asian Journal of Shipping and Logistics*, *32*(1), 23-32.

Notteboom, T. (2011). An application of multi-criteria analysis to the location of a container hub port in South Africa. *Maritime Policy & Management*, *38*(1), 51-79.

Notteboom, T. & Rodrigue, J. (2017). Re-assessing port-hinterland relationships in the context of global commodity chains. Ports, cities, and global supply chains, Routledge: 67-82.

Özceylan, E., Erbaş, M., Tolon, M., Kabak, M. & Durğut, T. (2016). Evaluation of freight villages: A GIS-based multi-criteria decision analysis. *Computers in Industry*, *76*, 38-52.

Pamučar, D., Gigović, L., Bajić, Z. & Janošević, M. (2017). Location Selection for Wind Farms Using GIS Multi-Criteria Hybrid Model: An Approach Based on Fuzzy and Rough Numbers. *Sustainability*, *9*(8), 1315.

Pekin, E. (2010). A GIS-based intermodal transport policy evaluation model, PhD thesis, Vrije Universiteit Brussel, Brussels.

Regmi, M. B. & Hanaoka, S. (2013). Location analysis of logistics centres in Laos. *International Journal of Logistics Research and Applications*, *16*(3), 227-242.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49-57.

Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega*, *64*, 126-130.

Rezaei, J. (2020). A concentration ratio for nonlinear best worst method. *International Journal of Information Technology & Decision Making*, *19*(3), 891-907.

Rodrigue, J. & Notteboom, T. (2009). The terminalization of supply chains: reassessing the role of terminals in port/hinterland logistical relationships. *Maritime Policy & Management*, *36*(2), 165-183.

Rodrigue, J. & Notteboom, T. (2013). Containerized freight distribution in North America and Europe. Handbook of Global Logistics, Springer: 219-246.

Rodrigue, J., Debrie, J., Fremont, A. & Gouvernal, E. (2010). Functions and actors of inland ports: European and North American dynamics. *Journal of Transport Geography*, *18*(4), 519-529.

Roso, V., Brnjac, N. & Abramovic, B. (2015). Inland intermodal terminals location criteria evaluation: The case of Croatia. *Transportation Journal*, *54*(4), 496-515.

Rožić, T., Ogrizović, D. & Galić, M. (2016). Decision making background for the location of inland terminals. *Pomorstvo, 30*(2), 141-150.

Simina, D., Patrick, S., Radu, C. & Others (2012). Economic benefits of developing intermodal transport in the European Union. *Annals of Faculty of Economics*, *1*(2), 81-87.

Song, D. & Dong, J. (2011). Effectiveness of an empty container repositioning policy with flexible destination ports. *Transport Policy*, *18*(1), 92-101.

Stević, Ž., Pamučar, D., Subotić, M., Antuchevičiene, J. & Zavadskas, E. K. (2018). The location selection for roundabout construction using Rough BWM-Rough WASPAS approach based on a new Rough Hamy aggregator. *Sustainability*, 10(8), 2817.

Tan, Z., Meng, Q., Wang, F. & Kuang, H. (2018). Strategic integration of the inland port and shipping service for the ocean carrier. *Transportation Research Part E: Logistics and Transportation Review*, *110*, 90-109.

Teye, C., Bell, M. G. H. & Bliemer, M. C. J. (2017). Urban intermodal terminals: The entropy maximising facility location problem. *Transportation Research Part B: Methodological*, *100*, 64-81.

Teye, C., Bell, M. G. & Bliemer, M. C. (2018). Locating urban and regional container terminals in a competitive environment: An entropy maximising approach. *Transportation Research Part B: Methodological*, *117*, 971-985.

Tsao, Y. & Thanh, V. (2019). A multi-objective mixed robust possibilistic flexible programming approach for sustainable seaport-dry port network design under an uncertain environment. *Transportation Research Part E: Logistics and Transportation Review*, *124*, 13-39.

Van den Berg, R. & De Langen, P. W. (2015). Assessing the intermodal value proposition of shipping lines: Attitudes of shippers and forwarders. *Maritime Economics & Logistics*, *17*(1), 32-51.

Van den Berg, R. & De Langen, P. W. (2015). Towards an 'inland terminal centred' value proposition. *Maritime Policy & Management*, *42*(5), 499-515.

Van Nguyen, T., Zhang, J., Zhou, L., Meng, M. & He, Y. (2020). A data-driven optimization of large-scale dry port location using the hybrid approach of data mining and complex network theory. *Transportation Research Part E: Logistics and Transportation Review*, *134*, 101816.

Wan, Y., Basso, L. J. & Zhang, A. (2016). Strategic investments in accessibility under port competition and inter-regional coordination. *Transportation Research Part B: Methodological*, *93*, 102-125.

Wei, J., Sun, A. & Zhuang, J. (2010). The selection of dry port location with the method of fuzzy-ANP. Advances in Wireless Networks and Information Systems. Berlin, Heidelberg, Springer Berlin Heidelberg. 72: 265-273.

Wilmsmeier, G., Monios, J. & Lambert, B. (2011). The directional development of intermodal freight corridors in relation to inland terminals. *Journal of Transport Geography*, *19*(6), 1379-1386.

Yaniv, I. (1997). Weighting and trimming: Heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes*, *69*(3), 237-249.

Yıldırım, B. F. & Önder, E. (2014). Evaluating potential freight villages in Istanbul using multi criteria decision making techniques. *Journal of Logistics Management*, *3*(1), 1-10.

5 Best-Worst Tradeoff Method

Liang, F., Brunelli, M. & Rezaei, J. Best-Worst Tradeoff Method. (Information Sciences, Under review, 2nd round)

Abstract

This study aims to develop a Multi-Attribute Decision-Making (MADM) method, the Best-Worst Tradeoff method, which draws on the underlying principles of two popular MADM methods (the Best-Worst Method (BWM) and the Tradeoff). The traditional Tradeoff method, which is based on the axiomatic foundation of multi-attribute value theory, considers the ranges of the attributes, but decision-makers/analysts find it hard to check the consistency of the paired comparisons when using this method. The traditional BWM, on the other hand, uses two opposite references (best and worst) in a single optimization, which not only frames the elicitation process in a more structured way, but helps decision-makers/analysts check the consistency. However, the BWM does not explicitly considers the attributes ranges in the pairwise comparisons. The method proposed in this study uses the "consider-the-opposite-strategy" and accounts for the range effect simultaneously. Specifically, the decision-maker considers the ranges of the attributes and provide two pairwise comparison vectors, then an optimization model is designed to determine the optimal weights of the attributes based on these two vectors. After that, consistency thresholds are constructed to check the consistency of the judgements. Finally, a case study is used to examine the feasibility of the proposed method.

5.1 Introduction

Nowadays, an increasing number of decisions are made in complex contexts in a host of different application domains. This ought to be supported by mathematically sound decision analysis methodologies. A number of these methodologies can be classified as multi-attribute decision-making (MADM), thanks to their capacity to handle problems where a multitude of, often conflicting, objectives arise (Greco et al., 2016). The common thread of these methods is the representation of the final value of each alternative, as a function of the degrees to which the same alternative satisfies a number of attributes, where each attribute level approximates the level of achievement of one of the objectives. In this context, it is often important to quantify the contribution of different attributes by means of weights (scaling constants) to aggregate the performances of alternatives with respect to the attributes into single values.

Various methods have been proposed to elicit the weights of criteria (Zardari et al., 2015; Dias et al., 2018) on the ground of the subjective preferences of experts. Some of the most popular methods are: the Analytic Hierarchy Process (AHP) (Saaty, 1977), Simple Multi-Attribute Rating Technique (SMART) (Edwards, 1977), Direct Rating method (Bottomley and Doyle, 2001), Swing (Von Winterfeldt and Edwards, 1986), the Best-Worst Method (BWM) (Rezaei, 2015; 2016), and the Tradeoff method (Keeney and Raiffa, 1976). In this study, we focus on the Tradeoff and BWM methods.

One of the obvious shortcomings of some methods is that the weight elicitation phase is developed *a priori*, on the basis of the perceived importance of the attributes alone. Conversely, in decision analysis, the weights of the attributes (or scaling constants) should be sensitive to the range of each attribute, i.e., if alternatives are very close to each other with respect to a particular attribute, that attribute would play a small role in discriminating between alternatives (von Nitzsch and Weber, 1993; Fischer, 1995). Methods that cannot account for the range of attributes values may lead to errors in the estimation of weights (Fischer et al., 1987; von Nitzsch and Weber, 1993; Pajala et al., 2019). According to previous studies (von Nitzsch and Weber, 1993; Fischer, 1995), even if the range of the attribute is mentioned, DMs often do not adjust their judgements on the weights properly, which means that methods that do not consider the ranges, like simple ranking or direct rating methods, should only be used with great care (von Nitzsch and Weber, 1993; Fischer, 1995). Although BWM encourages DMs to consider the range of criteria in advance, in practice, this is not done systematically. In this sense, methods like Swing (Von Winterfeldt and Edwards, 1986) and Tradeoff method (Keeney and Raiffa, 1976) which require DMs to provide their preference based on the range of attributes could handle this problem better than BWM could. How to take the range of attributes into consideration in BWM in an explicit and systematic way to avoid distortion or biases, is an important issue that requires further investigation.

The idea of taking the range effect into account, like the Tradeoff method and Swing do, can be incorporated into the BWM to remedy the potential distortion and biases. Compared to the Tradeoff method, Swing cannot make the consistency check, which is a serious shortcoming. As such, with regard to external validity, the Tradeoff method performs better than Swing (Zardari et al., 2015). However, the consistency check in the Tradeoff method may also be problematic. Keeney and Raiffa (1976) encouraged analysts to use the Tradeoff method to ask additional questions to increase the robustness of the results and identify possible inconsistencies. But how many and what additional questions should be asked remains unclear. In addition, analysts frequently fail to apply consistency checks when assessing value tradeoffs (Keeney, 2002). Although some researchers have tried to improve the Tradeoff method's procedure – one of the latest studies uses a flexible and interactive way to collect trade-off information from the DM (de Almeida et al., 2016) – the consistency estimation problem has not attracted enough attention.

By incorporating the philosophy of BWM, using two vectors of pairwise comparisons based on two opposite references (best and worst) within a single optimization model, our goal is to help mitigate the anchoring bias. This strategy, named "consider-the-opposite strategy", was initially developed by Bacon (Bacon, 1960) and has since been used in many psychological studies. Developing a parsimonious MADM method that incorporates the "consider-the-opposite strategy that can check the consistency systematically as well as consider the range effect based on the axiomatic foundation of Multi-Attribute Value Theory (MAVT) is the main objective and contribution of this study. To achieve this, the underlying principles of the Tradeoff method (Keeney and Raiffa, 1976) and the BWM (Rezaei, 2015; 2016) are adopted. More specifically, firstly, MAVT is used as a foundation to establish a trade-off for the objectives based on the range of the attributes. Secondly, two vectors of pairwise comparisons, Best-to-Others (BO) and Others-to-Worst (OW), are constructed to avoid that the DM is only anchoring on a fixed value or reference point, the way it is done in SMART and Swing. Thirdly, based on the BO and OW obtained from the DM, an optimization model is proposed to derive the weights of attributes. In addition, a cardinal consistency index and an ordinal consistency index are proposed to estimate the extent to which the preferences of a DM deviate from the perfect consistency, and thresholds will be proposed to decide whether or not the deviation is acceptable.

The remainder of the paper is organized as follows: In Section 5.2, the basic knowledge of MAVT and the procedure of the classical Tradeoff method are reviewed, while, in Section 5.3, the procedure of the BWT method is illustrated. Section 5.4 focuses on the consistency-checking, which includes the proposed consistency ratios and the thresholds for the BWT. A case study is used to illustrate the proposed method in Section 5.5, while the features of BWM are discussed in Section 5.6 and some conclusions are presented in Section 5.7.

5.2 Preliminaries

Of the many methods and theories developed to support MADM processes, MAVT is one of the most widely used, as well as the one for which Keeney and Raiffa (1976) proposed the well-known Tradeoff method. In the following sub-sections, MAVT and the classical Tradeoff method are discussed, after which the recently developed BWM method is introduced.

5.2.1 Multi-attribute value theory and the additive value function

A typical MADM problem consists of a non-empty finite set of m alternatives $A = \{A_1, A_2, \dots, A_m\}$ and a set of objectives that represent the goals of the DM. The satisfaction of the objectives is assumed to depend on a finite set of n attributes (also known as criteria) $C = \{C_1, C_2, \dots, C_n\}$. The set of possible levels that can be achieved by a generic alternative with respect to the jth attribute is denoted by X_j. Assuming that all the relevant attributes have been considered, each alternative can be associated with a consequence $\mathbf{x} = (x_1, x_2, \dots, x_n) \in X_1 \times X_2 \times \cdots \times X_n$, such that x_j indicates the level of the jth attribute achieved by the alternative.

Given the fact that, to our scope, an alternative can be fully described by its consequence, for the sake of simplicity, in the following section, we will consider consequences, $\mathbf{x}$, instead of alternatives. Moreover, according to MAVT, each attribute is rescaled and normalized into the interval $[0,1]$ thanks to a function $v_j: X_j \to [0,1]$. In fact, it is customary to assign values 0 and 1 to the least ($\underline{x}_j$) and most ($\overline{x}_j$) desirable attribute levels, respectively. Hence, $v_j(x_j) \in [0,1]$ can be interpreted as the value of the consequence evaluated with respect to the jth attribute. At this point, taking into account that the value of a consequence is a function of its attribute values, the main problem in MAVT is to find a function $u: [0,1]^n \rightarrow [0,1]$, which can correctly aggregate the n values $v_j(x_j)$ and represent the preferences of a DM.

The search for a simple and easily interpretable form for the function u is simplified by some known results stating that, under mild assumptions (i.e. mutual preference independence and measurability), the function u is additive. Namely,

$$u(v(\boldsymbol{x})) = u(v_1(x_1), \dots, v_n(x_n)) = \sum_{j=1}^{n} w_j v_j(x_j), \tag{5.1}$$

with $w_j \geq 0, \forall j$ and $w_1 + w_2 + \cdots + w_n = 1$.

At this point, what is left to do is find suitable values for the weights w_j. Among the various methods used to elicit weights, the Tradeoff method has a strong axiomatic foundation (Weber and Borcherding, 1993).

5.2.2 The classical Tradeoff method

The original Tradeoff procedure essentially consists of three parts. The goal of the first part is to obtain the preference relation of the attributes. Attributes are considered in pairs, and in each pair two hypothetical consequences are constructed and presented to DM for tradeoffs. These two consequences are only different with regard to the performance of the two attributes under consideration, and the performance of the other attributes is set to the worst level. In the first hypothetical consequence, the performance of the two attributes is set to their worst and best levels, respectively, and in the second consequence, they are set the other way around. The DM is asked to indicate which consequence is preferred. After having applied these pairwise comparisons to all the attributes, the relation of order on the set of weights can be obtained.

The second part is designed to obtain indifference relations. The DM still uses the hypothetical consequences created in the first part, but now, the DM is asked to manipulate the level of the more important attribute until indifference is reached. By iterating this procedure for a properly chosen set of $(n-1)$ comparisons between pairs of artificially constructed payoffs, indifference relations on these hypothetical paired consequences can be obtained.

The third part is intended to determine the weights. On the basis of the indifference relations obtained in the second part, together with the constraint $w_1 + w_2 + \cdots + w_n = 1$, the analyst can form a system of linear equations for which a unique set of weights exists and can be identified.

Despite the simplicity of the original Tradeoff method, Keeney and Raiffa (1976) questioned its robustness and consistency, by saying that "*it may be desirable to ask additional questions thereby getting an over-determined system of equations*". The same was argued by Eisenführ et al. (2010), who wrote that "*it is sensible not to limit ourselves to the determination of* $(n-1)$ *tradeoffs*". At this point, we face significant questions like: "how many additional tradeoffs should we assess?", and "which pairs should we choose to compare?" As such, it makes sense to propose a more structured and justified methods to deal with redundancy of preferences and their inconsistencies, which is the aim of this study.

5.2.3 The classical Best-Worst Method

The original BWM uses ratios of the relative importance of attributes in pairs estimated by a DM, from the two opposite anchored vectors, A^{BO} and A^{OW}. The basic steps of the original BWM can be summarized as follows (Rezaei, 2015; 2016):

Step 1. The set of attributes $\{C_1, C_2, \cdots, C_n\}$ is determined by the DM.

Step 2. The best (e.g. the most influential or important) and worst (e.g. the least influential or important) attributes are determined by the DM. The two attributes are shown by C_B and C_W, respectively.

Step 3. The preference of the best over all the other attributes is determined by the DM using a number from $\{1, 2, \dots, 9\}$. The obtained Best-to-Others vector is: $A^{BO} = (a_{B1}, a_{B2}, \cdots, a_{Bn})$, where a_{Bj} represents the preference of the best attribute C_B over attribute C_j, $j = 1,2, \cdots, n$.

Step 4. The preferences of all the attributes over the worst attribute are determined by the DM using a number from $\{1, 2, \dots, 9\}$. The obtained Others-to-Worst vector is: $A^{OW} = (a_{1W}, a_{2W}, \cdots, a_{nW})$, where a_{jW} represents the preference of attribute C_j over the worst attribute C_W, $j = 1, 2, \cdots, n$.

Step 5. The weights $(w_1^*, w_2^*, \cdots, w_n^*)$ are found by solving the following model:

$$
\begin{aligned}
\text{minimize} \quad & \max_j \left\{ \left| \frac{w_B}{w_j} - a_{Bj} \right|, \left| \frac{w_j}{w_W} - a_{jW} \right| \right\}, \\
\text{subject to} \quad & w_1 + w_2 + \cdots + w_n = 1, \\
& w_j > 0, \; j = 1,2, \dots, n.
\end{aligned}
\tag{5.2}
$$

5.3 Best-Worst Tradeoff method (BWT)

Based on the concepts of the traditional Tradeoff procedure and of the Best-Worst Method, a structured method, called Best-Worst Tradeoff method (BWT), is proposed in this section to obtain the weights of attributes and help check the consistency via a structured framework.

Step 1. Determine alternatives and attributes

We assume that m alternatives $A = \{A_1, A_2, \dots, A_m\}$, and n attributes $C = \{C_1, C_2, \dots, C_n\}$ have been determined by DM. Moreover, a different consequence vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is assigned to each alternative.

Step 2. Determine value function for each attribute

There are many alternative methods to elicit the attribute value functions v_j (Fishburn, 1967; Keeney and Raiffa, 1976; Dyer and Sarin, 1979; Greco et al., 2016). One of the most widely used methods is the mid-value splitting technique proposed by Keeney and Raiffa (1976).

Step 3. Identify Best and Worst attributes

Firstly, the DM needs to identify the "best" attribute C_B and the "worst" C_W. Hereafter, we will use the terms "best" and "worst", borrowed from the classical BWM, to indicate the attributes with the greatest and smallest weights, respectively. These two attributes will serve as the two yardsticks against which the other attributes will be compared, with the ultimate goal of avoiding the anchoring bias. According to the first part of the original Tradeoff method in Section 5.2.2, we need to create n hypothetical consequences $\mathbf{x}^j$ $(j = 1,2, \dots, n)$ to represent the best performance of the j^{th} attribute and the simultaneous worst performance of the other attributes (see Figure 5-1). The DM is asked to compare and rank the hypothetical consequences $\mathbf{x}^1, \dots, \mathbf{x}^j, \dots, \mathbf{x}^n$, so that C_B and C_W can be identified.

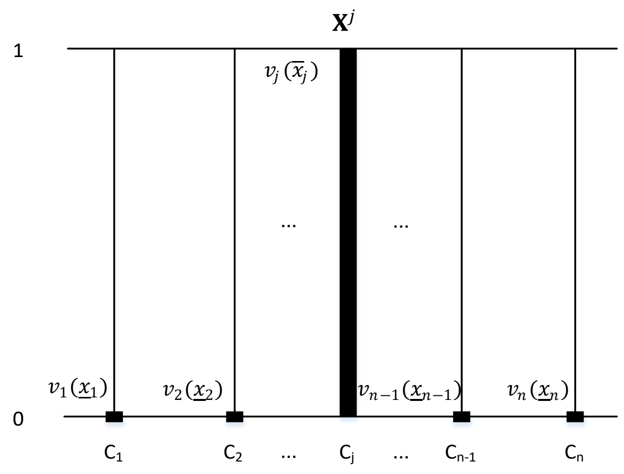


Figure 5-1. Hypothetical profile $\boldsymbol{x}^j$.

Step 4. Compare Best to others tradeoff

Assume that we are interested in comparing the best attribute to all the other attributes. Let us consider the instance of the comparison of the best attribute C_B and the k^{th} one C_k. For this scope we will need two auxiliary consequences: $\mathrm{x}^{B,k}$ and x^k. The first consequence is defined such that all attributes, except attribute C_B, achieve the lowest levels ($v_j(\underline{x}_j) = 0$), and the level of the best attribute is left to be determined. Assuming that C_k is the "other" attribute to be compared to the best, then the second consequence x^k has all the components at the lowest level of satisfaction ($v_j(\underline{x}_j) = 0$), except for the k^{th}, which, instead, has the highest ($v_j(\overline{x}_k) = 1$). Now, we wonder what degree of satisfaction of the best attribute in $\mathrm{x}^{B,k}$ would make the two consequences equally desirable, i.e. the value of $x_B^{B,k}$ for which,

$$\left(\underline{x}_1, \dots, x_B^{B,k}, \dots, \underline{x}_n\right) \sim \left(\underline{x}_1, \dots, \overline{x}_k, \dots, \underline{x}_n\right) \Leftrightarrow \mathbf{x}^{B,k} \sim \mathbf{x}^k \Leftrightarrow u(v(\mathbf{x}^{B,k})) = u(v(\mathbf{x}^k)), \tag{5.3}$$

where '$\sim$' indicates a relation of indifference between the two consequences.
Formally, assuming that we can use the additive representation, this corresponds to

$$u(v(\mathbf{x}^{B,k})) = w_B v_B\left(x_B^{B,k}\right) + \sum_{\substack{j=1\\ j\neq B}}^{n} w_j \underbrace{v_j(\underline{x}_j)}_{=0} = w_k \underbrace{v_k(\overline{x}_k)}_{=1} + \sum_{\substack{j=1\\ j\neq k}}^{n} w_j \underbrace{v_j(\underline{x}_j)}_{=0} = u(v(\mathbf{x}^k)).$$

which collapses into:

$$w_B v_B(x_B^{B,k}) = w_k. \tag{5.4}$$

At this point, asking for which value of $x_B^{B,k} \in X_B$, we obtain $\mathbf{x}^{B,k} \sim \mathbf{x}^k$ is equivalent to asking which $x_B^{B,k} \in X_B$ satisfies equation (5.4). The value $v_B(x_B^{B,k})$, which we denote as a_{kB}, is thus the DM's estimation of the ratio w_k/w_B. Now, we use the vector $(a_{1B}, a_{2B}, \dots, a_{nB})$ to collect the pairwise comparison values of $\mathbf{x}^{B,k}$ to $\mathbf{x}^k$ for all the k. Of course, its reciprocal $a_{Bk} = 1/a_{kB}$ corresponds to the DM's estimate of the ratio w_B/w_k. We use $A^{BO} = (a_{B1}, a_{B2}, \dots, a_{Bn})$ to indicate the Best-to-Others vector. The value a_{Bk} has a double interpretation as both the following conditions are equivalent and should, in theory, hold:

$$w_k a_{Bk} = w_B, \tag{5.5}$$

$$a_{Bk} = w_B/w_k. \tag{5.6}$$

According to Equation (5.5), a_{Bk} is the scalar to which one needs to multiply w_k to get w_B, while the latter (Equation (5.6)) stipulates that a_{Bk} is the ratio between the two weights w_B and w_k.

By considering the interpretation (5.5), if we consider all the $k \neq B$, we obtain the system of $(n-1)$ linear equations $w_B = a_{Bk} w_k,\ \forall k \neq B$. Similarly, we can obtain an equivalent system of equations (this time non-linear) if we consider the interpretation suggested by (5.6).

Example 1. Suppose a DM, who is going to buy a car, considers four attributes (C_1: $price$, C_2: $quality$, C_3: $safety$, C_4: $style$) to evaluate the alternatives. We further assume that C_1 is identified by the DM as the best attribute. In order to make the tradeoffs for the attributes (C_1 to C_2; C_1 to C_3; C_1 to C_4), we need six hypothetical consequences: $\mathbf{x}^{1,2} = \left(x_1^{1,2}, \underline{x}_2, \underline{x}_3, \underline{x}_4\right)$, $\mathbf{x}^{1,3} = \left(x_1^{1,3}, \underline{x}_2, \underline{x}_3, \underline{x}_4\right)$, $\mathbf{x}^{1,4} = \left(x_1^{1,4}, \underline{x}_2, \underline{x}_3, \underline{x}_4\right)$, $\mathbf{x}^2 = \left(\underline{x}_1, \overline{x}_2, \underline{x}_3, \underline{x}_4\right)$, $\mathbf{x}^3 = \left(\underline{x}_1, \underline{x}_2, \overline{x}_3, \underline{x}_4\right)$, $\mathbf{x}^4 = \left(\underline{x}_1, \underline{x}_2, \underline{x}_3, \overline{x}_4\right)$, where $\underline{x}_j$ and $\overline{x}_j$ represent the least and most desirable attribute levels, respectively. Then, the decision maker is asked to find the values of $x_1^{1,2}$, $x_1^{1,3}$, and $x_1^{1,4}$ such that the following three indifference relations hold:

$$\mathbf{x}^{1,2} \sim \mathbf{x}^2 \iff \left(x_1^{1,2}, \underline{x}_2, \underline{x}_3, \underline{x}_4\right) \sim \left(\underline{x}_1, \overline{x}_2, \underline{x}_3, \underline{x}_4\right),$$

$$\mathbf{x}^{1,3} \sim \mathbf{x}^3 \iff \left(x_1^{1,3}, \underline{x}_2, \underline{x}_3, \underline{x}_4\right) \sim \left(\underline{x}_1, \underline{x}_2, \overline{x}_3, \underline{x}_4\right),$$

$$\mathbf{x}^{1,4} \sim \mathbf{x}^4 \iff \left(x_1^{1,4}, \underline{x}_2, \underline{x}_3, \underline{x}_4\right) \sim \left(\underline{x}_1, \underline{x}_2, \underline{x}_3, \overline{x}_4\right).$$

These relations can be illustrated in Figure 5-2 ($\mathbf{x}^{1,2} \sim \mathbf{x}^2$), Figure 5-3 ($\mathbf{x}^{1,3} \sim \mathbf{x}^3$) and Figure 5-4 ($\mathbf{x}^{1,4} \sim \mathbf{x}^4$).

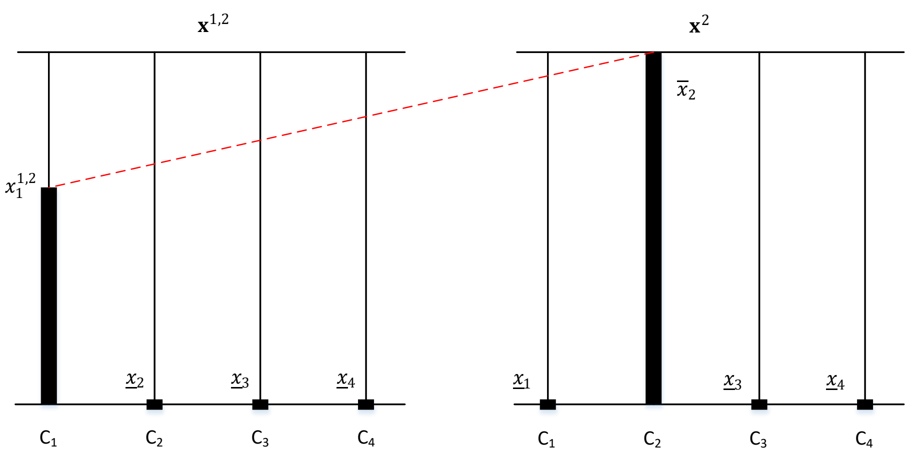


Figure 5-2. The DM states the level of the attribute C_1 which makes the consequences $\mathrm{x}^{1,2}$ and x^2 indifferent one to another.

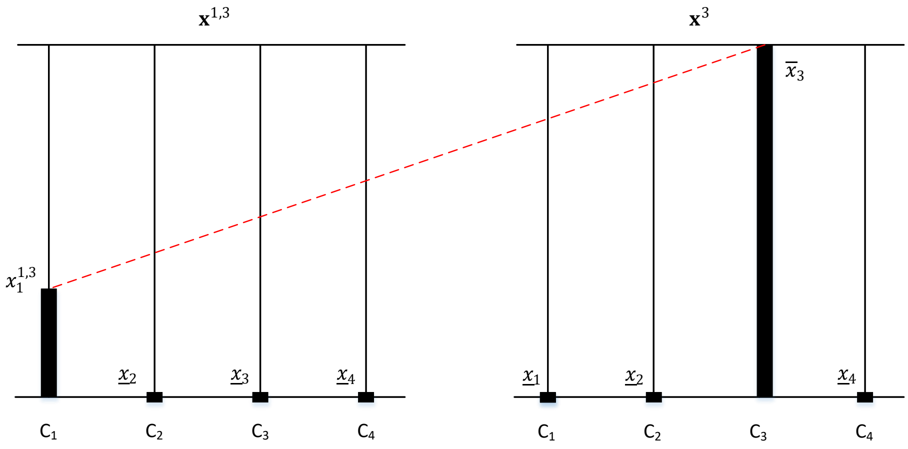


Figure 5-3. The DM states the level of the attribute C_1 which makes the consequences $\mathbf{x}^{1,3}$ and $\mathbf{x}^3$ indifferent one to another.

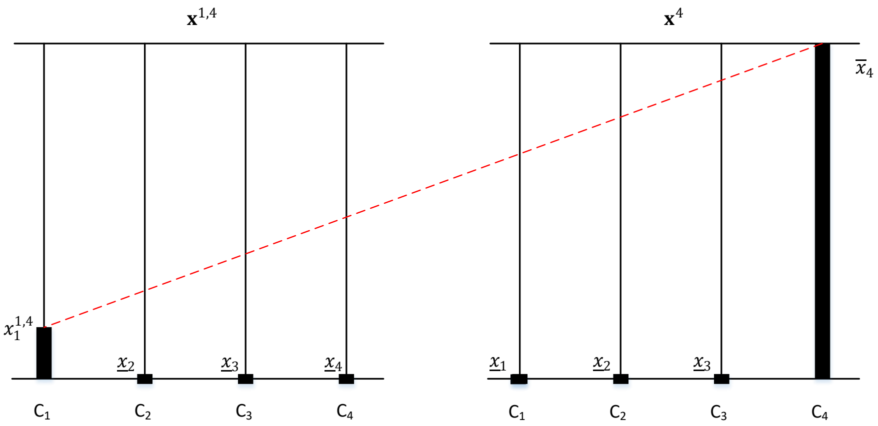


Figure 5-4. The DM states the level of the attribute C_1 which makes the consequences $\mathbf{x}^{1,4}$ and $\mathbf{x}^4$ indifferent one to another.

After having determined the value functions and $x_1^{1,2}$, $x_1^{1,3}$, $x_1^{1,4}$, we can obtain the relations:

$w_1 v_1(x_1^{1,2}) = w_2,\ w_1 v_1(x_1^{1,3}) = w_3,\ w_1 v_1(x_1^{1,4}) = w_4.$

As we know that the first criterion is the best, with the convention, $a_{B2} = 1/v_1(x_1^{1,2})$, $a_{B3} = 1/v_1(x_1^{1,3})$, $a_{B4} = 1/v_1(x_1^{1,4})$, one obtains:

$w_1 = w_2 a_{B2} = w_3 a_{B3} = w_4 a_{B4}.$

Step 5. Compare Others to Worst tradeoff

Similarly, the entire procedure can be repeated the other way round, this time to compare each attribute C_k $(C_k \neq C_W)$ to C_W, the worst attribute. In this case, we continue to use two auxiliary consequences, $\mathbf{x}^W$, and $\mathbf{x}^{k,W}$, but with different components. Now, $\mathbf{x}^W$ has all components at the lowest level, i.e. $v_j(\underline{x}_j) = 0$, except for the worst attribute, which reaches the highest level $(v_W(\overline{x}_W) = 1)$. Consequence $\mathbf{x}^{k,W}$, instead, has all the components at the lowest level, i.e. $v_j(\underline{x}_j) = 0$, except for the k^{th} attribute, which is left undetermined. When we assume $\mathbf{x}^W \sim \mathbf{x}^{k,W}$, i.e.:

$$u\big(v_1(\underline{x}_1^W), \dots, v_W(\overline{x}_W^{\mathrm{W}}), \dots, v_n(\underline{x}_n^W)\big) = u\big(v_1(\underline{x}_1^O), \dots, v_k(x_k^{\mathrm{O}}), \dots, v_n(\underline{x}_n^O)\big), \tag{5.7}$$

and use the additive value function, we obtain:

$$w_W \underbrace{v_W(\overline{x}_W)}_{=1} + \sum_{\substack{j=1 \\ j\neq W}}^{n} w_j \underbrace{v_j(\underline{x}_j)}_{=0} = w_k v_k(x_k) + \sum_{\substack{j=1 \\ j\neq k}}^{n} w_j \underbrace{v_j(\underline{x}_j)}_{=0},$$

which can be simplified into:

$$w_W = w_k v_k(x_k^{\mathrm{O}}). \tag{5.8}$$

By asking the DM what value of x_k makes the two consequences indifferent, we obtain $v_k(x_k)$, which we will call a_{Wk}. From this, thanks to reciprocity, i.e., $a_{kW} = 1/a_{Wk}$, all the values of the comparisons form the Others-to-Worst vector $A^{OW} = (a_{kW})$, where $k = 1,2,\dots,n$. We recover the following two interpretations for the value a_{kW},

$$w_W a_{kW} = w_k, \tag{5.9}$$

$$a_{kW} = w_k / w_W. \tag{5.10}$$

Implementing this procedure for all the paired comparison values a_{kW} of $\mathbf{x}^W$ to $\mathbf{x}^k$, we obtain the system of $(n-1)$ linear equations $w_k = a_{kW} w_W$, $\forall k \neq W$. Similarly, we can obtain an equivalent system of equations (this time non-linear) if we consider the interpretation suggested by (5.10).

Example 2. We continue with Example 1, and suppose C_4 is identified by the DM as the worst attribute. To make the tradeoffs between the attributes (C_1 to C_4; C_2 to C_{W4}; C_3 to C_4), we need to have 4 hypothetical consequences: $\mathbf{x}^{1,4} = (x_1, \underline{x}_2, \underline{x}_3, \underline{x}_4)$, $\mathbf{x}^{2,4} = (\underline{x}_1, x_2, \underline{x}_3, \underline{x}_4)$, $\mathbf{x}^{3,4} = (\underline{x}_1, \underline{x}_2, x_3, \underline{x}_4)$, $\mathbf{x}^4 = (\underline{x}_1, \underline{x}_2, \underline{x}_3, \overline{x}_4)$, where $\underline{x}_j$ and $\overline{x}_j$ represent the least and most desirable attribute levels, respectively. At this point the decision maker is asked to find the values of x_1, x_2, and x_3 such that the following indifference relations hold:

$$\mathbf{x}^{1,4} \sim \mathbf{x}^4 \iff (x_1, \underline{x}_2, \underline{x}_3, \underline{x}_4) \sim (\underline{x}_1, \underline{x}_2, \underline{x}_3, \overline{x}_4),$$

$$\mathbf{x}^{2,4} \sim \mathbf{x}^4 \iff (\underline{x}_1, x_2, \underline{x}_3, \underline{x}_4) \sim (\underline{x}_1, \underline{x}_2, \underline{x}_3, \overline{x}_4),$$

$$\mathbf{x}^{3,4} \sim \mathbf{x}^4 \iff (\underline{x}_1, \underline{x}_2, x_3, \underline{x}_4) \sim (\underline{x}_1, \underline{x}_2, \underline{x}_3, \overline{x}_4).$$

These relations can be simplified as $\mathbf{x}^{1,4} \sim \mathbf{x}^{2,4} \sim \mathbf{x}^{3,4} \sim \mathbf{x}^4$, and represented in Figure 5-5.

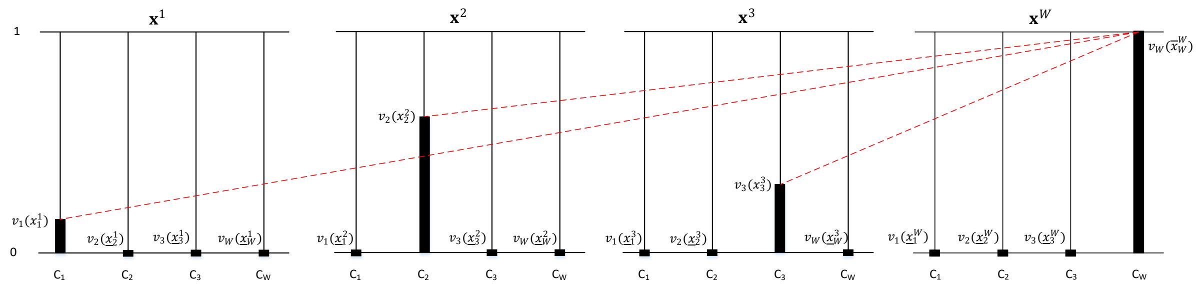


Figure 5-5. The DM states the levels of the attributes C_1, C_2, C_3, such that $\mathbf{x}^{1,4} \sim \mathbf{x}^{2,4} \sim \mathbf{x}^{3,4} \sim \mathbf{x}^4$.

After having determined the values of x_1, x_2, x_3, we can obtain the relations:

$$w_1 v_1(x_1) = w_2 v_2(x_2) = w_3 v_3(x_3) = w_4 \underbrace{v_4(\overline{x}_4)}_{=1}.$$

Remark. In our examples, the tradeoffs were made between fictitious alternatives where some attributes were set to their minimum levels. See, for instance, Figure 5-5. This suits the results obtained by Vetschera et al. (Vetschera et al., 2014) according to which alternatives with extreme values may improve the consistency of results. However, this should not be interpreted as binding. Values of irrelevant attributes in tradeoffs can be set to values other than their minimums, as long as the same values are present in both consequences and can therefore be canceled out.

Step 6. Find the optimal weights

After having obtained the pairwise comparison system, which we refer to as the set of judgements contained in vectors A^{BO} and A^{OW}, we can estimate the weights of the attributes. If we choose interpretations (5.5) and (5.9), we end up with the following system of linear equations in n variables:

$$\begin{cases} w_k a_{Bk} = w_B, \ \forall k \neq B \\ w_k = a_{kW} w_W, \ \forall k \neq W. \\ w_1 + w_2 + \cdots + w_n = 1 \end{cases} \tag{5.11}$$

Unless the judgements elicited from the DM are fully rational, such an equation system does not have a solution. Since subjective judgements are seldom rational, it is necessary to use some methods to determine good estimates for the weights.

Following the formulation of the BWM (Rezaei, 2015; 2016) we want to identify the weight vector that minimizes the greatest absolute violation of the equations. Considering the non-negativity and normality conditions for the weights, this corresponds to solving the following optimization problem:

$$\begin{array}{ll} \text{minimize} & \max_j \{ |w_j a_{Bj} - w_B|, |w_W a_{jW} - w_j| \} \\ \text{subject to} & w_1 + w_2 + \cdots + w_n = 1 \\ & w_j \geq 0, j = 1,2,\ldots,n \end{array} . \tag{5.12}$$

The optimization problem (5.12) can be equivalently rewritten as

$$\begin{array}{ll} \text{minimize} & \xi \\ \text{subject to} & |w_j a_{Bj} - w_B| \leq \xi, \quad \forall j \neq B \\ & |w_W a_{jW} - w_j| \leq \xi, \quad \forall j \neq W. \\ & w_1 + w_2 + \cdots + w_n = 1 \\ & w_j \geq 0, \ j = 1,2,\ldots,n \end{array} \tag{5.13}$$

This last formulation can be easily linearized and its solution yields the optimal weights.

If, conversely, one wants to privilege the interpretations (5.6) and (5.10) of the judgements a_{Bk} and a_{kW}, the system of (nonlinear) equations becomes

$$\begin{cases} a_{Bk} = w_B / w_k, & \forall k \neq B \\ a_{kW} = w_k / w_W, & \forall k \neq W \ , \\ w_1 + w_2 + \cdots + w_n = 1 \end{cases} \tag{5.14}$$

the optimization problem to minimize the maximum absolute discrepancy becomes,

$$\begin{array}{ll} \text{minimize} & \max\{|a_{Bj} - w_B/w_j|, |a_{jW} - w_j/w_W|\} \\ \text{subject to} & w_1 + w_2 + \cdots + w_n = 1 \\ & w_j > 0,\ j = 1,2,\ldots,n \end{array}, \tag{5.15}$$

which is equivalent to

$$\begin{array}{ll} \text{minimize} & \xi \\ \text{subject to} & |a_{Bj} - w_B/w_j| \le \xi \quad j \ne B \\ & |a_{jW} - w_j/w_W| \le \xi \quad j \ne W. \\ & w_1 + w_2 + \cdots + w_n = 1 \\ & w_j > 0,\ j = 1,2,\ldots,n \end{array} \tag{5.16}$$

5.4 Consistency measurement

As also indicated by Keeney and Raiffa (1976), it is important to keep inconsistencies at a "nominal" level. Keeney (2002) mentioned inconsistency as one of the causes that can lead to errors in measuring the scaling constants, so it is important to quantify and localize the degree of inconsistency of sets of preferences.

There are two kinds of consistency, ordinal consistency and cardinal consistency, and one common desideratum in decision-making analysis is that the judgements of the DM be as ordinally consistent and cardinally consistent as possible. To measure how DMs deviate from these consistency conditions, we propose two indices: the ordinal consistency ratio and cardinal consistency ratio for the BWT, inspired by Liang et al. (2020), Escobar et al. (2015) and Cavallo et al. (2016).

5.4.1 Ordinal consistency ratio

Definition 1 (Ordinal consistency): In the BWT, 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. In formal terms (Kendall, 1938):

$$(a_{Bk} - a_{Bj}) \times (a_{jW} - a_{kW}) > 0 \text{ or } (a_{Bk} = a_{Bj} \text{ and } a_{jW} = a_{kW}), \qquad \forall\, k \text{ and } j \tag{5.17}$$

Checking the violation of ordinal consistency is very important because it has a vital impact on the ranking of the attributes. In order to measure to what extent DMs violate the ordinal consistency, we need to define an Ordinal Consistency Ratio.

Definition 2 (Ordinal Consistency Ratio): The *Ordinal Consistency Ratio* OR of a pairwise comparison system is defined as:

$$OR = \max_j OR_j \tag{5.18}$$

where

$$OR_j = \frac{1}{n-1}\sum_{k=1}^{n} F\left(\left(a_{Bk} - a_{Bj}\right) \times \left(a_{jW} - a_{kW}\right)\right) \tag{5.19}$$

where $F(\gamma, \delta)$ is a step function, where $\gamma = a_{Bk} - a_{Bj}$, $\delta = a_{jW} - a_{kW}$, it is defined as:

$$F(\gamma, \delta) = \begin{cases} 1, & \text{if } \gamma \times \delta < 0, \\ 0.5, & \text{if } \gamma \times \delta = 0 \text{ and } (\gamma \neq 0 \text{ or } \delta \neq 0), \\ 0, & \text{otherwise.} \end{cases} \tag{5.20}$$

OR_j is called *local* ordinal consistency ratio[29], indicating the degree of consistency with respect to the j^{th} attribute. With this ordinal consistency ratio ($OR_j \in [0,1]$), we can determine if the j^{th} attribute has different rank (and to what extent) in the two vectors.

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

The rationale of OR_j formulation is that if attribute C_j overweighs attribute C_k, then the ordinal consistency should satisfy $a_{Bk} > a_{Bj}$ and $a_{jW} > a_{kW}$, i.e. $\left(a_{Bk} - a_{Bj}\right) \times \left(a_{jW} - a_{kW}\right) > 0$. If only one of $\left(a_{Bk} - a_{Bj}\right)$ and $\left(a_{jW} - a_{kW}\right)$ is equal to 0, we say that, in this situation, it violates weak ordinal relation, but if both are equal to 0, it is ordinal-consistent (Escobar et al., 2015; Cavallo et al., 2016). Overall, this approach is similar, but not identical, to Kendall's tau (Kendall, 1938).

5.4.2 Cardinal Consistency ratio

Definition 3 (Cardinal consistency): The preferences a_{Bj}, a_{jW} and a_{BW} elicited as in Section 5.3 are consistent if and only if:

$$a_{Bj}a_{jW} = a_{BW}, \ \forall j. \tag{5.21}$$

To measure the deviation from the perfect cardinal consistency, Liang et al. (2020) defined the following index:

Definition 4 (Cardinal Consistency Ratio): The Cardinal Consistency Ratio is formulated as follows:

$$CR = \max_j CR_j, \tag{5.22}$$

where,

$$CR_j = \begin{cases} \dfrac{\left|a_{Bj}a_{jW} - a_{BW}\right|}{a_{BW}a_{BW} - a_{BW}}, & a_{BW} > 1, \\ 0, & a_{BW} = 1. \end{cases} \tag{5.23}$$

---

[29] We opt for max instead of sum, because max could label a pairwise comparison system *in*consistent when there is at least one element that is not sufficiently consistent (for sufficiency, we usually use thresholds), which helps the analyst and the decision-maker locate the source of inconsistency for possible revision. The sum, on the other hand, looks at the whole system and not the individual pairwise comparisons. However, both max and sum could aggregate the local measures of inconsistency (Brunelli, 2016).

CR is the *global* consistency ratio for all attributes and represents the overall consistency of the preferences; CR_j represents the *local* consistency level associated with attribute C_j, with which we can locate the most inconsistent attribute.

5.4.3 Thresholds

Identifying inconsistent judgements and knowing how much they deviate from a fully consistent status is not enough. Instead, we need to know under what threshold the inconsistency is acceptable, which is why, drawing from the study by Liang et al. (2020), we propose a method to determine the thresholds of BWT based on the cardinal and ordinal consistency ratios.

If a DM is ordinally consistent, the rankings of the preferences obtained from the two judgement vectors (A^{BO} and A^{OW}) are the same, and only the intensities of preference may vary (Amenta et al., 2020). Therefore, in an ordinal sense, we can rely on the preferences provided by DM in this situation. Based on this idea, we can design a mechanism to find a suitable threshold.

Firstly, we use the Monte-Carlo method to estimate the probability distribution of CRs. We consider a number of attributes, n, ranging from 4 to 9 ($n = 2$ and $n = 3$ are excluded[30]). As $a_{BW} \in [1, \infty)$, which is a continuous set, in this study we consider only a subset of values, i.e. $a_{BW} \in \{2,3,4,5,6,7,8,9\}$. The choice of a discrete scale, with integers up to 9, is necessary to keep the Monte Carlo analysis computable, and is coherent with the standard approach used in the field of Design of Experiments (DoE) (Myers et al., 2016). We will then carry out a full factorial analysis on the $6 \times 8 = 48$ combinations (6 is the number of attributes and 8 the number of different values of a_{BW}). For each combination, 10,000 pairs of *ordinal-consistent* vectors (A^{BO} and A^{OW}) will be generated randomly and will be labeled as *acceptable* comparisons. Also, we generate 10,000 pairs of *ordinal-inconsistent* vectors, which will be categorized as the *unacceptable* group.

In light of the distributions of the CRs of the two groups, there are significant overlaps, which means that there is no value of CR that can split the sets of acceptable and unacceptable preferences. In probabilistic terms, we expect that there is a threshold below which the CRs can be part of the acceptable group as much as possible, and above which the CRs can be part of the unacceptable group as much as possible (for example, see Figure 5-6).

---

[30] Cases with $n = 2$ and $n = 3$ are excluded from the analysis, because in such cases, the ordinal-inconsistent situation does not appear.

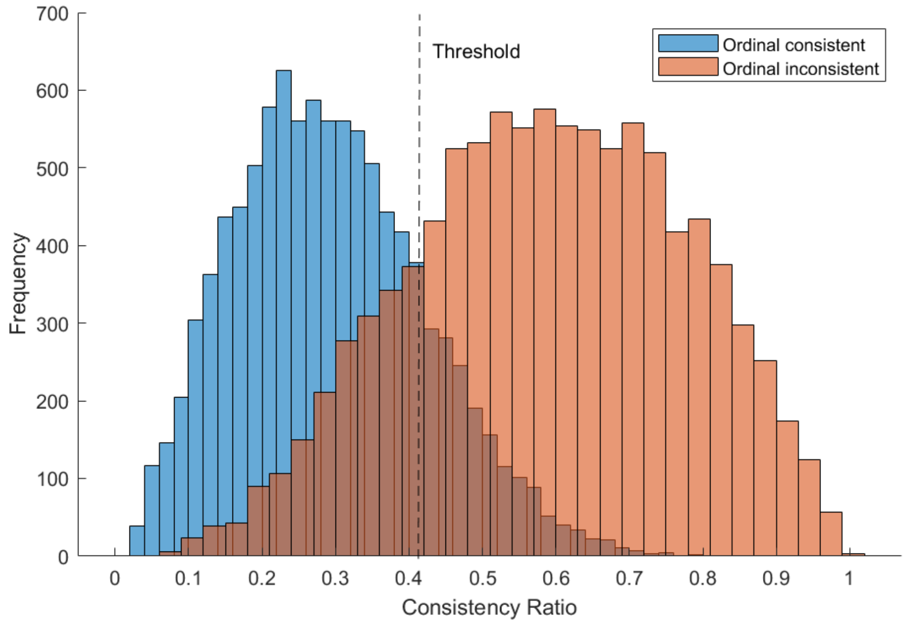


Figure 5-6. The distribution of CRs in the two groups (in case $n = 9$, $a_{BW} = 9$).

To achieve this, we can use the empirical cumulative distribution function, which can be defined as:

$$\hat{F}_N(CR_i) = \frac{1}{N}\sum_{i=1}^{N} I\{CR_i \leq CR_T\}, \tag{5.24}$$

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

$$I\{CR_i \leq CR_T\} = \begin{cases} 1, & if \quad CR_i \leq CR_T \\ 0, & \text{otherwise} \end{cases}, \tag{5.25}$$

where N is the number of sampled vectors A^{BO} and A^{OW}, CR_i is the consistency ratio of the i^{th} ($i \in \{1,2,\dots,N\}$) pair of vectors, $CR_T \in [0,1]$ is the possible threshold.

We denote the cumulative distribution of CRs in ordinal-consistent vectors (the Acceptable group) as $\hat{F}^A(CR_T)$, and the cumulative distribution of CRs from the Unacceptable group is denoted as $\hat{F}^U(CR_T)$. We accept the CR's within the threshold (which is $\hat{F}^U(CR_T)$), and reject the CR's beyond the threshold (which is $1 - \hat{F}^A(CR_T)$). In order to make the proportion of ordinal-inconsistent CR's that we accept as small as possible, and also make the proportion of ordinal-consistent CRs that we reject as small as possible, we need to calculate the threshold, CR_T, so that $\hat{F}^U(CR_T) = 1 - \hat{F}^A(CR_T)$. Figure 5-7 sketches the simulation algorithm used to calculate the thresholds.

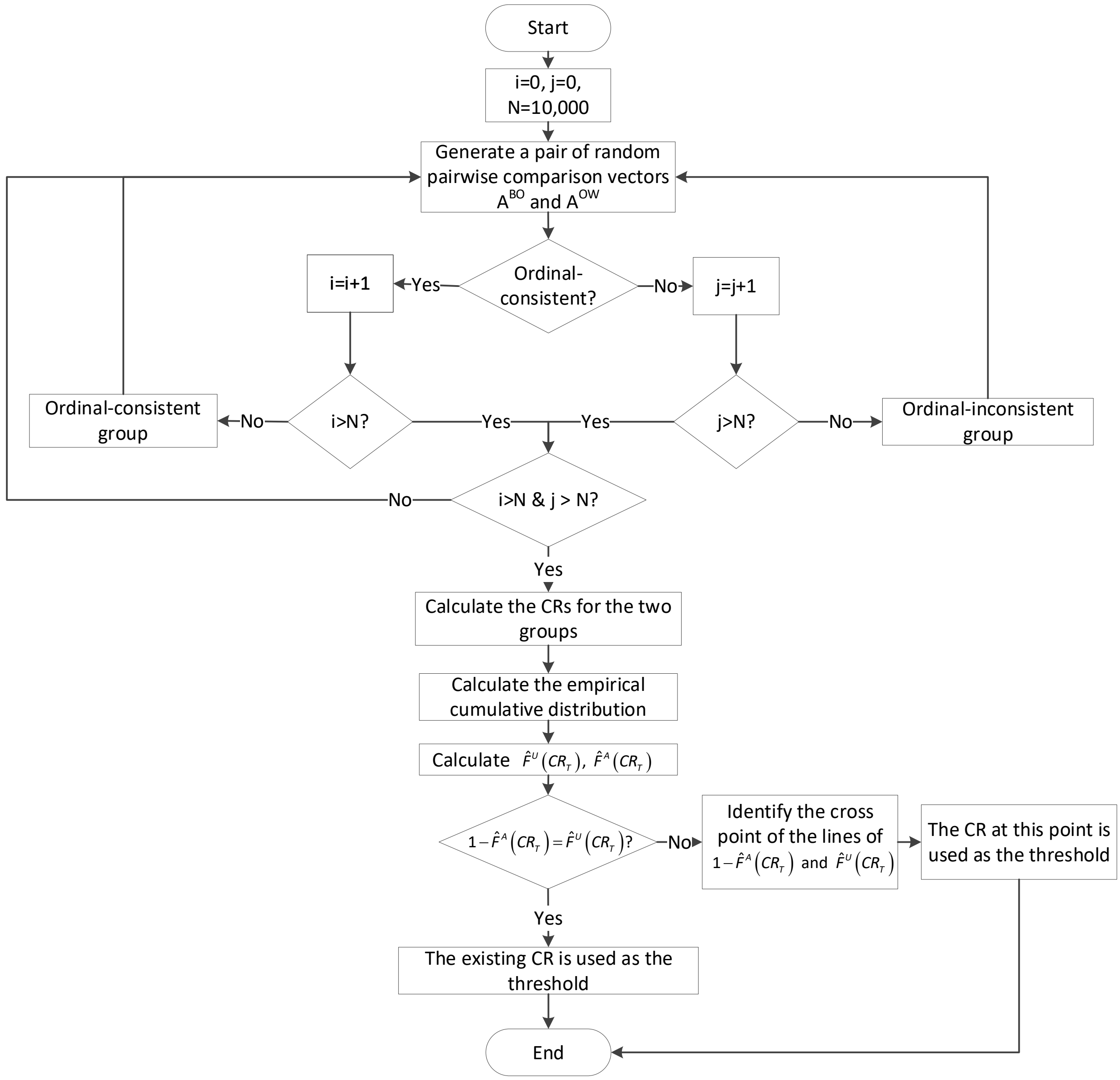


Figure 5-7. The simulation algorithm31 used to calculate the thresholds (adapted from Liang et al. (2020)).

Table 5-1 shows the thresholds for combinations of attributes range $n = \{4, \ldots, 9\}$ and integer values of a_{BW} from 2 to 9.

[31] The code for the simulation algorithm can be seen in: https://github.com/fuqi15/BWT

Table 5-1. The thresholds of the consistency ratios in different combinations.

a_{BW}	n					
	4	5	4	7	4	9
2	0.34	0.35	0.36	0.38	0.39	0.9
3	0.31	0.33	0.35	0.37	0.38	0.39
4	0.32	0.34	0.36	0.37	0.39	0.4
5	0.31	0.35	0.37	0.38	0.4	0.4
6	0.32	0.35	0.37	0.38	0.4	0.41
7	0.33	0.35	0.37	0.39	0.4	0.41
8	0.32	0.35	0.37	0.39	0.4	0.41
9	0.33	0.36	0.37	0.39	0.4	0.41

It is worth to mention that, unlike the original BWM, where the scale of values of a_{BW} is a discrete set $\{1,2,\ldots,9\}$, the BWT relies on a continuous scale, so that, in principle, a_{BW} can take any value greater than 1. For this reason, the proposal by Liang et al. (2020) must be readapted to this framework, but with two differences:

Firstly, in the calculation of the thresholds, all the real numbers in the interval $[1, a_{BW}]$ can be sampled, and not only the integers. The results of sampling from a continuous set substantially differs from the study by Liang et al. (2020), where entries were sampled from the discrete set $\{1,2,\ldots,a_{BW}\}$.

Secondly, to acknowledge the continuous nature of a_{BW} we seek for a continuous function (a *response surface* in the DoE terminology) to approximate the values in Table 5-1, and help find the thresholds also for non-integer values of a_{BW}. Formally, we consider the variables x and y to represent n and a_{BW}, respectively, and use the following quadratic fit,

$$Q(x,y) = \begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix}^T \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}^T \begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} + c, \tag{5.26}$$

where x_0, y_0, a_{11}, a_{12}, a_{22}, b_1, b_2, c are the parameters. By using a least squares minimization approach, the optimal parameters[32] yield a good fit. See Figure 5-8 for a graphical representation of the results[33].

[32] By minimizing the Euclidean norm, the optimal parameters are: $x_0 = 9.8584$, $y_0 = 7.8288$, $a_{11} = -0.0017$, $a_{12} = -0.0003$, $a_{22} = -0.0003$, $b_1 = 0.0058$, $b_2 = 0.0055$, $c = -0.4187$.

[33] It is worth mentioning that our decision to use a single quadratic function to identify thresholds for $n \in \{4,\ldots,9\}$ and $a_{BW} \in [2,9]$ was made for reasons of simplicity. In theory, we could have proposed different univariate quadratic fits for each single value of n.

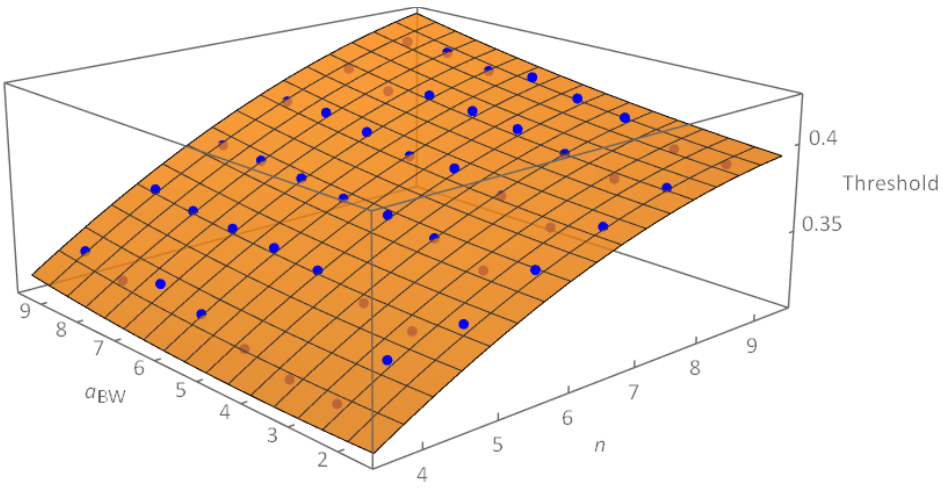


Figure 5-8. The graphical representation of the thresholds.

The adapted algorithm in this study considers a_{BW} as continuous space, which is closer to the reality. With these thresholds, we can now determine whether the consistency level (consistency ratio) of a DM is acceptable or not. Since $\mathrm{a_{BW}}$ could be a number with decimal digits, we suggest using the threshold of the approximate integer.

5.4.4 Improving consistency

When DMs provide their preferences, we need to firstly check their consistency. If their cardinal consistency level is not acceptable, or if they violate the ordinal consistency, we suggest the DMs to revise their judgment.

Usually, the consistency improving process is guided by a moderator, who helps the DMs to revise their preferences. The following steps describe the procedure of the consistency improving process, which is also sketched in Figure 5-9.

Step 1. Let $A^{BO(t)} = \left(a_{Bj}^{(t)}\right)$ and $A^{OW(t)} = \left(a_{jW}^{(t)}\right)$, $j = 1,2,\cdots,n$ be the original preferences provided by the DM. Let t indicate the iteration number.

Step 2. Compute the ordinal consistency ratio $OR^{(t)}$ and the cardinal consistency ratio $CR^{(t)}$ by using equations (5.17)-(5.23). Based on a_{BW} (if a_{BW} is not an integer, round it up to the nearest whole number) and n, check the corresponding consistency threshold CR_T in Table 5-1.

Step 3. Check the cardinal consistency. If $CR^{(t)} > CR_T$, the consistency level is not acceptable, go to Step 4 to revise the preferences. If $CR^{(t)} \leq CR_T$, go to Step 5 to check the ordinal consistency.

Step 4. Revise preferences and improve the consistency. The inconsistent preferences should be adjusted to cardinal-consistent, which should satisfy:

$$\frac{\left|a_{Bj}^{(t)}a_{jW}^{(t)}-a_{BW}\right|}{a_{BW}a_{BW}-a_{BW}} \leq CR_T, \tag{5.27}$$

Therefore, the DM should revise $a_{Bj}^{(t)}$ or $a_{jW}^{(t)}$ or both in the admissible ranges:

$$(a_{BW} - CR_T(a_{BW}a_{BW} - a_{BW}))/a_{jW}^{(t)} \leq a_{Bj}^{(t)} \leq (a_{BW} + CR_T(a_{BW}a_{BW} - a_{BW}))/a_{jW}^{(t)}, \quad (5.28)$$

$$(a_{BW} - CR_T(a_{BW}a_{BW} - a_{BW}))/a_{Bj}^{(t)} \leq a_{jW}^{(t)} \leq (a_{BW} + CR_T(a_{BW}a_{BW} - a_{BW}))/a_{Bj}^{(t)}, \quad (5.29)$$

and $a_{Bj}^{(t)}, a_{jW}^{(t)} \in [1, a_{BW}]$.

If the DM wants to revise his/her preferences in an acceptable range and also improve his/her consistency level[34], then the ranges of $a_{Bj}^{(t)}$ or $a_{jW}^{(t)}$ for adjustment are:

$$\min\left(a_{Bj}^{(t)}, a_{Bj}^{(t)} - \left(a_{Bj}^{(t)}a_{jW}^{(t)} - a_{BW}\right)/a_{jW}^{(t)}\right) \leq a_{Bj}^{(t)} \leq \max\left(a_{Bj}^{(t)}, a_{Bj}^{(t)} - \left(a_{Bj}^{(t)}a_{jW}^{(t)} - a_{BW}\right)/a_{jW}^{(t)}\right), \quad (5.30)$$

$$\min\left(a_{jW}^{(t)}, a_{jW}^{(t)} - \left(a_{Bj}^{(t)}a_{jW}^{(t)} - a_{BW}\right)/a_{Bj}^{(t)}\right) \leq a_{jW}^{(t)} \leq \max\left(a_{jW}^{(t)}, a_{jW}^{(t)} - \left(a_{Bj}^{(t)}a_{jW}^{(t)} - a_{BW}\right)/a_{Bj}^{(t)}\right). \quad (5.31)$$

Hint: If $a_{Bj}^{(t)} \times a_{jW}^{(t)} < a_{BW}$, then the DM needs to think about increasing $a_{Bj}^{(t)}$ or $a_{jW}^{(t)}$ or both, otherwise, decreasing the pairwise comparisons is recommended. In any case, the idea is to approach the condition $a_{Bj}^{(t)} \times a_{jW}^{(t)} \approx a_{BW}$.

After revision, set $t = t + 1$, and go back to Step 2.

Step 5. Check the ordinal consistency. If $OR^{(t)} = 0$, it means the given preferences are ordinal-consistent, go to next step. If $OR^{(t)} > 0$, it means the given preferences are ordinal-inconsistent, the DM should revise his/her judgments to reach full ordinal consistency, the local ordinal consistency ratio obtained by Equation (5.19) can help the DM to identify the most inconsistent attribute, then go back to Step 4.

Step 6. Output $A^{BO(t)}$, $A^{OW(t)}$, $OR^{(t)}$ and $CR^{(t)}$.

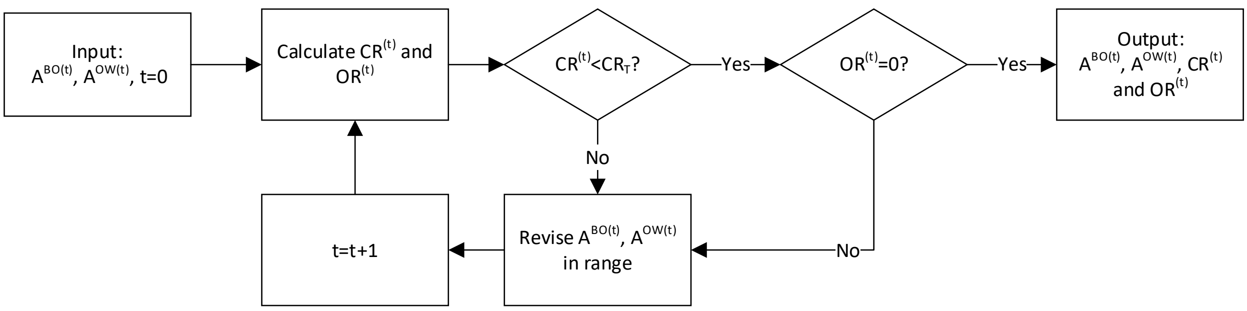


Figure 5-9. Consistency improving process.

[34] When the consistency level is not acceptable, the DM should revise his/her preferences. But if his/her preferences are ordinal-consistent and cardinal-consistent (acceptable consistency), then adjustments are not required, because the objective of elicitation is not achieving the optimal consistency level. However, if the DM tries to improve his/her consistency level, then the admissible ranges provide a way to reach that goal.

5.5 Case study

In this section, we illustrate the proposed BWT model by applying it to a port evaluation problem. The port performance measurement research helps ports anticipate and respond to possible future changes in port choice by shippers, freight forwarders and carriers. One of the studies examining this problem was conducted by Rezaei et al. (2019), from which we use the port leg-related services and facilities data (the attributes, alternatives and evaluation scores) recorded by that study to present the BWT procedure.

To evaluate this study, we contacted a competent expert: a program manager at Digital Port Solutions of Vopak (in the Netherlands). She is an expert in the port choice problem and is familiar with MADM methods. The interview was conducted online in three phases. In phase one, we introduced the problem to the expert; in phase two we identified the preferences of the expert with regard to the BWT procedure; in phase three we collected the modified values after checking the consistency. After the interview, we calculated the weights of attributes to produce the ranking of the alternatives. Below, we describe the nine steps of the BWT we used in this case study.

Step 1. Determine alternatives and attributes

Firstly, to evaluate the ports, the study by Rezaei et al. (2019) included six attributes: terminal handling charges, International Ship and Port Facility Security Code (ISPS), customs service (rated on a 1-to-7 scale), port reputation (rated on a 1-to-7 scale), satisfaction with terminal operations (rated on a 1-to-7 scale), and number of container terminals. Terminal handling charges and ISPS are *cost attributes* and all the others are *benefit attributes*. Seven ports were examined in the study. The scores of the alternative ports with respect to the attributes are presented in Table 5-2.

Table 5-2. The recorded scores for the seven ports (Rezaei et al., 2019).

Ports	Attributes					
	Terminal handling charges (€/TEU)	ISPS (€/unit)	Customs service	Port reputation	Satisfaction with terminal operations	Number of container terminals
Piraeus	106	11	4.2	3.8	3.4	3
Koper	145	11	5.12	5.24	5	1
Genoa	179	13	4.2	4.4	4.4	2
Antwerp	179	12	5.44	5	5.11	4
Rotterdam	202	13	5.5	5.93	5.29	6
Hamburg	223	16	5.56	6.06	5.41	4
Gdansk	103	14	4.6	5	4.4	2

Step 2. Determine value function for each attribute

Next, we need to determine the value functions of each attribute as stated in Section 5.3, and in this study, we adopt the mid-value splitting technique presented in the study of Keeney and Raiffa (1976), where the readers can find the detailed definitions. Here we simplify the assessment procedure as follows:

(1) For the jth attribute, we let $v(\underline{x}_j) = 0$ and $v(\overline{x}_j) = 1$.

(2) Determine the mid-value point (denoted as $x_{.5}$) of $[x_0, x_1]$, we let $v(x_{.5}) = 0.5$.

(3) Determine the mid-value point, $x_{.75}$, of $[x_{.5}, x_1]$, and make $v(x_{.75}) = 0.75$.

(4) Determine the mid-value point, $x_{.25}$, of $[x_0, x_{.5}]$, and let $v(x_{.25}) = 0.25$.

(5) To check the consistency of the preferences, the DM needs to be certain that $x_{.5}$ is the midvalue point of $[x_{.25}, x_{.75}]$; otherwise modification is necessary to guarantee the consistency.

(6) Use the points (x_k, k) for $k = 0, 0.25, 0.5, 0.75, 1$, to estimate the value function v_j.

Subsequently, the mid-value points for the attributes of this study are obtained and shown in Table 5-3. These points can be plotted in Figure 5-10 and interpolated by the v_1 - v_6 curves for C_1 to C_6, respectively.

Table 5-3. The mid-value points for each attribute.

	C_1	C_2	C_3	C_4	C_5	C_6
x	223	16	4.2	3.8	3.4	1
$x_{.25}$	194	14.8	4.6	4.5	4.2	2.25
$x_{.5}$	163	13.5	4.9	5	4.8	3.5
$x_{.75}$	127	12.25	5.3	5.7	5.2	4.75
$\overline{x}$	103	11	5.65	6.06	5.41	6

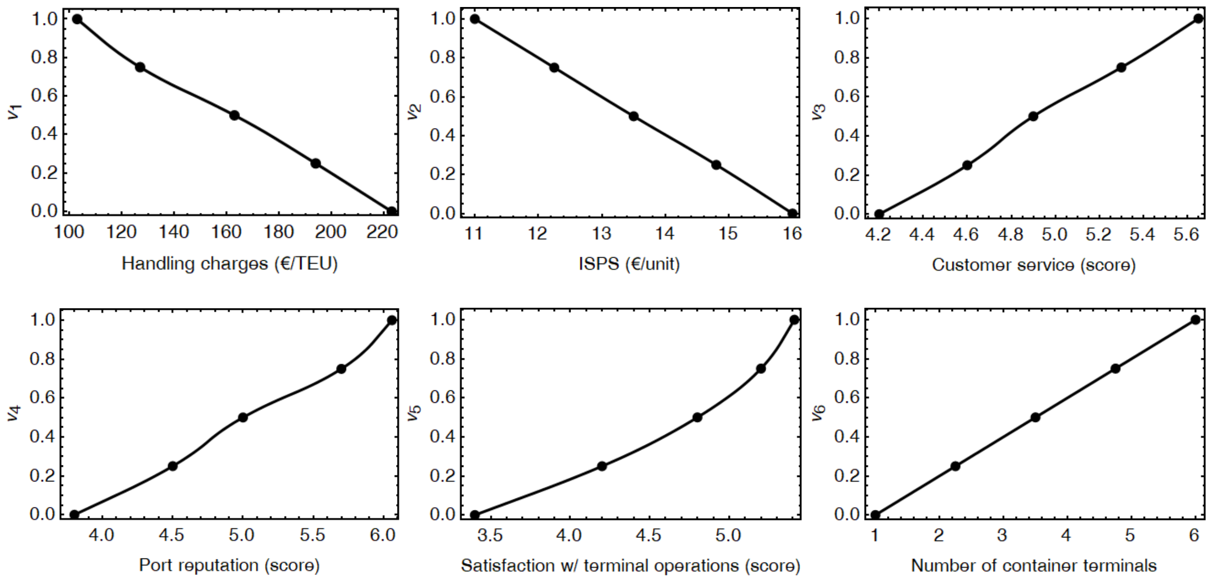


Figure 5-10. The assessed value functions for attributes C_1 to C_6 as second order interpolations of the values in Table 5-3.

Step 3. Identify the best and worst attributes

Following Section 5.3, the expert considered C_1 as the best attribute and C_3 as the worst.

Step 4. Determine the best to others tradeoffs

To tradeoff the best attribute C_1 to the other attributes, we generated the following hypothetical profiles based on the method described in Step 3 of Section 5.3 and asked the expert to provide the undetermined values $(x_1^{1,2}, x_1^{1,3}, x_1^{1,4}, x_1^{1,5}, x_1^{1,6})$ so that the paired profiles are indifferently preferred.

$$\mathbf{x}^{1,2} \sim \mathbf{x}^2 \iff \left(x_1^{1,2}, 16, 4.2, 3.8, 3.4, 1\right) \sim (223, \mathbf{11}, 4.2, 3.8, 3.4, 1),$$

$$\mathbf{x}^{1,3} \sim \mathbf{x}^3 \iff \left(x_1^{1,3}, 16, 4.2, 3.8, 3.4, 1\right) \sim (223, 16, \mathbf{5.56}, 3.8, 3.4, 1),$$

$$\mathbf{x}^{1,4} \sim \mathbf{x}^4 \iff \left(x_1^{1,4}, 16, 4.2, 3.8, 3.4, 1\right) \sim (223, 16, 4.2, \mathbf{6.06}, 3.4, 1),$$

$\mathbf{x}^{1,5} \sim \mathbf{x}^5 \iff \left(x_1^{1,5}, 16{,}4.2{,}3.8{,}3.4{,}1\right) \sim (223{,}16{,}4.2{,}3.8, \mathbf{5.41}, 1),$

$\mathbf{x}^{1,6} \sim \mathbf{x}^6 \iff \left(x_1^{1,6}, 16{,}4.2{,}3.8{,}3.4{,}1\right) \sim (223{,}16{,}4.2{,}3.8{,}3.4, \mathbf{6}).$

After assessment, the expert determined the following values:

$\left(x_1^{1,1}, x_1^{1,2}, x_1^{1,3}, x_1^{1,4}, x_1^{1,5}, x_1^{1,6}\right) = (103{,}200{,}210{,}180{,}190{,}185).$

Step 5. Determine the others to the worst tradeoff

In addition, the expert was asked to determine the values $(x_1, x_2, x_3, x_4, x_5, x_6)$ according to Section 5.3, so that the following indifference relations on the generated profiles can be satisfied:

$\mathbf{x}^{1,3} \sim \mathbf{x}^3 \iff (x_1, 16{,}4.2{,}3.8{,}3.4{,}1) \sim (223{,}16, \mathbf{5.56}, 3.8{,}3.4{,}1),$

$\mathbf{x}^{2,3} \sim \mathbf{x}^3 \iff (223, x_2, 4.2{,}3.8{,}3.4{,}1) \sim (223{,}16, \mathbf{5.56}, 3.8{,}3.4{,}1),$

$\mathbf{x}^{4,3} \sim \mathbf{x}^3 \iff (223{,}16{,}4.2, x_4, 3.4{,}1) \sim (223{,}16, \mathbf{5.56}, 3.8{,}3.4{,}1),$

$\mathbf{x}^{5,3} \sim \mathbf{x}^3 \iff (223{,}16{,}4.2{,}3.8, x_5, 1) \sim (223{,}16, \mathbf{5.56}, 3.8{,}3.4{,}1),$

$\mathbf{x}^{6,3} \sim \mathbf{x}^3 \iff (223{,}16{,}4.2{,}3.8{,}3.4, x_6) \sim (223{,}16, \mathbf{5.56}, 3.8{,}3.4{,}1).$

We received the following judgements from the expert:

$\left(x_1^1, x_2^2, x_3^3, x_4^4, x_5^5, x_6^6\right) = (210{,}15{,}5.56{,}5.2{,}4{,}3).$

By using the value functions from Step 2, the revised vectors A^{BO} and A^{OW} are:

$A^{BO} = (1{,}0.2{,}0.11{,}0.37{,}0.28{,}0.32),$

$A^{OW} = (0.11{,}0.21{,}1{,}0.54{,}0.18{,}0.4).$

Step 6. Check the consistency of the preferences

When we had the preferences A^{BO} and A^{OW}, we needed to check the consistency level and determine whether the preferences need to be revised. We first checked the ordinal consistency by using the method described in Section 5.4. Applying Equations (5.17)-(5.20), we found that $OR = 0.5$, which means that the judgements violated the ordinal consistency. To identify the locations of the violated values, we used the local ordinal consistency ratios, from Model (5.19), as shown in Table 5-4.

Table 5-4. The ordinal consistency check table.

	C_1	C_2	C_3	C_4	C_5	C_6	OR_j
C_1	0	0	0	0	0	0	0
C_2	0	0	0	1	0	1	0.4
C_3	0	0	0	0	0	0	0
C_4	0	1	0	0	1	1	0.6
C_5	0	0	0	1	0	1	0.4
C_6	0	1	0	1	1	0	0.6

In Table 5-4, the value "1" indicates that the judgements in two corresponding attributes are ordinal-inconsistent, and "0" means they are ordinal consistent. For example, the judgements in C_2 are not ordinal-consistent with C_4 and C_6. Based on Table 5-4 we could clearly identify which pairs of comparisons need to be modified.

Step 7. Modify the inconsistent preferences by repeating Steps 4 and 5

After we found and located inconsistency in the obtained preferences, we contacted the expert again to ask her to rethink about the judgements,. After careful consideration, the expert acknowledged some inconsistencies, and then modified the preferences by applying the consistency improving process in Section 5.4.4. According to formulas (5.27)-(5.31) of this process, we obtain the admissible ranges of A^{BO} and A^{OW} in Table 5-5.

Table 5-5. The admissible ranges for revision.

		C_1	C_2	C_3	C_4	C_5	C_6
Original ranges		[103,223]	[11,16]	[4.2,5.65]	[3.8,6.06]	[3.4,5.41]	[1,6]
Acceptable ranges	A^{BO}	[103,193.23]	[11,15.34]	[4.4,5.65]	[4.16,6.06]	[3.96,5.41]	[1.57,6]
	A^{OW}	[103,210]	[11,15.31]	[4.61,5.65]	[4.16,6.06]	[3.80,5.41]	[1.57,6]
Improving ranges	A^{BO}	[103,210]	[13.28,15.05]	[5.65,5.65]	[4.41,4.7]	[4.31,5.07]	[2.43,2.62]
	A^{OW}	[103,210]	[12.87,15]	[5.65,5.65]	[4.52,5.2]	[4,4.28]	[2.71,3]

Based on the acceptable ranges, the expert revised some values of the Others-to-Worst tradeoffs, and left the Best-to-Others tradeoffs unchanged:

$\left(x_1^{1,1}, x_1^{1,2}, x_1^{1,3}, x_1^{1,4}, x_1^{1,5}, x_1^{1,6}\right) = (103,200,210,180,190,185).$

$(x_1, x_2, x_3, x_4, x_5, x_6) = (210,14,5.56,4.2,4,1.8).$

By using the value functions from Step 2, we obtained the new A^{BO} and A^{OW}:

$A^{BO} = (1,0.2,0.11,0.37,0.28,0.32),$

$A^{OW} = (0.11,0.41,1,0.13,0.18,0.16).$

Then we checked the consistency of the modified preferences again using the ordinal and cardinal consistency indices proposed in Section 5.4. Now the ordinal consistency ratio $OR = 0$, which means that the judgements are fully ordinal-consistent; the cardinal consistency ratio $CR = 0.18$, which is less than the threshold 0.37 when we refer to Table 5-1 (in this case we have 6 attributes and $a_{BW} = \frac{1}{a_{WB}} = \frac{1}{0.11} \approx 9$), which means the judgements can be accepted.

Step 8. Calculate the optimal weights of attributes

The new set of values finally satisfies the ordinal consistency, while the cardinal consistency ratios are below the consistency threshold. Next, we can apply the BWT model in Section 5.3 to obtain the optimal weights for the attributes. In this study, we adopted the linear model (5.13), and obtained the optimal weights as follows.

$w = (0.39, 0.1, 0.03, 0.18, 0.14, 0.16).$

Step 9. Rank the alternatives

Finally, we use the additive value function (5.1) to aggregate the weights and the assessment values of alternatives (after being normalized by the value functions presented in Figure 5-10), and the aggregated values of alternatives are obtained as Table 5-6:

Table 5-6. The normalized values, aggregated values and the final ranking

Ports	Normalized value						Aggregated value	Ranking
	C_1	C_2	C_3	C_4	C_5	C_6		
Piraeus	0.97	1	0	0	0	0.4	0.54	6
Koper	0.61	1	0.65	0.55	0.59	0	0.74	3
Genoa	0.37	0.6	0	0.21	0.32	0.2	0.32	7
Antwerp	0.37	0.8	0.85	0.5	0.67	0.6	0.79	2
Rotterdam	0.18	0.6	0.89	0.9	0.85	1	0.87	1
Hamburg	0	0	0.93	1	1	0.6	0.73	4
Gdansk	1	0.41	0.25	0.5	0.32	0.2	0.68	5

Therefore, according to the ranking, the port of Rotterdam is the most favorable option.

5.6 Discussion

There are several advantages in the BWT proposed in this study. Combined with the case study, we try to discuss it and compare it with other methods from the perspectives of anchoring bias, consistency check, computational complexity and the completeness of information.

5.6.1 Anchoring bias analysis

The anchoring bias is a cognitive bias which explains people's tendency towards the first piece of information they receive when they are evaluating something (Tversky and Kahneman, 1974). Such bias is persistent in several MADM methods, especially those with a single anchor like SMART, Swing, and Tradeoff (see, for instance, Buchanan and Corner (1997), Montibeller and Von Winterfeldt (2015), Rezaei (2021)). A recent study in this area, conducted by Rezaei (2021), shows that respondents tend to provide larger scores than their actual values to the other attributes when they are using a larger anchor (for example in the SWING method, where respondents start with identifying the most important attribute assigning it a score equal to 100), whereas in methods with a small anchor (like in the SMART method where respondents start with identifying the least important attribute assigning it a 10) the respondents assign scores to attributes lower than their actual ones. The preferences used to obtain weights in the classical Tradeoff method could be equivalent to one of the two vectors that we used in BWT, Best-to-Others (A^{BO}) or Others-to-Worst(A^{BO}). For the vector A^{BO}, its anchor is the best attribute, which is similar to the SWING method; and for the vector A^{OW}, its anchor is the worst attribute, which is similar to the SMART method.

Taking the case study in Section 5.5 as an example, if we obtain the weights that are only based on the vector A^{BO} or A^{OW}, as we can see in Figure 5-11, the weight of the best attribute ($C1$) obtained from A^{BO} is obviously larger than that from A^{OW}, and the weight of the worst attribute ($C3$) obtained from A^{BO} is also larger than that from A^{OW}. From this perspective, if we only consider one vector to calculate weights, as the classical Tradeoff method does, then we would encounter the anchoring bias.

One of the advantages of BWT is to remedy this anchoring bias. By combining the two opposite reference attributes (best and worst), the potential anchoring bias is mitigated, as we can see from the BWT line, which is located between lines BO and OW in Figure 5-11.

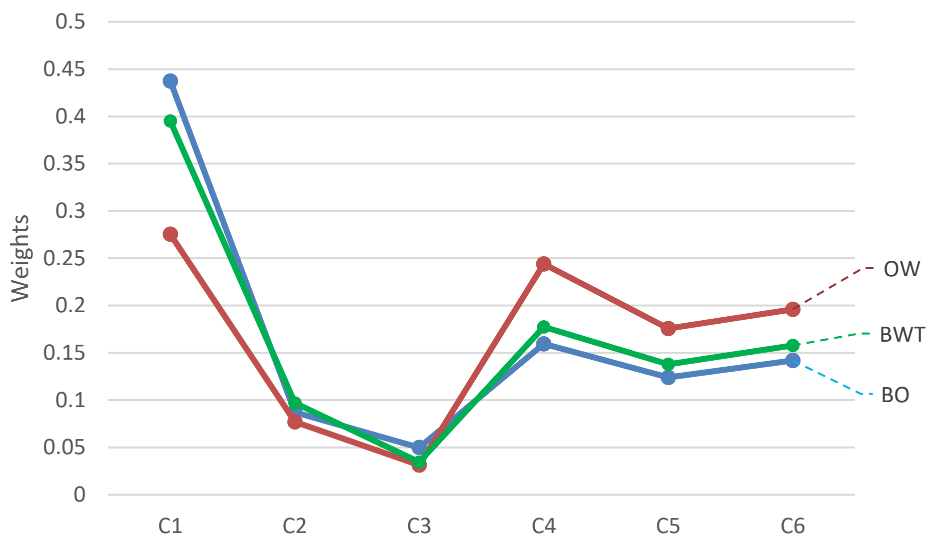


Figure 5-11. The weights obtained separately by vectors A^{BO}, A^{OW} and BWT.

5.6.2 Consistency check

Experimental studies have found that the classical Tradeoff method has higher inconsistency rate than Swing and Ratio method, and 67% of the subjects who applied Tradeoff elicitation procedure shown inconsistency [19,35]. Based on these studies, the FItradeoff method proposed by de Almeida et al. (2016; de Almeida et al., 2021) considers that the DM is not able to specify the tradeoff values and this information cannot be obtained in a consistent way from the DM.

The elicitation of indifference relations may not be as easy as for Swing or Ratio methods, but according to the successful applications of the Tradeoff method and our experiments, not only the elicitation is possible, but often also not difficult. Since the classical Swing and Ratio method are not able to check consistency of preferences (neither ordinal nor cardinal), then, from the perspective of validation (Zardari et al., 2015), it may be preferable to use the Tradeoff method. Although an extension of the FItradeoff method has considered the consistency checking and revision, it is for ordinal inconsistency and it still lacks detailed procedures for the revision of preferences (de Almeida et al., 2021). Therefore, solving the consistency problem is necessary for the classical Tradeoff method.

The BWT method proposed in this study assumes that a DM can specify his/her preferences and these preferences can be adjusted to an acceptable level. Based on these assumptions, we have developed a systematic consistency check and improvement process. With this process, a DM can locate the inconsistencies and visualize the acceptable adjustment and improvement ranges. It makes the consistency revision process, which is absent in the classical BWM and Tradeoff method, possible and easy.

For example, in the case study of Section 5.5, the consistency checking and improving process helped the expert identify her inconsistencies by using the local ordinal inconsistency ratios in Table 5-4, and showing the admissible ranges for adjustment in Table 5-5. Although the original preferences were ordinal-inconsistent, and the cardinal consistency ratios were in the acceptable range, it was still suggested to revise the preferences to be fully ordinal consistent. The expert only revised some of her A^{OW} judgments within the acceptable ranges. After this revision, the ordinal and cardinal consistency were acceptable, with no need for further adjustments.

5.6.3 Computational complexity

In its basic form, the original Tradeoff method requires a minimum of $(n-1)$ pairs of comparisons and, with this number of comparisons, it does not allow to check the consistency. To that end, on the other extreme, it could consider all the possible combinations of comparisons, which would result in $n(n-1)$ pairs (bidirectional), or $n(n-1)/2$ pairs (unidirectional) (Linares et al., 2016). The newly proposed FItradeoff method (de Almeida et al., 2021), requires $3(n-1)$ comparisons, but it is also worth mentioning that it uses inequalities, which, from the cognitive point of view, are less demanding questions than those asked in the classical Tradeoff method. The proposed BWT method requires $(2n-3)$ comparisons, which is linear with respect to the number of attributes, and, when a large number of attributes is used, the number of comparisons remains tractable. Figure 5-12 represents the number of comparisons required by various methods. In this sense, the BWT appears to strike a fair balance between having an acceptable level of redundancy in the questioning process, necessary for the evaluation of inconsistency, and the cognitive burden required from the DM, and presents a good scalability for larger problems. In addition, although we know inconsistent preferences are practically unavoidable, it is still difficult to identify which (and how much) judgements contribute to that inconsistency. Therefore, it is important to have a module within in BWT which enables the DM to check the consistency in a systematic way.

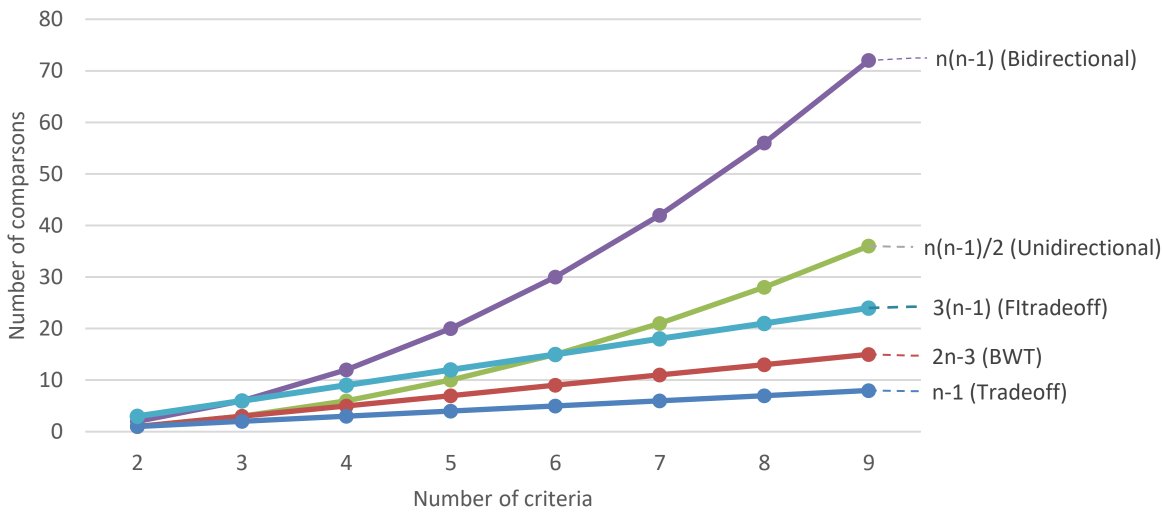


Figure 5-12. Relation between number of attributes and number of comparisons required by questioning techniques.

5.6.4 Methods with complete information V.S. methods with incomplete information

This section discusses the BWT with respect to information availability. In real decision contexts, it may be difficult, for a DM, to provide indifference tradeoff information. Therefore, to ease the preference elicitation process, a number of researches proposed methods requiring partial and/or incomplete information as input (Kirkwood and Sarin, 1985; Salo and Hämäläinen, 1992; de Almeida et al., 2016; de Almeida et al., 2021). Often, the incomplete information takes the form of weak order relations and inequalities instead of indifference relations and equalities. When we compare methods using incomplete information and methods using complete information, we shall consider three aspects: the availability of the complete information, the cost of eliciting preferences, and the consistency check of the comparisons.

The availability of the complete information. As the applications of classical Tradeoff method have proved that indifference relations can be elicited (Keeney et al., 1990), so, similarly, complete information can also be obtained in BWT. Of course, some DMs cannot, or may not be willing to, provide complete or certain information, but there are also DMs who are familiar with the situation and can provide certain/complete information. If a DM can only provide incomplete information, then methods like FItradeoff (de Almeida et al., 2021) could be a good option. On the contrary, if complete information is available, using the BWT method has the advantage of helping mitigate DMs' anchoring bias and measure (and improve) the inconsistency of preferences.

The cost of eliciting preferences. One major argument in favor of using incomplete information is that eliciting complete information demands higher cognitive effort. As observed in the literature (Weber and Borcherding, 1993), using inequalities seems easier, with respect to the required cognitive effort, but how many inequality questions demand the same effort as a given set of equality questions? For example, in the FItradeoff method (de Almeida et al., 2021), the minimum acceptable number of comparisons, in the form of inequalities, are $3(n-1)$, which is 3 times more than the classical Tradeoff method, and n more comparisons than BWT. However, given the different nature of the elicited information, it is hardly possible to attach a quantification of the cognitive effort required by each method. Moreover, such a measure of cognitive effort may not only depend on the method and the problem, but also on the DM involved in the process.

The consistency check of the comparisons. As Albert Einstein said, *everything should be made as simple as possible, but not simpler*. We try to make the elicitation procedure as simple as possible, but the minimum condition is that the consistency of these elicited preferences can be examined. The existing methods either lack a phase of consistency check, or they can only check ordinal consistency and seldom they are of any help in guiding the analyst and the DM through a consistency improvement process (de Almeida et al., 2021). One of the advantage of BWT is that it can check both ordinal and cardinal consistency. Moreover, it can help DMs to improve their consistency level with acceptable ranges and improving ranges for references which can reduce the cognitive effort required from DMs.

DMs could choose to provide complete or partial/incomplete information, and use different methods to deal with these two types of information, and there is no right or wrong choosing one way or the other. To us, methods using incomplete information are complements to methods using complete information, especially when the cost of obtaining complete information is too high, if not even impossible.

5.7 Conclusion and future research

In this study, we developed a multi-attribute decision-making method called BWT that can be viewed as an attempt to combine the merits of the traditional BWM and the Tradeoff method, without losing the characteristics that have made the two methods popular. More specifically, with the BWT, we can elicit weights in a more structured way using the prescriptive MAVT approach, which considers the attribute range effect, and at the same time have a guided choice of the attributes to be compared with a check of the consistency of the preferences.

From the point of view of the original BWM, the BWT is better at eliciting preferences, obtaining weights, and complies with the theory of MAVT. From the point of view of MAVT, the BWT, represents a scheme for questioning (and testing the consistency of) DMs in the elicitation process. We want to emphasize that our proposal to include thresholds should not be interpreted as a strict acceptance/rejection rule, but as an effort to increase the intelligibility of the inconsistency index CR.

In addition, the BWT model may make another potential contribution. The traditional Tradeoff method may suffer from anchoring bias (or scale compatibility bias) and loss aversion bias, if the preferences obtained from DMs depend on a measuring stick (anchor) (Tversky and Kahneman, 1974; Bleichrodt and Pinto, 2002; Lahtinen and Hämäläinen, 2016; Lahtinen et al., 2020). By introducing the consider-the-opposite-strategy, the use of two opposite reference attributes (best and worst) could reconcile the possible anchoring bias when eliciting judgements from the DMs (Rezaei, 2020).

While the method has been devised for situations where a decision-maker/analyst has full information about the alternatives including the ranges of attributes, in situations where a decision-maker has no full information about the alternatives, considering a nominal range (Pajala et al., 2019) could help applying the method, however in those situations the findings should be interpreted more carefully.

We leave it to future studies to examine whether the anchoring bias affects the ultimate decisions resulting from using the Tradeoff method and how BWT helps remedy the bias (if it is able to). Besides, while conducting the surveys, the experts sometimes were hesitant with regard to providing the precise judgements when we asked the tradeoff questions, so future studies could examine how incomplete and partial information can be considered within the BWT method.

Acknowledgment

The authors are very grateful to the constructive comments provided by the editor and three anonymous reviewers on the initial version of this article. The authors also gratefully acknowledge financial support for PhD study from the China Scholarship Council (No. 201708440305).

References

Amenta, P., Lucadamo, A. & Marcarelli, G. (2020). On the transitivity and consistency approximated thresholds of some consistency indices for pairwise comparison matrices. *Information Sciences*, *507*, 274-287.

Bacon, F. (1960). The new organon and related writings. New York, Liberal Arts Press.

Bleichrodt, H. & Pinto, J. L. (2002). Loss aversion and scale compatibility in two-attribute trade-offs. *Journal of Mathematical Psychology*, *46*(3), 315-337.

Borcherding, K., Eppel, T. & Von Winterfeldt, D. (1991). Comparison of weighting judgments in multiattribute utility measurement. *Management science*, *37*(12), 1603-1619.

Bottomley, P. A. & Doyle, J. R. (2001). A comparison of three weight elicitation methods: Good, better, and best. *Omega*, *29*(6), 553-560.

Brunelli, M. (2016). Recent advances on inconsistency indices for pairwise comparisons—A commentary 1. *Fundamenta Informaticae*, *144*(3-4), 321-332.

Buchanan, J. T. & Corner, J. (1997). The effects of anchoring in interactive MCDM solution methods. *Computers & Operations Research*, *24*(10), 907-918.

Cavallo, B., D'Apuzzo, L. & Basile, L. (2016). Weak consistency for ensuring priority vectors reliability. *Journal of Multi-Criteria Decision Analysis*, *23*(3-4), 126-138.

de Almeida, A. T., de Almeida, J. A., Costa, A. P. C. S. & de Almeida-Filho, A. T. (2016). A new method for elicitation of criteria weights in additive models: Flexible and interactive tradeoff. *European Journal of Operational Research*, *250*(1), 179-191.

de Almeida, A. T., Frej, E. A. & Roselli, L. R. P. (2021). Combining holistic and decomposition paradigms in preference modeling with the flexibility of FITradeoff. *Central European Journal of Operations Research*, *29*(1), 7-47.

Dias, L. C., Morton, A. & Quigley, J. (2018). Elicitation. Switzerland, Springer International Publishing.

Dyer, J. S. & Sarin, R. K. (1979). Measurable multiattribute value functions. *Operations research*, *27*(4), 810-822.

Edwards, W. (1977). How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, *7*(5), 326-340.

Eisenführ, F., Weber, M. & Langer, T. (2010). Rational decision making. New York, Springer.

Escobar, M. T., Aguarón, J. & Moreno-Jiménez, J. M. (2015). Some extensions of the precise consistency consensus matrix. *Decision Support Systems*, *74*, 67-77.

Fischer, G. W. (1995). Range sensitivity of attribute weights in multiattribute value models. *Organizational Behavior and Human Decision Processes*, *62*(3), 252-266.

Fischer, G. W., Damodaran, N., Laskey, K. B. & Lincoln, D. (1987). Preferences for proxy attributes. *Management Science*, *33*(2), 198-214.

Fishburn, P. C. (1967). Methods of estimating additive utilities. *Management science*, *13*(7), 435-453.

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

Keeney, R. L. (2002). Common mistakes in making value trade-offs. *Operations Research*, *50*(6), 935-945.

Keeney, R. L. & Raiffa, H. (1976). Decision analysis with multiple conflicting objectives. New York, John Wiley & Sons.

Keeney, R. L., Von Winterfeldt, D. & Eppel, T. (1990). Eliciting public values for complex policy decisions. *Management Science*, *36*(9), 1011-1030.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, *30*(1–2), 81-93.

Kirkwood, C. W. & Sarin, R. K. (1985). Ranking with partial information: A method and an application. *Operations Research*, *33*(1), 38-48.

Lahtinen, T. J. & Hämäläinen, R. P. (2016). Path dependence and biases in the even swaps decision analysis method. *European Journal of Operational Research*, *249*(3), 890-898.

Lahtinen, T. J., Hämäläinen, R. P. & Jenytin, C. (2020). On preference elicitation processes which mitigate the accumulation of biases in multi-criteria decision analysis. *European Journal of Operational Research*, *282*(1), 201-210.

Liang, F., Brunelli, M. & Rezaei, J. (2020). Consistency issues in the best worst method: Measurements and thresholds. *Omega*, *96*, 102175.

Linares, P., Lumbreras, S., Santamaría, A. & Veiga, A. (2016). How relevant is the lack of reciprocity in pairwise comparisons? An experiment with AHP. *Annals of Operations Research*, *245*(1-2), 227-244.

Montibeller, G. & Von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Analysis*, *35*(7), 1230-1251.

Myers, R. H., Montgomery, D. C. & Anderson-Cook, C. M. (2016). Response surface methodology: process and product optimization using designed experiments, John Wiley & Sons.

Pajala, T., Korhonen, P. & Wallenius, J. (2019). Judgments of importance revisited: What do they mean? *Journal of the Operational Research Society*, *70*(7), 1140-1148.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49-57.

Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega*, *64*, 126-130.

Rezaei, J. (2020). A concentration ratio for nonlinear best worst method. *International Journal of Information Technology & Decision Making*, *19*(3), 891-907.

Rezaei, J. (2021). Anchoring bias in eliciting attribute weights and values in multi-attribute decision-making. *Journal of Decision Systems*, *30*(1), 72-96.

Rezaei, J., van Wulfften Palthe, L., Tavasszy, L., Wiegmans, B. & van der Laan, F. (2019). Port performance measurement in the context of port choice: an MCDA approach. *Management Decision*, *57*(2), 396-417.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*(3), 234-281.

Salo, A. A. & Hämäläinen, R. P. (1992). Preference assessment by imprecise ratio statements. *Operations Research*, *40*(6), 1053-1061.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Vetschera, R., Weitzl, W. & Wolfsteiner, E. (2014). Implausible alternatives in eliciting multi-attribute value functions. *European Journal of Operational Research*, *234*(1), 221-230.

von Nitzsch, R. & Weber, M. (1993). The effect of attribute ranges on weights in multiattribute utility measurements. *Management Science*, *39*(8), 937-943.

Von Winterfeldt, D. & Edwards, W. (1986). Decision analysis and behavioral research. Cambridge, Cambridge University Press.

Weber, M. & Borcherding, K. (1993). Behavioral influences on weight judgments in multiattribute decision making. *European Journal of Operational Research*, *67*(1), 1-12.

Zardari, N. H., Ahmed, K., Shirazi, S. M. & Yusop, Z. B. (2015). Weighting methods and their effects on multi-criteria decision making model outcomes in water resources management. New York, Springer International Publishing.

6 Conclusion

Reaching a decision is the end of an analysis, but also the start of a story. A good decision could be half the success, but a bad one could turn the entire journey into a nightmare. This can be seen every day, in the business of governments, companies, organizations and individuals. Thinking twice might be of little help, but thinking systematically could significantly improve the quality of the decision-making process. As a systematically structured decision-making method, BWM helps DMs to elicit their preferences more consistently and arrive at a better decision. Nonetheless, there are several issues that still need to be tackled in BWM. This thesis tries to improve BWM in terms of consistency, uncertainty, consensus and range sensitivity, with the aim of making it more comprehensive and practical.

Firstly, in this thesis, we try to verify the reliability of the provided preferences in BWM, which involves measuring the inconsistency and uncertainty in the DM's judgments. Next, a group BWM is developed to handle the MCDM problem taking multiple stakeholders with different sets of criteria into consideration. Finally, a study incorporates the merits of the classical Tradeoff method to tackle the range sensitivity issue in BWM. The results of the studies in response to the research questions are presented below.

6.1 Conclusions for Study 1: Checking the (in)consistency of preferences in BWM

- *How to develop a BWM model-independent (in)consistency measurement?*
- *How to take the ordinal (in)consistency into account?*
- *How to determine the (in)consistency thresholds?*

The original BWM can only provide the inconsistency information after solving the optimization problems, which could lead to different inconsistency results if we use different optimization models. To cope with this problem, this study proposes a model-independent (input-based) (in)consistency measurement that is able to provide a better feedback to the DM. Thanks to this (in)consistency measurement, it is easy for a DM to identify their most inconsistent judgments, which can be very useful in the preference revision process. It has several desirable properties and can provide more valuable information than the original one (output-based (in)consistency measurement), for instance the location of the most inconsistent preferences. In addition, the original BWM has not considered ordinal (in)consistency measurement, which is considered to be essential in checking the consistency of a DM. Therefore, in this study, an ordinal (in)consistency measurement was developed to complement the existing cardinal (in)consistency measurement. With this ordinal (in)consistency measurement, the possible contradictions could be explicated even when the cardinal consistency of the judgments is considered to be good enough. The proposed ordinal consistency ratio can help a DM to identify and correct their judgments to meet the ordinal consistency condition. Finally, the existing BWM studies lack a mechanism to provide a meaningful interpretation to the consistency ratios. Establishing consistency thresholds can help us to judge acceptable and unacceptable consistency levels. Hence, in this study, we construct the consistency threshold for the input- and output-based consistency ratios in different scales and different numbers of criteria. With the help of Monte-Carlo simulations, the probability distributions of the cardinal consistency ratios can be balanced in order to make the portion of the cardinal consistency ratios that violate ordinal consistency to be accepted is as small as possible, and the portion of the cardinal consistency ratios that satisfy ordinal consistency to be rejected is as small as possible. Based on these thresholds, the preferences provided by a DM can now be assessed as to whether they can be accepted or not.

6.2 Conclusions for Study 2: Managing uncertain information in BWM

- *How to capture DM's ambiguity information?*
- *How to elicit weights based on the uncertain preferences?*

In the original BWM, it is a assumed that DMs are certain about their judgments. However, in reality, due to the uncertainty of the decision problems and due to limited human cognition, the judgements are affected by bounded rationality. First, in order to capture the uncertain information, a belief structure is introduced to the BWM, a concept adopted from Dempster-Shafer theory that uses the "degrees of belief" to express the extent to which a DM believes a specific proposition to be true. With the help of the belief structure, we can capture more comprehensive data as well as preferences with both discord and non-specificity. Then, to handle this form of information, a belief-based BWM is developed in Chapter 3. The basic idea behind the belief-based BWM is that preferences with greater belief should be valued more, and preferences with lower belief should be valued less. Therefore, the belief-based BWM not only enables a DM to provide their basic belief assignments more flexibly, but also takes the belief levels of the preferences into account, making the decision rely more on the stronger beliefs and less on the weaker beliefs.

Furthermore, checking the reliability of the preferences is very important, since inconsistency and uncertainty usually lead to unreliable results. Thus, after determining the weights, to measure the reliability of the provided preferences, an (in)consistency measurement and an uncertainty measurement are proposed for the belief-based BWM. A method designed to measure the reliability degree of a DM's judgments to incorporate their inconsistency and uncertainty levels is then developed. Finally, to demonstrate the applicability and feasibility of the proposed method, a real-life case study is carried out on the assessment of the large-scale infrastructure project criteria system in Indonesia. As we asked for feedback from the DMs of the project for confirmation, they indicated that the belief-based BWM can handle the deficiencies of the existing methods, and can be a solution for their future projects.

6.3 Conclusions for Study 3: Reaching consensus in group BWM

- *How to accommodate different criteria sets in group BWM?*
- *How to reach consensus with interval weights?*

In the existing group BWM studies, the DMs (actors, experts or stakeholders) are asked to provide their preferences using the same set of criteria, which is not always done in real-world problems. That is to say, BWM does not have a solution for a heterogeneous group with different sets of criteria, which is why, in this study we propose a consensus model to fill this gap. By using this consensus model, firstly, a number of stakeholders with different objectives are identified and then each stakeholder can evaluate the problem and identify their own set of criteria, which may be different from those identified by the other stakeholders. Next, a nonlinear BWM is used to prioritize the importance of the criteria identified by the various stakeholders and obtain the interval weights, on the basis of which the aggregated value of each alternative is obtained by an additive value function for each stakeholder. Because the alternative values are located at intervals, after combining with the interval weights, traditional methods like averaging the interval centers tend to omit the ranges and the overlaps of the intervals, which is why a consensus model is proposed in this study and used to aggregate the group assessments, the aim being to eliminate outliers and overlapping areas from the final aggregated values as much as possible, because they represent the consensual opinions of the experts. The resulting interval values can then be ranked, allowing us to select the most desirable alternative. The developed multi-stakeholder BWM model is applied to a real-life inland terminal selection case study that was initiated by shipping line Maersk. The company's managers confirm that the method proposed in this study is very intuitive and reasonable, and the analysis of this study has provided the company with a valuable reference for its strategic selection.

6.4 Conclusions for Study 4: Accounting range sensitivity of criteria in BWM

- *How to account ranges (sensitivity) of criteria in BWM?*

The traditional BWM implies that DMs should consider the alternatives while providing their preferences with regard to the importance of certain criteria. However, in practice, DMs directly assess the criteria without considering the alternatives in systematic way. This insensitivity to the range of criteria may cause some distortion or biases in the elicitation of weights. To deal with this problem, we introduced a Tradeoff method into BWM, based on the multi-attribute value theory, which explicitly takes the ranges of criteria into account. Combined with the "consider-the-opposite-strategy" idea of the BWM, a new method called Best-Worst Tradeoff (BWT) method is proposed. To be specific, a DM is asked to provide two pairwise comparison vectors based on the values of the (ranges of the) criteria. The two vectors will be used as input for the proposed optimization model, which is designed to determine the optimal weights of the criteria. Additionally, to measure the inconsistency of the pairwise comparisons, a cardinal consistency ratio and an ordinal consistency ratio are proposed. On top of that, a table of consistency thresholds is constructed, with which we can judge whether the provided preferences are sufficiently rational. The combination of the BWM and the Tradeoff method allows the proposed BWT method to combine the merits of the other two methods, without losing the characteristics that have made them popular. In other words, we can now use the prescriptive MAVT approach and the consider-the-opposite-strategy to account for the criteria ranges and keep the inconsistency of the provided preferences under control. Finally, this method is applied to a performance evaluation project in the Netherlands. After analysis, we found that BWT performs better compared to the existing methods with regard to consistency checking, anchoring bias and computational complexity.

6.5 Reflections

In the course of these studies, I had the impression that our brains are not designed for certainty and precision. In our questionnaires, DMs are asked to describe their judgment on the importance of some criterion by using a rating scale, for instance a Likert scale, and they can only provide a "rough" number. Things could be even fuzzier when we use ratio scales in pairwise comparison methods like BWM and AHP. Although the scale indicates very clearly what the numbers represent, for example "2" represents "equally to slightly more important", "9" represents "extremely more important", the DM simply cannot naturally link the two criteria with a number. There may be two reasons to explain that difficulty. First, there is no universal standard regarding the importance measurement. It could be hard to understand the meanings of the numbers, and even the meaning of "importance". Everything depends on someone's perceptions, and those are exclusive. Second, the preferences could be changing. The value of a product to someone depends on their demand, which depends on their living situation. As the situation changes, so does the value of the product. Similarly, someone's perception of importance also changes with their situations. This is also why including the alternatives is necessary when we elicit preferences from a DM, because the ranges of the criteria serve as anchors, determining the "situation".

Using direct rating on importance is already difficult when it comes to obtaining a clear answer from a DM, never mind using ratio scales. The precision of the connection between the linguistic interpretations and the numbers is a major question mark, and the question whether or not a DM understands the meanings of the ratios is another issue. These issues require more investigation. However, even when we assume that the connection between numbers and the linguistic terms is reasonable and the DM understands its meaning, there is still one issue, which is a natural characteristic of perceptions: ambiguity (uncertainty). To capture this uncertainty, we could use linguistic terms, interval numbers, rough numbers, fuzzy numbers, or the belief structures we used in the belief-based BWM to complement fuzzy numbers. These kinds of information appear to help DMs to express their hesitations in an understandable way, but the accuracy is doubtful. It is a paradox. We are trying to use certainty to replace uncertainty. However, we have to try, because of the nature of human beings, we have the desire to try and bring everything under control, even if we may never succeed.

Uncertainty could be one of the major sources of the so-called inconsistency of a DM. If the preferences (the judgments of importance in the weighting method) of a DM cannot be interpreted precisely, how could it be fully consistent (we are talking about cardinal consistency here)? Mathematically speaking, it could be easy to determine whether or not the preferences of a DM are perfectly consistent compared to the transitivity condition. However, it is hard to be fully cardinal-consistent for DMs, especially when the number of criteria is large, which will increase the feelings of uncertain. Thus, it is not realistic to ask DMs to revise their preferences to try and be fully cardinal-consistent, and minor inconsistency can be acceptable (or we could say it is within the acceptable range of error/misunderstanding). However, the question then becomes what range is acceptable? Telling good (acceptable) from bad (unacceptable) needs a standard, but where is this standard (if there exists one)? We could simply use the rule of thumb standard, for instance AHP uses 10% as the threshold, but is this a reasonable boundary? Intuitively, the error margins should be increased as the number of criteria increases, which means that using a fixed 10% threshold is apparently not suitable. In this thesis, we use ordinal consistency as a reference point, based on which we constructed the thresholds. The reason is that we consider ordinal consistency as the minimum requirement for a DM to provide their preferences. For more details about the procedure, see Chapter 2. It appears to be a reasonable way to construct the thresholds, but we do not claim that it is the right or only way. What we have developed is just one way, because we are fully aware that defining acceptable reference standards is difficult, it requires knowledge, power, time and consensus.

How to reach consensus is another interesting subject. In our daily lives, the most efficient way to reach a consensus may be to follow the authorities. But if we try to ask everybody’s opinion, it could be tedious and very difficult to arrive at a group decision. In Study 2 (Chapter 3) and Study 3 (Chapter 4), we proposed two different models to aggregate the preferences of the group DMs. In Study 2, the aggregation is based on the weights of the DMs, but in Study 3, the aggregation is based on the overlaps of the preferences. The former one focuses on the DMs, the latter one focuses on the preferences. Most of the aggregation methods of the group MCDM methods derive from these two approaches, focusing either on DMs or on preferences. How to determine the weights of DMs and how to measure the preferences make the methods different.

When we proposed BWT, some reviewers thought it complicated the problem, since the questions posed in BWT are more difficult for DMs to answer than the ones we ask in the original BWM. This is partially true, but it is worth noticing that (i) the difficulty of the questions in BWT should not be compared to that of the questions asked when using BWM; they should instead be compared to the questions asked when using the traditional Tradeoff, whose difficulty level is comparable; (ii) If we compare the difficulty of the questions involved (of the same level in our method and the original Tradeoff) to the questions we ask in BWM, it is true that these questions are more difficult (it is perhaps better to say that they need a greater cognitive effort), but such a difficulty is for a valid purpose, which is to incorporate the attribute ranges. In other words, we could also argue that this issue itself is a ‘tradeoff’ problem: a tradeoff between difficulty and reliability. As discussed in Study 4, taking the ranges into account could produce more reliable weights. Having said that, a DM facing a real-world decision-making problem, depending on the situation at hand, could decide whether they want to go for a simpler method like BWM or a more sophisticated one like BWT. We think this is the case with all decision-making problems where the choice of method itself is a decision-making problem.

6.6 Limitations and future research

The consistency thresholds we have constructed in Study 1 only indicate the acceptable level of cardinal consistency ratios, not that of ordinal consistency ratios. It could be interesting to examine the threshold for the level of ordinal consistency violation. Also, conducting psychological experiments to determine whether the constructed thresholds are reasonable in practice could add valuable insight as well. Furthermore, as our proposed method to construct the consistency thresholds is merely one alternative, it could be worthwhile to see where there are simpler and more intuitive consistency threshold constructing models available.

In Study 2, we determine the weights of DMs based on their inconsistency and uncertainty, which could be too biased, because we have neglected other dimensions, like expertise, position and peer valuation. Thinking back, it is debatable whether the inconsistency and uncertainty of a DM can fully reflect their reliability. In fact, an experienced expert could be very uncertain about the solution and very inconsistent on providing their preferences, because they have considered many different dimensions. Laypeople could be very certain and consistent in their preferences, because they know little about the situation (Dunning-Kruger effect, a hypothetical cognitive bias stating that people with low ability at a task overestimate their ability). Perhaps the reliability of an expert is more closely related to their working years, academic qualifications or the number of projects in which they have been involved. However, if we think deeper, reliability is about correctness, credibility, so judging based on the number of successful decisions (projects) of an expert (DM) could be the most important criterion and it could be interesting to take a closer look at the determination.

Traditionally, we take the average of all DMs' preferences or the majority rule. However, in many situations, the preferences or judgments of a DM can be ambiguous, or they could be a set or range of judgements, making the overlapping options or ranges of the provided preferences of the DMs the preferred consensus. Identifying these overlapping areas and integrating them is the consensus model we proposed in this thesis. However, the consensus model we constructed in Study 3 is only used to aggregate the overall values of alternatives of each DM, instead of aggregating the weights of criteria of each DM, because different sets of criteria are applied by different DMs. In future research, it will be interesting to develop a consensus model to aggregate the weights of different sets of criteria before calculating the overall values of alternatives. Additionally, it will be interesting to use a unit-sum constraint for the interval weights aggregation that is lacking in our study.

We had high expectations of the BWT in practice, when we came up with the idea described in Study 4, because it combines the advantages of BWM and the Tradeoff method, enabling us to determine DM consistency, consider the range sensitivity and reconcile the potential anchoring bias. However, in practice , we found that it is difficult for experts to compare the values of one criterion to another, not only because of the different units make such a comparison difficult to make, but also because experts sometimes cannot provide certain tradeoff values. The cognitive effort required in this method gives experts assessment fatigue, especially when there are many criteria, which means it is better to allow the experts to provide uncertain preferences. As such, in the future, partial information BWT, interval BWT, fuzzy BWT, belief-based BWT can complement the BWT proposed in this thesis.

BWM is a practical method that has been applied to various fields, and we believe it can and will be applied to any fields that involved multi-criteria decision-making. Nonetheless, to adapt to different situations, it is still necessary to investigate the related theories and methodologies, for example probability theory, evidence theory, fuzzy set theory and other MCDM methods. Furthermore, examining the source of uncertainty and inconsistency can help us to understand our behavior better, allowing us to improve the reliability of the elicited preferences. On top of that, it could be also interesting to study dynamic BWM model so as to capture the changing perceptions of a DM. After all, it is a method relied on humans.

Overall, from the perspectives of consistency, uncertainty, consensus and range sensitivity, this thesis has significantly contributed to the establishment of the BWM. It is gratifying to see the proposals have been adopted increasingly by many researchers, companies and policy-makers, and we believe that these methods will be applied in more and more different fields. As with all the newly developed methods, our proposals are not perfect, and more works has to be done to test and improve the proposed models in practice.

Summary

It is our choices that make us who we are. To lead a better life, we have to make better decisions. Nowadays, decisions are increasingly made in complex contexts, in a host of different application domains. Because of that, we need more reliable decision analysis methodologies to improve our decisions. The ability to deal with multi-dimensionality is one of the critical requirements of the decision analysis methods that help us make better decisions. Multi-Criteria Decision-Making (MCDM) is one of the most popular approaches when it comes to formulating and solving decision-making problems, best-known for its ability to handle problems where a multitude of, often conflicting, criteria arise. As one of the latest MCDM methods, the Best-Worst Method (BWM) has been studied substantially and applied increasingly to various fields since its introduction, thanks to its simplicity, flexibility and general applicability.

Despite its popularity, some significant issues of BWM have not yet been systematically investigated in existing literature, including: (i) the inconsistency in the preferences provided by Decision-Makers (DMs), (ii) the uncertain information embedded in the DMs' judgements, (iii) problems in reaching a consensus in group decision-making, and (iv) the range sensitivity in an MCDM problem that is not taken into account in BWM. The main objective of this thesis is to develop an approach to measure, check and improve inconsistency, to develop an approach to incorporate judgments uncertainty, to develop a method to reach consensus and to incorporate range sensitivity in the BWM.

Inconsistency: Several measures have been proposed to deal with the inconsistency issue: Firstly, since the feedback of the existing (in)consistency measurements of BWM depend on the optimization models used to elicit the weights, an input-based (as well as model-independent) (in)consistency measurement is proposed in this thesis that enables DMs to obtain the consistency level before choosing an optimization model. Secondly, to complement the cardinal (in)consistency measurement of the original BWM, an ordinal (in)consistency measurement is developed that enables DMs to check their violation level involving ordinal consistency. Thirdly, this thesis establishes (in)consistency thresholds for BWM, providing the

DMs a meaningful interpretation, so that they can determine when their judgments should be revised and when they can be accepted.

Uncertainty: The original BWM can only handle judgments under certainty, although some extensions of BWM have tried to include the ambiguous judgments of DMs, most have only considered fuzziness. A belief-based BWM proposed in this thesis makes it possible to capture the DMs’ uncertain judgments with discord and non-specificity. Besides, excessive uncertainty and inconsistency may lead to unreliable results, on account of this, a reliability index is designed in this thesis to provide a way to monitor the reliability of DMs’ judgments.

Consensus: Firstly, a consensus framework considering heterogeneous group of DMs (experts or stakeholders) with different sets of criteria is developed in this thesis to complement the existing group BWM models, using the same set of criteria for all the group members. In addition, since the weights of criteria for each DM obtained from the nonlinear BWM are usually intervals, aggregating those weights in the traditional ways (e.g., averaging) could neglect the overlaps of the intervals. The proposed consensus model considers the overlaps of the interval weights, in order to reach the best agreement among the group members.

Range sensitivity: The original BWM does not systematically take the ranges of criteria into account, which could lead the resulting range insensitivity to biases. Combining the “consider-the-opposite-strategy” inherent in the procedure of preference elicitation of the BWM and the principles of the Tradeoff method, we proposed a Best-Worst Tradeoff (BWT) method, which makes it possible not only to incorporate the range of the criteria, but to mitigate the potential anchoring biases while eliciting preferences as well. Moreover, a(n) (in)consistency measurement framework adapted to the BWT is proposed so that DMs who use this method can check their consistency levels, and use them as a feedback to improve the consistency of the judgments.

To examine the practicality and feasibility of the proposed methods, the belief-based BWM is applied to an evaluation of large infrastructure projects in Indonesia, the consensus model is utilized to analyze an inland terminal location selection project in Germany, and the BWT framework is used in a port performance evaluation case study in the Netherlands. The (in)consistency thresholds developed in this thesis have already been used by many other scholars, and they will become a standard part of any BWM application.

To summarize, the main contributions of this thesis are: (i) providing a model-independent method for measuring the (in)consistency of preferences provided by DMs, constructing thresholds to help DMs to judge the acceptance of the resulted consistency ratios, (ii) introducing belief-structure into BWM to capture the uncertainty or hesitation of the DMs’ judgments and proposing a model to check the reliability of the DMs, (iii) accommodating different criteria sets in group BWM and building a model to facilitate DMs to reach consensus with interval weights, and (iv) developing an MCDM method, the BWT, to combine the merits of the traditional BWM and the Tradeoff method, which accounts the range sensitivity of criteria to avoid distortion or biases. Overall, this thesis will contribute significantly to the establishment of the BWM.

Samenvatting

Onze keuze bepalen wie wij zijn. Om een beter leven te leiden, moeten we betere keuzes maken. Vandaag de dag nemen we meer en meer beslissingen in complexe contexten in een verscheidenheid aan verschillende toepassingsdomeinen. Dat betekent dat we betere beslissingsanalysemethodes nodig hebben om onze beslissingen logischer te maken. Het vermogen om met multi-dimensionaliteit om te gaan is één van de essentiële vereisten van de beslissingsanalysemethodes die ons in staat stellen betere beslissingen te maken. Multi-Criteria Decision-Making (MCDM) is één van de populairste methodes om beslissingsproblemen te formuleren en op te lossen, en een methode die met name bekend staat op het vermogen met problemen om te gaan die worden gekenmerkt door een veelvoud aan, vaak tegenstrijdige, criteria. Een van de meer recent ontwikkelde MCDM methodes, the Best-Worst Method (BWM), is al vaak onderzocht en wordt meer en meer toegepast in verschillende gebieden sinds de ontwikkeling ervan, dankzij zijn eenvoud, flexibiliteit en algemene toepasbaarheid.

Ondanks zijn populariteit zijn er een paar belangrijke onderdelen van BWM die in de bestaande literatuur nog niet systematisch zijn onderzocht, waaronder: (i) de inconsistentie in de voorkeuren van beslissers (Decision-Makers – DMs), (ii) de onzekere informatie die is verankerd in de oordelen van DMs, (iii) het probleem om een consensus te bereiken in groepsbeslissingen, (iv) de bereikgevoeligheid in een MCDM-probleem die in BWM niet wordt meegenomen. De belangrijkste doelstelling van dit proefschrift is om een manier te ontwikkelen om inconsistentie te meten, controleren en verbeteren, om een manier te ontwikkelen om onzekerheid in de oordelen van DMs mee te nemen, om een methode te ontwikkelen om een consensus te bereiken en om bereikgevoeligheid in BWM op te nemen.

Inconsistentie: Er zijn meerdere manier voorgesteld om met het probleem van inconsistentie om te gaan: ten eerste wordt er in dit proefschrift een input-based methode voorgesteld om (in)consistentie te meten (die tevens model-onafhankelijk is), aangezien de feedback van de bestaande (in)consistentiemetingen van BWM afhankelijk zijn van de optimalisatiemodellen die worden gebruikt om de gewichten vast te stellen, waardoor DMs in staat zijn om de gewenste consistentie te bereiken voordat ze een optimalisatiemodel kiezen. Ten tweede wordt er, om de kardinale (in)consistentiemetingen van de oorspronkelijke BWM te complementeren, een ordinale (in)consistentiemeting ontwikkeld die DMs in staat stelt om na te gaan in hoeverre er wordt afgeweken van ordinale consistentie. Ten derde heeft dit proefschrift (in)consistentiedrempels voor BWM bepaald, zodat DMs op een zinvolle manier kunnen nagaan wanneer hun oordelen moeten worden herzien en wanneer ze acceptabel zijn.

Onzekerheid: De oorspronkelijke BWM kan alleen oordelen onder onzekerheid verwerken en, hoewel sommige uitbreidingen van BWM hebben geprobeerd om de ambigue oordelen van DMs mee te nemen, betreft dat in de meeste gevallen uitsluitend fuzziness. Een op overtuiging gebaseerde BWM die in dit proefschrift wordt voorgesteld maakt het mogelijk om de onzekere oordelen van DMs met onenigheid en niet-specificiteit te vangen. Bovendien kunnen excessieve onzekerheid en inconsistentie leiden tot onbetrouwbare resultaten, dus hebben we in dit proefschrift een betrouwbaarheidsindex ontwikkeld om de betrouwbaarheid van de oordelen van DMs te monitoren.

Consensus: Om te beginnen is er in dit proefschrift een consensusraamwerk ontwikkeld dat heterogene groepen DMs (experts of stakeholders) met verschillende verzamelingen van criteria die de bestaande groeps-BWM-modellen complementeert, die dezelfde verzameling criteria hanteert voor alle groepsleden. Daarnaast kan, vanwege het feit dat de gewichten van criteria van verschillende DMs van de non-lineaire BWM meestal bestaan uit intervallen, het aggregeren van die gewichten op traditionele manieren (d.w.z. via gemiddelden) de overlap van die intervallen over het hoofd kunnen zien. Het voorgestelde consensusmodel neemt het overlappen van de intervalgewichten in ogenschouw, teneinde de optimale overeenstemming tussen de groepsleden te realiseren.

Bereikgevoeligheid: De oorspronkelijke BWM neemt de bereiken van de criteria niet op een systematische manier in ogenschouw, wat kan leiden tot verschillende soorten bias door de resulterende bereikongevoeligheid. Wij combineren de "overweeg-het-tegenovergestelde-strategie" die inherent is aan de manier waarop in BWM voorkeuren worden verkregen met de principes van de Tradeoff-methode, in wat wij een Best-Worst Tradeoff (BWT) methode noemen, die het niet alleen mogelijk maakt om het bereik van de criteria mee te nemen, maar die tevens de mogelijke verankeringsbias bij het verkrijgen van oordelen kan verminderen. Bovendien stellen we een raamwerk voor het meten van (in)consistentie voor, zodat DMs die deze methode hanteren hun inconsistentieniveau kunnen zien, die wordt gebruikt als feedback om de consistentie van de oordelen te verbeteren.

Om te beoordelen hoe praktisch en haalbaar de voorgestelde methodes zijn, wordt de op overtuiging gebaseerde BWM toegepast op een evaluatie van grote infrastructuurprojecten in Indonesië, wordt het consensusmodel gebruikt om in Duitsland de optimale locatie voor een binnenlandse terminal te bepalen en wordt het BWT-raamwerk toegepast in de evaluatie van een haven in Nederland. De (in)consistentiedrempels die we in dit proefschrift hebben ontwikkeld zijn inmiddels al gebruikt door vele andere wetenschappers en ze worden een standaard onderdeel van elke BWM-toepassing.

Samengevat zijn de belangrijkste bijdragen van dit proefschrift: (i) het ontwikkelen van een model-onafhankelijke methode om de (in)consistentie van oordelen die door DMs worden verstrekt te meten en het instellen van drempels om DMs te helpen te beoordelen in welke mate de resulterende consistentieratio's acceptabel zijn, (ii) het introduceren van een overtuigingsstructuur in BWM om de onzekerheid of aarzeling van de oordelen van DMs mee te nemen en het voorstellen van een model van de betrouwbaarheid van de DMs te controleren, (iii) het bieden van de mogelijkheid om verschillende verzamelingen criteria in groeps-BWM te gebruiken en het bouwen van een model om DMs te helpen met gewichtsintervallen een consensus te bereiken, en (iv) het ontwikkelen van een MCDM-methode, BWT, om de voordelen van de traditionele BWM en de Tradeoff-methode aan elkaar te koppelen op een manier die de bereiksgevoeligheid van criteria meeneemt en vervorming of bias tegengaat. Al met al draagt dit proefschrift op een significante manier bij aan de verdere verspreiding van de BWM.

About the author

Fuqi Liang was born in Guangdong, in December 1989. He studied economic commerce at Shenzhen University and graduated in 2013 with a bachelor degree. After that he continued his master in the same school and majored in management science and engineering. From 2017 to 2021, he was a Ph.D. student at Delft University of Technology's Section of Transport & Logistics, under the supervision of Dr. Jafar Rezaei and Dr. Matteo Brunelli. His current research interests focus on modelling human decision-making.

List of Publications

Liang, F., Verhoeven, K., Brunelli, M. & Rezaei, J. (2021). Inland terminal location selection using the multi-stakeholder best-worst method. *International Journal of Logistics Research and Applications*, 1-23.

Liang, F., Brunelli, M., Septian, K. & Rezaei, J. (2021). Belief-Based Best Worst Method. *International Journal of Information Technology & Decision Making*, 20(01), 287-320.

Liang, F., Brunelli, M. & Rezaei, J. (2020). Consistency issues in the best worst method: Measurements and thresholds. *Omega*, 96, 102175.

Qin, Q., Liang, F.*, Li, L. & Wei, Y. (2017). Selection of energy performance contracting business models: A behavioral decision-making approach. *Renewable and Sustainable Energy Reviews*, 72, 422-433.

Qin, Q., Liang, F., Li, L., Chen, Y. & Yu, G. (2017). A TODIM-based multi-criteria group decision making with triangular intuitionistic fuzzy numbers. *Applied Soft Computing*, 55, 93-107.

Liang, F., Brunelli, M. & Rezaei, J. Best-Worst Tradeoff Method. (*Information Sciences,* Under review, 2nd round)

Liang, F., Qin, Q., Li, L., & Chen, Y. W. TODIM-based consensus model for group decision making with triangular intuitionistic fuzzy environment. (Under review)

Liang, F., Beukema, K. Spatial-placement of infrastructure for drone parcel delivery: A case study in the Netherlands. (Working paper)

“Everything should be made as simple as possible, but not simpler.”

—— Albert Einstein ——

ISBN: 978-94-6366-484-4