UNIVERSITAT
RAMON LLULL

Preference Disaggregation: Towards an Integrated Framework

Mohammad Ghaderi

http://hdl.handle.net/10803/404257

C.I.F. G: 59069740 Universitat Ramon Llull Fundació Rgtre. Fund. Generalitat de Catalunya núm. 472 (28-02-90)

UNIVERSITAT RAMON LLULL

DOCTORAL THESIS

Title	Preference Disaggregation: Towards an Integrated Framework
Presented by	Mohammad Ghaderi
Centre	ESADE Business School
Department	Marketing Management, Operations Management and Innovation and Information Systems Management
Directed by	Dr. Núria Agell Dr. Francisco Ruiz

To the memory of Dr. Eric Jacquet-Lagrèze (1947-2017)

Acknowledgments

First and foremost I must acknowledge my supervisor, Dr. Núria Agell, for the extensive support that she generously provided me during these years. You have not only been a great encouraging and knowledgeable academic supervisor, but also a wonderful friend that I could always come to whenever I needed to talk about anything. I am grateful for all the trust that you put in my capabilities, and always inspiring me to follow my ambitions. They say, PhD life is a heavily stressful one; I would like to add: unless you have a wonderful supervisor.

I would like to express my deep appreciation to my co-supervisor, Dr. Francisco Ruiz for his academic supports and guidances. I am grateful to the enormous support from Pilar Gállego, ESADE PhD programme manager, and Cristina Costa from Ramon Llull University. Your kind support and constant availability for helping me with even my non-academic paperworks has been so heartwarming. I am thankful to Professors Juan Manuel Batista, Jatinder Singh, Ioana Schiopu, Cecilio Angulo, Skander Esseghaier, Xari Rovira, and everyone whom I had the chance to work closely with during these years at ESADE. I am thankful to Professor Roman Słowiński for his warm hospitality during my stay at Laboratory of Intelligent Decision Support Systems at Poznan, and to Milosz Kadziński for his time and helpful comments on my works.

My special thanks to my family for their emotional support and encouragement. Thank you to my friends Alexander, Mehdi, Germán, Stefan, Menna, Albert, Amer, Julia, Saadi, Joonho, Solange, and all my colleagues from PhD programme. I acknowledge the financial support received from the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya and by European Social Funds (2014FI_B - 00637, 2015FI_B1 - 00105, 2016FI_B2 - 00058).

Thank you to my students. You taught me that high expectations and powerful relationships co-exist. You helped me understand that teaching is about more than great lesson planning and content. Teaching is about consistency and purpose ... and fun!

Thank you all.

Abstract

Preference disaggregation aims at capturing preference models by decomposing indirect preference information that are in form of holistic choices or judgments. From a multiple criteria decision aiding perspective, such information is taken as input to an inference procedure that yields to a preference model based on all the, usually conflicting, points of view that together form a basis for the judgments. Studying human judgments and choices has received increasing attention in the last few years from several disciplines, including behavioral science (decision analysis, preference disaggregation), artificial intelligence (preference learning), and economics and marketing (choice modeling). The three streams, although originated from different philosophies, are converging rapidly to a comprehensive understanding of human preferences, that is the main element of decisions and actions. This doctoral dissertation sheds light on this phenomenon by introducing an integrated analytical framework that allows capturing preferences of a complex form by observing holistic choices, decisions, and judgments.

Contents

List of Figures

List of Tables

As soon as questions of will or decision or reason or choice of
action arise, human science is at a loss.

Noam Chomsky

1

Introduction

Alice is passionate about art and sun. She recently spent a week in Barcelona to enjoy the great art museums and warm sunny days that the city has to offer. During her trip, she stayed in a hotel near the beach. At the end of her trip, as she usually does, Alice writes online reviews about some of the places she has visited and the hotel that she was staying in. In her overall review of the hotel, which is the number of stars that she gives to the hotel, Alice considers several factors such as comfort and cleanness, ease of access to the central area, price, the staff friendliness, as well as her overall feelings and impression of the hotel. Her overall assessment of

the hotel can change by some, probably less visible, factors such as the music that she has heard in the lobby, the decoration and colors of the hotel, the ambiance of the restaurant in hotel, and several other factors that Alice may or may not be aware of, but they change her feelings about the hotel in a direct or indirect way. Alice evaluates other services and products in a similar manner. Alice is not the only person whose decisions and judgments are based on multiple evaluation criteria, but the same is true for all consumers in the marketplace.

When facing a decision situation such as choosing a product or a service amongst all the other available alternatives, consumers typically consider several criteria to evaluate the existing alternatives and to eventually make a final decision. Some of these criteria are visible to the consumer who directly takes them into account, whereas some others can operate in an indirect way and at a subconscious level. Whether direct or not, both types of criteria form the preferences of the consumer which result in her final decision. Decomposing the holistic preferences based on the evaluation criteria and understanding how each of these criteria contributes to the final decisions and choices made by the Decision Maker (DM) is critical in marketing and consumer behavior studies. This is the focus in preference disaggregation paradigm based on the Multiple Criteria Decision Aiding (MCDA) perspective.

1.1 Objective of the Thesis

Modeling of consumer preferences among multiattribute alternatives has been one of the major activities in consumer research (Green and Srinivasan 1978, 1990; Gustafsson et al. 2013; Rao et al. 2014). The main objective of this dissertation research is to develop techniques to unfold the choices made by consumers in order to capture their preferences and predict their choices. Such techniques enhance our understanding of consumers preferences as the basis of their decisions, as well as the decision making strategies that they pursue to make a choice. This

dissertation is broadly motivated by two prevalent trends: developing efficient methodologies for processing preference information of realistic size, and the increasing need in marketing and branding contexts for analytical frameworks that enable us to explore the cognitive process of decision making, to elicit complex preferences, and to identify the main factors influencing consumers' choices and the mechanism under which they operate to shape consumers' preferences. This thesis addresses these questions from a MCDA perspective, a subfield of behavioral operational research, and following a preference disaggregation approach.

1.2 Overarching Theoretical Framework

Research on understanding human preferences has received increasing attention in several disciplines, including psychology (Kahneman et al. 1982; Keeney and Raiffa 1993; Tversky and Kahneman 1985), economics and marketing (Ben-Akiva and Lerman 1985; Rossi and Allenby 2003; Train 2003), artificial intelligence (De Gemmis et al. 2009; Fürnkranz and Hüllermeier 2010), and behavioral operational research (Figueira et al. 2016). These paradigms, however, view "preferences" from a different perspective, employ different approaches in analyzing them, and seek for different implications of such understanding. According to Lichtenstein and Slovic, preferences can be conceived of as an individual's attitude towards a set of objects, typically reflected in an explicit decision-making process (Lichtenstein and Slovic 2006). In economics and marketing, however, preferences are viewed slightly different than the psychological perspective by Lichtenstein and Slovic. From this point of view, preferences reflect satisfaction, enjoyment, or utility that is gained by some alternatives/items, and form a basis for judgments and choice. Studying preferences in economics and marketing mainly addresses analysis of demand and consumer's willingness to pay, product development, brand preference studies, and quality improvements in multiple dimensions. The implications in artificial intelligence, however,

direct towards recommender systems and human-computer interactions. From a behavioral research perspective, understanding preferences are of interest for decision support systems and decision aiding practices. The three paradigms, although differ in their theoretical approach, prevalent contexts, methodologies (statistical modeling versus mathematical programming versus machine learning techniques), even terminology (decision maker versus consumer versus user or agent), but are converging rapidly to gain a comprehensive understanding of human judgments and choices.

The assessment of alternatives in real-world decision making requires consideration of a variety of criteria. With an indirect elicitation of preferences, the preference model is constructed in order to identify a rational basis underlying the provided holistic judgments (Jacquet-Lagrèze and Siskos 2001). Within MCDA, various preference models can be used for this purpose, see, e.g., (de Almeida 2007; Kadziński et al. 2015; Sarabando and Dias 2010; Słowiński et al. 2002; Soylu 2011; Vetschera and de Almeida 2012; Zheng et al. 2014), but the most prevailing one is Multiattribute Value Theory (MAVT) (Keeney and Raiffa 1993). MAVT is built on utility theory which was developed during the 1940s and 1950s by Von Neuman and Morgenstern (Von Neumann and Morgenstern 1944) and Savage (Savage Leonard 1954). In general, utility is a real number representing the preferability of an item. Utility theory, although initially developed for normative decision making, has been one of the main methodological streams of multiple criteria analysis (Zopounidis and Pardalos 2010). MAVT is often found to be appropriate for practical decision support due to the high interpretability of numerical scores that can be decomposed into per-criterion marginal values and easily explained due to low amount of inter-criteria parameters (Greco et al. 2012; Sarabando and Dias 2010).

In decision making involving multiple criteria, the main attempt concerns the way by which the final decision should be made. In this spirit, several aggregation procedures have been developed to identify the *best* alternative amongst the available solutions, by simultaneously

4

considering several conflicting points of view. Goal programming (Charnes and Cooper 1961) has largely impacted the philosophy and evolution of this paradigm. A comprehensive taxonomy of such procedures is provided in Hwang and Yoon (2012). These procedures mainly follow a normative approach that is based on the axioms of rationality and optimality, assuming that the decision behavior is purposive and goal directed (Einhorn and Hogarth 1981). Following this view, therefore, the aim is to find the decisions that *must* be made in order to ensure the efficient use of means to attain certain ends (Fishburn 1970; Tversky 1972; Von Neumann and Morgenstern 1944). When such goals conflict, the notion of optimality does not make much sense as the problem is not maximizing or minimizing an explicit criterion, but it seeks for compromises between the conflicting goals that reflect one's values. Therefore, in presence of multiple conflicting criteria, the notion of *optimality* is replaced by *consistency* with one's values (Einhorn and Hogarth 1981).

The main interest in many decision making situations, however, is in the opposite direction (Jacquet-Lagrèze and Siskos 2001). Given the final decisions made by the DM, how is it possible to find the rational basis through which the decisions are made? The actual behavior of a DM often involves several cognitive biases and limitations, henceforth deviating from predictions derived from normative approaches (Simon 1957). Descriptive theories of human judgments and choices are developed along this direction by analyzing the decisions that are already made by the DM in order to discover, understand, and explain the rational behind them (Kahneman and Tversky 1984; Slovic et al. 1977; Tversky and Kahneman 1975; Von Winterfeldt and Edwards 1993). According to Tsoukiàs, the main difference between normative and descriptive approaches is the sources that they derive their model of rationality from. According to this view, normative approaches derive rationality models from the norms that are established a priori (e.g. maximizing expected utility), whereas source of rationality model for the descriptive approaches is by observing how DMs make decisions (Tsoukiàs 2007). Dis-

aggregation techniques in MCDA aim at understanding judgments and choices of a DM by decomposing the holistic preferences over the set of relevant criteria that form the basis for the preferences.

Following MAVT perspective in MCDA, preferences of a DM are captured by constructing an additive value function. Additive value functions represent the preferences of the DM by computing an overall score for each alternative (Vetschera et al. 2014). This comprehensive value indicates the alternative's quality from all relevant points of view considered jointly. Following this perspective, the most representative preference disaggregation approach in MCDA is the family of UTA methods (Jacquet-Lagrèze and Siskos 1982).

1.2.1 State of the art

Multiple criteria decision aiding is a rapidly growing field of study which has increasingly attracted attention in recent years from various fields of applications. The main goal of MCDA is to achieve a set of criteria aggregation procedures that enables the development of decision support models considering preferences and evaluations of the DM (Doumpos and Zopounidis 2013). A criterion can be viewed as a tool which allows one to represent consequences related to a point of view so as to be able to establish partial preferences (Bouyssou 1990). Tsoukiàs formally defines a criterion as a preference relation, that is a binary relation, on the set of alternatives, or a function representing the criterion (Tsoukiàs 2007). In fact, a criterion is a function that measures performances of the alternatives on each of their characteristics (Doumpos and Zopounidis 2013). For a decision aiding purpose, the family of criteria is needed to be consistent, which means i)it needs to address every important point of view (exhaustive), ii)the partial preferences have to be consistent with the comprehensive preference, i.e. if one alternative is preferred over another according to all points of view, the same should

hold at the comprehensive level, and finally iii)the criteria should not be redundant (Bouyssou 1990).

Analytical frameworks based on MCDA, being mathematically sound and practically flexible, have provided powerful toolkits to effectively tackle complex decision problems of various kinds. These frameworks have been successfully applied in different contexts such as finance, sustainability, health care, urban design, customer satisfaction measurement, supply chain management, and marketing (Angilella et al. 2016; Bisdorff et al. 2015; Doumpos et al. 2015; Grigoroudis and Siskos 2002, 2009; Kadziński et al. 2016; Spronk et al. 2016).

Following the framework suggested by (Pardalos et al. 2013), four main streams can be distinguished in MCDA research:

- multiobjective mathematical programming,

- multiattribute utility/value theory,

- outranking relations approach,

- preference disaggregation.

The aim of multiobjective mathematical programming is to discover the set of non-dominated solutions/alternatives - usually a set of large size - in presence of conflicting objectives/criteria (Ehrgott 2006; Koopmans et al. 1951). Multiattribute utility/value theory attempts to model the preferences of a DM by means of a value function (certain environment) or a utility function (decision under uncertainty) (Keeney and Raiffa 1993). The principles of outranking relations have been established by Bernard Roy by introducing the family of ELECTRE methods. An outranking relation, according to Roy, is a binary relation such that alternative a outranks alternative b if there are sufficient arguments in favor of this assertion, while there are not many arguments against it. Outranking relations are used to model the preferences of the DM by

constructing preference, indifference, and incomparability relations (Figueira et al. 2005; Roy 1968, 1991). Finally, preference disaggregation, the main focus of this dissertation, aims at inferring preference models from indirect preference information. This approach is discussed in more details in the next section.

1.2.2 Preference Disaggregation

The disaggregation paradigm in MCDA formally started with the introduction of the UTA method in 1982 (Jacquet-Lagrèze and Siskos 1982). Preference disaggregation, the focus of this dissertation, aims to identify the aggregation model behind the indirect and holistic preference information from the DM (Jacquet-Lagrèze and Siskos 1982). Additive value function is the most common form of preference model used in disaggregation analysis. Ordinal regression techniques are employed to infer a value function, or a set of value functions in case of robust ordinal regressions (Greco et al. 2008), compatible with the decision examples provided by the DM. In this regards, decision examples can take different forms such as ordering of reference alternatives, pairwise comparisons, assigning alternatives to a set of predefined ordered categories, and choice of the best alternatives. Moreover, several other types of preference information, such as ordering of the evaluation criteria, intensities of preferences, or desired ranks of some alternatives, can be incorporated in disaggregation analysis to enrich the inference procedure.

Disaggregation approach is of interest because it requires a low level of cognitive effort and interaction time from the DM. With its flexibility in processing several types of preference information, its ability to address different types of decision problems such as ranking, sorting, and choice, and its data driven approach that yields generic outcomes without relying on prior assumptions about the DM's shape of preferences, preference disaggregation can be a powerful approach in marketing to study consumers behavior.

1.3 Main Research Problems and Contributions

This chapter presents the main motivations of this dissertation, as well as the main contributions to the addressed research problems.

Most of the existing preference disaggregation methods assume a monotonic relationship between input attributes and preferences. The few methods currently addressing non-monotonicity demand excessive computational effort or require extra information from the DM. Developing an efficient methodology for learning non-monotonic preference models is therefore of great importance.

The assumption of monotonicity is widely used because it seems reasonable for many criteria (such as price of a product, level of risk, security, safety and ease of use of a service, required time and effort for accomplishing a task). However, this is not the case for many other attributes. In a medical context, for instance, attributes such as sugar and cholesterol levels in blood are clear examples of non-monotonic attributes. In finance, there are many indicators used to depict the financial performance of a firm that experts believe must be controlled within a specific range. For instance, a large value for the *cash to total assets ratio* implies that the firm is losing many profitable investment opportunities, whereas a low value indicates a low capacity of the firm to cope with operating expenditures (Despotis and Zopounidis 1995). In marketing, non-monotonic preferences are widely prevalent in product design. Examples include screen size for a smartphone, sweetness of a chocolate, or any design attribute based on ideal point model (Rao 2014).

This dissertation mainly addresses the following three problems:

- As a motivating example, the thesis elaborates on the case of non-monotonic preferences in the context of branding by investigating the impact of brand color on customers perception of a brand (Ghaderi et al. 2015). In this regard, the main question is **how**

can MCDA be used for exploring the cognitive role that color plays in determining brand perception? This question will be addressed by proposing an analytical framework and analyzing the secondary survey data combined with the data that is independently collected by measuring the colors of the well-known beauty and care brands that have a single dominating color (**chapter 2**).

- Inspired by the example of brand color, the second research problem is **how can we capture preferences of non-monotonic form?** The assumption of monotonicity limits the applicability of the MCDA methods in many fields, therefore it is crucial to overcome this limitation. The analytical framework to elicit such complex preference models still needs to be tractable and efficient in terms of computational demand, in order to ensure its applicability to data sets of realistic size. These requirements inspired the research presented in **chapter 3**. The problem is addressed by developing an analytical framework based on a Linear Programming (LP) approach. The framework is discussed through an illustrative example, a comparison with some benchmark methods, and a comprehensive experimental analysis (Ghaderi et al. 2017).

- The preference disaggregation frameworks in MCDA, including the one introduced in this dissertation, involves the choice of some parameters which might affect the performance of the inference process. The quality of the outcomes of a disaggregation analysis, in addition to the parametrization of the problem, could depend also on the settings of the decision problem. In this respect, the next questions addressed in this dissertation is **how and to what extent the parametrization procedures and problem settings can change outcome quality of the inference model? Which procedures should be followed in different settings in order to ensure a desired level of quality?** To address these questions, **chapter 4** will argue how expressiveness of the inferred prefer-

ence model, as well as the robustness of the recommendations resulting from the derived model (the two main dimensions of the outcome quality), are influenced by different characteristics of a decision problem and by different procedures for selection of the characteristic points on the scale of evaluation criteria (Kadziński et al. 2017).

1.4 Structure of the Thesis

This thesis is structured as a compendium of publications, organized as follows.

Chapter 2 addresses the challenging question in branding literature about how colors can shape customer perceptions of a brand. It will be demonstrated how the problem can be formulated by employing a MCDA framework. By introducing an analytical framework, a real data set of customers perceptions of dozens of cosmetics brands are analyzed. The results demonstrate a strong connection between perceptions of a brand and its color. It will also be shown how this analytical framework can be employed to create an intuitive perception-color map in order to help brand managers to better position their brands in customers minds. To this aim, forty main attributes of a brand image are mapped on a color space, based on the findings of the analysis.

Chapter 3, inspired by chapter 2, introduces a framework for inferring non-monotonic additive preference models from a set of indirect pairwise comparisons. Moreover, the applicability and effectiveness of the proposed methodology is demonstrated through an extensive experimental analysis covering a broad class of decision problem settings.

Chapter 4 investigates the impact of the parametrization of a decision problem, as well as the settings of a decision problem, on the outcome quality of the elicitation process. The quality of the outcomes is defined based on the expressiveness of a preference model derived from the supplied indirect preference information, and robustness of the recommendations derived from

the inferred preference model. Amongst others, particularly special attention will be given to the impact of various procedures for selection of characteristic points, which define shape of marginal value functions in a preference model. In the thesis, a new procedure for selection of characteristic points will be introduced. The results from experimental analysis confirm that the introduced procedure demonstrates overall better performance than the existing ones, and its performance improves when more preference information is becoming available.

Chapter 5 presents a general discussion by addressing theoretical, as well as practical implications, and concludes the thesis. Finally, the references of the thesis are provided in the end.

1.5 MAIN OUTPUTS OF THE THESIS

The three central chapters are already published in Pattern Recognition Letters, European Journal of Operational Research, and Computers & Operations Research journals. An overview of the three central chapters is provided in Table 1.5.1. The three papers have been presented in several conferences at different stages of their development and have evolved based on the constructive feedbacks received at each presentation. The list of conferences is provided in Table 1.5.2.

Table 1.5.1: This table shows the central chapters of the thesis and the journals they are published in.

Chapter	Article Title	Authors	Journal	Publishing Date	Journal Metrics
2	Understanding the Impact of brand color on brand image: a preference disaggregation approach. A linear programming approach	Ghaderi, M., Ruiz, F., Agell, N.	Pattern Recognition Letters	Volume 67, issue 1, pages 107-115, 2015	ABS 3 SJR Q1, Impact Factor: 1.586
3	Learning non-monotonic additive value functions in multiple criteria decision aiding.	Ghaderi, M., Ruiz, F., Agell, N.	European Journal of Operational Research	Volume 259, issue 3, pages 1073-1084, 2017	ABS 4, SJR Q1, Impact Factor: 2.679
4	Expressiveness and robustness measures for the evaluation of an additive value function in multiple criteria preference disaggregation methods: an experimental analysis.	Kadziński, M., Ghaderi, M., Wąsikowski, J., & Agell, N.	Computers & Operations Research	Accepted, May 2017	ABS 3, SJR Q1, Impact Factor 1.988

Table 1.5.2: This table shows the conferences in which elements of this dissertation have been presented.

Conference Name	Date	Location
85th Meeting of EURO Working Group on Multiple Criteria Decision Aiding (**EWG-MCDA 85**)	April 20-22, 2017	Padova, Italy
28th European Conference on Operational Research {(EURO)}	July 3-6, 2016	Poznan, Poland
82nd Meeting of EURO Working Group on Multiple Criteria Decision Aiding (**EWG-MCDA 82**)	September 24-26, 2015	Odense, Denmark
23rd International Conference on Multiple Criteria Decision Making (**MCDM 23**)	August 2-7, 2015	Hamburg, Germany
1st International Conference on Decision Support System Technology: Big Data Analytics For Decision Making (**ICDSST**)	May 27-29, 2015	Belgrade, Serbia
81st Meeting of EURO Working Group on Multiple Criteria Decision Aiding (**EWG-MCDA 81**)	March 26-28, 2015	Annecy, France
The 17th International Conference of the Catalan Association of Artificial Intelligence (**CCIA**)	October 22-24	Barcelona, Spain
22nd International Conference on Multiple Criteria Decision Making (**MCDM 22**)	June 17-21	Malaga, Spain

2

Understanding the Impact of Brand Colour on Brand Image: a Preference Disaggregation Approach[1]

What is the role that colour plays in perception of a brand by customers? How can we explore the cognitive role that colour plays in determining brand perception? To answer these questions we propose a preference disaggregation method based on *multi-criteria decision aid*. We identify the criteria aggregation model that underlies the global preference of a brand with

respect to each brand image attribute. The proposed method is inspired by the well-known UTASTAR algorithm, but unlike the original formulation, it represents preferences by means of non-monotonic value functions. The method is applied to a database of brands ranked on each brand image attribute. For each brand image attribute, non-monotonic marginal value functions from each component of the brand colour are obtained separately. These functions contain the fitness between each colour component and each brand image attribute, in an understandable manner.

2.1 INTRODUCTION

Colour is one of the key ingredients of brands which plays an important role in the purchase decisions of customers. As an aesthetic stimuli, colour can shape consumer preferences and alter perceptions by communicating meaningful messages (Zaltner 1975). As an essential element of a brand, colour can signal quality (Lohse and Rosen 2001), affect perception of quality (Chebat and Morrin 2007), contribute to brand recognition and brand image (Lightfoot and Gerstman 1998), and affect brand personality (Labrecque and Milne 2012). Colour, in addition, intrinsicly communicates the desired image (Bottomley and Doyle 2006) and is considered a strategic tool for marketers and brand managers for differentiating brands from competitors, signalling product attributes, and grabbing customer attention (Schindler 1986).

Colour operates via two mechanisms: sensory and cognitive. In the sensory mechanism, colour helps retrieve information in blurry conditions, by distinguishing, for example, an object from its background. In the cognitive mechanism, colour helps perception by playing a diagnostic role and characterising the object that is being represented (an orange sunset and the blue of the sea have specific meaning). As brand image is characterised by the perception of the

customers, brand colour influences brand image through the cognitive mechanism (Wu et al. 2014).

The importance of colour to the marketers is not limited to brand colours. Studies support a significant impact of packaging colours on customer intention to buy and perceived quality. Hoegg and Alba found that colour cues dominate taste cues. In their experiment using orange juice, participants perceived a significantly greater difference in the taste of two identical samples with different colours, than two different samples with the same colour (Hoegg and Alba 2007). Garber et al, in their experimental study, found that colour affects identification and flavour perception of both congruently and incongruently coloured beverages (Garber et al. 2000).

The common practice for understanding colour trends in industry is based on the opinion of field experts, whose judgements are based on past experiences and are difficult to substitute by analytic models. In this paper, we explore the relationship between brand colour and customer perception of brand image in an understandable and interpretable manner. To this end, we propose a preference disaggregation method based on a multi-criteria decision analysis (MCDA) framework. The aim of this approach is to analyse the holistic preferences of a set of alternatives in a multi-criteria setting in order to identify the criteria aggregation model that underlies global preferences, and represent the existing preferential system using a set of marginal value functions.

To address this paradigm, several methods have been proposed in the literature considering different forms of comprehensive preferences and various tasks, for instance UTA (UTilités Additives) (Jacquet-Lagrèze and Siskos 1982), Pairwise comparisons UTA (Doumpos 2012), UTADIS (UTilités Additives DIScriminantes) (Devaut et al. 1980), fuzzy UTASTAR (Patiniotakis et al. 2011) and many others (Doumpos 2012; Kliegr 2009). Most of these methods assume a monotonic relationship between preferences and attribute levels. However, as the

relationship between brand perception and colour attributes, for example colour hue, is not necessarily monotone, we introduce a new method based on UTASTAR, that is applicable in non-monotonic settings.

The paper is organised as follows. A brief introduction on colour measurement is provided in the next section. An overview of the theoretical framework of preference disaggregation is then presented. Because we are focusing on the ranking problem, the section contains a review of the most widely used UTA variant, UTASTAR, and some of the non-monotonic UTA-like methodologies for the ranking problem. In Section 2.4, the proposed methodology is introduced, followed by an illustrative example to make a comparison with the UTA-NM method. In Section 2.5 the method is applied to a comprehensive set of brand image attributes, in order to explore the impact of brand colour on brand image. Finally, we conclude the paper and present possible future directions.

2.2   Colour coordinates and colour spaces

Several numeric specifications for colour definition can be found in the literature. We refer the interested reader to the recent study in (Falomir et al. 2015). The most classic and internationally accepted of these are based on tristimulus values or coordinates. The most known of these is RGB, proposed by the *Commission International de l'Eclairage* (CIE) in 1931. RGB uses additive colour mixing and describes what type of light (red, green or blue) needs to be emitted to produce a given colour. The RGB colour model is implemented in different ways, depending on the capabilities of the system used. By far the most common is the 24-bit implementation. This model is thus limited to a range of $256 \times 256 \times 256 \approx 16.7$ million colours. It is a convenient colour model for computer graphics, but it can be unintuitive in practice. The specification of a desired colour can be difficult for untrained people (for example, selecting brown using an RGB

vector can be difficult). HSV is another colour space which was developed to approximate the way humans perceive colours. For this reason, in marketing studies HSV colour space is widely used. In this single-hexcone model of colour space, hue (H) of a colour refers the pure colour it resembles and demonstrates its position on the colour wheel, where it starts from 0 for red, and continues to 60 for yellow, 120 for green, and ends up at 360 or the starting position. Saturation (S) refers to the intensity of the pure colour. In other words, it describes the purity of the colour with respect to white. The value of 100 means a very vivid colour, while 0 means the least purity, where too much white dominates the colour. Value (V) measures the brightness of the colour where 100 means a totally bright and 0 means a totally dark colour. Most colour researchers in marketing focus only on colour hue and usually do not consider the other two attributes. Geometrical representation of the two colour systems RGB and HSV is presented in Fig. 2.2.1.
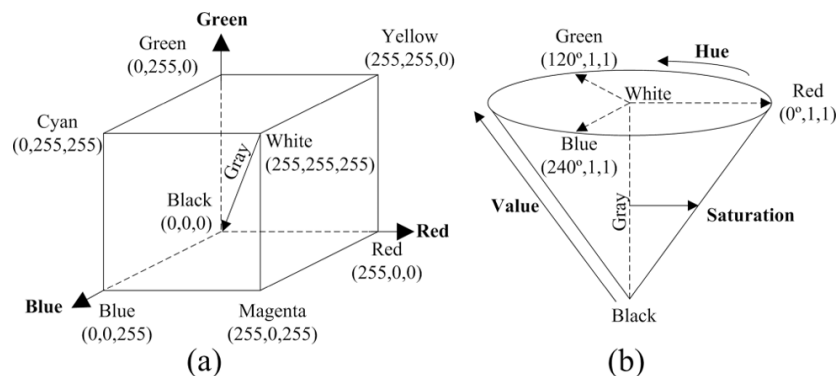
Figure 2.2.1: RGB (a) and HSV (b) coordinates

2.3 PREFERENCE DISAGGREGATION METHODOLOGIES

UTA (Utilités Additives) is one of the most representative preference disaggregation methods. It was first introduced by Jacquet-Lagrèze and Siskos as a linear programming (LP) model to capture the preferential system of the decision maker (DM) through nonlinear (piecewise linear) monotonic additive value functions (Jacquet-Lagrèze and Siskos 1982). The aim of

the UTA method is to reproduce, through a set of value functions, the ranking made by the DM over the set of alternatives by minimising the level of ranking errors. Ranking errors are generally defined as the distance between the global values of two consecutive alternatives that are ranked incorrectly. However, the definition of the error slightly differs in the variants of UTA. The method leads to a simple LP model where the optimal solution can be easily obtained.

Several extensions of UTA method have been introduced in the MCDA literature since then, incorporating variations on the original algorithm and considering different forms of global preference and optimality criteria. In most of the extensions of UTA method, the input attributes are normally expected to be monotone with respect to the preferences. The assumption of monotonicity is widely used, and it seems reasonable for many criteria, such as price, risk level, security, safety, comfort, required time, and effort. However, this is not the case for many other attributes, such as colour coordinates. In this paper, we propose an extension for UTA method to handle non-monotone preferences suitable for addressing the problem of understanding the impact of brand colour on brand image. In the following subsections, we present the most representative UTA method for ranking (UTASTAR) and briefly introduce some variants of the method which attempts to consider non-monotonic attributes.

2.3.1 UTASTAR method

Suppose that $G = \{g_1, g_2, \cdots, g_m\}$ is a set of criteria to evaluate a set of preordered alternatives $A = \{a_1, a_2, \cdots, a_N\}$ in which a_1 is the most and a_N is the least preferred alternative in the ranking list. Each criterion is defined as a function $g_i : A \rightarrow \mathbb{R}$, where $g_i(a_n) = x_i^n$. The value x_i^n is the performance of the alternative a_n over the criterion g_i. Given a weak ordering (ranking) over the set of alternatives specified by the DM, the aim of the UTASTAR algorithm is to represent the underlying preference model of the given ranking through estimating a set of

monotonic additive value functions (as consistent as possible with the preferential structure of the DM). Specifically, the UTA method estimates a set of marginal value functions $v_i : g_i \rightarrow [0, 1]$ to be aggregated in an additive manner in order to estimate the comprehensive value associated with each alternative. Finally, alternatives are ranked based on the comprehensive values. The formulation of the UTASTAR method involves defining a_i characteristic points and henceforth $a_i - 1$ subintervals $[g_i^0, g_i^1], [g_i^1, g_i^2], \cdots, [g_i^{a_i-2}, g_i^{a_i-1}]$ on the ith criterion, in which g_i^0 and $g_i^{a_i-1}$ are the minimum and maximum performance levels over the ith scale, respectively. The marginal value at a characteristic point g_i^l on criterion i is expressed as in equation (2.1).

$$v_i(g_i^l) = \sum_{j=1}^{l} (v_i(g_i^j) - v_i(g_i^{j-1})) = \sum_{j=1}^{l} v_{ij} \tag{2.1}$$

where $v_{ij} \equiv v_i(g_i^j) - v_i(g_i^{j-1}) \geq 0$ due to the monotonicity of the criteria.

The marginal value for an alternative a_n whose performance on the ith scale is $x_i^n \in [g_i^l, g_i^{l+1}]$ is obtained by linear interpolation between $v_i(g_i^l)$ and $v_i(g_i^{l+1})$ as follows:

$$v_i(x_i^n) = \sum_{j=1}^{l} v_{ij} + \frac{x_i^n - g_i^l}{g_i^{l+1} - g_i^l} \cdot v_{i,l+1} \tag{2.2}$$

The comprehensive value of an alternative a_n is obtained by the sum of all the marginal values, as in equation (2.3).

$$V(a_n) = \sum_{i=1}^{m} v_i(x_i^n) \tag{2.3}$$

The UTASTAR linear programming problem is provided in (2.4).

$$min\ z = \sum_{n=1}^{N}(\sigma^+(a_n) + \sigma^-(a_n))$$

subject to

$$V'(a_n) - V'(a_{n+1}) \geq \delta \text{ iff } a_n \succ a_{n+1}, \forall n = 1, 2, \cdots, N-1$$

$$V'(a_n) - V'(a_{n+1}) = 0 \text{ iff } a_n \sim a_{n+1}, \forall n = 1, 2, \cdots, N-1 \quad\quad (2.4)$$

$$\sum_{i=1}^{m} \sum_{j=1}^{a_i - 1} v_{ij} = 1$$

$$V'(a_n) = V(a_n) - \sigma^+(a_n) + \sigma^-(a_n)$$

$$v_{ij}, \sigma^+(a_n), \sigma^-(a_n) \geq 0, \forall i, j, n$$

in which $\sigma^+(a_n)$ and $\sigma^-(a_n)$ are the overestimation and underestimation error terms, respectively. The term δ is a parameter (a small value), and the first two constraints represent the preorder relations provided by the DM. The third constraint ensures that the maximal shares of the criteria in the comprehensive value of the alternatives sum up to 1, and the objective function minimises the deviation of the estimated value function from the preferential model of the DM. By solving this model, the marginal value function over each criterion scale will be constructed based on the expression in (2.1).

2.3.2 NON-MONOTONIC UTA-LIKE ALGORITHMS

The input attributes in the UTASTAR method are normally expected to be monotone with respect to the preferences. However, this is not a reasonable requirement for colourimetric components. Obviously, no one can expect a monotonic relationship between a colour preference degree and its degree of greenness, or hue. Therefore, an improvement in the UTASTAR algorithm for handling non-monotonic preferences is of a great importance in this setting.

Although several attempts have been made in the literature to overcome the mentioned shortcoming (Despotis and Zopounidis 1995; Doumpos 2012; Eckhardt and Kliegr 2012;

Kliegr 2009), all are computationally intensive, or require extra information from the DM. One way to address non-monotone preferences is to divide the range of the criteria into intervals so that the preferences are monotonic in each interval, and then treat each interval separately. Following this idea, in the approach of Despotis and Zopounidis, it is assumed that each marginal value function is non-decreasing from the starting point of the range to a middle point, and it is non-increasing from this middle point to the end of the range (Despotis and Zopounidis 1995). This middle point corresponds to the most preferable value of the criterion. The main drawback of this method is that the exact value function shape and the most preferable value need to be known beforehand. Kliegr proposed another non-monotone methodology called UTA-NM, which relaxes the monotonicity condition of the UTASTAR algorithm, that in theory allows any shape for the marginal value function (Kliegr 2009). To avoid the over-fitting problem, UTA-NM simultaneously minimise the sum of the errors and the complexity of the model expressed by the number of changes in the sign of the marginal value functions. The method suffers from severe performance issues. Even for very small toy problems, tens of binary variables were involved, causing the method to be computationally infeasible for real-world problems. In another paper, Eckhardt and Kliegr propose *local preferences transformation*, a heuristic attribute preprocessing algorithm that transforms arbitrary input attributes into a space approximately monotone with respect to user preferences, thus making it suitable for UTA (Eckhardt and Kliegr 2012). Finally, non-monotonic additive value functions, introduced by Doumpos in 2012 (Doumpos 2012), consider a broader class of non-monotonic value functions that leads to a nonlinear integer programming problem, which is difficult to solve with data sets of realistic size. Thus, an evolutionary approach is employed, based on the differential evolution algorithm.

2.4 Proposed methodology

The method we introduce here, inspired by the UTA methodology, is fast and tractable. The general idea is to relax sign constraint in the decision variables that represent difference of value levels between two consecutive breakpoints. Therefore, marginal value function can change the monotonicity at any breakpoint.

This may lead to two problems: the first is the over-fitting problem in the case that monotonicity changes arbitrarily many times. This potential problem is handled simultaneously in two ways. Firstly, we defined a small, but reasonable, number of breakpoints. The breakpoints are constructed so that each sub-interval contains the same number of data points and hence the same amount of information. Secondly, the slope of the marginal value function in each sub-interval is controlled by defining upper and lower bounds for the associated decision variables. The bounds for each decision variable is defined with respect to the length of the corresponding sub-interval. The longer the sub-interval, the wider the bound. This constraint not only controls the over-fitting problem, but also increase interpretability of the extracted value functions.

The second problem is about normalisation. By normalisation, we mean that the minimum and maximum global values must be equal to zero and one, respectively. Fixing the minimum and maximum global values is essential for obtaining the relative importance of the criteria. The challenge is that we cannot predict where the maximum value will be achieved on each criterion scale in order to force the sum to be one. Furthermore, we do not know the attribute level corresponding to the minimum marginal value on each criterion to set them equal to zero. To solve this problem, an iterative approach is followed. Whenever the maximum global value is less than one, its value is forced to be increased in the next iteration, by adding a new constraint considering the performance level corresponding to the highest marginal value in the

current stage. The added constraint is applied in the next iteration, and will be removed from the LP model in the following iterations, because it does not have to be necessarily satisfied in the final solution. Whenever the maximum global utility is greater than one, a restrictive constraint is imposed to ensure that the global utility of the attribute levels corresponding to the highest marginal utility in the current stage will not have a value greater than one in all the following iterations. Furthermore, to satisfy another condition of normalisation (namely, minimum global utility being zero), a penalisation term is added to the objective function to penalise any deviation.

2.4.1 CHARACTERISTIC POINTS DEFINITION

Defining the breakpoints is an important step in all the UTA-like methodologies as it directly affects the number of decision variables. We define the breakpoints based on the idea of equal frequency intervals. This means that we expect equal numbers of distinct performance values in each sub-interval of the criterion, except for the two ends of the criterion scale. Let us denote by h_i the number of distinct performance levels of alternatives over the ith criterion, and by c_i the desired frequency in each sub-interval of the ith scale. It is easy to show that the number of decision variables corresponding to the ith criterion is equal to $[h_i/c_i] + 1$, in which $[x]$ is the largest integer number less than or equal to x. Considering that the number of distinct performance levels might be much higher for some criteria than others, defining the same value for all the c_i variables leads to associating many decision variables with the former, and few with the latter. This leads to over-fitting on the former criteria and inaccurate results on the latter, resulting in dramatically different degrees of freedom for the different estimated value functions. To overcome this issue, we suggest that c_i be a function of h_i, and we propose $c_i = \left[\sqrt{h_i}\right]$. Following this method, we expect that all the criteria will have almost the same degree of freedom.

2.4.2 Initial solution

Following UTASTAR notation, the marginal utilities for each criterion $i \in \{1, 2, \cdots, m\}$ are represented as in equation (2.1), except for a new type of decision variable that has been added, $v_{i,0}$, which allows any level of value within the range $[0, 1]$ for the lowest possible performance over the criterion scale. The marginal value at a breakpoint g_i^l on criterion i is expressed as:

$$v_i(g_i^l) = v_{i,0} + \sum_{j=1}^{l}(v_i(g_i^j) - v_i(g_i^{j-1})) = \sum_{j=0}^{l} v_{ij} \tag{2.5}$$

and the marginal value for an alternative a_n whose performance on the ith scale is $x_i^n \in [g_i^l, g_i^{l+1}]$ is obtained by linear interpolation between $v_i(g_i^l)$ and $v_i(g_i^{l+1})$, as follows:

$$v_i(x_i^n) = \sum_{j=0}^{l} v_{ij} + \frac{x_i^n - g_i^l}{g_i^{l+1} - g_i^l} \cdot v_{i,l+1} \tag{2.6}$$

The comprehensive value is obtained by the formula in (2.3). No normalisation constraint is imposed in the initial solution, and sign constraint over decision variables are relaxed. However, some constraints are imposed to obtain a solution as close as possible to the feasible solution. The first issue to be considered here is having a non-negative estimated marginal value over any characteristic point. Suppose that vector $V_i = (v_{i,0}, v_{i1}, \cdots, v_{i,a_i-1})$ demonstrates the decision variables corresponding to the marginal value of the ith criterion. The following set of

26

constraints then guarantees that the estimated marginal value at any point on a criterion scale is non-negative:

$$
\left.\begin{aligned}
& v_{i,0} \geq 0 \\
& v_{i,0} + v_{i_1} \geq 0 \\
& v_{i,0} + v_{i_1} + v_{i_2} \geq 0 \\
& \quad \cdots \\
& \quad \cdots \\
& \quad \cdots \\
& v_{i,0} + v_{i_1} + v_{i_2} + \cdots + v_{i,a_i-1} \geq 0
\end{aligned}\right\} \quad \text{for } i = 1, 2, \cdots, m \tag{2.7}
$$

The following set of constraints also guarantees that the estimated marginal value at any point on a criterion scale is less than 1:

$$
\left.\begin{aligned}
& v_{i,0} \leq 1 \\
& v_{i,0} + v_{i_1} \leq 1 \\
& v_{i,0} + v_{i_1} + v_{i_2} \leq 1 \\
& \quad \cdots \\
& \quad \cdots \\
& \quad \cdots \\
& v_{i,0} + v_{i_1} + v_{i_2} + \cdots + v_{i,a_i-1} \leq 1
\end{aligned}\right\} \quad \text{for } i = 1, 2, \cdots, m \tag{2.8}
$$

Note that (2.7) and (2.8) can be written in a more compact way using the $a_i \times a_i$ lower triangular matrix A_i with $a_{kp} = 1$ for elements where $k \geq p$. Then (2.7) and (2.8) can be written as:

$$
A_i V_i \geq 0, \forall i \tag{2.9}
$$

27

$$A_i V_i \leq 1, \forall i \tag{2.10}$$

It is important to bear in mind that the normalisation condition is not guaranteed in the initial solution because the maximum of the estimated comprehensive value is not necessarily equal to 1.

Finally, the following set of constraint limits the slope of the value function at any interval.

$$\left| \frac{v_{ij}}{g_i^j - g_i^{j-1}} \right| \leq \frac{1}{g_i^0 - g_i^{a_{i-1}}}, \forall i = 1, 2, \cdots, m, \forall j = 1, 2, \cdots, a_{i-1} \tag{2.11}$$

The linear format of the above constraints is presented as follows.

$$\left. \begin{array}{l} v_{ij} \leq \frac{g_i^j - g_i^{j-1}}{g_i^0 - g_i^{a_{i-1}}} \\[2mm] -v_{ij} \leq \frac{g_i^j - g_i^{j-1}}{g_i^0 - g_i^{a_{i-1}}} \end{array} \right\} \forall i = 1, 2, \cdots, m, \forall j = 1, 2, \cdots, a_{i-1} \tag{2.12}$$

The LP model of the initial solution is presented in (2.13).

$$min \; z = \sum_{n=1}^{N} \left(\sigma^+(a_n) + \sigma^-(a_n) \right)$$

subject to

$V'(a_n) - V'(a_{n+1}) \geq \delta \text{ iff } a_n \succ a_{n+1}, \forall n = 1, 2, \cdots, N-1$

$V'(a_n) - V'(a_{n+1}) = 0 \text{ iff } a_n \sim a_{n+1}, \forall n = 1, 2, \cdots, N-1$

set of constraints in (2.7)

set of constraints in (2.8)

set of constraints in (2.12)

$V'(a_n) = V(a_n) - \sigma^+(a_n) + \sigma^-(a_n)$

$v_{ij} \; URS, i = 1, 2, \cdots, m, j = 0, 1, \cdots, a_i - 1$

$\sigma^+(a_n), \sigma^-(a_n) \geq 0, n = 1, 2, \cdots, N$

$$\tag{2.13}$$

in which URS means UnRestricted in Sign variable.

In the outcome achieved by solving the model in (2.13), let the breakpoints with maximum and minimum marginal value on the ith criterion scale be g_i^* and g_{i*}, respectively.

$$g_i^* = \arg\max_j v_i(g_i^j) \tag{2.14}$$

$$g_{i*} = \arg\min_j v_i(g_i^j) \tag{2.15}$$

Furthermore, let us assume that f^* denotes the sum of the overestimation and underestimation errors in the optimal solution of the model in (2.13). By storing this information, the iterative part of the algorithm can be started as explained in the following section.

2.4.3   ITERATIVE PART

The missing piece in the aforementioned model is the normalisation to ensure that the maximum achievable comprehensive value is equal to 1. Because the comprehensive value is the sum of m marginal values and the maximum marginal value of each criterion might occur at any breakpoint of the criterion scale, $\prod_i a_i$ possible combinations of decision variables exist to constitute the maximum comprehensive value. The general idea is to detect the combinations of decision variables that have the potential to cause the maximum comprehensive value to exceed 1 and restrict them by adding a new constraint. Another possibility is that the maximum affordable comprehensive value is less than 1. In this case, we impose a new constraint to enforce an increase in the maximum comprehensive value by a small number, ε, in the next iteration and subsequently remove this constraint. The contribution in the objective function is to introduce two types of penalisation for deviating from the normalised solution. Suppose that we already

have the solution from iteration t and we want to move to the next iteration, $t + 1$. The idea constitutes the following three aspects:

1. If the maximum comprehensive value achieved in the iteration t, V_t^* is greater than 1, a new constraint will be added to the model that considers the position of the maximum value in the marginal value functions. The added constraint will be kept in all the subsequent iterations. Let us denote by v_i^t the extracted marginal value function in the tth iteration and g_i^{t*} the breakpoint on the ith criterion with the highest marginal value. The constraint that has to be added and kept in all of the subsequent iterations is as follows.

$$\sum_{i=1}^{m} v_i^t(g_i^{t*}) \leq 1 \tag{2.16}$$

   Going forward, we call these types of constraints 'restrictive constraints'.

2. If V_t^* is less than 1, a new constraint will be added to the model in the next iteration, $t + 1$, considering the location of the breakpoints corresponding to the maximum value in each criterion. The constraint that has to be added in iteration $t + 1$ is as follows.

$$\sum_{i=1}^{m} v_i^t(g_i^{t*}) \geq V_t^* + \varepsilon \tag{2.17}$$

   in which ε is a very small real number so that $\varepsilon \in (0, 1 - V_t^*)$. The added constraint will be imposed only in the next iteration, and will be removed later. We refer to this type of constraint as an 'incremental constraint'.

3. Two types of penalties are defined and considered in the objective function, one for the case that the maximum comprehensive value deviates from 1 and another for the case

that the minimum comprehensive value deviates from 0, all based on the solution obtained in the last iteration. For the first type, the penalty is proportional to the distance between the maximum comprehensive value and 1, $\mid \sum_{i=1}^{m} v_i^t(g_i^{t*}) - 1 \mid$. Based on the imposed constraints, explained above, we know that in the case that $\sum_{i=1}^{m} v_i^t(g_i^{t*})$ exceeds 1 in iteration t, a new constraint will be imposed in the iteration $t + 1$ that forces this term to have a value less than 1. Therefore, this penalisation term can be rewritten as $1 - \sum_{i=1}^{m} v_i^t(g_i^{t*})$. The second penalisation factor is proportional to the distance of the lowest comprehensive value and 0, $\mid \sum_{i=1}^{m} v_i^t(g_{i*}^t) - 0 \mid = \sum_{i=1}^{m} v_i^t(g_{i*}^t)$, in which g_{i*}^t denotes the breakpoints on the ith criterion with the lowest marginal value. To prevent penalties from dominating the two error terms in the objective function, the coefficients of the penalty terms are defined as a certain percentage of the sum of the error values in the optimal solution of the last iteration. Therefore, penalty terms in the objective function of iteration $(t + 1)$ are multiplied by the coefficients $p_{max} \cdot f^t$ and $p_{min} \cdot f^t$ in which p_{max} and p_{min} are real positive numbers, and f^t is the sum of the error terms in the optimal solution of the last iteration. The underlying logic behind these two penalisation terms is that the position of the maximum and minimum marginal values over criterion breakpoints will change only if it leads to a significant decrease in the error term values.

The LP model of iteration $(t + 1)$ is as follows in (2.18).

$$min \ z = \sum_{n=1}^{N} (\sigma^+(a_n) + \sigma^-(a_n)) + p_{max} f^t (1 - \sum_{i=1}^{m} v_i^t (g_i^{t*}))$$

$$+ p_{min} f^t \sum_{i=1}^{m} v_i^t (g_{i*}^t)$$

Subject to

$$V'(a_n) - V'(a_{n+1}) \geq \delta \ \text{iff} \ a_n \succ a_{n+1}, \forall n = 1, 2, \cdots, N - 1$$

$$V'(a_n) - V'(a_{n+1}) = 0 \ \text{iff} \ a_n \sim a_{n+1}, \forall n = 1, 2, \cdots, N - 1$$

$$A_i V_i \geq 0, \forall i = 1, 2, \cdots, m$$

$$A_i V_i \leq 1, \forall i = 1, 2, \cdots, m$$

for all the iterations $k \leq t$ with V_k^* greater than 1 $\qquad\qquad (2.18)$

$$\sum_{i=1}^{m} v_i^k (g_i^{k*}) \leq 1$$

only if V_t^* less than 1

$$\sum_{i=1}^{m} v_i^t (g_i^{t*}) \geq V_t^* + \varepsilon$$

$$V'(a_n) = V(a_n) - \sigma^+(a_n) + \sigma^-(a_n)$$

$$v_{ij} \ URS, i = 1, 2, \cdots, m, j = 0, 1, \cdots, a_i - 1$$

$$\sigma^+(a_n), \sigma^-(a_n) \geq 0, n = 1, 2, \cdots, N$$

The algorithm can be summarised by the following steps:

Step 0: Define the appropriate breakpoints on each criterion scale and represent the marginal value of each alternative in terms of the decision variables v_{ij}.

Step 1: Set iteration:=0, solve the LP model (2.13) and find g_i^*, g_{i*} and f^0 (iteration=0).

Step 2: Set iteration:=iteration+1. Delete the incremental constraint, if any. Keeping all of the restrictive constraints that were previously added to the model, add the new restrictive

32

constraint (2.16) to the model if the maximum comprehensive value of the previous iteration exceeds 1. If the maximum comprehensive value of the previous iteration is less than 1, add the incremental constraint (2.17) to the next iteration.

Step 3: Check if the normalisation condition is satisfied (i.e. if the maximum marginal values add up to 1 and the minimum comprehensive value is 0. If both conditions are satisfied, go to step 4. If not, go back to step 2.

Step 4: Represent the marginal value function of each criterion by the v_{ij} variables achieved in the last iteration. Calculate the value of each alternatives by (2.6). Rank the alternatives based on the estimated values.

2.4.4 An illustrative example to compare with UTA-NM

In order to illustrate the method, a typical example based on car characteristics is employed. In this example, we assume a set of marginal value functions over a set of three criteria as DM tacit knowledge and we calculate the rank of alternatives based on them. Without prior knowledge over the marginal values and considering only the ranking, we then analyse the extent to which the captured set of marginal value functions are really aligned with the ones previously assumed.

The three criteria are *price*, *maximum speed* and *personal capacity* and the considered alternatives are 28 different cars. The assumed marginal value functions over each of the three criteria and its maximal shares in the comprehensive values that the DM tacitly assigns to each of the criteria are depicted in Fig. (2.4.2).

The model parameters are set such that both proposed method and UTA-NM have the same number of decision variables for each specific criterion, and thus the same degree of freedom for the associated value function. The extracted marginal values and the maximal shares over the criteria for the proposed method and UTA-NM are provided in Fig. (2.4.3) and (2.4.4),
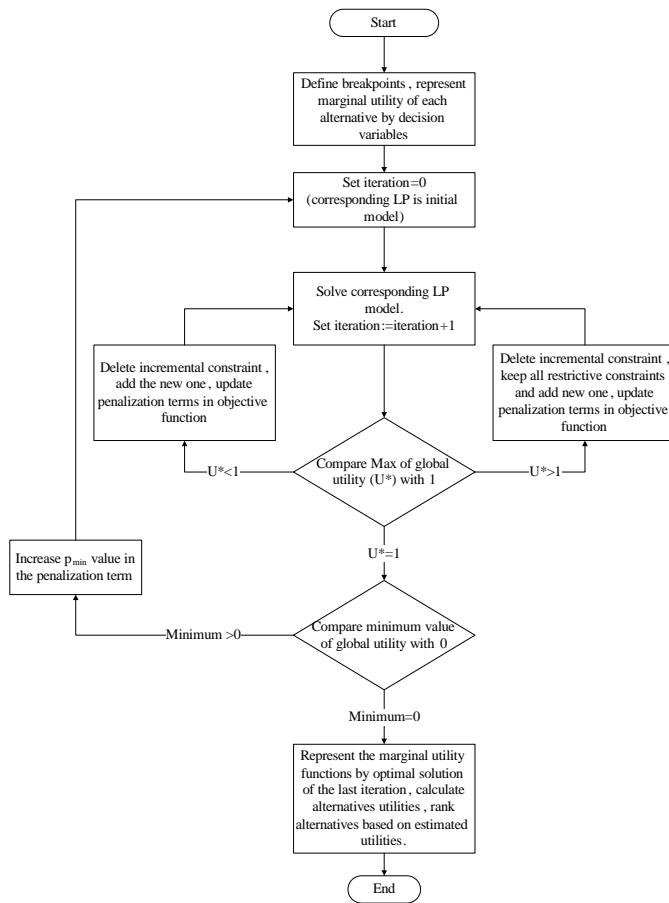
Figure 2.4.1: Proposed Method Flowchart
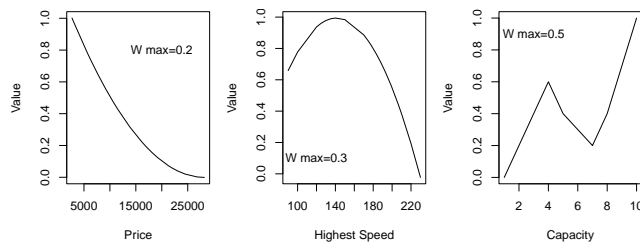
Figure 2.4.2: Assumed value functions for the illustrative example of car characteristics

respectively. It is important to highlight that using the proposed method, the final solution

was obtained after 54 iterations in less than 1 second (using a 64-bit OS on a 2.53 GHz Intel

Core2Duo using MATLAB R2012b), while using UTA-NM exceeded 15 seconds.
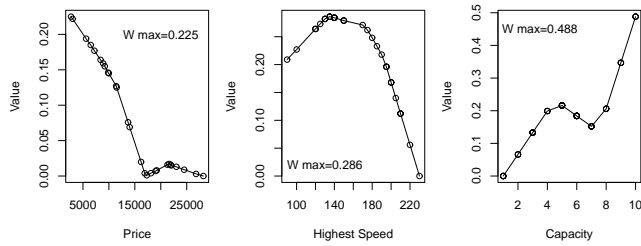
Figure 2.4.3: Extracted value functions for the proposed method in the illustrative example of car characteristics
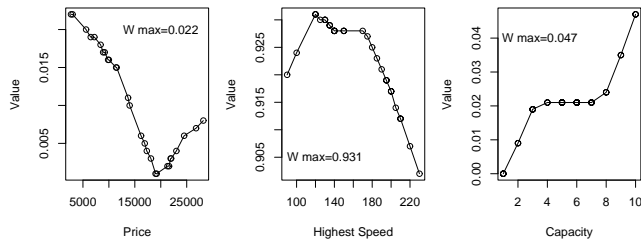
Figure 2.4.4: Extracted value functions by UTA-MN method in the illustrative example

From these results, it can be deduced that by using the proposed method both the marginal value function shapes and criteria weights accord with the preferential system of the DM. However, it can be seen that UTA-NM was not successful in estimating the assumed value functions, also the maximal shares are far from the expected values.

To assess the strength of the methods in reproducing the ranking given by the DM, the Kendall τ measure has been used. In the proposed method a value of 90.0% is obtained, while, in UTA-NM the obtained value was 83.1%.

2.5   BRAND COLOUR AND BRAND IMAGE

Our focus in this study is to explore the contribution of brand colour in brand image in an understandable way. Several studies examined impacts of brand colour on various aspects of the brand. In a scenario-based experiment, Babin et al. found that effects of colour on behavioural

intentions are mediated by the cognitive reactions they create (Babin et al. 2003). As by definition, the concept of brand image is based on the perception of consumers, we expect a strong association between brand image and brand colour.

Although studies show that all three colour components influence brand personality (Labrecque and Milne 2012), most colour research in marketing focuses only on colour hue, and usually ignores the other two attributes: saturation and value (Labrecque et al. 2013). However, in this study, we consider all three attributes together. Furthermore, as studies show that the influence of colour differs across product categories (Lohse and Rosen 2001), we analyse only brands from a particular sector, namely beauty products.

2.5.1 EXPERIMENT DESCRIPTION AND DATASET

Our data comes from a survey conducted by Young and Rubicam's BrandAsset Valuator consulting group. The dataset contained many measurements of several aspects of brand and was published recently (Lovett et al. 2014). In their quarterly survey (ten quarters from 2008 to the second quarter of 2010) a representative sample of the U.S. population, $17,000$ individuals, were asked about 250 brands. The survey measures a broad range of perceptions and attitudes of brands. In their survey, 40 different attributes of brand image (arrogant, energetic, chic, etc.) are included, and each respondent is asked to check whether (s)he can associate the brand with each of these attributes. For each attribute, the dataset contains the percentage of respondents who associated this attribute with the brand. In our experiment, we converted all the percentages into a ranking. Hence we considered the relative position of brands with respect to each brand image attribute. The brands at the top of the ranking with respect to each brand image attribute are deemed to have a strong association with that particular attribute in customer minds. Furthermore, we only considered single-coloured brands, as the interaction among colours is not the topic of this study. We used 34 single-coloured beauty brands. Finally, we measured the

HSV colour component for each brand logo and added these to the dataset. A small portion of the dataset is presented in table 2.5.1 for illustration, where only 8 brands and 3 brand image attributes together with the brand colour components are presented.

Table 2.5.1: Brand colour components and rankings with respect to brand image items

Brand name	Colour components			Rankings in image items		
	H	S	V	Daring	Fun	Trendy
Always	214	82	66	4	34	22
Caress	45	50	100	18	21	18
Charmin	203	78	86	28	27	29
Clairol	208	98	53	3	8	4
Colgate	358	88	93	8	14	27
CoverGirl	324	10	20	2	1	2
Crest	199	99	76	23	3	21
Dial Soap	205	100	71	7	23	25
...	...	...	...	...	...	...

2.5.2 EXPERIMENTAL RESULTS

For each of the brand image items, the association with the brand colour is analysed using the proposed methodology. The set of extracted value functions represents the colour patterns with respect to that brand image item. The value functions are used to calculate the utility of each brand from the perspective of that particular brand image item. The greater the utility of a brand, the greater is the likelihood of a strong connection between the brand and the brand image attribute. Finally, the brands were ranked on their utilities. The obtained ranking was compared with the initial ranking from the data in order to measure the accountability of the extracted value functions. The Kendall τ measure was used for this purpose.

The extracted value functions for the brand image attribute *fun* is presented in the Fig. (2.5.1). By setting $P_{min} = 0.6$ and $P_{max} = 0.01$, the results are obtained in 1252 iterations. The Kendall τ measure between the initial ranking and the ranking from the extracted value

functions is equal to 0.63. This indicates a strong association between the two rankings. So we can conclude that brand colour significantly explains the perception of the brand to be *fun*. Under the null hypothesis, when there is no dependency between the two rankings, and when the number of elements in the ranking list is sufficiently large, namely larger than 10, the τ measure follows a normal distribution with the mean equal to zero and variance equal to $\frac{2(2n+5)}{9n(n-1)}$ in which n is the number of elements in the ranking list (Long and Cliff 1997). In our experiment n is equal to 34 which is the number of brands in the list. The statistical test shows that τ is significantly positive with p-value<0.001.
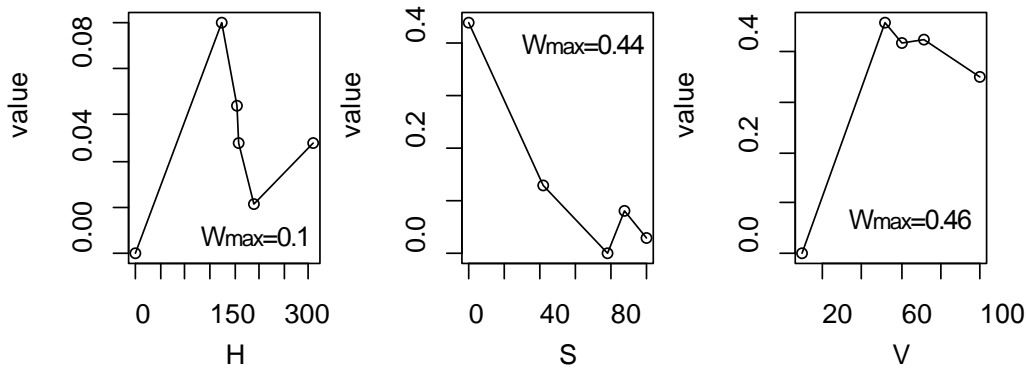
Figure 2.5.1: Value functions for the item fun

The maximal contribution of each H, S, and V colour component into the comprehensive utility of the brand with respect to the item *fun* is 0.10, 0.44, and 0.46, respectively. This indicates that the colour hue plays the least role in the perception of a brand as fun. The value functions demonstrate that brands with the colour hue green, less saturated, and moderately bright tend to be perceived as fun.

The same analysis has been conducted for all the other 39 brand image attributes. The Kendall τ measure of each analysis and colour component weights are given in Table 2.5.2.

As it can be seen from Table 2.5.2, for 29 of the 40 brand image attributes the Kendall τ measure is significant at the 0.001 level of significance. This reveals a significant contribution of

brand colour in the perception of brand by customers. More interestingly, comparison among the weights of colour components shows that colour hue is usually the least important component. Colour hue has the least weight in 39 of the 40 brand image attributes, and only for the image attribute *kind* does it have the second highest weight at the top of component V. Statistical tests show that colour hue (H) weight is significantly less than colour saturation (S) weight ($t(39) = -27.66$ and $p - value < 0.001$), and colour value (V) weight ($t(39) = -19.90$ and $p - value < 0.001$), while there is no significant difference between colour saturation and colour value weights ($t(39) = -1.06$ and $p - value = 0.29$). This indicates that the two colour components S (how pure or whitened is the colour) and V (how dark or bright is the colour) play much more of a role in determining the customer perception of brand image than colour hue. Finally, it is important to highlight that customer perception of a brand as *intelligent* is not influenced by the brand colour.

From each set of the extracted marginal value functions, it is possible to determine the H,S, and V values which lead to the highest utility with respect to each particular brand image attribute. For example, from the extracted value functions for the brand image *fun* in the Fig. (2.5.1), it can be seen that $(V, H, S) = (174.5, 0, 52)$ leads to the highest possible utility. Therefore, a brand manager can choose the corresponding colour to be perceived as *fun* by customers. We did the same analysis for all the attributes which are significant at 0.001 level, and obtained their position in the HSV colour space. The resulted brand image-colour map is presented in Fig. (2.5.2).

The map clearly describes the colour space by the brand image attributes. It also demonstrates the interrelation of brand image items from the brand colour perspective.

2.6   Conclusion and future work

This paper presents a disaggregation methodology based on the UTA method that enables the use of non-monotonic additive models in ranking and other multi-criteria decision problems. The main difference between the proposed methodology and existing non-monotonic methods is that our method is capable of obtaining marginal value functions and the relative importance of attributes (maximal shares in the comprehensive values) following an LP approach. Marginal value functions obtained by the proposed method are free in shape. Over-fitting is prevented by appropriate breakpoints definition and value functions slope restriction.

The proposed method does not require further information regarding the shape of the value functions, nor the most desirable value of each attribute. The only information it requires from the DM is a weak ordering over a set of alternatives. The results from the illustrative example and the experiment shed light on the usefulness and effectiveness of the proposed method.

The proposed method is applied to a real data set of brand image to delineate the role of brand colour in brand perception. The results support a significant contribution of all three colour components in almost all brand image attributes. We also find that colour value and saturation dominate the colour hue role in brand perception by customers.

As future work, we are interested in analysing several product categories to study how the association between brand colour and brand image changes across industries. It would also be interesting to compare the colour pattern for each brand image attribute across product categories. It would be particularly interesting to look for a universal rule of brand perception by brand colour regardless of the product category.

Table 2.5.2: Extracted weights of the colour components and Kendall τ of the reproduced ranking (single star means significant at 0.01 and double star means significant at 0.001)

Brain image attributes	Weights			Kendall τ
	H	S	V	
Arrogant	0.192	0.425	0.373	0.383**
Authentic	0.161	0.416	0.414	0.318*
Best brand	0.120	0.455	0.424	0.540**
Care-free	0.184	0.407	0.408	0.451**
Cares for customers	0.155	0.400	0.438	0.487**
Charming	0.111	0.441	0.438	0.344*
Daring	0.290	0.357	0.353	0.380**
Down to Earth	0.156	0.397	0.441	0.455**
Energetic	0.188	0.373	0.434	0.333*
Friendly	0.147	0.410	0.443	0.430**
Fun	0.097	0.441	0.462	0.629**
Gaining In Popularity	0.149	0.395	0.455	0.365*
Glamorous	0.120	0.442	0.430	0.526**
Good value	0.081	0.469	0.443	0.480**
Healthy	0.106	0.401	0.492	0.316*
Helpful	0.163	0.419	0.414	0.469**
Healthy	0.106	0.401	0.492	0.316*
High Performance	0.195	0.401	0.401	0.390**
Independent	0.207	0.366	0.425	0.419**
Intelligent	0.156	0.400	0.436	0.258
Kind	0.323	0.440	0.228	0.590**
Obliging	0.181	0.386	0.426	0.415**
Original	0.155	0.383	0.458	0.383**
Prestigious	0.246	0.429	0.323	0.458**
Progressive	0.148	0.378	0.467	0.533**
Restrained	0.187	0.418	0.394	0.451**
Rugged	0.172	0.468	0.353	0.619**
Sensuous	0.120	0.430	0.443	0.601**
Simple	0.232	0.427	0.331	0.554**
Social	0.108	0.446	0.446	0.358*
Socially Responsible	0.111	0.431	0.451	0.487**
Straightforward	0.168	0.407	0.427	0.554**
Stylish	0.176	0.385	0.433	0.501**
Traditional	0.158	0.407	0.425	0.458**
Trendy	0.097	0.430	0.463	0.326*
Trustworthy	0.105	0.435	0.453	0.601**
Unapproachable	0.208	0.389	0.399	0.376*
Up To Date	0.137	0.425	0.429	0.465**
Upper class	0.174	0.409	0.417	0.298*
Visionary	0.121	0.435	0.435	0.326*
Worth more	0.111	0.442	0.442	0.412**
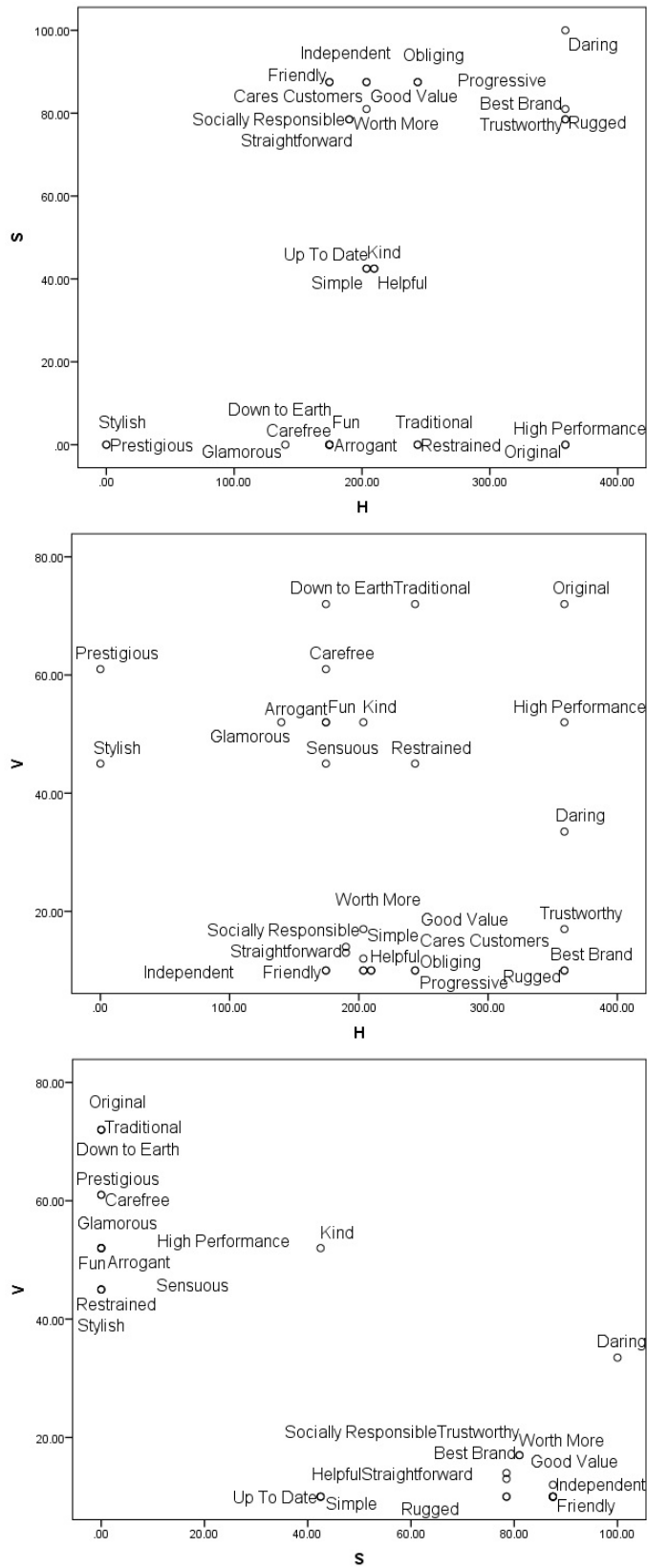
Figure 2.5.2: Colour map of the brand image attributes

3

A Linear Programming Approach for Learning

Non-Monotonic Additive Value Functions in Multiple

Criteria Decision Aiding[1]

A new framework for preference disaggregation in multiple criteria decision aiding is introduced. The proposed approach aims to infer non-monotonic additive preference models from

a set of indirect pairwise comparisons. The preference model is presented as a set of marginal value functions and the discriminatory power of the inferred preference model is maximized against its complexity. To infer a value function that is compatible with the supplied preference information, the proposed methodology leads to a linear programming optimization problem that is easy to solve. The applicability and effectiveness of the new methodology is demonstrated in a thorough experimental analysis covering a broad range of decision problems.

3.1 Introduction

Preference disaggregation techniques are a subfield of multiple criteria decision aiding (MCDA) and cover techniques that are used to infer the judgment policy of a decision maker (DM) by analyzing a given preferential structure (Jacquet-Lagrèze and Siskos 2001). The aim of these techniques is to elicit the DM's preferential system and judgment policy, and to identify the criteria aggregation model that underlies the final decision (de Almeida et al. 2016; Doumpos and Zopounidis 2004a). Disaggregation methods are of interest because they require a low level of cognitive effort and interaction time from the DM. In contrast to many other MCDA methods, disaggregation methods do not expect input from the DM in form of the parameters of a preference model. The DM is not required to specify any of the parameters in the model development process, nor explain the rationale behind his/her decisions (Dias et al. 2002). Most of the existing preference disaggregation methods assume a monotonic relationship between input attributes and preferences. The few methods currently addressing non-monotonicity demand excessive computational effort or require extra information from the DM. Developing an efficient methodology for learning non-monotonic preference models is therefore of great importance.

The assumption of monotonicity is widely used because it seems reasonable for many criteria

(such as price of a product, level of risk, security, safety and ease of use of a service, required time and effort for accomplishing a task). However, this is not the case for many other attributes (such as desired temperature, weight and volume of a product, or attributes related to the composition or organoleptic properties). In a medical context, for instance, attributes such as sugar and cholesterol levels in blood are clear examples of non-monotonic attributes. In finance, there are many indicators used to depict the financial performance of a firm that experts believe must be controlled within a specific range. For instance, a large value for the *cash to total assets ratio* implies that the firm is losing many profitable investment opportunities, whereas a low value indicates a low capacity of the firm to cope with operating expenditures (Despotis and Zopounidis 1995). Thus, dealing with non-monotonic preferences is important both from theoretical and practical perspectives. The monotonicity constraint limits applicability of the MCDA methods in many fields, and it is important to overcome this limitation.

UTA (UTilités Additives) is one of the most representative preference disaggregation methods. It was first introduced in (Jacquet-Lagrèze and Siskos 1982) and follows a linear programming (LP) approach. The aim of UTA is to infer a non-linear (piecewise linear) monotonic additive value function which can reproduce the ranking of a set of alternatives supplied by the DM. Several extensions of the UTA method have since been introduced in the MCDA literature that incorporate variations in the original algorithm, consider different forms of input preference information and optimality criteria, and consider the robustness of inferences. For instance: Pairwise comparisons UTA (Doumpos and Zopounidis 2004b); UTASTAR (Siskos and Yannacopoulos 1985); UTADIS (UTilités Additives DIScriminantes) I, II, III (Devaud et al. 1980; Doumpos and Zopounidis 1998; Jacquet-Lagrèze 1995; Zopounidis and Doumpos 1999b); stochastic UTA (Siskos 1983), Fuzzy UTASTAR (Patiniotakis et al. 2011); ACUTA (Bous et al. 2010); UTA^{GMS} (Greco et al. 2008); GRIP (Generalized Regression with Intensities of Preference) method as an extension of the UTA^{GMS} (Figueira et al. 2009); selection of

representative value functions in ranking and choice problems as an extension of the UTA$^{\text{GMS}}$ and GRIP (Kadziński et al. 2012b); and in sorting problems as an extension of UTADIS$^{\text{GMS}}$ (Greco et al. 2011a); RUTA (Kadziński et al. 2013); MCHP as a general framework in case of a hierarchical structure in the family of evaluation criteria (Corrente et al. 2015; Corrente and Greco 2012); and others (Beccacece et al. 2015; Corrente et al. 2016b; Kadziński et al. 2014, 2012a; Özpeynirci et al. 2017). Interested readers are referred to (Siskos et al. 2016) for a thorough review of UTA methods.

The highlighted preference disaggregation methods have many applications in numerous fields, for example: global e-governance evaluation (Siskos et al. 2014); predicting business failure (Zopounidis and Doumpos 1999a); assessing the environmental impact of cities (Kadziński et al. 2016); assessing investing risk (Doumpos et al. 2001) and country risk (Doumpos and Zopounidis 1997); finance (Zopounidis et al. 2015); healthcare (Doumpos et al. 2015); evaluating R&D projects (Jacquet-Lagrèze 1995); energy planning (Diakoulaki et al. 1999); job evaluation problems (Spyridakos et al. 2001); branding (Ghaderi et al. 2015) and many other applications.

In most of the applications mentioned above, the input attributes are normally expected to be monotonic with respect to the preferences. Several studies have been conducted in the literature to deal with cases where the assumption of monotonicity is no longer valid. One way to address non-monotonicity in a preference model is to divide the range of criteria into intervals so that the preferences are monotonic in each interval and then treat each interval separately. Following this idea, in the method introduced by Despotis and Zopounidis, a quadratic shape for the marginal value functions is admitted (Despotis and Zopounidis 1995). The method requires as input the performance value at which the monotonicity direction changes. Another non-monotonic methodology, UTA-NM, relaxes the monotonicity condition in the UTASTAR algorithm, allowing, in theory, any shape for marginal value functions (Kliegr 2009). The

complexity of the inferred marginal value functions, defined by the number of changes in monotonicity direction, is minimized. The method entails severe performance issues: tens of binary variables are involved even for very small instances, thus making the method computationally infeasible for real-world problems. In another study, Eckhardt and Kliegr proposed *local preference transformation*, a heuristic attribute preprocessing algorithm that transforms arbitrary input attributes into a space that is approximately monotonic with respect to user preferences, thus making it suitable for UTA (Eckhardt and Kliegr 2012). Finally, the methodology introduced in (Doumpos 2012) considers a broader class of non-monotonic value functions. However, this methodology leads to a non-linear integer programming problem that is difficult to solve with realistically sized data sets. Thus, an evolutionary approach based on the differential evolution algorithm is employed.

This paper introduces a new methodology for inferring non-monotonic additive preference models from a set of indirect pairwise comparisons. The main idea is to relax monotonicity constraints for possibly non-monotonic criteria, and then optimizing the complexity of the preference model against its discriminatory power. The complexity of value function is controlled by minimizing variation in the slope of the marginal value functions. A mapping is used to represent the inferred value function in a standard form, while the preference orders among alternatives are preserved. The methodology leads to a linear fractional programming (LFP) problem which can easily be transformed into an LP problem, and so can efficiently deal with datasets of realistic size. In this paper, the input preferences are assumed to be in the form of pairwise comparisons: however; the methodology can be adopted to deal with other types of preference disaggregation problems with other preferential structures (such as ranking, preference intensities, evaluations on a subset of criteria, sorting, and rank-related requirements).

The paper is organized as follows. In Section 3.2 an overview of the preference disaggregation approach following the UTA framework is presented. Several extensions of UTA to

non-monotonic setting are discussed in this section. In section 3.3, the new methodology is introduced, followed by an illustrative example and a comparison with two well-known non-monotonic UTA-like methods: Despotis UTA and UTA-NM. To assess the performance of the new methodology, a thorough experimental analysis over a broad range of decision problems is conducted in the next section. Section 3.5 introduces an extension of the proposed methodology that enables the analyst to manipulate properties of the inferred preference model when necessary. Finally, we conclude the paper and suggestions for future research are highlighted.

3.2 REMINDER ON PREFERENCE DISAGGREGATION

Throughout the paper, the following notation is used:

- $A = \{a_1, a_2, \cdots, a_N\}$ is a finite set of alternatives described over a set of criteria, and $A^R \subset A$ is the set of reference alternatives on which the DM provides some preference information.

- $G = \{g^1, g^2, \cdots, g^M\}$ is the family of evaluation criteria, $g^m \colon A \to X^m$, in which X^m is the evaluation scale of the mth criterion, and $g^m(a_n) = x_n^m$ is the performance of alternative a_n over the criterion g^m. Without loss of generality, in this section we assume that the greater $g^m(a_n)$, the better is the alternative a_n on criterion g^m.

- $G^K \subseteq G$ is the set of all possibly non-monotonic criteria.

- $g_0^m, g_1^m, \cdots, g_{a_m}^m$ are $a_m + 1$ breakpoints to divide the evaluation scale of mth criterion into a_m subintervals $[g_0^m, g_1^m], [g_1^m, g_2^m], \cdots, [g_{a_m-1}^m, g_{a_m}^m]$. Moreover, g_0^m and $g_{a_m}^m$ are the minimum and maximum performance levels for the mth scale, respectively.

The DM is expected to provide a set of holistic pairwise judgments in the form of partial preorder $\succsim$ on A^R. For the two alternatives $a, b \in A^R$, $a \succsim b$ means that a is at least as good as b. The indifference, $\sim$, is the symmetric part of $\succsim$, and the strict preference relation, $\succ$, is the asymmetric part. A set of marginal value functions $v^m \colon X^m \to [0, 1]$ is estimated for the

set of evaluation criteria so that the preference information provided by the DM is reproduced. Each marginal value function associates a numerical score to each alternative which is its value from the perspective of corresponding criterion. The numerical scores are aggregated additively to obtain the comprehensive value. The marginal value at a breakpoint g_l^m, $1 \leq l \leq a_m$ for criterion m is expressed as follows:

$$v^m(g_l^m) = \sum_{j=1}^{l}(v^m(g_j^m) - v^m(g_{j-1}^m)) = \sum_{j=1}^{l} v_j^m \tag{3.1}$$

where $v^m(g_l^m)$ is the value associated with the lth breakpoint on the mth criterion, and $v_j^m = v^m(g_j^m) - v^m(g_{j-1}^m)$.

The marginal value corresponding to the performance $x_n^m \in [g_{l-1}^m, g_l^m]$ is obtained by linear interpolation between $v^m(g_{l-1}^m)$ and $v^m(g_l^m)$, as follows:

$$v^m(x_n^m) = \sum_{j=1}^{l-1} v_j^m + \frac{x_n^m - g_{l-1}^m}{g_l^m - g_{l-1}^m} \cdot v_l^m. \tag{3.2}$$

The preference information provided by the DM is translated into the following set of constraints:

$$\left.\begin{array}{l} V(a) - V(b) \geq \varepsilon, \text{ if } a \succ b, \text{ for } a, b \in A^R, \\ V(a) - V(b) \geq 0, \text{ if } a \succsim b, \text{ for } a, b \in A^R, \\ V(a) - V(b) = 0, \text{ if } a \sim b, \text{ for } a, b \in A^R, \end{array}\right\} E_{DM}^{A^R} \tag{3.3}$$

in which $V(a) = \sum_{m=1}^{M} v^m(a)$ is the comprehensive value of $a \in A^R$, and ε is a positive parameter to discriminate comprehensive values of pairs belonging to strict preference relations.

The set of all value functions compatible with the preferences given by the DM, denoted by $\mathcal{U}$, is defined by the following set of constraints:

$$
\left.
\begin{aligned}
& E_{DM}^{A^R}, \\
& v^m(g_o^m) = 0, \ m = 1, \cdots, M, \\
& \sum_{m=1}^{M} v^m(g_{a_m}^m) = 1, \\
& v_j^m \geq 0, \ j = 1, \cdots, a_m, \ m = 1, \cdots, M
\end{aligned}
\right\} \ E(\mathcal{U})
\tag{3.4}
$$

in which the last three constraints define conditions on normalization and monotonicty of value functions. In order to obtain the most discriminant value function representing the DM preferences, the following LP problem has to be solved:

$$
\textit{Maximize } \varepsilon, \ \textit{s.t. } E(\mathcal{U}).
\tag{3.5}
$$

Let us denote by ε^* the optimal solution of the above LP problem. There exists at least one value function compatible with the preferences provided by the DM if $E(\mathcal{U})$ is feasible and $\varepsilon^* > 0$. Otherwise, some comparisons of reference alternatives will not be reproduced. The reason for such incompatibility could be the assumptions made about the preference model of the DM, i.e. monotonic and additive value function, or inconsistency among the judgments made by the DM. The DM will be asked to revise some comparisons in order to obtain a compatible value function, or to pursue the analysis and accept some level of incompatibility. A detailed discussion on dealing with incompatibility can be found in (Greco et al. 2008).

3.2.1 Non-Monotonic UTA-Like Methods

A brief overview of the studies extending the UTA method to a non-monotonic setting is presented in this section. The first extension of the UTA method to a non-monotonic setting

was proposed by (Despotis and Zopounidis 1995). This method follows an LP approach to infer a non-monotonic additive value function from a given ranking over a set of alternatives. In the model, a specific form of non-monotonicity is assumed. The marginal value functions are assumed to be quadratic in shape, and the DM is required to specify, for each criterion, the performance value at which the monotonicity direction changes.

In a more general approach in (Doumpos 2012), a broader class of non-monotonic additive value functions is considered. The value functions are allowed to be quasiconvex, quasiconcave, or s-type, and inferred directly from a set of decision examples without requiring any additional information from the DM. The major technical difference between this method and UTASTAR is in the formulation of the value functions. The marginal value at a breakpoint g_l^m is defined as follows.

$$v^m(g_l^m) = \sum_{j=1}^{l} y_j^m v_j^m + d^m \qquad (3.6)$$

The decision variable $y_j^m \in \{-1, 1\}$ defines the monotonicity of the value function over the subinterval $[g_{j-1}^m, g_j^m]$ in the mth criterion. The parameter d^m is used to ensure that the marginal values are non-negative. To avoid arbitrary marginal value functions, the variables y_j^m are introduced to admit only quasiconvex, quasiconcave or s-type forms. Finally, in a similar way to the original UTA method, comprehensive values are normalized in the interval $[0, 1]$. This formulation leads to a mixed integer non-linear programming (MINLP) problem, which is classified as an NP-hard (non-deterministic polynomial-time hard) problem, and therefore, its solution is computationally intensive. In the original paper, an evolutionary approach, based on the differential evolution algorithm, is employed to solve the problem.

The method termed UTA-NM (Kliegr 2009), also removes the monotonicity condition imposed by the UTASTAR algorithm and, in contrast to UTASTAR, accepts negative values

for the variables v_j^m. To ensure normalisation, however, many binary variables are added to the model and this adds considerably to its complexity. To prevent overfitting and highly complex marginal value functions, every change in the monotonicity is penalised. However, the penalisation term introduced entails an excessive computational cost since even trivial tasks take tens of seconds to complete.

Finally, a heuristic preprocessing algorithm, *local preference transformation*, transforms arbitrary input attributes to approximately monotonic attributes, thus making it suitable for UTA (Eckhardt and Kliegr 2012). The definition of the transformation function depends on the type of the input attribute, which can be either nominal or cardinal (ordinal attributes are not addressed in the original paper). In both nominal and cardinal cases, some ratings are assigned to the alternatives based on their weak order. For nominal attributes, the rating of a performance level is calculated as the average rating of all the alternatives with that performance level. For cardinal attributes, a univariate linear regression is used to find the relationship between the cardinal input performances as independent, and ratings as dependent variables. The first shortcoming of this algorithm is that monotonic attributes are not guaranteed after local preference transformation (Eckhardt and Kliegr 2012). Another issue is that the determination coefficient of the univariate linear regression model is too small in presence of highly non-monotonic preferences. As a result, the relationship is not reliable. The third drawback is the need to assign arbitrary rating (but compatible with the ranking given by the DM) for the alternatives. Different ratings usually lead to different marginal value functions.

3.3 PROPOSED METHODOLOGY

As in the approaches of (Doumpos 2012) and (Kliegr 2009), the non-monotonicity of value functions in our proposed method is permitted by accepting negative values for the decision

variables v_j^m. However, the complexity of the value functions is controlled by minimizing variation in slope. Controlling variation in slope has been used before in (Doumpos and Zopounidis 2002) for the determination of the minimum number of criteria subintervals, and in (Greco et al. 2011b) to obtain parsimonious preference models.

Due to the possible non-monotonicity in the marginal value functions, the least and the most preferred performance levels are not known a priori. The resulting value functions, therefore, are not initially normalized. To represent the value function in a standard form, a transformation is employed. The ordinal relations among the alternatives are preserved under the transformation. In the next subsection, before providing the model formulation, we introduce the transformation for normalizing the estimated value function.

3.3.1 Transformation

Suppose that $V = \{v^1, v^2, \cdots, v^M\}$ is the set of marginal value functions in which $v^m \colon X^m \to \mathbb{R}$ is the marginal value function over the mth criterion, and the value function is not necessarily in a UTA-like standard format. UTA-like methods commonly infer an unweighted form of the additive value functions (Siskos et al. 2005):

$$
\begin{aligned}
V(\cdot) &= \sum_{m=1}^{M} v^m(\cdot), \\
\sum_{m=1}^{M} v_m(g^{m*}) &= 1, \\
v_m(g_*^m) &= 0 \qquad \forall m = 1, \cdots, M
\end{aligned}
\tag{3.7}
$$

in which g^{m*} and g_*^m denote performance levels corresponding to the greatest and the least marginal values, respectively. In non-monotonic setting, these two values do not necessarily coincide with the extremes of the evaluation scale, i.e. $g_{a_m}^m$ and g_o^m. We are looking for a transformation T that maps the functional space $\mathcal{V}$ to the functional space $\mathcal{V}_\mathcal{T}$, in which for $V \in \mathcal{V}$ and $V_T \in \mathcal{V}_\mathcal{T}$, $V_T = T(V)$, and $\mathcal{V}_\mathcal{T}$ is the space including comprehensive value

functions in the UTA standard format represented in (3.7). The mapping is also required to preserve the ordinal relations among the alternatives, i.e. to satisfy the following three conditions:

$$\sum_m v^m(a) > \sum_m v^m(b) \iff \sum_m T(v^m)(a) > \sum_m T(v^m)(b),$$

$$\sum_m v^m(a) \geq \sum_m v^m(b) \iff \sum_m T(v^m)(a) \geq \sum_m T(v^m)(b), \qquad (3.8)$$

$$\sum_m v^m(a) = \sum_m v^m(b) \iff \sum_m T(v^m)(a) = \sum_m T(v^m)(b).$$

The mapping $T : V \to V_T$, defined as follows, satisfies all the requirements specified in (3.7) and (3.8).

$$v_T^m = T(v^m) = \frac{v^m - v^m(g_*^m)}{\sum_m (v^m(g^{m*}) - v^m(g_*^m))}. \qquad (3.9)$$

In the above mapping, the minimum value of $v_T^m = T(v^m)$ is 0 at g_*^m, by its definition. In addition, $\sum_m v_T^m(g^{m*}) = \sum_m \frac{v^m(g^{m*}) - v^m(g_*^m)}{\sum_m (v^m(g^{m*}) - v^m(g_*^m))} = 1$. Moreover, the ordinal relations among the alternatives are preserved under this transformation, i.e. $\sum_{m=1}^{M} v^m(a) > \sum_{m=1}^{M} v^m(b) \iff \sum_{m=1}^{M} v_T^m(a) > \sum_{m=1}^{M} v_T^m(b)$. The proof is straightforward and provided below.

$$\sum_{m=1}^{M} v^m(a) > \sum_{m=1}^{M} v^m(b) \iff$$

$$\iff \sum_{m=1}^{M} (v^m(a) - v^m(g_*^m)) > \sum_{m=1}^{M} (v^m(b) - v^m(g_*^m)) \iff$$

$$\iff \sum_{m=1}^{M} \underbrace{\frac{v^m(a) - v^m(g_*^m)}{\sum_{m=1}^{M} (v^m(g^{m*}) - v^m(g_*^m))}}_{v_T^m(a)} > \sum_{m=1}^{M} \underbrace{\frac{v^m(b) - v^m(g_*^m)}{\sum_{m=1}^{M} (v^m(g^{m*}) - v^m(g_*^m))}}_{v_T^m(b)}.$$

The proof for the cases of $"\geq"$ and $"="$ in (3.8) is similar to the above.

3.3.2 Model Formulation

The value functions are assumed to be piecewise linear, and so featuring a constant slope for each subinterval. The slope of the mth marginal value function for the jth subinterval is equal to $v_j^m/(g_j^m - g_{j-1}^m)$. Therefore, the change in the slope at the interior breakpoint g_j^m is $v_{j+1}^m/(g_{j+1}^m - $

$g_j^m) - v_j^m / (g_j^m - g_{j-1}^m)$. The slope variation is minimized across all the $\sum_{m=1}^{M}(a_m - 1)$ interior breakpoints. Let us denote by γ_j^m the non-negative variable which specifies the difference in slope for the two subintervals $[g_{j-1}^m, g_j^m]$ and $[g_j^m, g_{j+1}^m]$. The following set of constraints restricts variation in slopes:

$$\left.\begin{array}{l} \frac{v_{j+1}^m}{g_{j+1}^m - g_j^m} - \frac{v_j^m}{g_j^m - g_{j-1}^m} \leq \gamma_j^m, \\[2mm] \frac{v_j^m}{g_j^m - g_{j-1}^m} - \frac{v_{j+1}^m}{g_{j+1}^m - g_j^m} \leq \gamma_j^m, \\[2mm] j = 1, \cdots, a_m - 1, \ m = 1, \cdots, M. \end{array}\right\} E_{slope}^{A^R} \qquad (3.10)$$

The above constraint ensures that the absolute difference in slopes of the marginal value function over the two subintervals $[g_{j-1}^m, g_j^m]$ and $[g_j^m, g_{j+1}^m]$ should be less than or equal to the variable γ_j^m that is going to be optimized. Moreover, in order to bound the solution space, the following set of constraints is employed to restrict the marginal values at any breakpoint to the interval between 0 and 1:

$$\left.\begin{array}{l} \left.\begin{array}{l} v^m(g_0^m) \geq 0, \\[1mm] v^m(g_0^m) + v_1^m \geq 0, \\[1mm] \vdots \\[1mm] v^m(g_0^m) + v_1^m + \cdots + v_{a_m}^m \geq 0, \end{array}\right\} E_{low}^{A^R} \\[6mm] \left.\begin{array}{l} v^m(g_0^m) \leq 1, \\[1mm] v^m(g_0^m) + v_1^m \leq 1, \\[1mm] \vdots \\[1mm] v^m(g_0^m) + v_1^m + \cdots + v_{a_m}^m \leq 1, \end{array}\right\} E_{up}^{A^R} \end{array}\right\} E_{bound}^{A^R} \qquad (3.11)$$

in which the sets of constraints in $E_{low}^{A^R}$ and $E_{up}^{A^R}$ guarantee that the marginal value at any point is non-negative and less than 1, respectively. The choice of 0 and 1 as lower and upper bounds are rather arbitrary and any other two values could be chosen to avoid an unbounded solution

space. The set of all possibly non-monotonic value functions representing the preference model of the DM, denoted by $\mathcal{U}^{NM}$, is defined by the following set of constraints:

$$\left.\begin{array}{l} E^{A^R}_{DM}, \ E^{A^R}_{slope}, \ E^{A^R}_{bound}, \\[2mm] v^m(g^m_0), v^m_j \geq 0, \ j = 1, \cdots, a_m, \ \forall m : g^m \in G/G^K, \\[2mm] \gamma^m_j \geq 0, \ j = 1, \cdots, a_m - 1, \ m = 1, \cdots, M, \\[2mm] \varepsilon \geq 0. \end{array}\right\} \quad E(\mathcal{U}^{NM}) \qquad (3.12)$$

Unknowns in the above set of constraints are $v^m(g^m_0)$, v^m_j, γ^m_j, and ε. The second line of constraints ensures monotonicity of those criteria that are known to be monotonic a priori. The other decision variables, $v^m(g^m_0)$ and v^m_j for the non-monotonic criteria, are unrestricted in sign to allow non-monotonicity.

To find an instance of a possibly non-monotonic value function compatible with the supplied pairwise comparisons, we are considering two objectives simultaneously. First, the parameter ε that is going to be maximized to obtain a value function with the most discriminatory power. Second, the term $\sum_{m,j} \gamma^m_j$ which is going to be minimized in order to obtain the most parsimonious value function. The two objectives are conflicting because the more complex the value function, the better it can discriminate the alternatives belonging to the strict preference relation. To account for the both objectives, several approaches can be followed. However, one desired property of the optimal solution, subject to (3.12), is that it remains optimal in the transformed solution space under the transformation (3.9). One natural expectation is that $V^* \succ_{obj} V \iff V^*_T \succ_{obj} V_T$, where $V^* \succ_{obj} V$ reads as $''$the value function V^* is better than V according to the objective function$''$, $V^*_T = T(V^*)$, and $V_T = T(V)$. One possible approach to account for objectives, is to introduce weights for the two objectives and minimize $\frac{1}{\eta} \sum_{m,j} \gamma^m_j - \varepsilon$, in which η determines the relative importance of each objective. Another possibility is to follow a lexicographic approach by first maximizing ε, then keeping ε at its optimal

level by adding a new constraint and minimizing $\sum_{m,j} \gamma_j^m$. In our formulation, we construct the objective function in a fractional form - that is, we minimize the ratio $\sum_{m,j} \gamma_j^m / \varepsilon$. Therefore, to obtain one possibly non-monotonic instance of value functions compatible with the supplied preference information, the following LFP problem should be solved:

$$\text{Minimize } \frac{\sum_{m,j} \gamma_j^m}{\varepsilon}, \text{ s.t. } E(\mathcal{U}^{NM}). \tag{3.13}$$

The above LFP problem can be easily transformed into an LP problem using the Charnes-Cooper transformation (Charnes and Cooper 1962).

Following this formulation, it can be shown that for any instance of the solution space, the value of the objective function remains unchanged under the transformation (3.9), hence the optimal solution in the original optimization problem (3.13) remains optimal amongst the solution instances after the transformation. To prove this, for any two alternatives $a_i, a_j \in A^R$, for which $a_i \succ a_j$, the relation between differences in comprehensive values before and after transformation is as follows:

$$V_T(a_i) - V_T(a_j) = \sum_m \frac{(v^m(a_i) - v^m(g_*^m)) - (v^m(a_j) - v^m(g_*^m))}{\sum_m (v^m(g^{m*}) - v^m(g_*^m))} \tag{3.14}$$

$$= \sum_m \frac{v^m(a_i) - v^m(a_j)}{\sum_m (v^m(g^{m*}) - v^m(g_*^m))} \tag{3.15}$$

$$= \frac{V(a_i) - V(a_j)}{\sum_m (v^m(g^{m*}) - v^m(g_*^m))}. \tag{3.16}$$

This means that the difference in comprehensive values of alternatives is rescaled with respect to the constant value in the denominator. Therefore, the same pair of alternatives entails

minimal difference in comprehensive values before and after transformation. This implies that $\varepsilon_T^* = \varepsilon^*/\sum_m(v^m(g^{m*}) - v^m(g_*^m))$ in which ε^* is the value of ε in the optimal solution of (3.13), and ε_T^* is the minimal difference in comprehensive values of alternatives according to the value function in the standard form. Similarly, it can be shown that the variation in slope at any interior breakpoint is not influenced by the shift in marginal value functions, i.e. subtracting $v^m(g_*^m)$, but only by the rescaling, and hence $(\gamma_j^m)_T = \gamma_j^m/\sum_m(v^m(g^{m*}) - v^m(g_*^m))$, in which $(\gamma_j^m)_T$ is the variation in the slope of the mth standardized marginal value function at j^{th} breakpoint. Therefore, $\sum_{m,j}(\gamma_j^m)_T/\varepsilon_T = \sum_{m,j}\gamma_j^m/\varepsilon$. This implies that the value function corresponding to the minimum value of the objective function in (3.13) holds the smallest ratio of "total variation in slope divided by minimal differences in comprehensive values", amongst all the value functions in the standard form.

3.3.3 ILLUSTRATIVE EXAMPLE

The applicability of the proposed methodology is illustrated using a synthetic example in this section. In this example, $N = 10$ alternatives are evaluated based on $M = 2$ criteria and $p = 9$ pairs of alternatives are selected randomly for comparison by the hypothetical DM. The performances of the alternatives are presented in table 3.3.1.

The supplied pairwise comparisons are presented in Figure 3.3.1, in which an arrow from a_i to a_j implies $a_i \succ a_j$. According to the supplied preferences, the alternative a_1 is not in the reference set A^R.

Discretization of criteria scales is performed by following general value functions in which every distinct performance value is considered as a breakpoint. Therefore, the breakpoints for the first criterion, g^1, are $g_0^1 = -9.4, g_1^1 = -9.2, g_2^1 = -5.2, g_3^1 = -4.4, g_4^1 = -3.2, g_5^1 = -0.4, g_6^1 = 3.8, g_7^1 = 4.6$ and $g_8^1 = 6.5$. The marginal values at each of these breakpoints are $v^1(-9.4)$, $v^1(-9.4) + v_1^1, v^1(-9.4) + v_1^1 + v_2^1, ..., v^1(-9.4) + v_1^1 + v_2^1 + \cdots + v_8^1$. These marginal values are

Table 3.3.1: Randomly generated performances of 10 alternatives evaluated by 2 criteria

alternatives	g^1	g^2
a_1	0.5	4.5
a_2	-4.4	-1.3
a_3	-0.4	-0.7
a_4	-3.2	-4.7
a_5	4.6	-4.2
a_6	-5.2	9.1
a_7	-9.2	4.9
a_8	6.5	-0.5
a_9	3.8	6.2
a_{10}	-9.4	-8.7
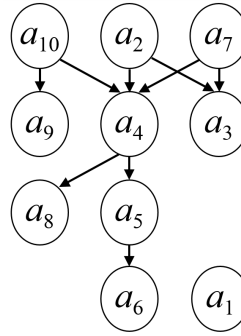
Figure 3.3.1: Pairwise comparisons of the hypothetical alternatives

the unknowns that are going to be estimated. Breakpoints and marginal values for the second criterion are defined in a similar way. To obtain the comprehensive value of an alternative, the marginal values need to be summed. For example, the comprehensive value of the alternative a_4 is defined as $V(a_4) = v^1(-3.2) + v^2(-4.7) = v^1(-9.4) + v_1^1 + v_2^1 + v_3^1 + v_4^1 + v^2(-8.7) + v_1^2$. In the next step, the constraints representing the supplied preferences $\left(E_{DM}^{A^R}\right)$, are created. For

example, according to the supplied preferences $a_2 \succ a_4$. The constraint according to this pairwise comparison is as follows:

$$
\begin{aligned}
a_2 \succ a_4 \rightarrow & V(a_2) - V(a_4) \geq \varepsilon \rightarrow \\
& \left(v^1(-9.4) + v_1^1 + v_2^1 + v_3^1 + v^2(-8.7) + v_1^2 + v_2^2 + v_3^2 \right) \\
& -\left(v^1(-9.4) + v_1^1 + v_2^1 + v_3^1 + v_4^1 + v^2(-8.7) + v_1^2 \right) \geq \varepsilon \rightarrow \\
& -v_4^1 + v_2^2 + v_3^2 \geq \varepsilon.
\end{aligned}
\tag{3.17}
$$

The remaining constraints in E_{DM}^{AR} are created in a similar way.

The next step is to create the constraints for controlling variation in slope of the marginal value functions. For example, the constraint associated with the first interior breakpoint of the first criterion, $g_1^1 = -9.2$, is as following:

$$
\begin{aligned}
\frac{v_2^1}{-5.2-(-9.2)} - \frac{v_1^1}{-9.2-(-9.4)} &\leq \gamma_1^1, \\
\frac{v_1^1}{-9.2-(-9.4)} - \frac{v_2^1}{-5.2-(-9.2)} &\leq \gamma_1^1,
\end{aligned}
\tag{3.18}
$$

and, as another example, the constraint associated with the third interior breakpoint of the second criterion, $g_3^2 = -1.3$, is as following:

$$
\begin{aligned}
\frac{v_4^2}{-0.7-(-1.3)} - \frac{v_3^2}{-1.3-(-4.2)} &\leq \gamma_3^2, \\
\frac{v_3^2}{-1.3-(-4.2)} - \frac{v_4^2}{-0.7-(-1.3)} &\leq \gamma_3^2.
\end{aligned}
\tag{3.19}
$$

The remaining constraints for all the interior breakpoints are defined similarly.

Finally, to bound the solution space, the constraints in $E^{A^R}_{bound}$ must be added. For the sake of illustration, $E^{A^R}_{low}$ for the first criterion, g^1, is provided here:

$$v^1(-9.4) \geq 0,$$
$$v^1(-9.2) + v^1_1 \geq 0,$$
$$v^1(-9.4) + v^1_1 + v^1_2 \geq 0, \tag{3.20}$$
$$\vdots$$
$$v^1(-9.4) + v^1_1 + v^1_2 + \cdots + v^1_8 \geq 0.$$

To obtain one instance of the value functions representing DM preferences, the following LFP problem should be solved:

$$Minimize \ \frac{\gamma^1_1 + \cdots + \gamma^1_7 + \gamma^2_1 + \cdots + \gamma^2_7}{\varepsilon}$$

$$s.t.$$

$$E^{A^R}_{DM}, E^{A^R}_{slope}, E^{A^R}_{bound},$$
$$v^1(-9.4), v^1_1, v^1_2, \cdots, v^1_8 \geq 0, \tag{3.21}$$
$$v^2(-8.7), v^2_1, v^2_2, \cdots, v^2_8 \geq 0,$$
$$\gamma^1_1, \cdots, \gamma^1_7, \gamma^2_1, \cdots, \gamma^2_7 \geq 0,$$
$$\varepsilon \geq 0.$$

By solving the above optimization problem, estimation of the decision variables corresponding to the two marginal value functions are presented in table 3.3.2.

The value obtained for ε^* is 0.058. Moreover, according to the results, $\gamma^1_4 = 0.253$, $\gamma^2_3 = 0.480$ and the estimated value for all the remaining γ^m_j variables is equal to 0. The estimated comprehensive values for the alternatives are presented in the second column of table 3.3.3.

According to the inferred value function, $a_{10} \succ a_2 \sim a_7 \succ a_4 \succ a_3 \succ a_5 \succ a_8 \succ a_1 \succ a_6 \succ a_9$, which is compatible with the supplied preference information.

Table 3.3.2: Estimated values for the decision variables related to the marginal value functions

	g^1		g^2
$v^1(-9.4)$	0.505	$v^2(-8.7)$	0.950
v^1_1	-0.005	v^2_1	0.034
v^1_2	-0.095	v^2_2	0.004
v^1_3	-0.019	v^2_3	0.025
v^1_4	-0.029	v^2_4	-0.011
v^1_5	-0.022	v^2_5	-0.004
v^1_6	-0.033	v^2_6	-0.100
v^1_7	-0.006	v^2_7	-0.024
v^1_8	-0.015	v^2_8	-0.053

Table 3.3.3: Estimated comprehensive values for the alternatives

alternatives	comprehensive values according to initial value function	comprehensive values according to transformed value function
a_1	1.234	0.318
a_2	1.399	0.714
a_3	1.337	0.565
a_4	1.341	0.576
a_5	1.284	0.437
a_6	1.226	0.299
a_7	1.399	0.714
a_8	1.279	0.425
a_9	1.176	0.180
a_{10}	1.455	0.848

The current value function is not in a standard form because $v^1(g^{1*}) + v^2(g^{2*}) = v^1(-9.4) + v^2(-1.3) = 0.505 + 1.013 = 1.518 \neq 1$. Moreover, $v^1(g^1_*) = v^1(6.5) = 0.280 \neq 0$, and $v^2(g^2_*) = v^2(9.1) = 0.821 \neq 0$. To this end, the marginal value at every breakpoint for the first criterion will be subtracted by $v^1(g^1_*) = 0.280$ and divided by $v^1(g^{1*}) - v^1(g^1_*) + v^2(g^{2*}) - v^2(g^2_*) = 0.416$. Similarly, the second marginal value function will be subtracted by 0.821 and divided by 0.416. The inferred marginal value functions before and after the transformation are presented in Figure 3.3.2.
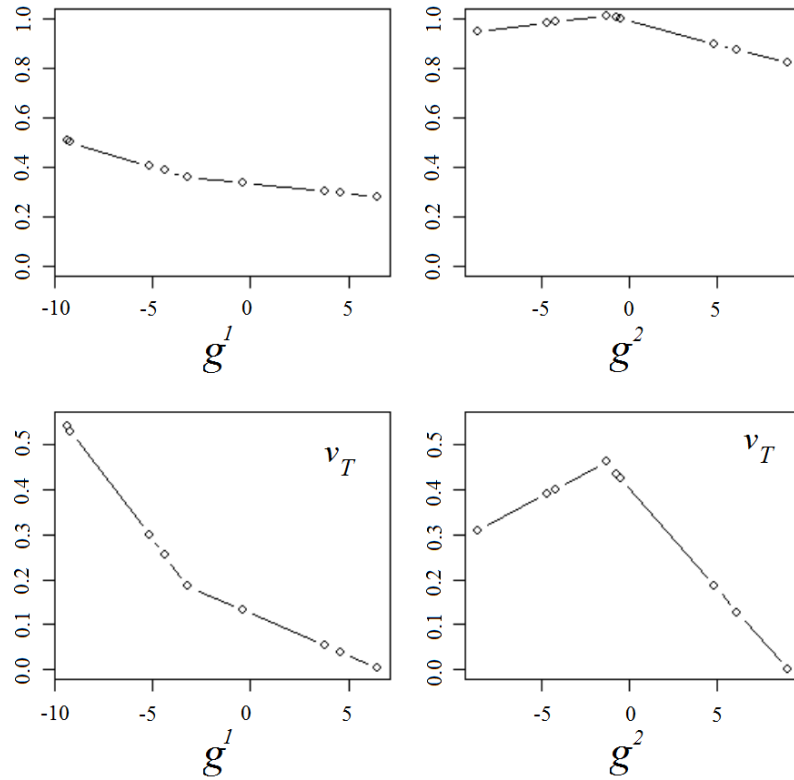
Figure 3.3.2: Inferred marginal value functions before transformation (top row) and after transformation (bottom row)

Comprehensive values of the alternatives according to the transformed value function are presented in the third column of table 3.3.3. These values yield to the same ranking for the alternatives that was obtained before. Therefore, the orders among the alternatives are preserved under the transformation. Moreover, the minimal differences in the post-transformation comprehensive values of the alternatives belonging to the strict preference relation is $\varepsilon_T^* = \varepsilon/0.416 = 0.138$.

3.3.4 A Comparison with Despotis UTA and UTA-NM

The comparison presented in this Section is performed using data provided by (Despotis and Zopounidis 1995), presented in Table 3.3.4. It consists of fifteen firms evaluated on three finan-

Table 3.3.4: Multicriteria evaluation of firms by the DM

Firm	Cash to total assets (g^1)	Long term debt and stockholder's equity to fixed assets (g^2)	Total liabilities to total assets (g^3)	rank
a	3.80	2.40	60.70	1
b	5.84	1.96	63.70	2
c	0.04	1.14	64.26	6
d	4.89	2.92	55.04	1
e	0.57	1.72	64.70	3
f	16.70	2.32	53.29	4
g	3.16	4.10	23.90	9
h	25.42	3.35	59.03	7
i	17.99	1.34	73.84	10
j	3.98	3.26	84.95	8
k	0.76	2.74	84.44	11
l	24.16	2.83	70.51	9
m	2.53	2.54	81.05	5
n	35.06	9.56	61.08	12
o	0.72	0.97	99.67	13

cial ratios. The DM ranks the firms by considering their financial performances. The results obtained by the proposed methodology, Despotis UTA (Despotis and Zopounidis 1995) and UTA-NM (Kliegr 2009) are presented and compared.

The value functions inferred by Despotis UTA, UTA-NM, and the proposed methodology are presented in Figures 3.3.3, 3.3.4, and 3.3.5, respectively.

According to the results by Despotis UTA, the first and the third criterion are in quadratic shape, and the second criteria is not relevant with a nearly zero value for the entire scale.

Similar to Despotis UTA, the marginal value function for the first criterion, g^1, estimated by UTA-NM is in quadratic shape. However, in the marginal value function by UTA-NM, there is a sudden fall from the performance value of 16.70 to 17.99, while the marginal value function by Despotis UTA is smoother and easier to interpret. There is a similar problem with the
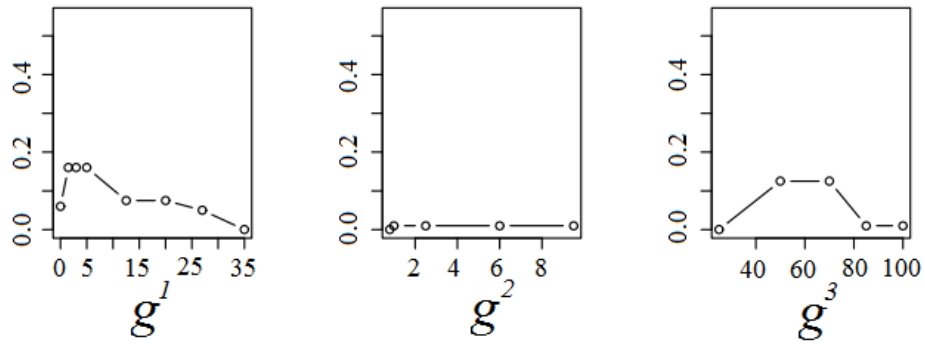
Figure 3.3.3: Estimated marginal value functions in (Despotis and Zopounidis 1995)
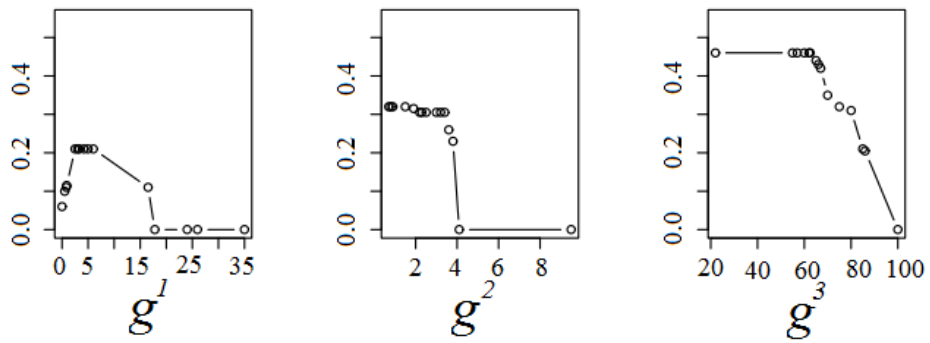
Figure 3.3.4: Estimated marginal value functions by UTA-NM

second marginal value function estimated by UTA-NM. Finally, unlike Despotis UTA, the third marginal value function estimated by UTA-NM is monotonic.
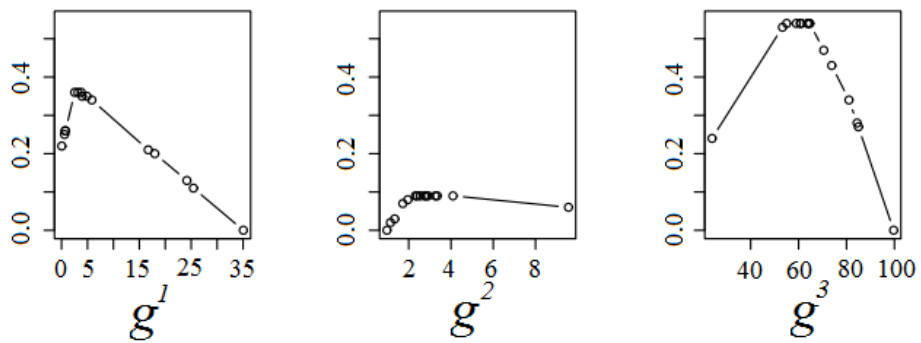
Figure 3.3.5: Estimated marginal value functions by the proposed methodology

Similar to Despotis UTA, the marginal value functions estimated by the proposed method

for the first and the third criterion are quadratic and smooth, and the second criterion is relatively irrelevant.

Also, the ranking of alternatives according to the value function derived by each of the three methodologies is plotted in Figure 3.3.6.
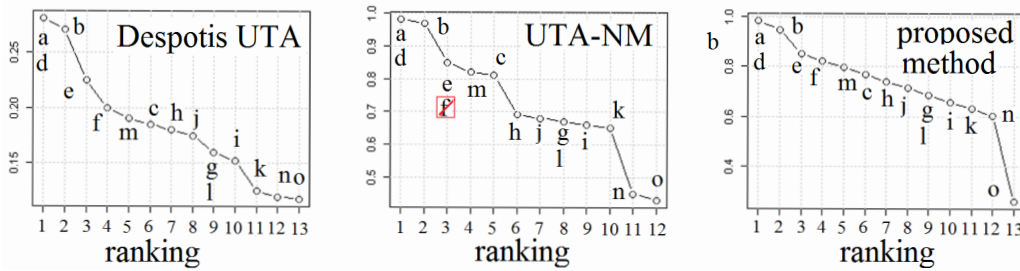
Figure 3.3.6: Comprehensive values versus ranking of alternatives obtained by each of the three methdologies

The marginal value functions estimated by the proposed methodology, as well as those estimated by Despotis UTA, are in complete agreement with the input ranking (Kendall's $\tau = 100\%$). However, unlike Despotis UTA, our method does not need any a priori knowledge about the location of the point at which monotonicity direction changes.

Kendall's τ measure for the UTA-NM result is 98.1% (alternative f is misranked) which is close to the accuracy level of the other two methods. However, UTA-NM needs to solve a complex mixed integer programming problem that incurs excessive computational cost.

The three methods obtain almost similar marginal value functions for the first criterion. The second criterion is estimated to be relatively irrelevant by Despotis UTA and the proposed methodology, but monotonic with an importance greater than 0.3 by UTA-NM. For the third criterion, Despotis UTA and the proposed methodology yield to similar results, but UTA-NM infers a monotonic value function.

3.3.5 MORE ON SLOPE-VARIATION RESTRICTION

The variable γ_j^m is used to control the variation in slope at the interior breakpoint g_j^m. The optimal values of the variables γ_j^m, however, are not easy to interpret. For this reason, in this section we go into more detail to explain how this parameter is related to the change in slope angle, which is more tangible. Suppose that θ_1 is the left slope angle and θ_2 is the right slope angle at a breakpoint g_j^m, as shown in Figure 3.3.7. The constraints in $E_{slope}^{A^R}$ imply that for a given θ_1 and γ_j^m, the slope angle θ_2 falls within the following range:

$$\arctan\left(-\gamma_j^m + \tan\theta_1\right) \leq \theta_2 \leq \arctan\left(\gamma_j^m + \tan\theta_1\right) \tag{3.22}$$

in which $\tan\theta_1 = \frac{v_j^m}{g_j^m - g_{j-1}^m}$ and $\tan\theta_2 = \frac{v_{j+1}^m}{g_{j+1}^m - g_j^m}$. Therefore, variation in slope angle, i.e. $|\theta_1 - \theta_2|$, at the breakpoint g_j^m is determined by the value of θ_1 (or θ_2) and γ_j^m. In other words, for a given γ_j^m, the slope angle varies depending on the steepness of the left slope, and for a given slope, the slope angle varies depending on the size of γ_j^m. In Figure 3.3.8, the upper and lower bounds of admissible θ_2 values at a breakpoint are computed for $\gamma_j^m = 1$ and $-90° \leq \theta_1 \leq 90°$. In this figure, the solid segment is the bisector $\theta_2 = \theta_1$.
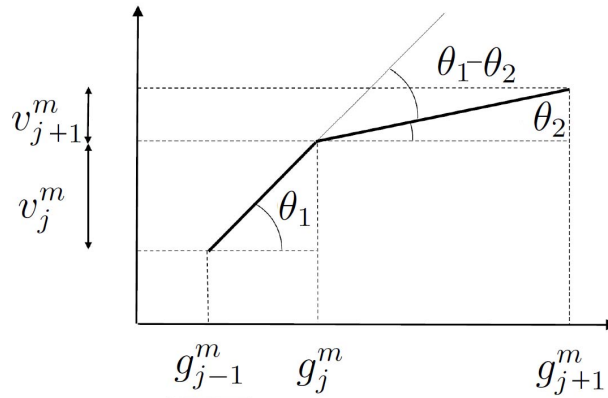
Figure 3.3.7: Change of slope angle at a breakpoint
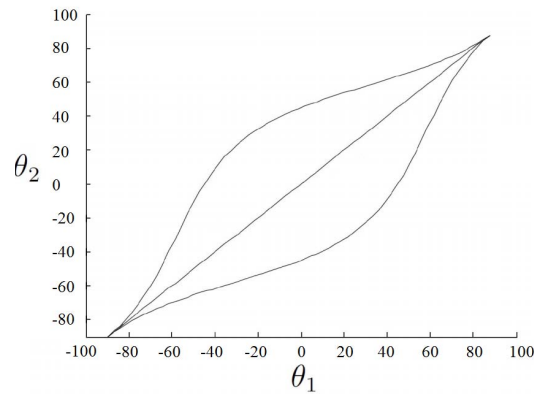
Figure 3.3.8: Range of admissible θ_2 for different θ_1 values and $\gamma = 1$

From Figure 3.3.8, it can be seen that for a fixed value of γ_j^m, slope angle variation is larger when θ_1 is close to 0, while the variation becomes smaller when the slope is steeper. This implies that radical changes in slope angles are prevented. On the other hand, the amount of variation in slope angle for a given θ_1 depends on the value of γ_j^m. In Figure 3.3.9, the ranges of acceptable θ_2 given θ_1 for different values of γ_j^m are presented.
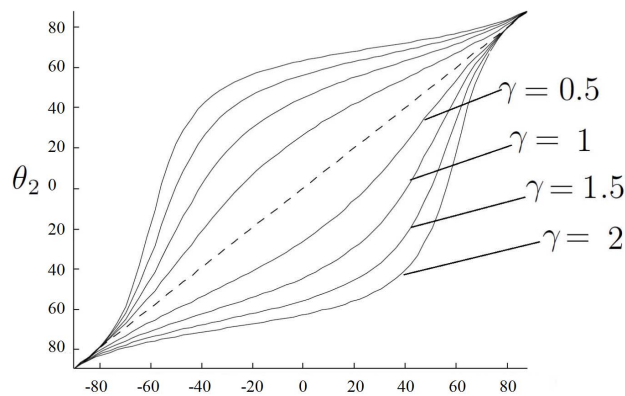
Figure 3.3.9: The role of γ in tolerance: greater γ values provide wider tolerance ranges

From the two Figures 3.3.8 and 3.3.9, it can be seen that the maximum variation in slope angle occurs when θ_1 is sufficiently close to 0. By focusing only on this range of θ_1, that is the range with maximum variation in the slope, the parameter γ_j^m can be interpreted as a

determinant of the upper bound of slope angle variation. In other words, for a given γ_j^m we can say how much, at most, the slope angle can change. To substantiate values of γ_j^m, maximal variation in slope angle is plotted against different values of γ_j^m in Figure 3.3.10. This figure helps to attach physical meanings to the values of γ_j^m. For example, it shows that in $\gamma_j^m > 2$ the upper bound of slope angle variation is near $90°$, hence 2 is considered as a "large" value for γ_j^m because at this level, radical changes in slope can occur. Moreover, it can be seen that the relationship between γ_j^m and $\max |\theta_1 - \theta_2|$ is linear for $\gamma_j^m \leq 1$, and for each 0.1 units of γ_j^m, an amount slightly less than $6°$ will be added to the upper bound of slope angle variation.
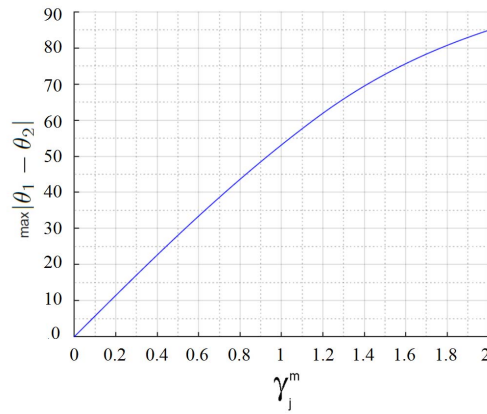
Figure 3.3.10: Upper bound of $|\theta_1 - \theta_2|$ for different γ_j^m values

3.4   EXPERIMENTAL ANALYSIS

To further examine performance of the proposed methodology, it is tested over a broad range of decision problems.

3.4.1   SETTINGS

We characterize each decision problem with a set of parameters: number of criteria (M); number of alternatives (N) that some of them will be evaluated by a simulated DM; number of

pairwise comparisons (p) made by the DM; and finally, the complexity level in judgment policy of the simulated DM (D). A larger value of D signifies a more complex value system used by the simulated DM. Later in this section we will explain how this complexity is modeled. The values considered for the parameters defining each decision problem are presented in Table 3.4.1.

Table 3.4.1: Different decision problem settings considered in the experiment

N	M	p	D
$\{6,8,10,12,14\}$	$\{2,3,4,5\}$	$\{4,6,8,10,12,$ complete ranking$\}$	$\{1,2,3,4\}$

The experiment has been conducted as follows:

Step 1: The performances of alternatives for criteria are generated randomly from a uniform distribution in the range $[-10, 10]$.

Step 2: Then p pairs of alternatives are randomly selected to be compared by the simulated DM. To guarantee that there exists at least one compatible value function for each decision setting, we generated polynomial marginal value functions with D degrees in the following form:

$$v(x) = p_0 \prod_{d=1}^{D} (\frac{x}{10} + p_d), \tag{3.23}$$

in which $p_0, p_1, \cdots, p_D \in [-1, 1]$ are random coefficients generated from a uniform distribution. Moreover, in order to better define complexity of the marginal value function by D, the coefficients are selected such that $p_0 \neq 0$ and $p_1 \neq p_2 \neq \cdots \neq p_D$. The value function is obtained by adding the marginal value functions, and the selected pairs of alternatives are evaluated according to the generated value function.

Step 3: For the generated decision problem, the optimization problem (3.13) is solved to infer a set of marginal value functions compatible with the provided pairwise comparisons. The

transformation (3.9) is then used to normalize the value function. In all decision problems in the experiment, discretization of criteria is performed by defining a breakpoint for each distinct performance value.

To increase reliability and provide statistically invariant results, each decision setting is repeated 100 times, i.e. for each configuration 100 different decision problems are generated. Overall, 48000 decision problems are solved. To measure quality of the results obtained by solving each decision problem, three different measures are used:

- γ, maximum value of $(\gamma_j^m)_T$ across all the interior breakpoints and all the criteria, i.e. $\max_{m,j}(\gamma_j^m)_T$, averaged over the 100 repetitions of the same configuration,

- $\mathcal{MC}$, maximum number of changes in monotonicity direction across all the criteria (i.e. the extreme value across the marginal value functions), averaged over the 100 repetitions of the same configuration,

- ε_T^* minimal difference in standardized comprehensive values of alternatives, as an indicator of discriminatory power of the inferred preference model, averaged over the 100 repetitions of the same configuration.

3.4.2 RESULTS

Summaries of the results are presented in Table 3.4.2. In this table, mean values of γ and $\mathcal{MC}$ show that the complexity of value functions are well-controlled. Maximum value of 0.17 for the variable γ implies that the upper bound of change in slope angle is less than $10°$, according to Figure 3.3.10. Moreover, maximum value of 2.46 for the variable $\mathcal{MC}$ implies a nearly three times change in monotonicity, which can happen in the scenarios where 4 degree polynomials are used to simulate DM preferences. Moreover, the results show that ε_T^* has a large mean value with a narrow confidence interval at 99% level of confidence. A large mean value of ε_T^* implies

that the inferred preference models have high discriminatory power, and the narrow confidence interval indicates that ε_T^* is not greatly affected by the simulation design parameters.

Table 3.4.2: Summaries of experimental results

	Mean	Minimum	Maximum	99% Confidence Interval	
γ	0.0250	0.0	0.17	0.021	0.029
$\mathcal{MC}$	0.270	0.0	2.46	0.225	0.316
ε_T^*	0.104	0.01	0.25	0.098	0.111

The impact of simulation design parameters, i.e. number of alternatives, criteria, pairwise comparisons, and degree of the polynomials simulating DM preferences, on the three variables γ, $\mathcal{MC}$, and ε_T^* is presented in Figure 3.4.1.

The figure shows that ε_T^* decreases when the number of pairwise comparisons increases; however, it is not influenced by the remaining design parameters. In fact, when more pairwise comparisons are provided by the DM, it becomes more difficult to separate the alternatives. Complexity of inferred value functions, on the other hand, increases when a greater number of alternatives or pairwise comparisons are presented in the decision problem. In this case, in order to better separate the alternatives and maintain discriminatory power, the value function needs to hold a more complex form. However, a larger number of criteria provides greater degrees of freedom for the value function in order to separate the alternatives while maintaining the level of ε_T^*. Therefore the complexity level of the value function decreases while its discriminatory power is maintained. We need to emphasize that in our experiment the criteria are assumed to be independent. The results might be different in the presence of correlated criteria and noisy data. This requires a separate investigation using a comprehensive set of real data.

Further analysis demonstrated an interaction effect between p and M. The analysis shows that for higher values of M, the association between the complexity in the inferred value func-
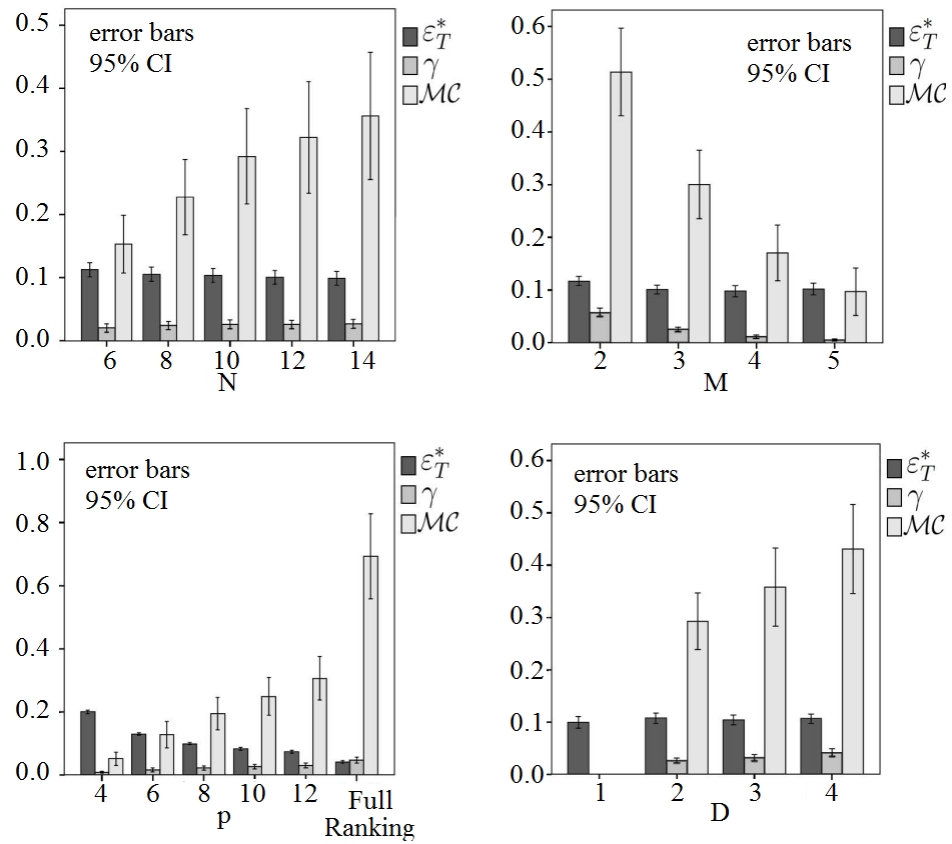
Figure 3.4.1: The impact of numbers of alternatives (N), criteria (M), pairwise comparisons (p), and degree of polynomials employed to simulate DM preferences (D) on γ, $\mathcal{MC}$, and ε_T^*

tions, $\mathcal{MC}$, and the amount of supplied preferences, p, diminishes. In other words, the effect of p on $\mathcal{MC}$ is weaker for higher values of M. This is indicated in Figure 3.4.2.

Most importantly, Figure 3.4.1 shows that complexity of value functions increases by the degree of polynomials. The figure demonstrates that when linear preference models are employed to simulate DM preferences ($D = 1$), in all the corresponding 12000 decision problems the inferred value functions are monotonic ($\mathcal{MC} = 0$) and linear ($\gamma = 0$). By increasing D, the complexity in judgment policy of the DM will be reflected in the inferred value functions.

In Figure 3.4.3, the two variables measuring complexity, γ and $\mathcal{MC}$, are standardized with respect to their mean and standard deviation, and are plotted against degrees of polynomials.
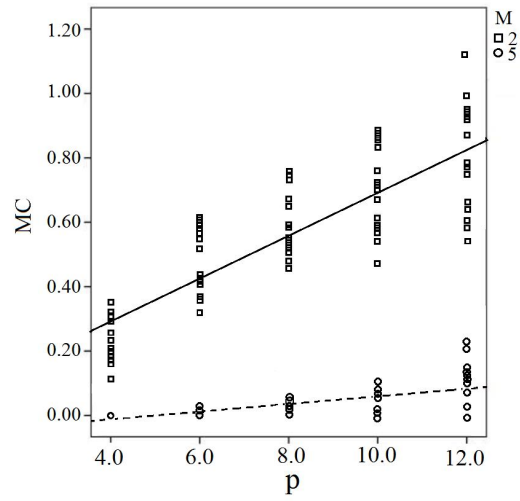
Figure 3.4.2: The effect of p on $\mathcal{MC}$ diminishes by increasing M

The plot shows that both measures similarly reflect the complexity inherent in a preferential system of the DM.
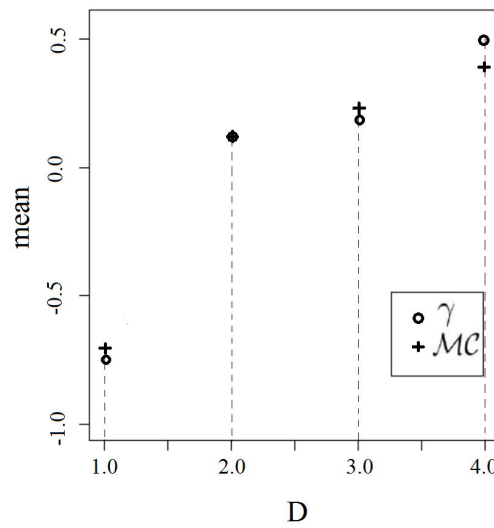
Figure 3.4.3: Average value of standardized γ and $\mathcal{MC}$ versus different degrees of polynomials employed to simulate DM preferences

3.5   An Extension of the Proposed Methodology

The two notions of slope variation and the discriminatory power of value function are conflicting. The more complex the value function, the higher the degrees of freedom it has to better separate the pairs of alternatives belonging to the strict preference relation. The experimental results show that the proposed methodology makes a good balance between these two notions by minimizing the ratio of variation in slope to the discriminatory power of the value function. In some cases, however, the analyst might prefer to further decrease the complexity of the estimated value function at the cost of its discriminatory power. In other words, in some cases the minimal difference in comprehensive values, i.e. ε_T^*, might be sufficiently large that it could seem logical to the analyst to obtain a simpler value function with less variation in slope, at the cost of decreasing the discriminatory power of the inferred compatible value function. To make this possible, the objective function in (3.13) can be adjusted as will be explained here. The objective function at (3.13) minimizes the ratio $\sum_{m,j} \gamma_j^m / \varepsilon$. Minimizing this function is equivalent to minimizing the logarithmic difference between the two terms, i.e. minimizing $\log(\sum_{m,j} \gamma_j^m) - \log(\varepsilon)$. Therefore the objective function at (3.13) can be considered as a weighted average of the logarithmic transformation of slope variation and separating threshold terms, where the two objectives have equal weights. In order to further decrease the variation in slope, at the cost of decreasing separating threshold, the objective function at (3.13) should be replaced by the following objective function in which the importances of the two objectives are adjusted:

$$w \cdot \log(\sum_{m,j} \gamma_j^m) - (1 - w) \cdot \log(\varepsilon), \tag{3.24}$$

in which $0 < w < 1$ denotes the importance of the first objective. The new optimization

problem is convex, but no longer linear. For $w = 0.5$, the problem will be equivalent to the LFP problem presented at (3.13). The results from experimental analysis in section 3.4 show that this value of w makes a good trade-off between the two objectives, decreasing complexity and increasing discriminatory power of the value function. However, if the analyst decides to sacrifice some discriminatory power to obtain a simpler value function, then the weights should be adjusted by setting $w > 0.5$. The following example illustrates the use of this argument.

3.5.1 Illustrative Example

In this section, we demonstrate how the extended version of the proposed methodology can be used. To this aim, we use the example presented in section 3.3.3. In that example, the inferred value function has a high discriminatory power ($\varepsilon_T^* = 0.14$), and one of the two inferred marginal value functions is non-monotonic. The extended methodology can be utilized to further decrease complexity of marginal value functions at the cost of discriminatory power. The value of ε_T^* is big enough for this purpose.

The value functions obtained by LFP model and for two different values of w by extended model are presented in Figure 3.5.1.

Also the ranking of alternatives resulted from the LFP model (3.13) and the two w values of the extended methodology is presented in Figure 3.5.2.

The marginal value function in Figure 3.5.1 and the numerical results in table 3.5.1 demonstrate that how by increasing w the inferred value function becomes more parsimonious, but with less discriminatory power.
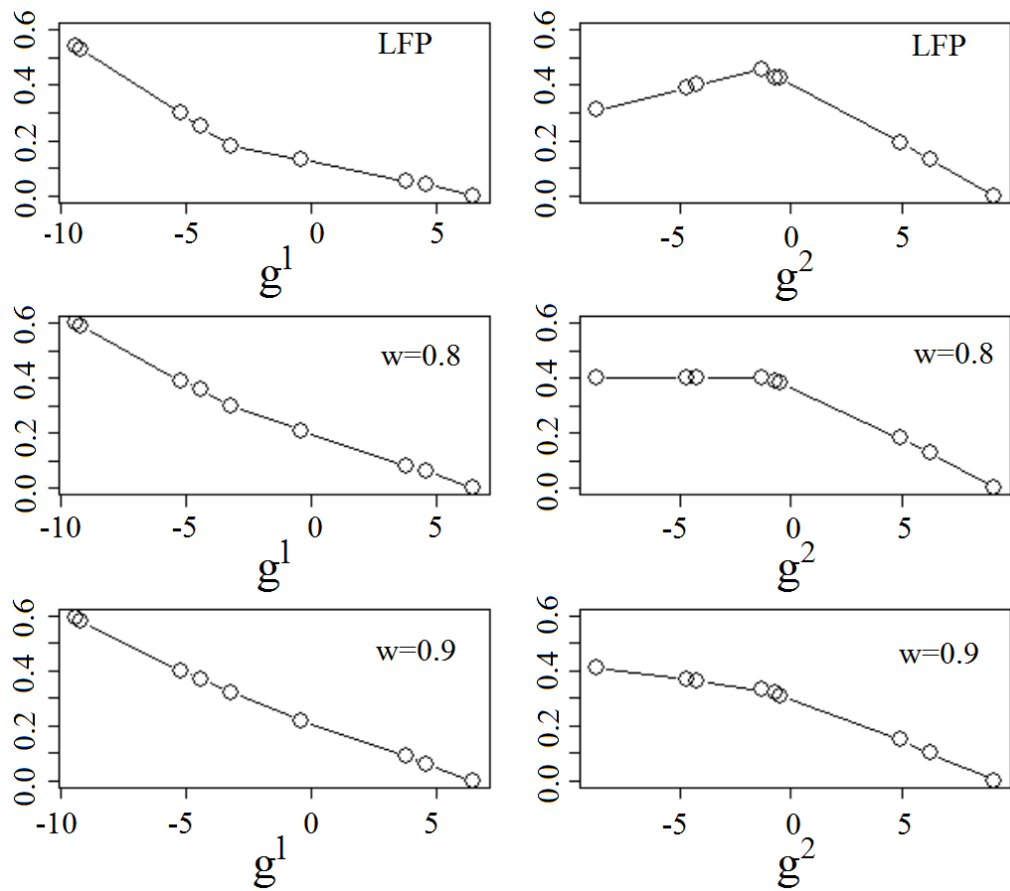
Figure 3.5.1: The inferred value functions using the LFP model (top), and extended method with $w = 0.8$ (middle) and $w = 0.9$ (bottom)
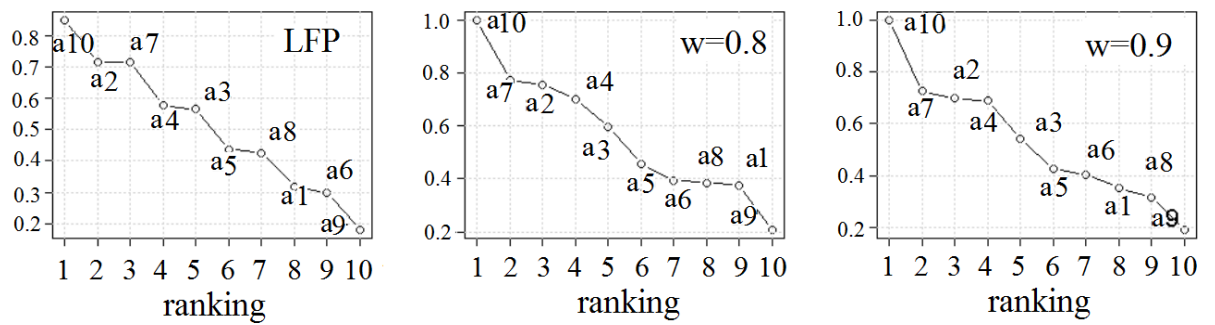
Figure 3.5.2: Ranking of alternatives obtained by the LFP model (left), the extended model with $w = 0.8$ (middle), and $w = 0.9$ (right)

Table 3.5.1: Results obtained by the LFP model (left column), the extended model with $w = 0.8$ (middle column), and $w = 0.9$ (right column)

	LFP	$w = 0.8$	$w = 0.9$
ε_T^*	0.14	0.05	0.01
$\max_j \gamma_j^m$	$\{0.04, 0.06\}$	$\{0.01, 0.02\}$	$\{0.01, 0.01\}$
$\sum_{m,j} \gamma_j^m$	0.10	0.07	0.04
MC	$\{0, 1\}$	$\{0, 0\}$	$\{0, 0\}$

3.6 Conclusions

The paper introduces a new framework for inferring non-monotonic additive value functions from a set of indirect pairwise comparisons. The complexity of the estimated value function is controlled by minimizing variation in the slope of the marginal value functions. The discriminatory power of the value function, on the other hand, is maximized to better separate the pairs of alternatives with strict preference relation, according to the preferences supplied by the DM. If some criteria are known to be monotonic a priori, the proposed methodology incorporates this information in the process of inferring the preference model. An extension of the methodology is introduced that enables the analyst to manipulate, in cases that might be reasonable, properties of the inferred preference model.

To demonstrate how the proposed methodology works, an illustrative example using synthetic data is used. The proposed methodology is also compared with Despotis UTA and UTA-NM, two well known non-monotonic UTA-like methods, using the data presented in (Despotis and Zopounidis 1995). Unlike Despotis UTA, in the proposed methodology the performance value at which the monotonicity direction changes does not needed to be specified. Unlike UTA-NM, which incurs excessive computational cost due to the mixed integer programming approach, our method needs to solve an LP problem, and is therefore applicable to realistically sized datasets.

To assess performance of the proposed methodology in different settings, a thorough experimental analysis is conducted. Considering different aspects characterizing a decision problem, $48,000$ decision problems are simulated and the quality of the inferred compatible preference model by the proposed methodology for each decision problem is evaluated using three measures. The results show that the upper bound for the 99% confidence interval of the measure $\mathcal{MC}$ is less than 0.5. This implies that in most of the cases, either all the marginal value functions are linear, or the most complex inferred marginal value function holds a quadratic form, while polynomials of up to 4 degrees were used to simulate DM preferences. In addition, in all the cases that a linear value function is employed to simulate DM preferences, the inferred compatible value function is also linear. The results also show that the complexity in the inferred value function is well controlled, and the inferred value functions have high discriminatory power. The narrow confidence interval for the measure of discriminatory power indicates that this property is not very sensitive to the characteristics of the decision problem. More detailed analysis shows that the amount of supplied preference information is the only parameter affecting the discriminatory power of the inferred value function. We emphasize again that these findings are results of this experiment, where the generated data are not noisy and the criteria are assumed to be independent, and might not apply to a general situation. This requires a separate investigation using real datasets.

Future work includes applying the proposed method to a real decision aiding problem in the presence of non-monotonic preferences, for example, dealing with aesthetic dimensions such as appearance, color and packaging shapes of a product in a marketing context. Another interesting and important avenue for future research is to test the results of the experimental analysis on larger and real data with more alternatives, as well as, inconsistent data and cases where the criteria are not independent. Extending the proposed methodology to the case of hierarchical structure of the family of evaluation criteria (Corrente et al. 2015; Corrente and Greco 2012),

as well as handling interacting criteria (Greco et al. 2014) are important directions for future research. Another important topic for future research is to test the performance of the methodology in the presence of ordinal criteria. Since the proposed methodology speaks about the slope of marginal value functions, difference between performances needs to be interpretable. It is interesting to see if quantifying ordinal scales differently will affect the performance of the methodology, and if yes, which aspects of the proposed method are more sensitive to this issue. Finally, it is also interesting to reformulate the methodology to address sorting problems.

4

Expressiveness and robustness measures for the evaluation of an additive value function in multiple criteria preference disaggregation methods: an experimental analysis[1]

An additive value function is one of the prevailing preference models in Multiple Criteria Decision Aiding (MCDA). Its indirect elicitation through pairwise questions is often applied due to

---

[1]This chapter is published in Computers & Operations Research, authored by Kadziński M., Ghaderi M., Wąsikowski J., & Agell N., ABS 3, SJR Q1, Impact Factor 1.988 (Kadziński et al. 2017)

lowering the cognitive effort on the part of a Decision Maker (DM). A practical usefulness of this approach is influenced by both expressiveness of the assumed model and robustness of the recommendation computed with its use. We experimentally evaluate the above characteristics in view of using an additive value function in the preference disaggregation context. The simulation results are quantified with the following four measures: (1) the share of decision scenarios for which a set of compatible value functions is non-empty, (2) the minimal difference between comprehensive values of reference alternatives compared pairwise by the DM, (3) the number of pairs of alternatives for which the necessary preference relation confirmed by all compatible functions holds, and (4) the number of non-trivial certain inferences which cannot be derived directly from the preference information. We discuss how these measures are influenced by the settings with different numbers of alternatives, criteria, pairwise comparisons, and performance distributions. We also study how the results change when applying various procedures for selection of the characteristic points which define the shape of per-criterion marginal value functions. In this regard, we compare four existing discretization algorithms with a new supervised technique proposed in this paper. Overall, we indicate that expressiveness and robustness are contradictory objectives, and a compromise between them needs to be reached to increase the usefulness of an additive value model in the preference disaggregation methods.

4.1   Introduction

The assessment of alternatives in real-world decision making requires consideration of a variety of criteria. This problem can be effectively tackled with Multiple Criteria Decision Aiding (MCDA) (Greco et al. 2016), which constitutes a framework to support Decision Maker (DM in structuring a decision problem and suggesting a recommendation about the decisions at stake (Zopounidis 2000). Such a recommendation derives from the comprehensive evaluation

of the considered alternatives by performing some kind of aggregation of their performances on multiple criteria. Within MCDA , various preference models can be used for this purpose, see, e.g., (de Almeida 2007; Kadziński et al. 2015; Sarabando and Dias 2010; Słowiński et al. 2002; Soylu 2011; Vetschera and de Almeida 2012; Zheng et al. 2014), but the most prevailing one is Multi-Attribute Value Theory (MAVT) (Keeney and Raiffa 1993).

Additive value functions represent preferences of the DM by computing an overall score for each alternative (Vetschera et al. 2014). Such comprehensive value indicates the alternative's quality from all relevant points of view considered jointly. MAVT is often found appropriate for practical decision support due to the high interpretability of numerical scores that can be decomposed into per-criterion marginal values and easily explained due to low amount of inter-criteria parameters (Greco et al. 2012; Sarabando and Dias 2010).

Elicitation of an additive value function can be conducted in a direct or indirect way. Direct elicitation requires specification of the preference model parameters and, thus, may involve some greater cognitive effort from the DM (Kadziński and Tervonen 2013). For lowering the elicitation effort, questions concerning exemplary decisions are often employed. As noted in (Jacquet-Lagrèze and Siskos 2001), with an indirect elicitation of preferences, the preference model is constructed so that to find a rational basis underlying the provided holistic judgments. In the context of multiple criteria ranking, preference disaggregation based on pairwise preference questions has been considered in the UTA-like methods (Corrente et al. 2013, 2016a; Figueira et al. 2009; Greco et al. 2008; Jacquet-Lagrèze and Siskos 1982, 2001; Siskos and Grigoroudis 2010; Siskos et al. 2005).

Indirect elicitation of preferences raises two practical concerns, which are not present when using the direct questioning. The first one is related to the potential incompatibility of pairwise comparisons provided by the DM with an assumed preference model (Mousseau et al. 2003). In case of incompatibility, the inferred model is not able to restore all supplied holistic

judgments. The other concern derives from the incompleteness of pairwise comparisons of reference alternatives and indetermination in the definition of a single compatible value function. That is, when the preference information provided by the DM can be fully reconstructed with an additive value function, there is typically more than a single compatible function (Bous et al. 2010; Corrente et al. 2013, 2014; Figueira et al. 2009; Greco et al. 2008; Kadziński and Michalski 2016). When applied on the set of alternatives, these functions may lead to different rankings. This, in turn, raises concerns about the stability of the suggested recommendation.

In this perspective, the usefulness of MAVT-based preference disaggregation methods in practical decision aiding is much influenced by two characteristics: expressiveness of the underlying preference model and robustness of the recommendation computed with its use. Expressiveness is related to the ability of a model to reconstruct preference information provided by the DM. The more expressive the model is, the greater the variety and quantity of decision policies it can be used to represent. Robustness refers to the analysis of uncertainties and imperfections observed in the actual decision support processes. As noted in (Kadziński and Tervonen 2013), investigation of robustness of the provided conclusions consists in verifying whether they are valid for all acceptable or for the most plausible instances of an assumed preference model. Obviously, the greater the robustness of the recommendation that can be derived with a particular preference model after taking into account uncertainties in the indirect inference of its parameters, the better.

This paper aims to experimentally evaluate the expressiveness of an additive value function and robustness of the recommendation computed with its use in a preference disaggregation setting. First, we discuss some measures that can be used to quantify the results of such an investigation. When it comes to the expressiveness, we refer to the percentage of randomly generated decision scenarios where consistency between the provided pairwise comparisons and an assumed preference model is observed. In case the random preference statements are

guaranteed to be non-conflicting, we also measure the minimal difference between comprehensive values of reference alternatives compared pairwise by the DM. As far as robustness is concerned, it is quantified with the number of pairs for which the necessary preference relation $\succsim^N$ confirmed by all compatible additive value functions holds (Greco et al. 2008). In this regard, we additionally distinguish the non-trivial inferences by counting the pairs related by $\succsim^N$ that is not part of the transitive closure and is not discoverable with a condition for the necessary preference inference based on a single preference statement (Spliet and Tervonen 2014).

Secondly, we verify how these four measures are affected by different values of the parameters characterizing a decision situation. These include the numbers of criteria, alternatives, and pairwise comparisons, as well as the distribution of alternatives' performances. A special attention is paid to investigating the impact of a parameterization of per-criterion marginal value functions. Indeed, they can be either piece-wise linear with different pre-defined numbers of characteristic points (i.e., the points where a slope angle of the function can change) (Siskos et al. 2005), or general with all unique performances corresponding to the characteristic points (Greco et al. 2008).

Thirdly, we present different methods for selecting the characteristic points. Apart from considering a state-of-the-art procedure with equal interval widths between all adjacent pairs of characteristic points (Siskos et al. 2005), we adapt for this purpose some other discretization techniques such as equal frequency binning, k-means clustering (MacQueen 1967), or kernel-based discretization (Biba et al. 2007). We also propose a new supervised algorithm which can be used for the context-specific discretization in case pairwise comparisons are provided at the method's input. The latter technique uses the characteristic points to separate as many pairs of performances corresponding to the reference alternatives compared pairwise by the DM as

possible. These five procedures are also compared in terms of their impact on the expressiveness and robustness measures.

Although there exist some previous studies investigating expressiveness (Pirlot et al. 2010) or robustness (Spliet and Tervonen 2014), our research is unique in a sense of providing a view on the trade-off between these two measures. Moreover, it is also richer in terms of systematically investigating the impact of a broad set of parameters describing both a decision situation and a value function itself.

The organization of the paper is the following. In the next section, we remind the existing MAVT-based disaggregation methods. In Section 4.3, we present five procedures that can be applied for selection of the characteristic points, and illustrate how they work. Section 4.4 is devoted to the measures that can be used for assessing expressiveness of an additive value function and robustness of the recommendation computed with its use. In Section 4.5, we report the outcomes of our experimental study. The last section concludes the paper, indicating avenues for future research.

4.2   REMINDER ON PREFERENCE DISAGGREGATION VALUE-BASED METHODS

We use the following notation (Corrente et al. 2013):

- $A = \{a_1, \ldots, a_i, \ldots, a_n\}$ - a finite set of n alternatives;

- $A^R = \{a^*, b^*, \ldots\} \subseteq A$ - a finite set of reference alternatives for which the DM accepts to provide incomplete pairwise comparisons;

- $G = \{g_1, \ldots, g_j, \ldots, g_m\}$ - a finite set of m evaluation criteria, $g_j : A \to \mathbb{R}$;

- $X_j = \{g_j(a_i), \ a_i \in A\}$ - the set of deterministic evaluations (performances) on g_j; we assume, without loss of generality, that the greater $g_j(a_i)$, the better;

- $x_j^1, \ldots, x_j^{n_j(A)}$ - the ordered values of X_j, $x_j^k < x_j^{k+1}$, $k = 1, \ldots, n_j(A) - 1$, where $n_j(A) = |X_j|$ and $n_j(A) \leq n$.

Preference model. To model preferences of the DM, we use an additive value function:

$$U(a) = \sum_{j=1}^{m} u_j(a) \in [0, 1], \tag{4.1}$$

where the marginal functions u_j are defined through a set of γ_j characteristic points g_j^s, $s = 1, \ldots, \gamma_j$. The marginal values assigned to the characteristic points need to satisfy the monotonicity constraints.

Preference information. We expect the DM to provide a set B^R of holistic pairwise judgments on A^R in form of a weak preference relation $(a^* \succsim b^*)$. Strict preference $\succ$ is the asymmetric part of $\succsim$, and indifference $\sim$ is its symmetric part. Moreover, for each criterion g_j, one needs to select between considering a piece-wise linear value function with γ_j characteristic points or a general value function with the characteristic points defined by all performances in X_j. In the previous case, $g_j^1 = x_j^1$ and $g_j^{\gamma_j} = x_j^{n_j(A)}$. Moreover, $\gamma_j = 2$ indicates a linear function. When using the general functions, $\gamma_j = n_j(A)$ and $g_j^s = x_j^s$, $s = 1, \ldots, \gamma_j$.

Preference disaggregation. The pairwise comparisons provided by the DM are represented with set $\mathcal{U}$ of compatible value functions being able to reconstruct them. This set is defined

with the following constraints:

$$
\left.
\begin{aligned}
& U(a^*) \geq U(b^*) + \varepsilon, \;\; \text{if } a^* \succ b^* \text{ for } a^*, b^* \in A^R, \\[4pt]
& U(a^*) = U(b^*), \;\; \text{if } a^* \sim b^* \text{ for } a^*, b^* \in A^R, \\[4pt]
& U(a^*) \geq U(b^*), \;\; \text{if } a^* \succsim b^* \text{ for } a^*, b^* \in A^R, \\[4pt]
& u_j(g_j^1) = 0, \; \sum_{j=1}^{m} u_j(g_j^{\gamma_j}) = 1, \\[4pt]
& \text{for all } j = 1, \ldots, m \text{ and } s = 2, \ldots, \gamma_j : \\[4pt]
& u_j(x_j^s) \geq u_j(x_j^{(s-1)}),
\end{aligned}
\right\} \; E(\mathcal{U})
\qquad (4.2)
$$

where ε is a variable whose value will be subsequently optimized to check if $\mathcal{U}$ is non-empty.

For piece-wise linear value functions, the marginal values corresponding to $x_j^k \in [g_j^{s-1}, g_j^s)$, $k = 1, \ldots, n_j(A)$, are defined with linear interpolation:

$$
u_j(x_j^k) = u_j(g_j^{s-1}) + (u_j(g_j^s) - u_j(g_j^{s-1}))(x_j^k - g_j^{s-1})/(g_j^s - g_j^{s-1}). \qquad (4.3)
$$

Consistency check. To verify whether $\mathcal{U}$ is non-empty, the following Linear Programming (LP) model needs to be solved:

$$
Maximize \; \varepsilon, \;\; \text{s.t. } E(\mathcal{U}). \qquad (4.4)
$$

Let us denote with ε^* the optimal solution of problem (4.4). If $E(\mathcal{U})$ is feasible and $\varepsilon^* > 0$, there exists at least one compatible value function able to reproduce the pairwise comparisons provided by the DM, i.e., $\mathcal{U} \neq \emptyset$. Otherwise, if $E(\mathcal{U})$ is infeasible or $\varepsilon^* \leq 0$, then $\mathcal{U}$ is empty, which indicates inconsistency between the holistic judgments and an assumed form of the additive model.

Certainty check. Each compatible value function determines a complete ranking of alternatives. When using all functions in $\mathcal{U}$ on A, the order of some alternatives can be ambiguous. The certain part of the recommendation confirmed by all compatible value functions is defined with a necessary weak preference relation $\succsim^N$ (Greco et al. 2008):

$$a \succsim^N b \Leftrightarrow a \succsim b \text{ for all } U \in \mathcal{U}. \tag{4.5}$$

To verify whether this relation holds for a pair $(a, b) \in A \times A$, the following LP program needs to be solved:

$$\textit{Maximize } \varepsilon, \text{ s.t. } E^N(a, b), \tag{4.6}$$

where $E^N(a, b) = \{E(\mathcal{U})\} \cup \{U(b) \geq U(a) + \varepsilon\}$. Let us denote with $\varepsilon_N^*(a, b)$ the optimal solution of problem (4.6). Assuming the set of compatible value functions is non-empty, $a \succsim^N b$ if $\varepsilon_N^*(a, b) \leq 0$.

4.3 Selecting Characteristic Points for Marginal Value Functions

When using piece-wise linear value functions, the extreme characteristic points correspond to the worst and the best performances. However, if $\gamma_s > 2$, the selection of $\gamma_s - 2$ intermediate points is arbitrary. In this section, we discuss five procedures that can be applied for this purpose, and illustrate how they work.

4.3.1 Procedures for Selection of the Characteristic Points

Equal width binning (EWB). The equal width binning technique is an unsupervised algorithm that divides the range $[x_j^1, x_j^{n_j(A)}]$ into $\gamma_j - 1$ equal sub-intervals with the endpoints:

$$g_j^s = x_j^1 + (x_j^{n_j(A)} - x_j^1)(s - 1)/(\gamma_j - 1), \ s = 1, \ldots, \gamma_j. \tag{4.7}$$

This procedure has been traditionally used in the UTA-like methods (Siskos et al. 2005).

Equal frequency binning (EFB). When accounting for the distribution of performances, we construct $\gamma_j - 1$ sub-intervals so that the number of performances assigned to each of them is the same, i.e., $n_j(A)/\gamma_j$. If such a perfect distribution is not attainable, i.e., $n_j(A)/\gamma_j$ is not an integer, we balance the distribution so that the numbers of performances assigned to different bins differ by at most one. In such a case, the sub-intervals corresponding to relatively worse evaluations accommodate more performances. For example, with $\gamma_j = 4$ and 10 unique performances, the three sub-intervals would contain, respectively, 4, 3, and 3 observations. The characteristic point separating two adjacent sub-intervals is selected so that its distances from the best performance in the left bin and the worst performance in the right bin are equal.

Discretization using k-means clustering (KMC). The set of $n_j(A)$ unique performances can be divided into $\gamma_j - 1$ sub-intervals using k-means clustering algorithm with $k = \gamma_j - 1$ (MacQueen 1967). The underlying motivation is to represent similar performances (contained in the same group) using a single piece of a marginal function, and to use the characteristic points for separating these groups. Thus, analogously as in EFB, we set the characteristic point separating two adjacent groups as a mid-point between the extreme performances observed in these groups. Since a slope angle of the marginal value function is allowed to change in the char-

acteristic point, using KMC it should be possible to better emphasize the difference between marginal values assigned to the performances in various groups.

The algorithm of separating the performances into k groups is initialized with $k = \gamma_j - 1$ randomly generated values in $[x_j^1, x_j^{n_j(A)}]$, playing the role of so called centroids. Then, it iteratively performs the following steps: (1) for each performance, calculate its distance from the current centroids, (2) reassign each performance to a group with the centroid that is nearest to it, and (3) compute new centroids by averaging all performances assigned to a particular group. The algorithm is stopped when no single performance is reallocated in the following two iterations.

Discretization using kernel density estimation (KDE). The aim of a kernel density estimation is to construct a set of sub-intervals that best fit the data using the non-parametric density estimators (Biba et al. 2007). In each iteration, the algorithm selects the best cut-point for splitting one of the current intervals. At the beginning, just a single interval corresponding to the entire performance range is considered. The $N = n_j(A) - 1$ candidates are the middle points between adjacent performances. The evaluation T of each candidate derives from applying two density functions p and f:

$$Score(T) = \sum_{i=1}^{k} (p(x_i) - f(x_i)) + \sum_{i=k+1}^{N} (p(x_i) - f(x_i)), \qquad (4.8)$$

where the left (right) part of the external sum refers to the performances that fall into the left (right) interval. Moreover, f is a simple binning density function:

$$f(x_i) = m_{ins}/(wN), \qquad (4.9)$$

where m_{ins} is the number of instances in a single bin (left or right), and w is a bin-width. Finally,

p is a kernel density function that differentiates the instances in the same bin based on their distances to x_i:

$$p(x_i) = \frac{1}{hN} \sum_{j=1}^{N} K\left(\frac{x_i - x_j}{h}\right), \tag{4.10}$$

where h is a bandwidth, K is a kernel function, and x_j represents another point in the same bin as x_i. Following (Biba et al. 2007), we assumed $w = h$ and defined K as a Gaussian kernel function:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}. \tag{4.11}$$

Finally, an interval with the largest score T is selected for splitting. The algorithm is repeated until $\gamma_j - 2$ splits are conducted.

Spliting supervised by pairwise preference relations (SSP). Let us present a new supervised algorithm for selecting the characteristic points. It aims to select the characteristic points so that to separate the maximal number of performances corresponding to the pairs of reference alternatives for which the DM indicated a strict preference relation. The underlying motivation is to involve more decision variables (corresponding to the marginal values assigned to the characteristic points) in the linear constraints translating these pairwise comparisons. We predict this to be beneficial for discriminating the comprehensive values of reference alternatives.

Let $P_j = \{p_j^1, \ldots, p_j^l, \ldots, p_j^L\}$ denote a set of all candidates for the characteristic points defined as the mid-points between each pair of adjacent performances. Then, $v_l \in \{0, 1\}, l = 1, \ldots, L$, is a binary variable which is instantiated with one if p_j^l is selected as a characteristic point. Let $B_\succ^R$ represent all pairs of indices (q, r) corresponding to the reference alternatives $a_q, a_r \in A^R$, such that $a_q^* \succ a_r^*$. Then, the number of selected characteristic points between $g_j(a_q)$ and $g_j(a_r)$ is:

$$f_{q,r} = \sum_{l\,:\,min\{g_j(a_q), g_j(a_r)\} \leq p_j^l \leq max\{g_j(a_q), g_j(a_r)\}} v_l. \tag{4.12}$$

Thus, to maximize the overall number of performances corresponding to the pairs of reference alternatives related by $\succ$ which are separated by the characteristic points, we have to solve the following Mixed-Integer Linear Programming (MILP) problem:

$$\text{Maximize } F = \sum_{(q,r) \in B_\succ^R} f_{q,r}, \text{ s.t. } \sum_{l=1}^{L} v_l = \gamma_j - 2. \tag{4.13}$$

4.3.2   ILLUSTRATIVE STUDY

In this section, we illustrate the use of procedures presented in Section 4.3.1 for selection of the characteristic points in the context of an exemplary decision problem. It involves 14 alternatives, $A = \{a_1, \ldots, a_{14}\}$, evaluated in terms of 4 criteria, $G = \{g_1, \ldots, g_4\}$. All performances were randomly generated from the uniform distribution (see Table 4.3.1) so that the dominance relation was empty. For each criterion, the number of characteristic points was set to 5; thus, each marginal value function is composed of 4 linear pieces. We also generated the following 8 pairwise comparisons provided by a hypothetical DM:

$$a_5 \succ a_{11}, \; a_6 \succ a_3, \; a_6 \succ a_4, \; a_9 \succ a_{13}, \; a_{10} \succ a_4, \; a_{10} \succ a_{11}, a_{13} \succ a_8, \; a_{14} \succ a_{12}. \tag{4.14}$$

In Figure 4.3.1, we present the marginal value functions for all criteria with the characteristic points selected by five different methods. The presented functions are the most discriminant ones, obtained by solving problem (4.4). To support understanding of the logic underlying each algorithm, we marked the performances of alternatives with the vertical lines, and represented the performances of reference alternatives compared pairwise by the DM as the extreme points of the horizontal lines in the upward part of the figures.

As desired, EWB selected the characteristic points so that to construct four linear pieces with
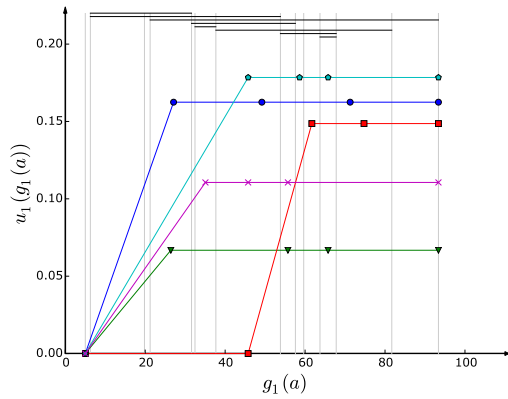
(a) marginal function on g_1: u_1
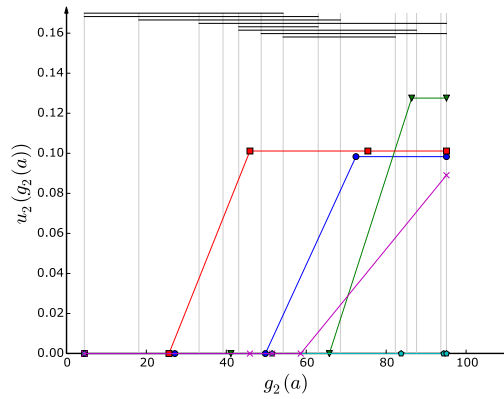
(b) marginal function on g_2: u_2
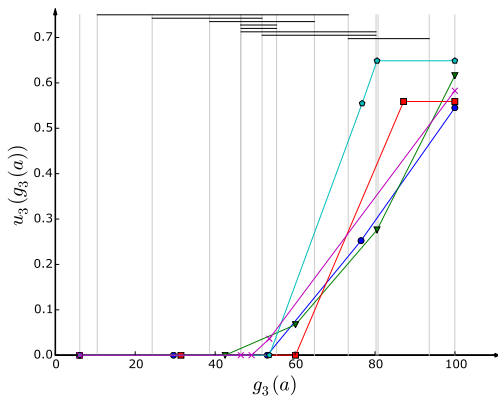
(c) marginal function on g_3: u_3
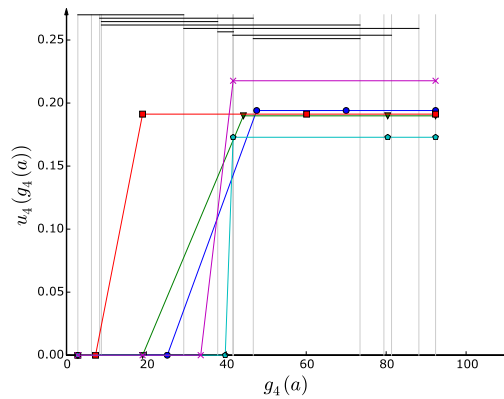
(d) marginal function on g_4: u_4

Figure 4.3.1: Marginal value functions for g_1, g_2, g_3, and g_4; different colors (see online version) and markers represent five methods for selection of the characteristic points: blue (circle) – EWB; green (triangle down) – EFB, red (square) – KMC; azure (pentagon) – KDE; purple (cross) – SSP. Gray vertical lines represent performances of alternatives; extreme points of the black horizontal lines represent the performances of reference alternatives compared pairwise by the DM.

Table 4.3.1: Randomly generated performances for 14 alternatives evaluated in terms of 4 criteria.

	g_1	g_2	g_3	g_4
a_1	5.01	93.69	99.97	79.44
a_2	19.81	39.18	6.12	92.36
a_3	57.58	87.60	46.35	8.32
a_4	6.28	63.01	46.49	73.46
a_5	63.73	82.30	24.25	41.80
a_6	31.59	43.11	55.37	46.75
a_7	59.65	85.19	80.75	6.26
a_8	81.69	48.72	10.51	88.19
a_9	32.45	33.19	93.57	2.86
a_{10}	53.83	4.49	80.25	8.76
a_{11}	67.79	54.23	51.74	37.86
a_{12}	93.37	18.12	38.63	81.34
a_{13}	37.68	95.09	73.30	29.36
a_{14}	21.23	68.56	64.84	41.62

equal widths. For example, since $x_2^1 = 4.49$ and $x_2^{14} = 95.09$, the width of each sub-interval is equal to 22.65. Obviously, the results provided by EWB depend neither on the distribution of performances nor on the provided preference information. On the contrary, the former aspect is accounted by EFB. The characteristic points selected by this approach divide the performance range into sub-intervals accommodating approximately the same number of performances. Indeed, for each marginal value function, 4 or 3 unique performances correspond to its first two and last two linear pieces, respectively.

Furthermore, the k-means clustering constructs sub-intervals accommodating similar performances irrespective of their number. Thus, some of the linear pieces generated with this method accommodate a significant share of performances (see, e.g., the first piece for u_1), while for some other pieces the number of respective performances is significantly lower (see, e.g., the third piece for u_1). In all cases, the characteristic points have been selected as the midpoints between extreme performances mapped to the neighboring linear pieces. For example, since the greatest performance mapped to the first piece for u_1 is 37.68 and the least perfor-

mance mapped to the second piece is 53.83, the selected characteristic point corresponds to $(37.68 + 53.83)/2 = 45.755$.

When it comes to KDE, it tries to find the best cut-points by accounting for the density induced in the sub-intervals by the current cut and kernel density of constructed intervals. As a result, the method selects characteristic points corresponding to the performances with high density, thus, often separating pairs of performances which are very close to each other. Finally, SSP constructs the characteristic points so that to separate as many pairs of performances of reference alternatives compared pairwise by the DM as possible. For example, for u_1 and u_2 the numbers of such pairs separated by the three intermediate characteristic points are equal to 12 and 20, respectively.

4.4 Assessing Expressiveness and Robustness in Preference Disaggregation Value-based Methods

4.4.1 Measures for Assessing Expressiveness of an Additive Value Function and Robustness of the Recommendation Computed with its Use

The aim of our study was to investigate expressiveness of an additive value function and robustness of the recommendation computed with its use in a preference disaggregation setting. On one hand, expressiveness is related to the ability of a preference model to reconstruct indirect preference information provided by the DM. On the other hand, robustness refers to the impact of indetermination in the definition of a single compatible value function on the recommendation delivered by the method. To quantify the expressiveness, we used the following two measures:

- The percentage $\%(\mathcal{U} \neq \emptyset)$ of randomly generated decision scenarios for which the

pairwise comparisons provided by the DM are fully consistent with an assumed preference model, i.e., when there exists at least one compatible value function. The greater $\%(\mathcal{U} \neq \emptyset)$, the greater flexibility of an additive value function in representing holistic judgments derived from different decision policies or DMs' value systems.

- The minimal difference ε^* between comprehensive values of reference alternatives related by a strict preference relation by the DM in case all supplied pairwise comparisons can be reproduced. This measure reflects the ability of an additive value function for discriminating reference alternatives related by $\succ$. For a single decision situation, its value is derived from the optimal solution of problem (4.4). Note that different values of ε^* are comparable. Being defined on the conjoint interval scale of the marginal value function (Krantz et al. 1971; Wakker 1989), they have the meaning of intensity.

As far as robustness is concerned, it is quantified with the following two measures:

- The number $| \succsim^N |$ of pairs for which the necessary preference relations (Greco et al. 2008) holds refers to the definition of robust conclusions which need to be confirmed by all value functions compatible with the provided pairwise comparisons (Kadziński and Tervonen 2013). Obviously, the greater $| \succsim^N |$, the richer is the most certain part of the recommendation delivered with the preference disaggregation method.

- The number $| \succsim_I^N |$ of pairs for which the inferred necessary preference relation does not directly follow the provided preference information (called also "the number of non-trivial necessary inferences"). Spliet and Tervonen (Spliet and Tervonen 2014) formulated a theorem on "a single statement inference" which indicates the conditions under which the necessary preference relation for a given pair of alternatives would hold referring only to the performances of alternatives. We wish $| \succsim_I^N |$ to indicate the number

of pairs for which $\succsim^N$ is instantiated with a contribution of the employed preference model rather than based on the performances only. Thus, $\mid \succsim^N_I \mid$ neither includes inferences derived from the aforementioned theorem nor these which can be obtained from the transitivity of the preference relation (indeed, if $a \succsim b$ and $b \succsim c$, then $a \succsim^N c$).

4.4.2 SIMULATION DESIGN

The experimental study has been conducted in the following way:

Step 1: Generate a simulated decision problem with a pre-defined number of criteria m and alternatives n. The performances are generated randomly from the $(0, 100)$ range. To ensure that the dominance relation is empty, we redraw the performance matrices for which this condition is not satisfied. In this way, we do not violate the dominance principle with any pairwise comparison generated in Step 2.

Step 2: Simulate the DM's ranking or r pairwise comparisons stating the truth of a strict preference relation $\succ$. For the verification of consistency of pairwise comparisons with an assumed model, such holistic preference information is drawn from a randomly generated permutation of alternatives. For investigating the remaining measures, to guarantee that there exists at least one compatible value function for all considered scenarios, we (1) generate linear marginal value functions with uniformly distributed weights, (2) determine a ranking derived from their use, and (3) draw a pre-defined number of pairwise comparisons from this ranking.

Overall, we considered 5500 different problem settings (see Table 4.4.1). They are distinguished by the numbers of alternatives (n; ranging from 6 to 14), criteria (m; ranging from 2 to 5), pairwise comparisons (r; ranging from 2 to 8, or a complete ranking of all alternatives) and characteristic points for each criterion (γ_j; ranging from 2 (linear function) to 6, or general function (G)), distribution of the performances (uniform or skew normal[2]), and a method

---

[2]To generate the random performances from the skew normal distribution, we used the "rsn" function from

Table 4.4.1: Different problem settings considered in the experimental analysis.

char. points	criteria	numbers of alternatives	pairwise comparisons	performance distribution	discretization method
$\{2, G,$ $3, 4, 5, 6\}$	$\{2, 3, 4,$ $5, 6\}$	$\{6, 8, 10,$ $12, 14\}$	$\{2, 4, 6, 8$ complete ranking$\}$	$\{$uniform, skew normal$\}$	$\{$EWB, EFB, KMC, KDE, SSP$\}$

employed for selection of the characteristic points (EWB, EFB, KMC, KDE, or SSP). Note that the discretization methods are used only if $\gamma_j \in \{3, 4, 5, 6\}$ (see Table 4.4.1).

Step 3: For the analysis of consistency between the DM's pairwise comparisons and an assumed preference model, solve problem (4.4), and note whether a set of compatible value functions is non-empty. For computing the minimal difference between comprehensive values of reference alternatives compared pairwise by the DM, solve (4.4) and note the optimal solution ε^*. Then, for each pair of alternatives, verify it is related by the necessary preference $\succsim^N$ by solving problem (4.6). Finally, distinguish the non-trivial necessary inferences by analyzing the transitivity of the preference information and the consistency with the single inference theorem.

For each problem setting, the above scenario has been repeated 100 times. This number of repetitions was verified to provide statistically invariant results.

Technical details. The experiments were implemented in R, and for solving LP problems, we used Rglpk library (Theussl and Hornik 2015). The computations were performed on the c4.8xlarge machine with 36 cores of Intel® Xeon® Processor (E5-2666 v3) purchased from Amazon EC2 service. A total execution time was about 4 days ($\sim$ 140 days on a single core).

---

the SN package `http://azzalini.stat.unipd.it/SN` with the following parameters: location $\xi = 5$, scale $\omega = 2$, and shape $\alpha = 5$. The implementation of this package is based on (Azzalini 1985).

4.5 Results

In this section, we report the outcomes of our experimental study concerning expressiveness of an additive value function and robustness of the recommendation computed with its use. We refer to the four measures defined in Section 4.4. Their values are averaged over different experimental runs so that to show how expressiveness and robustness depend on the parameters characterizing a decision problem (such as the number of alternatives, criteria, and pairwise comparisons, or performance distribution) or a preference model (such as the number of characteristic points or the method for their selection).

In Section 4.5.1, we compare linear, general, and piece-wise linear marginal value functions with different numbers of characteristic points. For clarity of presentation, in this part we select the characteristic points only with EWB. This procedure has been traditionally used in the UTA-like methods. In Section 4.5.2, we deepen this comparison by discussing the results of a statistical analysis concerning four different settings distinguished by the low and high numbers of criteria and characteristic points as well as the low and high numbers of alternatives and pairwise comparisons. In Section 4.5.3, we focus on the analysis of piece-wise linear functions, while highlighting the impact of five different methods for selection of the characteristic points. In Section 4.5.4, we discuss the trade-off between expressiveness and robustness, where in Section 4.5.5, we check how the results are influenced by either poor or rich preference information.

EXPRESSIVENESS

In this section, we focus on two measures for assessing how expressiveness of an additive value function is affected by different characteristics of a decision situation. In Figure 4.5.1, we refer to the percentage $\%(\mathcal{U} \neq \emptyset)$ of randomly generated decision situations for which the pairwise comparisons provided by the DM are fully consistent with an assumed preference model, while Figure 4.5.2 accounts for the minimal difference ε^* between comprehensive values of reference alternatives related by a strict preference by the DM.

First, when increasing the number of alternatives or the number of pairwise comparisons, the expressiveness of an additive value function decreases. In the former case, this is due to involving greater number of alternatives in the provided pairwise comparisons, whereas when more pairwise comparisons are supplied quite naturally the chances that they are all reproduced or strongly discriminated become smaller.

On one hand, the impact of the number of alternatives on $\%(\mathcal{U} \neq \emptyset)$ and ε^* is rather marginal. For example, for all shapes of marginal value functions, the differences between average $\%(\mathcal{U} \neq \emptyset)$ or ε^* values observed for 6 and 14 alternatives are less than 10% and 0.05, respectively. On the other hand, the influence of additional pairwise comparisons is significant. For example, the linear value functions exhibit about 60% difference in $\%(\mathcal{U} \neq \emptyset)$ for the problems with 2 and 8 pairwise comparisons, while the difference between ε^* values obtained for 2 and 4 comparisons is more than 0.2. When referring to ε^*, the more general the considered marginal value functions, the more sharp the observed decrease in the expressiveness.

As a result, the functions involving greater number of characteristic points are relatively more discriminant when preference disaggregation involves few pairwise comparisons.
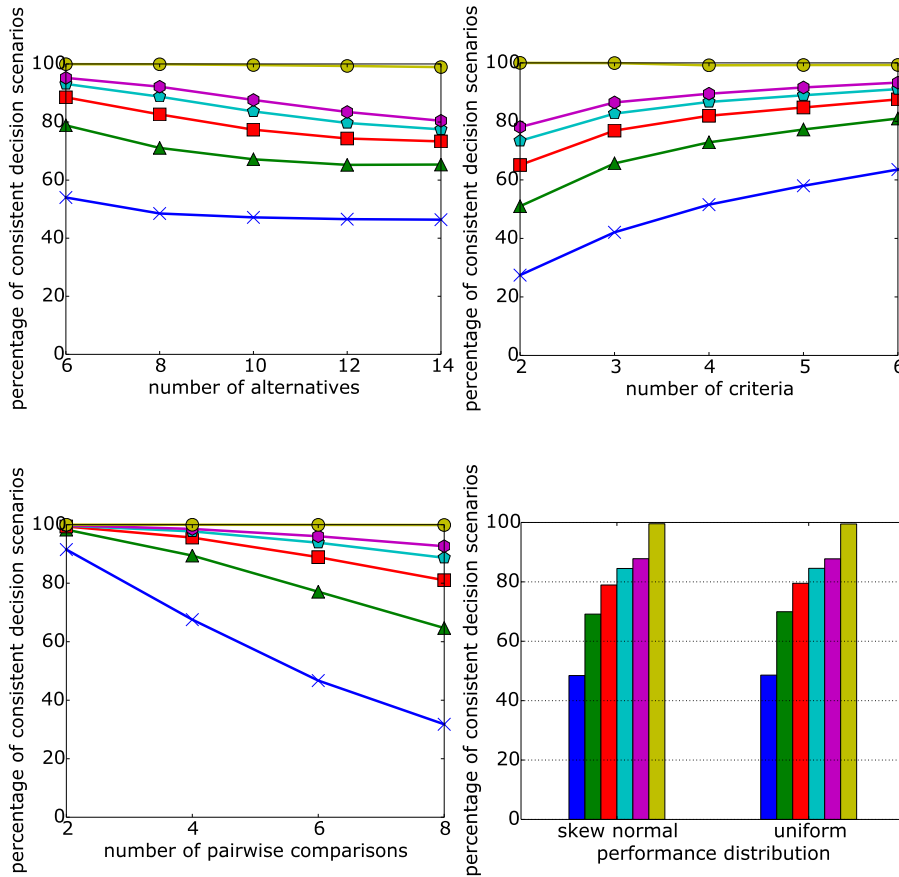
Figure 4.5.1: The percentage $\%(\mathcal{U} \neq \emptyset)$ of randomly generated decision situations for which the pairwise comparisons provided by the DM are fully consistent with an assumed preference model, for different numbers of alternatives, criteria, pairwise comparisons, and types of performance distribution. The series represent different shapes of marginal value function with the characteristic points selected with EWB: blue (X marker) – linear functions (2 characteristic points), green (triangle) – 3 ch.p., red (square) – 4 ch.p., azure (pentagon) – 5 ch.p., purple (hexagon) – 6 ch.p., yellow (circle) – general.

On the contrary, when increasing the number of criteria, the flexibility of an additive model in finding at least one compatible value function or discriminating the pairs of reference alternatives increases. In general, for all considered shapes of marginal value functions, the underlying

trend is represented with a concave function. That is, the differences in expressiveness observed between the adjacent numbers of criteria become less and less when more criteria are involved in the specification of a decision problem. Precisely, the greatest increase in the expressiveness can be observed when passing from 2 to 3 criteria, while the difference observed between 5 and 6 criteria is the least.

Furthermore, the results obtained for $\%(\mathcal{U} \neq \emptyset)$ for both considered performance distributions are very similar. When it comes to ε^*, with the performances distributed uniformly its average value is greater by ~ 0.02 than in case of using the skew normal distribution for generating the performances. This suggests that the type of performance distribution does not influence or has very marginal influence on the expressiveness of an additive value function.

Finally, for all considered decision scenarios, the expressiveness of an additive value function increases with the number of characteristic points. On one hand, the practical usefulness of linear value functions is very limited when the DM compares several pairs of alternatives which are evaluated in terms of few criteria. On the other hand, general value functions prove to be very flexible in representing preference information of the DM. Consequently, if preferences of the DM do not violate the dominance and are not contradictory, it is very likely that they could be reproduced when all unique performances correspond to the characteristic points. Indeed, for our experiment, the number of consistencies was 100% or very close to it for all considered scenarios. However, when the number of alternatives or pairwise comparisons is small or the number of criteria is great, the marginal functions with $4 - 6$ characteristic points provide a satisfactory level of expressiveness.

The differences observed between the adjacent numbers of characteristics points decrease substantially. That is, when passing from linear function to 2-piece-wise linear, the gain in expressiveness is meaningful (e.g., $\sim 23\%$ more consistencies for decision problems with 10 alternatives). On the contrary, moving from 5 to 6 characteristic points does not add much
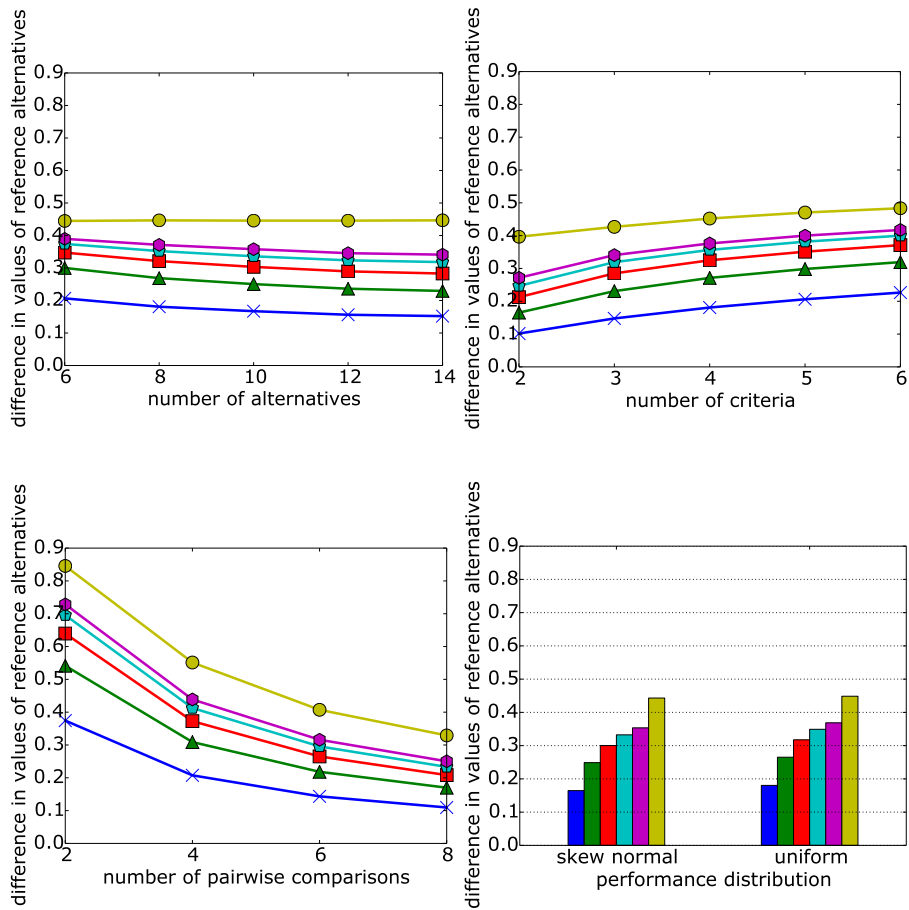
Figure 4.5.2: The minimal difference ε^* between comprehensive values of reference alternatives related by a strict preference by the DM, for different numbers of alternatives, criteria, pairwise comparisons, and types of performance distribution. The series represent different shapes of marginal value functions with the characteristic points selected with EWB: blue (X marker) – linear functions (2 characteristic points), green (triangle) – 3 ch.p., red (square) – 4 ch.p., azure (pentagon) – 5 ch.p., purple (hexagon) – 6 ch.p., yellow (circle) – general.

to the flexibility of the model. Although this general trend is likely to be true also for greater problem sizes, the scales of observed improvements (e.g., $\sim 5\%$ when passing from 5 to 6 points) are influenced by the considered numbers of alternatives and pairwise comparisons.

Robustness

In this section, we discuss how the two measures used for assessing robustness of the recommendation computed with an additive value function are influenced by different characteristics of a decision problem/model. In Figure 4.5.3, we present the average number $\mid \succsim^N \mid$ of pairs for which the necessary preference relation holds for scenarios involving different numbers of alternatives, criteria, pairwise comparisons, and types of performance distribution.

When increasing the numbers of alternatives and pairwise comparisons, the numbers of pairs related by the necessary preference exhibit a clearly increasing trend which is represented by a convex or concave function, respectively. Obviously, the greater the number of alternatives, the more pairs for which the necessary relation can be instantiated, while with more pairwise comparisons, the set of compatible value functions becomes more constrained, which makes the necessary inference easier.

When increasing the number of criteria, the number of necessary inferences decreases substantially. Thus, involving less criteria for evaluating the alternatives is advantageous for the robustness of the computed recommendation. On the contrary, $\mid \succsim^N \mid$ seems to be independent from the performance distribution. Indeed, the numbers of pairs for which the necessary relation holds are almost the same with the performances generated from the uniform or skew normal distribution.

The most interesting conclusions can be derived from the comparison of robustness measures for different shapes of marginal functions. Clearly, the less the number of characteristic points, the more robust the recommendation. In this regard, the results obtained with the linear

value functions are meaningfully more robust than the outcomes computed with the piece-wise linear ones. Indeed, this model has confirmed its ability to generalize the provided pairwise comparisons for pairs of non-reference alternatives. The quantified advantage of using linear value function over piece-wise linear ones ranges from 2 to 17 (4 to 16) pairs related by the necessary preference for different numbers of alternatives (criteria). Thus, the more alternatives or pairwise comparisons and the less criteria, the more beneficial it is to use a linear model.

When moving to the piece-wise linear functions, one can observe a sudden decrease in the robustness of the recommendation. As a general rule, when adding more characteristic points, the differences between the values of robustness measures decline substantially. Thus, the use of 3 or 4 characteristic points is still beneficial when compared to the general functions. However, the series obtained for 5 or 6 points overlap to a large extent with these for the general functions for different numbers of alternatives, pairwise comparisons, and criteria (in case these are greater than two). In fact, for these three shapes, the average numbers of pairs related by the necessary preference for 6 alternatives, 6 criteria, or 2 pairwise comparisons are exactly the same.

The difference in the robustness when using various shapes of marginal value function can be easily justified when referring to Figure 4.5.4. It reveals the numbers $\mid \succsim_I^N \mid$ of necessary inferences which do not directly follow from the provided pairwise comparisons and/or performances of alternatives. Obviously, the absolute numbers for $\mid \succsim_I^N \mid$ are less than for $\mid \succsim^N \mid$. Thus, although the general trends remain the same as in Figure 4.5.3, they are less sharp when increasing the number of alternatives or pairwise comparisons.

Interestingly, none of the test instances contained any non-trivial necessary inference when using the general functions. This is consistent with the results presented in (Spliet and Tervonen 2014). Consequently, the general additive value models are unlikely to be useful for decision support in contexts where the DM provides holistic pairwise statements, and then
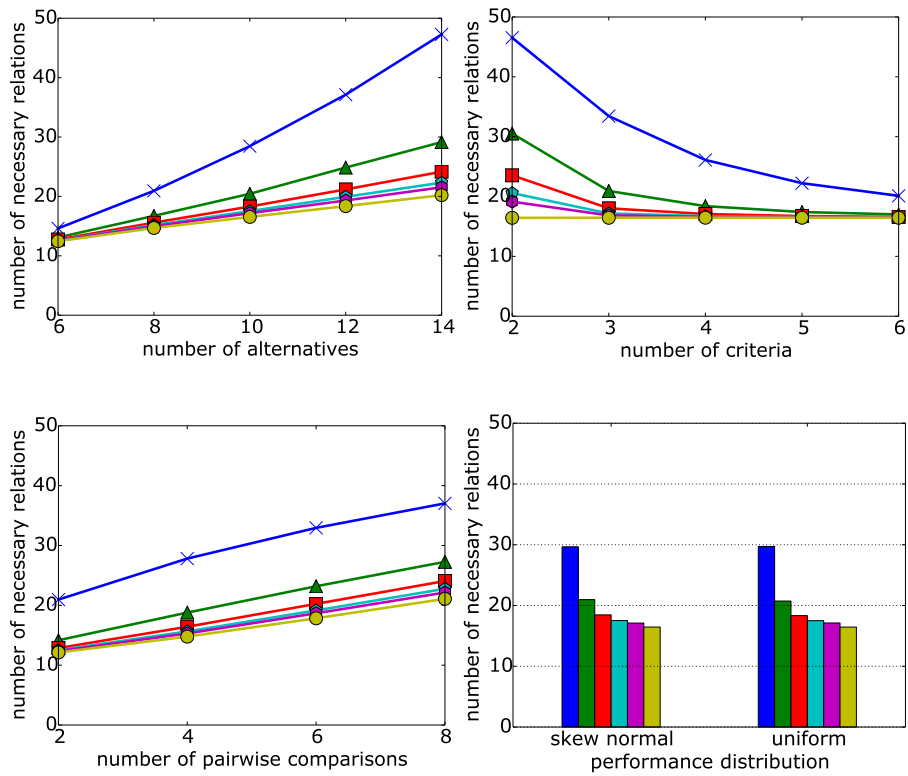
Figure 4.5.3: The average number $| \succsim^N |$ of pairs related by the necessary preference for different numbers of alternatives, criteria, pairwise comparisons, and types of performance distribution. The series represent different shapes of marginal value function with the characteristic points selected with EWB: blue (X marker) − linear functions (2 characteristic points), green (triangle) − 3 ch.p., red (square) − 4 ch.p., azure (pentagon) − 5 ch.p., purple (hexagon) − 6 ch.p., yellow (circle) − general.

the outcomes of robustness analysis are presented to him/her to stimulate the reaction in the following iteration. Since all necessary inferences observed in our experiment for the general functions may be derived only from the analysis of alternatives' performances, the same conclusions hold also for the value models involving less characteristic points. Consequently, the joint analysis of $\left| \succsim^N \right|$ and $\left| \succsim_I^N \right|$ proves what is the contribution of the employed preference model in the inference of robust conclusions.

Our results indicate that the analysis with piece-wise linear functions with a limited number of characteristic points may be useful for practical decision aiding, allowing to derive the non-trivial conclusions that would be confirmed by all compatible value functions. In particular, for problems involving the maximal number of alternatives or minimal number of criteria, we were able to find on average 27 and 30 necessary preference relations that were neither obtained from the transitive closure nor from the single inference theorem. Moreover, the trends observed for these functions are clearly non-linear. The results indicate that when at least five criteria are used to describe alternatives, then the linear functions are the only sensible option.

The values of $\left| \succsim_I^N \right|$ observed for linear functions are meaningfully greater than for the marginal functions with 3 characteristic points. Still, these functions may derive numerous non-obvious certain conclusions when several alternatives or pairwise comparisons and few criteria are involved. Finally, the robustness of additive models with > 4 characteristic points is again very low, and their practical usefulness in the preference disaggregation setting is questionable.

4.5.2 Statistical Analysis of the Simulation Results

In this subsection, we analyze the results of our experimental study in a hierarchical regression model. This supports comprehension of the impact that different parameters of a decision problem/model have on the expressiveness and robustness measures. Figures presented in the e-Appendix exhibit the trends for four measures discussed in the previous section additionally

Figure 4.5.4: The average number $|\succsim_I^N|$ of necessary inferences which do not directly follow the provided preference information, for different numbers of alternatives, criteria, pairwise comparisons, and types of performance distribution. The series represent different shapes of marginal value functions with the characteristic points selected with EWB: blue (X marker) – linear functions (2 characteristic points), green (triangle) – 3 ch.p., red (square) – 4 ch.p., azure (pentagon) – 5 ch.p., purple (hexagon) – 6 ch.p., yellow (circle) – general.
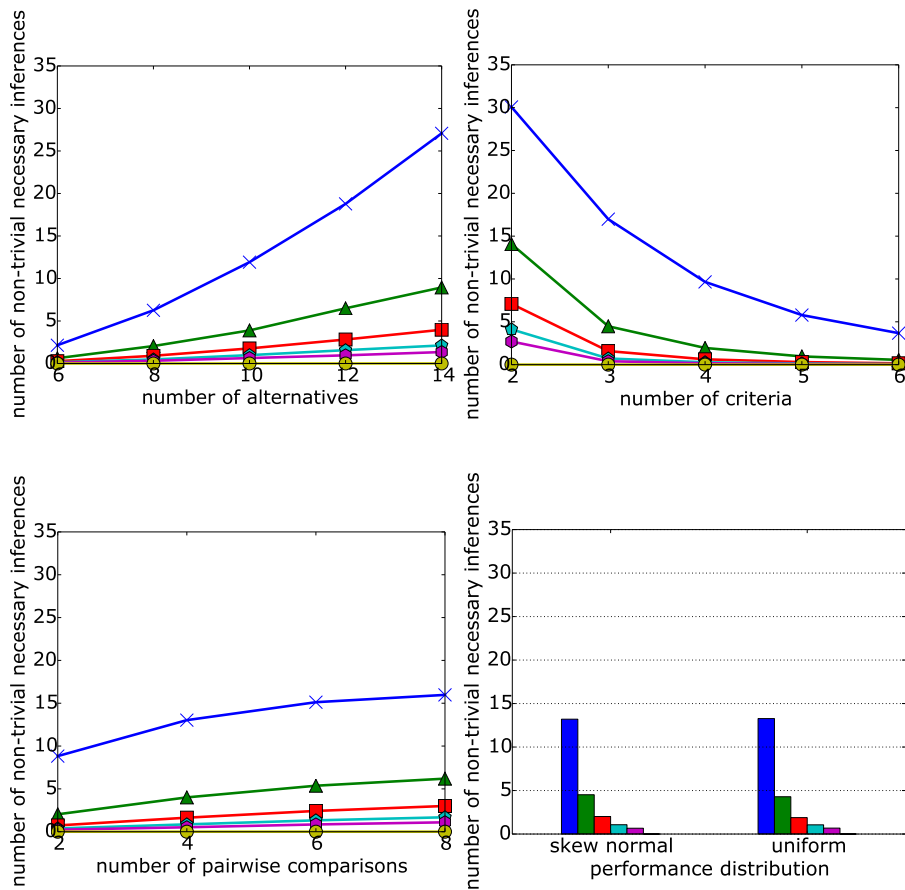
enriching the mean values with the confidence intervals. These confirm that a variance in the expressiveness and robustness measures changes across different levels of the model parameters. For example, the variance in the number of pairs related by the necessary preference increases with the number of alternatives and decreases with the number of criteria.

The heterogeneity of variances raises some concerns about reliability of the regression analysis results. Indeed, the test of significance is severely sensitive to this property. However, the method of least squares is producing unbiased estimates even in the presence of heterogeneous variances (Hayes and Cai 2007). Therefore, interpretations about the relative importance of different parameters in view of their impact on the values of expressiveness and robustness measures are possible. Nonetheless, the predictive ability of the regression model is limited when the estimates by the least square method are derived in the presence of heteroscedasticity. Hence, to get more accurate results from the regression analysis, we have divided the considered problem settings into four groups with more homogeneous variances, and we analyzed the models derived for each of these groups individually.

The separation into the four groups has been conducted as follows. Figures 4.5.1-4.5.4 indicate that the numbers of criteria and characteristic points have a similar impact on the four measures of expressiveness and robustness, even though the magnitudes of their influence might vary from one parameter to another. The same observation can be made for the numbers of alternatives and pairwise comparisons. Thus, we have grouped these parameters under the two dimensions, called "model flexibility" (for the numbers of criteria and characteristic points) and "problem constructivity" (for the numbers of alternatives and pairwise comparisons). For each dimension, we have defined two levels, deemed as low and high, which overall gives four different settings. For the "model flexibility", the low level corresponds to 2 or 3 criteria and characteristic points, whereas for the high level the respective numbers are 5 and 6. For the "problem constructivity", the low level corresponds to 6 or 8 alternatives and 2 or 4 pairwise

comparisons, whereas the high level involves 12 or 14 alternatives and 6 or 8 pairwise comparisons.

The results of a hierarchical regression analysis are provided in Tables 4.5.1 and 4.5.2 for the low and high levels of "model flexibility", respectively. Note that both tables refer to the low and high levels of "problem constructivity". For each considered setting, the parameter with the greatest impact on a given measure has been marked in bold.

Let us emphasize that when interpreting the coefficients derived from the regression analysis, it is important to consider the range of variables (measures). For instance, $\%(\mathcal{U} \neq \emptyset)$ varies between 0 and 100, while ε^* takes values in the range $[0, 1]$. Thus, for example, when considering the setting with low "problem constructivity" and high "model flexibility", the impact of the number of pairwise comparisons r on ε^* is meaningful, while being negligible for $\%(\mathcal{U} \neq \emptyset)$ even though the coefficient associated with r is larger in magnitude for $\%(\mathcal{U} \neq \emptyset)$ when compared to ε^* (-0.28 vs. -0.15).

Table 4.5.1: Results of a hierarchical regression analysis for the low level of "model flexibility" (m - number of criteria, γ_j - number of characteristic points, n - number of alternatives, r - number of pairwise comparisons; an empty cell means indicates "no impact").

			constant	m	γ_j	n	r	adjusted R-square	mean	standard deviation
problem constructivity	low	$\%(\mathcal{U} \neq \emptyset)$	25.46	13.66	**22.07**		-12.37	89%	77.67	19.18
		ε^*	0.04	0.09	**0.13**		-0.09	95%	0.31	0.13
		$\lvert \succsim^N \rvert$	13.19	-4.29	**-5.48**	3	2.03	89%	15.83	5.36
		$\lvert \succsim^N_I \rvert$	15.32	-4.3	**-5.46**	2.05		83%	5.26	4.46
	high	$\%(\mathcal{U} \neq \emptyset)$	-44.7	20.96	**32.08**		-7.69	96%	34.09	21.32
		ε^*	-0.04	**0.05**	0.05		-0.02	94%	0.10	0.04
		$\lvert \succsim^N \rvert$	54.64	-19.61	**-24.67**	7.06	2.91	96%	56.08	18.08
		$\lvert \succsim^N_I \rvert$	64.22	-19.61	**-24.67**	6.18		96%	33.87	17.54

The following conclusions can be derived from the analysis of results presented in Tables 4.5.1 and 4.5.2:

- with the increase in "model flexibility" and the decrease in "problem constructivity",

Table 4.5.2: Results of a hierarchical regression analysis for the high level of "model flexibility" (m - number of criteria, γ_j - number of characteristic points, n - number of alternatives, r - number of pairwise comparisons; an empty cell means indicates "no impact").

			constant	m	γ_j	n	r	adjusted R-square	mean	standard deviation
problem constructivity	low	$\%(\mathcal{U} \neq \emptyset)$	98.18	0.29	0.13		-0.28	78%	99.63	0.37
		ε^*	0.89	0.02	0.02		**-0.15**	100%	0.65	0.16
		$\mid \succsim^N \mid$	-0.32			0.96	**1.41**	100%	10.60	1.73
		$\mid \succsim_I^N \mid$	0.18	-0.02	-0.02	0.01	0.01	64%	0.03	0.02
	high	$\%(\mathcal{U} \neq \emptyset)$	92.91	0.81	**1.27**		-0.96	91%	97.65	1.29
		ε^*	0.33	0.02	0.02		**-0.03**	88%	0.31	0.04
		$\mid \succsim^N \mid$	-0.82			0.87	**1.7**	100%	22.39	1.94
		$\mid \succsim_I^N \mid$	0.99	**-0.13**	-0.11	0.02	0.04	91%	0.19	0.10

the mean values of the expressiveness and robustness measures, respectively, increase or decrease (note that this is consistent with what has been discussed in Section 4.5.1);

- when changing the level of "model flexibility" from low to high, the impact of the accounted parameters on the four studied measures becomes meaningfully smaller, except for the impact of the number of pairwise comparisons on ε^* and $\mid \succsim^N \mid$ (which is still rather negligible in magnitude);

- the measures of robustness ($\mid \succsim^N \mid$ and $\mid \succsim_I^N \mid$) generally become more sensitive to the parameter values with the increase in the level of "problem constructivity";

- for the low level of "model flexibility", the number of characteristic points is the most influential parameter, followed by the number of criteria; this emphasizes the importance of appropriate selection of the characteristic points when few criteria are considered; for the high level of "model flexibility", these two parameters do not have any impact on the number of pairs $\mid \succsim^N \mid$ related by the necessary preference;

- the number of alternatives has a negligible impact on the measures of expressiveness ($\%(\mathcal{U} \neq \emptyset)$ and ε^*);

- the number of pairwise comparisons has a negligible impact on $|\succsim_I^N|$ when considering the low level of "model flexibility"; nonetheless, when the preference model is more flexible its relative impact on ε^* and $|\succsim^N|$ is the greatest among all parameters.

4.5.3 The Impact of Methods for Selecting the Characteristic Points on the Expressiveness and Robustness Measures

In this section, we verify the impact of methods for selecting the characteristic points on the expressiveness and robustness measures. We skip the discussion on how these two characteristics change with the increase in the numbers of alternatives, criteria, pairwise comparisons, characteristic points, or different performance distribution. These trends are, in general, the same as for EWB (see Section 4.5.1). Instead, we focus on the comparison of results obtained with five procedures: EWB, EFB, KMC, KDE, and SSP. Hence, we neglect the outcomes obtained with the linear or general value functions, because the discretization methods cannot be applied in these two contexts.

Expressiveness

In this section, we focus on two measures for assessing how expressiveness of an additive value function is influenced by different procedures for selecting the characteristic points. In Figure 4.5.5, we present the averaged trends for five different parameters of a decision problem or a preference model.

Although the difference between five different methods are small, some regularities can be observed for all considered dimensions. The best results have been obtained with EFB. This suggests that balancing the number of performances which are mapped to each linear piece of a marginal value function is beneficial for increasing the expressiveness of a model. The advantage of EFB over the remaining methods is particularly visible for greater problem sizes

or numbers of characteristic points. In particular, EFB allowed to reproduce from 5% to 10% more complete rankings of alternatives than other discretization methods.

Intermediate performance can be attributed to SSP and KDE. With smaller problem sizes or lower numbers of characteristic points and pairwise comparisons, their expressiveness is similar to EFB. The advantages of using SSP are clearly visible for different numbers of pairwise comparisons. The supervised preference-based character of this algorithm allows to obtain better results than the unsupervised procedures.

However, with the increase in the numbers of alternatives, criteria, or characteristic points, the advantage of SSP and KDE over the two worst methods, i.e., EWB and KMC, is decreasing. In any case, using either an equal width or a clustering algorithm for the construction of linear pieces is not beneficial for the model's expressiveness. The ranking of all procedures is best visible for different distribution types. The slight differences concern the comparison of SSP and KDE. The previous is better for the uniform distribution, whereas the latter is more advantageous with the skew normal distribution.

The expressiveness for the case when there exists at least one value function compatible with the provided holistic statements is measured with ε^* (see Figure 4.5.6). One can note some differences in the performance of different procedures when compared with $\%(\mathcal{U} \neq \emptyset)$. The most evident one is that SSP can be indicated as an overall best performer. For all decision scenarios, this procedure was able to discriminate comprehensive values of reference alternatives compared pairwise by the DM by at least 0.02 more. The greater the problem size or the more flexible the model, and the less the number of pairwise comparisons, the greater the advantage of SSP over the remaining methods. The exceptions can be indicated when a complete ranking of the alternatives is considered and there are only three characteristic points. In the former case, the flexibility of all methods is very limited by such a rich input preference
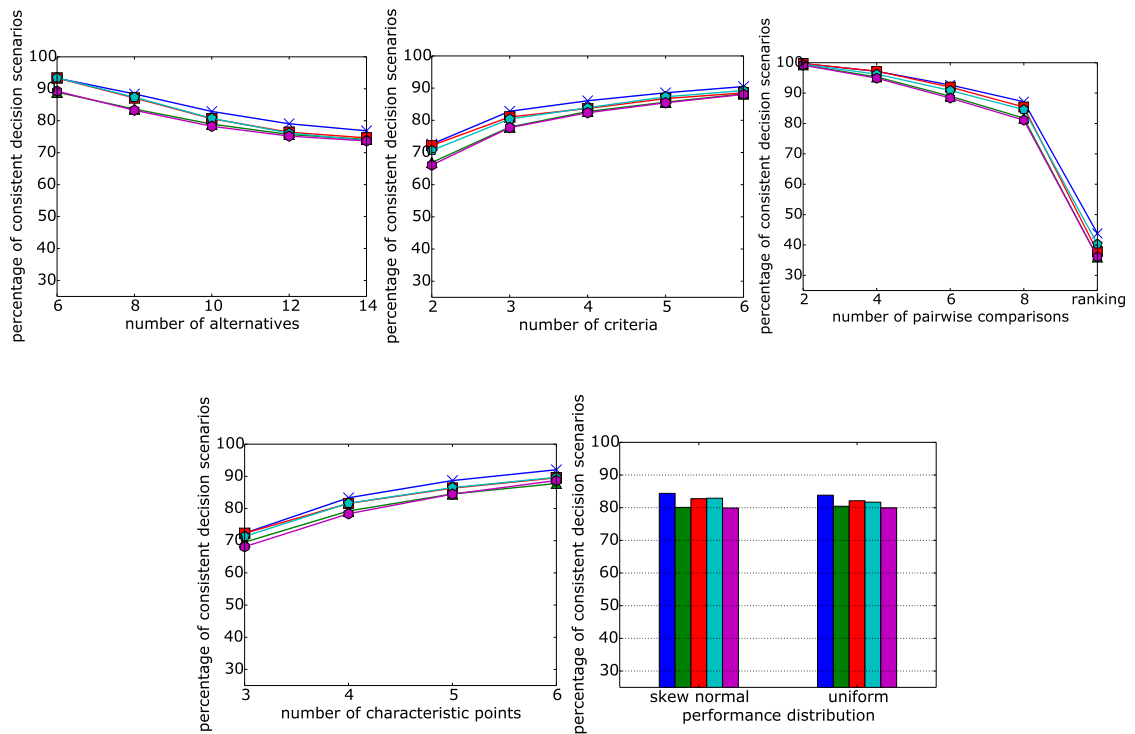
Figure 4.5.5: The percentage $\%(\mathcal{U} \neq \emptyset)$ of randomly generated decision situations for which the pairwise comparisons provided by the DM are fully consistent with an assumed preference model, for different numbers of alternatives, criteria, pairwise comparisons, characteristic points, and types of performance distribution. The series represent different types of methods for selecting the characteristic points: blue (X marker) – EFB; green (triangle) – EWB; red (square) – SSP; azure (pentagon) – KDE; purple (hexagon) – KMC. For clarity, the minimum value on y-axis has been set to 25%.

information, and, thus, they attain similar results. In the latter case, the impact of different methods is marginal, because they decide upon a selection of just a single intermediate point.

The best procedure in terms of $\%(\mathcal{U} \neq \emptyset)$, i.e., EFB, is ranked second with respect to an average ε^* value. However, its advantage over EWB or KDE is rather marginal. The clearly worst performer in terms of this measure (by about 0.01) is KMC.

Some interesting conclusions can be derived from the analysis of different performance distributions. First, for the uniform distribution EWB performs better than for the skew normal one. This can be easily justified, since EWB constructs the sub-intervals with equal widths, which is suitable for dealing with the uniformly distributed performances, but not for other distributions. In the same spirit, KMC performs relatively worse for the skew normal distribution. One of the assumptions of the k-means algorithm is that all clusters should not have varying variance. This assumption is violated in case of the skew normal distribution.

Robustness

As explained in Section 4.5.1, only the analysis of non-trivial necessary inferences allows judging the impact of the parameterization of an employed preference model on its robustness. Thus, in this section, when comparing different methods for selection of the characteristic points, we focus on the number $|\succsim_I^N|$ of pairs related by the necessary preference which cannot be derived directly from the provided preference information and performances (see Figure 4.5.7).

Although the differences observed for the various methods are relatively small, they still reveal that the method for selection of the characteristic points does influence the robustness of conclusions that can be computed with the use of an additive value function. For example, with all other parameters equal, when considering the decision scenarios with 14 alternatives or 2 criteria, the choice of one discretization method over another may bring on average two addi-
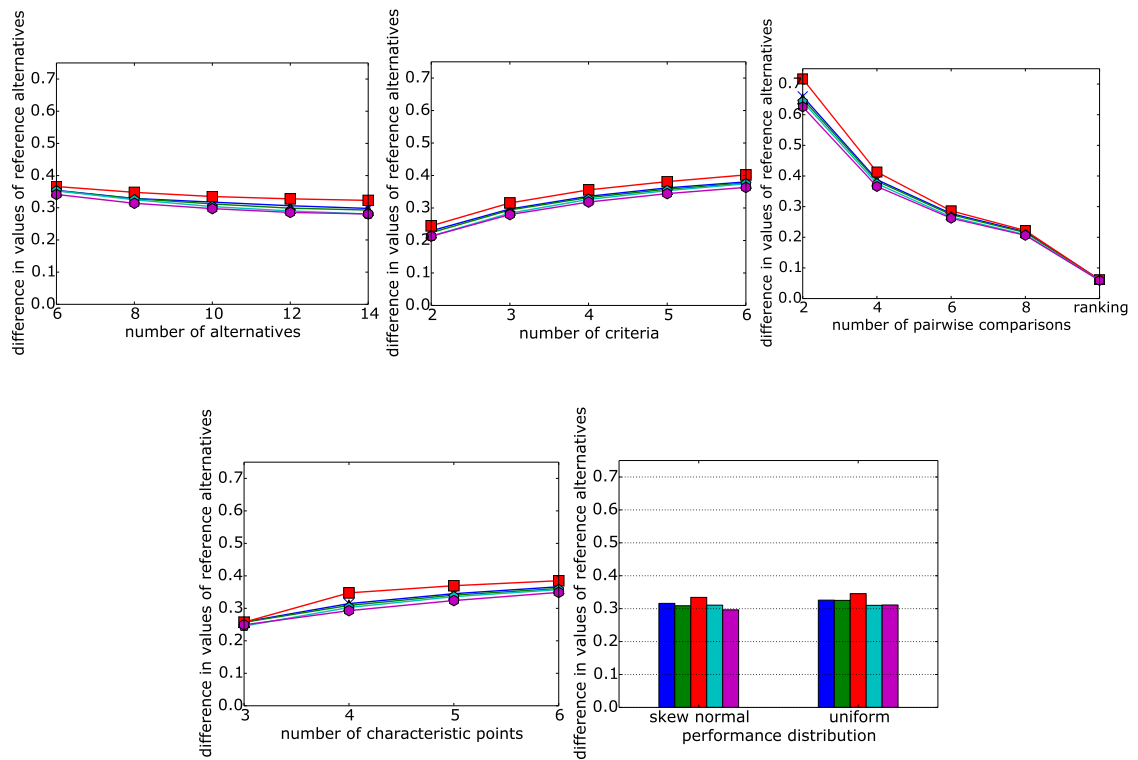
Figure 4.5.6: The minimal difference ε^* between comprehensive values of reference alternatives related by a strict preference by the DM, for different numbers of alternatives, criteria, pairwise comparisons, characteristic points, and types of performance distribution. The series represent different types of methods for selecting the characteristic points: blue (X marker) – EFB; green (triangle) – EWB; red (square) – SSP; azure (pentagon) – KDE; purple (hexagon) – KMC.

tional necessary inferences, and for all conducted tests the best method brings one additional necessary relation more than the worst one. In general, greater differences between the five considered methods can be observed for the scenarios with a great number of alternatives and low numbers of criteria and characteristic points.

The best robustness of conclusions is offered by KMC. This is due to a relatively low flexibility of the underlying model which maps similar performances on the same linear piece of a marginal value function. As a result, the space of compatible value functions is smaller, which makes the necessary inference easier. There are, however, two exceptions in this regard. The robustness of a model based on KDE is increasing relatively faster with the growth of the number of alternatives than for the other algorithms. Consequently, KDE is the best performer when a large number of alternatives is involved. Moreover, KDE proves its superiority in terms of robustness when the performances are uniformly distributed. On the contrary, it is much worse for the problems with a skew normal distribution of performances, allowing to instantiate the necessary relation for ~ 1 pair less than in case of a uniform one. This confirms sensitiveness of KDE with respect to the performance distribution.

The SSP procedure is an intermediate performer. Its gap to the best procedures is less for problems with a relatively greater size, many characteristic points, and numerous pairwise comparisons. These characteristics indicate scenarios when there is greater potential for SSP to construct the sub-intervals and select the characteristic points which better discriminate the provided preference information. Finally, irrespective of the considered dimension, the worst performance in terms of the robustness can be attributed to EFB. This indicates that the space of compatible value models for this method is the greatest in all scenarios, which makes the necessary inference more challenging.
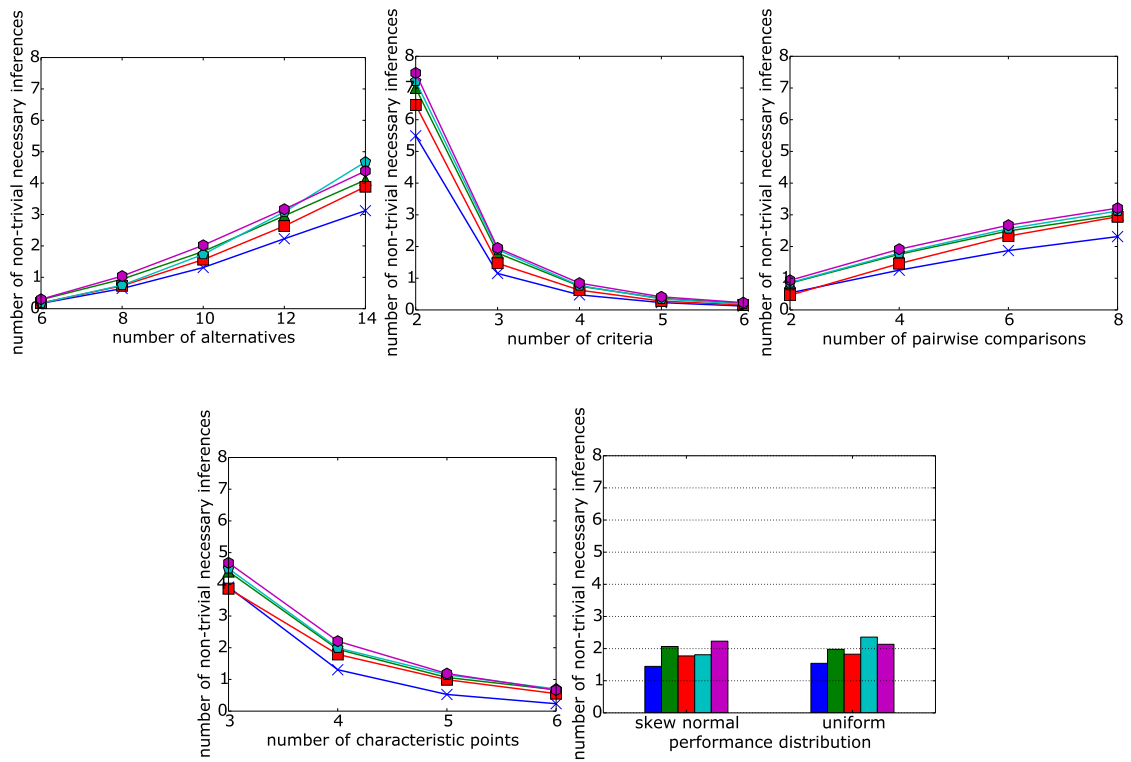
Figure 4.5.7: The average number of inferred necessary preference relations $|\succsim_I^N|$ which do not directly follow the provided preference information, for different numbers of alternatives, criteria, pairwise comparisons, characteristic points, and types of performance distribution. The series represent different types of methods for selecting the characteristic points: blue (X marker) – EFB; green (triangle) – EWB; red (square) – SSP; azure (pentagon) – KDE; purple (hexagon) – KMC.

4.5.4 Trade-off Between Expressiveness and Robustness

To better capture the trade-off between expressiveness and robustness for five methods for selection of the characteristic points, in Figure 4.5.8 we illustrate the trends for all four measures in terms of different γ_j values. Let us emphasize that although the differences observed between the performances of various methods are rather small, they are statistically significant (with only few obvious exceptions, when the performances of two methods are equal or very similar). To confirm this conclusion, in 4.7 we present p-values derived from the Wilcoxon signed-rank test for the comparison of all pairs of procedures.

The joint analysis of expressiveness and robustness confirms that these two characteristics are contradicting. On one hand, KMC is the best performer in terms of $|\succsim^N|$ and $|\succsim_I^N|$ and the worst performer in terms of $\%(\mathcal{U} \neq \emptyset)$ and ε^*. On the other hand, EFB proves its superiority when it comes to reproducing the holistic preference information, but fails to provide sufficiently many robust conclusions.

These different qualities are best combined by KDE and SSP. The former provides the second highest number of robust conclusions, while not falling much in expressiveness when more than three characteristic points need to be selected. The latter is the second best performer for $\%(\mathcal{U} \neq \emptyset)$ and significantly outperforms the remaining procedures with respect to ε^*, while not losing much in robustness for the scenarios with at least four characteristic points.

4.5.5 Performance of Procedures for Selection of the Characteristic Points in View of an Incremental Specification of Pairwise Comparisons

In this section, we consider the impact of different discretization methods on the expressiveness and robustness measures in view of the interactive preference construction paradigm (Corrente et al. 2013; Tsoukiàs 2007). In this setting, the DM is expected to provide her/his preference
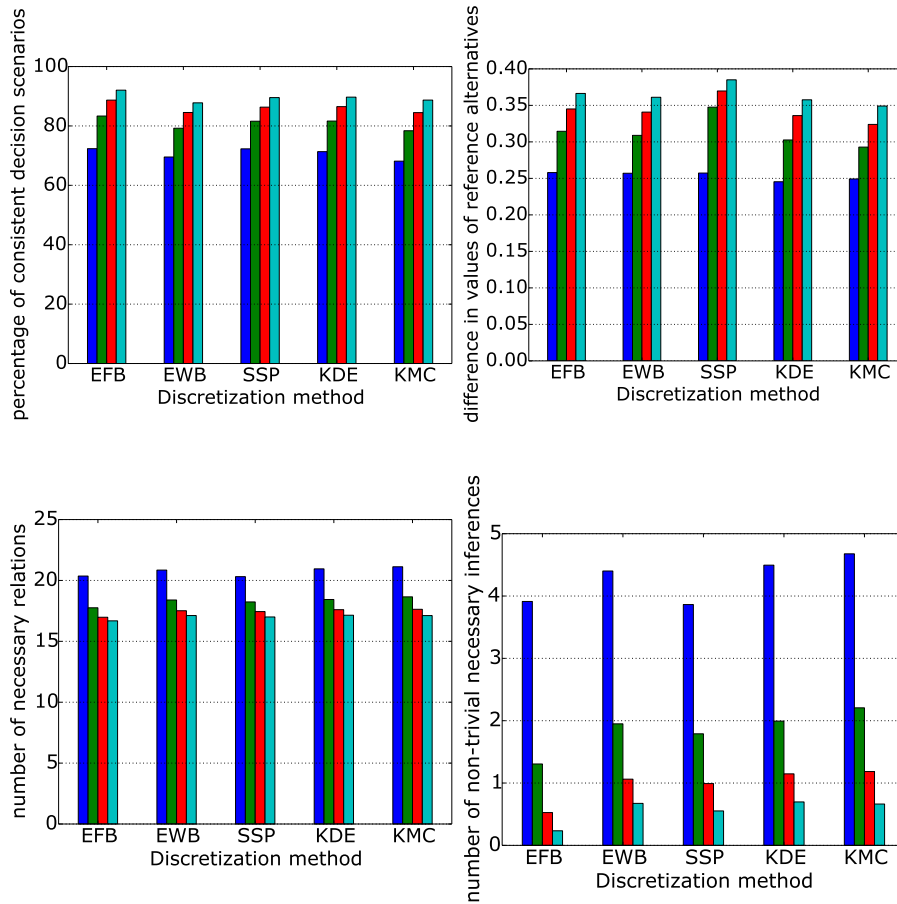
Figure 4.5.8: The average values of $\%(\mathcal{U} \neq \emptyset)$, ε^*, $| \succsim^N |$, and $| \succsim_I^N |$ for different discretization methods. Series represent different numbers of characteristic points: blue − 3 characteristic points (ch.p.); green − 4 ch.p.; red − 5 ch.p.; azure − 6 ch.p.

Table 4.5.3: Test of homogeneity of variances for the decision scenarios with poor preference information.

Variable	Levene statistic
$\%(\mathcal{U} \neq \emptyset)$	14.58**
ε^*	1.46
$\log\lvert \succsim^N \rvert$	0.47
$\log\lvert \succsim_I^N \rvert$	2.09

**$p < 0.001$

information incrementally. At an early stage of interaction, relatively few pieces of preference information are available. As the interaction process evolves, the DM learns more about her/his preferences, which leads her/him to providing more pairwise comparisons. In this regard, it is interesting to analyze the performance of different discretization procedures for two scenarios distinguished by either very poor or relatively rich preference information. In our analysis, these correspond to the test instances involving 2 or 8 pairwise comparisons, respectively.

Decision scenarios with poor preference information

Before conducting the analysis of variance for different measures in decision scenarios with poor preference information, we test the homogeneity of variances. The results presented in Table 4.5.3 indicate that all measures but $\%(\mathcal{U} \neq \emptyset)$ have homogeneous variances across the five groups corresponding to different discretization methods.

When it comes to three measures with homogeneous variances, the analysis suggests that the impact of discretization method is significant for ε^* $(F_{(4, 995)} = 15.95, p < 0.001)$ and $\lvert \succsim_I^N \rvert$ $(F_{(4, 972)} = 13.1, p < 0.001)$, while the results for $\lvert \succsim^N \rvert$ do not vary across different procedures $(F_{(4, 995)} = 0.47, p = 0.76)$[3]. To investigate the impact of different discretization methods on $\%(\mathcal{U} \neq \emptyset)$, these procedures are compared pairwise using Welch's

---

[3] For the test of significance, we used a logarithmic transformation of the robustness measures, since F-test is extremely sensitive to even small deviation from normality (Box 1953; Markowski and Markowski 1990).

t-test. The results indicate that SSP performs slightly, though not significantly, better than EFB, while significantly outperforming the remaining procedures.
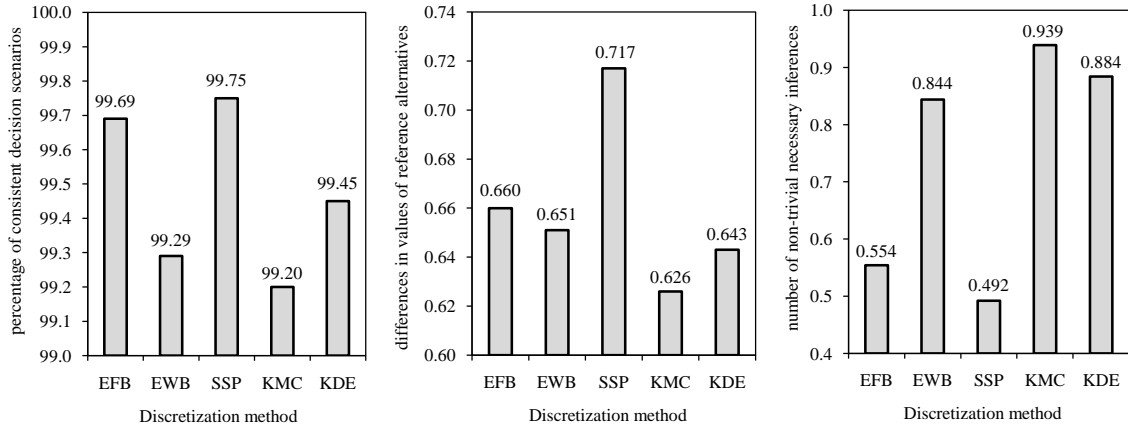
Figure 4.5.9: Mean plots of $\%(\mathcal{U} \neq \emptyset)$, ε^* and $|\succsim_I^N|$ for different discretization methods in poor preference information setting.

Figure 4.5.9 provides the mean plots for the three measures whose performances vary when using different discretization methods, i.e., $\%(\mathcal{U} \neq \emptyset)$, ε^*, and $|\succsim_I^N|$. They indicate that SSP performs the best in terms of both expressiveness measures ($\%(\mathcal{U} \neq \emptyset)$ and ε^*), while attaining the worst results for robustness. On the contrary, KMC attains the best outcomes in terms of robustness, while being outperformed by the remaining procedures when it comes to expressiveness. KDE and EWB maintain a reasonable compromise between these two characteristics in decision scenarios with poor preference information. Overall, the trade-off between expressiveness and robustness is observable for all methods.

DECISION SCENARIOS WITH RICH PREFERENCE INFORMATION

The tests of homogeneity of variances in decision scenarios with rich preference information indicate that for all measures there is no evidence supporting heterogeneity across the five groups corresponding to different discretization methods (see Table 4.5.4). The analysis of variances suggests that $|\succsim^N|$ does not depend on the choice of discretization method

Table 4.5.4: Test of homogeneity of variances for the decision scenarios with rich preference information.

Variable	Levene statistic
$\%(\mathcal{U} \neq \emptyset)$	1.66
ε^*	1.41
$\log\lvert \succsim^N \rvert$	0.51
$\log\lvert \succsim_I^N \rvert$	1.57

$(F(4, 995) = 0.43, p = 0.79)$, while the impact of discretization method is significant for $\%(\mathcal{U} \neq \emptyset)$ $(F(4, 995) = 2.63, p = 0.001)$, ε^* $(F(4, 995) = 2.48, p < 0.05)$, and $\lvert \succsim_I^N \rvert$ $(F(4, 976) = 3.93, p < 0.01)$.

The mean plots for the latter three measures are provided in Figure 4.5.10. While EFB is the only method which competes with SSP in terms of expressiveness, SSP outperforms EFB with respect to robustness. Further, KMC and KDE perform well in view of robustness, but attain poor results for expressiveness. Overall, the results indicate that SSP performs very well in the decision scenarios when rich preference information is available. In fact, it is the sole method which attains a reasonable compromise between expressiveness and robustness. Its relatively good performance for the test cases with numerous pairwise comparisons can be justified by its supervised character, i.e., accounting for the holistic judgments when selecting the characteristic points.

4.6 CONCLUSIONS

In this paper, we experimentally evaluated the expressiveness of an additive value function and the robustness of conclusions computed with its use in the preference disaggregation setting. We indicated how these characteristics are affected by different numbers of alternatives, criteria, pairwise comparisons, and performance distribution. A special attention was paid to investigating the impact of both the number of characteristic points which define the shape
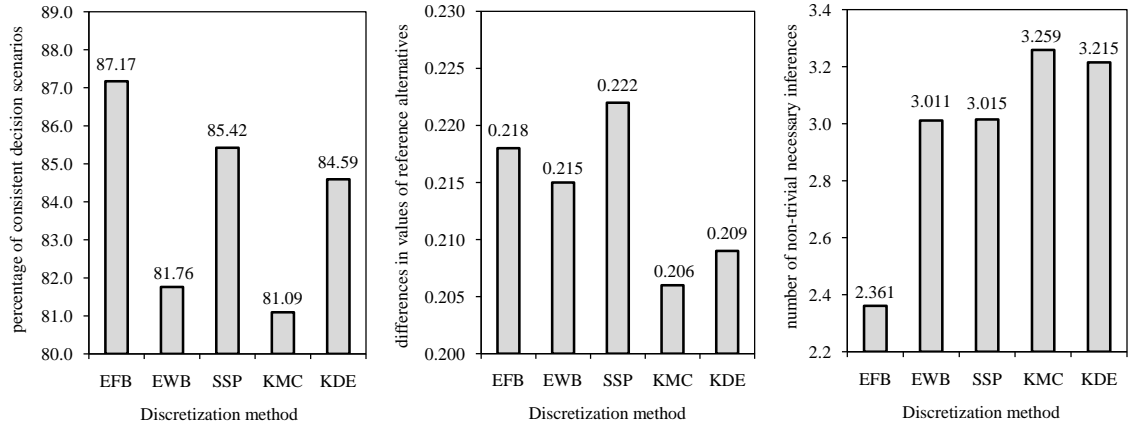
Figure 4.5.10: Mean plots of $\%(\mathcal{U} \neq \emptyset)$, ε^* and $|\succsim_I^N|$ for different discretization methods in rich preference information setting.

of per-criterion marginal value functions and applying different procedures for their selection. Our research outcomes were quantified with four measures: the share of decision scenarios for which the set of compatible value function is non-empty, the minimal difference between comprehensive values of reference alternatives compared by the DM, the number of pairs related by the necessary preference, and the number of non-trivial certain inferences which cannot be derived directly from the preference information.

The extensive experimental study indicated that the expressiveness of an additive value increased with the number of criteria and characteristic points, and decreased when more alternatives were considered or additional pairwise comparisons were provided. The impact of changes in these dimensions on the robustness was just the opposite. Furthermore, in most considered scenarios, the influence of performance distribution on the analyzed measures was negligible. Although the computational experiments were restricted in size to instances representative of real-life decision support problems, we expect our conclusions to be valid also for greater problems.

In most cases, the trend of the change in expressiveness or robustness was non-linear with respect to the change in the parameterization of a decision problem or a preference model. That

is, meaningfully greater changes were observed in the lower scale ranges of different dimensions (e.g., when passing from 2 to 3 criteria, from 2 to 3 characteristic points, or from 2 to 4 pairwise comparisons), whereas the differences observed in the upper scale ranges were rather marginal (e.g., when passing from 5 to 6 criteria, from 5 to 6 characteristic points, or from 6 to 8 pairwise comparisons).

We also compared five different methods for selection of the characteristic points. The results indicated that with all other parameter values equal, one may gain at least 5% in expressiveness and/or derive many additional non-trivial certain conclusions just by suitably selecting the discretization methods. The best compromise in expressiveness and robustness was reached by the newly proposed supervised algorithm which constructs the characteristic points so that to separate as many pairs of performances of reference alternatives compared pairwise by the DM. Its advantage was particularly evident when numerous pairwise comparisons were available.

Some of our findings agree with the conclusions from the previous studies. On one hand, we proved that linear value functions are not expressive enough for practical decision aiding. Indeed, Korhonen et al. (Korhonen et al. 2012) indicated earlier that human judgments are very often not fully consistent with a linear model. On the other hand, we confirmed that the general functions are not robust enough to be useful by themselves for decision support in preference disaggregation contexts (Spliet and Tervonen 2014). Still, we were able to prove that with piece-wise linear value functions we may often reproduce the holistic judgments of the DM, and derive the non-trivial robust conclusions.

To our best knowledge, our study is the first that captures expressiveness and robustness together. In this way, we confirmed that these two objectives are contradictory, and a compromise between these characteristics needs to be reached when selecting a specific parameterization of an additive model for real-world problems.

Our results indicate avenues for some practice-oriented methodological development. On

one hand, one should propose procedures for construction of a parsimonious preference model (Słowiński et al. 2013), i.e., the simplest model that is able to reconstruct the DM's pairwise comparisons, being adequate to the complexity of the indirect preference information. This idea derives from the rule that the simplest explanation is most likely the correct one. In this regard, one may, e.g., minimize the number of characteristic points or a deviation of marginal function from linearity. On the other hand, when using preference disaggregation methods in an interactive way, we may start by using the linear model, and then move to piece-wise linear or general functions once the simpler model is not expressive enough. In the context of interactive evolutionary multi-objective optimization, an analogous idea has been recently suggested in (Branke et al. 2016).

The experimental analysis of the use of an additive value function in preference disaggregation methods can be also extended in future studies. In particular, we may consider other measures capturing expressiveness and robustness. For example, the former may be quantified with the magnitude of inconsistency when all pairwise comparisons cannot be reproduced, whereas the latter may account for the entropy-like measures to reflect the variety of rankings that can be obtained with a set of compatible value functions (Ciomek et al. 2016), differentiation of alternatives' comprehensive values, or the stability of an inferred function as defined in (Bous et al. 2010). The measures considered in the experimental study can also refer to other characteristics such as, e.g., the similarity between the rankings generated with different discretization techniques or the "linearity" of the inferred marginal functions (Bous et al. 2010). Finally, the study can be extended to additive models that take into account positive and negative interactions among criteria (Angilella et al. 2014) or admit non-monotonic shape of marginal functions (Doumpos 2012; Ghaderi et al. 2015, 2017, 2014).

4.7 Appendix: Results of the Statistical Comparison of the Impact of Methods for Selection of the Characteristic Points on the Expressiveness and Robustness Measures

Table 4.7.1: p-values derived from the Wilcoxon signed-rank test for the comparison of all pairs of procedures for selection of the characteristic points. The results concern different numbers of characteristic points $\gamma_j \in \{3, 4, 5, 6\}$. p-values greater than 0.01 are in bold. If p-value is less than a pre-defined significance level (e.g., $a = 0.01$), the null hypothesis is rejected and we can deduce that the difference between a pair of compared methods is statistically significant.

char. points	$\%(\mathcal{U} \neq \emptyset)$				ε^*			
	3	4	5	6	3	4	5	6
(EFB & EWB)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(EFB & SSP)	**0.977**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(EFB & KDE)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(EFB & KMC)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(EWB & SSP)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(EWB & KDE)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(EWB & KMC)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(SSP & KDE)	0.000	**0.021**	0.008	0.000	0.000	0.000	0.000	0.000
(SSP & KMC)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(KDE & KMC)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

char. points	$\mid \succsim^N \mid$				$\mid \succsim_I^N \mid$			
	3	4	5	6	3	4	5	6
(EFB & EWB)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(EFB & SSP)	**0.036**	0.000	0.000	0.000	**0.385**	0.000	0.000	0.000
(EFB & KDE)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(EFB & KMC)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(EWB & SSP)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(EWB & KDE)	0.000	0.000	**0.156**	0.000	0.000	0.000	**0.687**	0.000
(EWB & KMC)	0.000	0.000	0.000	0.080	0.000	0.000	0.000	0.000
(SSP & KDE)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(SSP & KMC)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(KDE & KMC)	0.000	0.000	0.000	**0.029**	0.000	0.000	0.000	0.000

5

Conclusion and Future Research

This chapter provides an integrated discussion of the previous chapters. It highlights the theoretical contributions of the thesis to two main research paradigms: Decision Science, and Consumer Behaviour. Then managerial implications of this research will be addressed. Finally, limitations and future avenues for the research will be highlighted.

5.1 Implications for Theory

The three main chapters of the thesis collectively contribute to two main paradigms: decision science, and consumer research. The key concept relating the two paradigms, which are connected to some extent, yet also independent, is the notion of *preference*. The thesis provides an integrated framework for decision analysis from a multiple criteria perspective based on indirect preferences, together with a careful examination of the impact of different characteristics of a decision problem. The example of analytical investigation of the impact of brand colour on brand image demonstrates how a complex problem can be formulated and addressed from such perspective. In the following subsections, the theoretical contributions of the thesis in the two active research paradigms will be discussed.

5.1.1 Contribution to Decision Science

The novel preference disaggregation approach, introduced in the thesis, contributes to multiple criteria decision analysis literature by addressing the case of non-monotonicity in preferences. Compared to the existing methodologies, this framework is able to capture non-monotonic preferences without making any assumption on the shape of the preference model and without asking for extra information from the DM. In addition, this thesis contributes to decision science studies by performing a series of experimental analysis based on simulating a broad range of decision settings. Experimental analysis approach based on simulation has been used to enhance our understanding of how characteristics of a decision problem can influence the quality of the inference outcomes (chapter 4), as well as to test the performance of the newly introduced disaggregation framework (chapter 3). The framework of the simulation study can be utilized by scholars who are interested in experimental analysis of various types of decision problems. Moreover, the findings from chapter 4 can be used for better elicitation of

preferences in a decision aiding practice at several stages of interaction and in different settings of a decision problem.

5.1.2 Contribution to Consumer Behaviour literature

Modeling of consumer preferences among multiattribute alternatives has been one of the major activities in consumer research (Green and Srinivasan 1978). The classical approach for this purpose is based on two important works at 1970s, the behavior theory approach by Fishbein (Fishbein 1976) and the conjoint measurement of judgmental data by Green (Green and Rao 1971). Since then, conjoint analysis has received considerable academic and industry attention as a basis for the analysis of consumers' tradeoffs among multiattribute products (Green and Srinivasan 1990; Gustafsson et al. 2013; Rao et al. 2014). The framework introduced in this thesis addresses consumers' tradeoffs, relative importance of different attributes in the purchase decision, as well as the mechanism each attribute contribute in shaping preferences of the consumer. With respect to latter, the introduced framework has several advantages over the classical conjoint analysis approach due to its powerful capacity in capturing various preference models. Moreover, its flexibility in admitting various types of preference information as input to enrich the inference process facilitates modeling complex preferences with the least cognitive burden on consumer more than anytime ever. The framework introduced in this thesis can be viewed as an alternative or supplementary framework for conjoint analysis, especially in the cases where no prior knowledge on the shape of preferences is available.

5.2 Implications for Practitioners

This thesis has several implications for the practitioners. The framework introduced in this thesis can help practitioners in decision making process at complex environments where several

conflicting points of view should simultaneously be considered, especially when some of these evaluation criteria contain an ideal point.

In particular, modeling the questions on the design of a brand elements as a multiple criteria analysis problem, as shown in chapter 2, helps brand managers to leverage brand elements such as brand colour to promote a desired image of brand. The findings draw on the analysis of indirect perceptual information from customers and gives a clear picture of the current position of a brand in the customers' minds, and helps on the steps need to ba taken to in re-designing of a brand to achieve a desired position.

Disaggregation of preferences of complex form based on preference information of the simplest form (pairwise comparison) helps marketing managers to gain a better understanding of the needs of consumers in the marketplace which, in turn, results in delivering better products and services. Moreover, the analytical framework introduced in this thesis helps on predicting how a consumer will react to changes in some of the product attributes. This provides insights on the sensitivity of the target market to some changes in the features of a product or service. Such results can provide a basis for the decisions on product improvement, also allocation of resources in order to gain the highest value by investing on the attributes that the market care the most about. Also, by capturing non-linearity in preferences, the framework helps to identify the levels on attributes up to which the investment is beneficial and beyond that the gain in value is becoming rather marginal.

5.3 Discussion, Limitations and Future Research

The study of the brand colour and its impact on brand image based on an analytical framework, in chapter 2, supports the finding from mostly qualitative studies in this field. Given the hypothetically non-monotonic impact of colour on brand image, an iterative computational proce-

dure based on linear programming is developed for this purpose. The results show a significant contribution of all the three components of a brand colour, hue, saturation, and brightness, in almost all dimensions of the brand image. Interesting finding is that brightness of a colour and its saturation dominate the impact of colour hue (the pigment that creates the colour). The main objective of this study was to introduce an analytical framework for processing holistic information of customers' perceptions of a brand, drawing on theories from cognitive psychology that explain the mechanisms under which the color operates to shape or change the perceptions. The findings, however, are the subject of several limitations mainly because some other relevant factors, such as ethnicity and cultural background of the respondent, which could potentially be relevant are not controlled. In addition, the framework is able to address the brands with a single representative colour only, whereas most of the brands use a combination of several colours next to each other. Investigating the interaction among the colors requires a more sophisticated analytical framework which is an interesting and important subject for the future research.

Inspired by chapter 2, in chapter 3 a more comprehensive framework for inferring non-monotonic additive preference models from indirect preference information is introduced. The disaggregation framework aims at inferring possibly non-monotonic additive value models by balancing the expressiveness versus complexity of the inferred model. An extension of the methodology, in addition, is introduced which enables the analyst to further simplify the inferred preference model at the cost of its expressiveness, by changing an exogenous parameter. The performance of the framework is tested based on a comprehensive experimental analysis. The disaggregation framework, however, is limited to handling cardinal criteria only, since it controls the complexity in the preference model by restricting the change in the slope of the inferred marginal value functions. Ordinal criteria, however, are present in many decision situations. The proposed framework, although for practical purpose is applicable to qualitative criteria, but from a theoretical point of view restricting the slope of value function for such cri-

teria does not make much sense. Extending the methodology towards a general framework for capturing non-monotonic preference models in presence of both quantitative and qualitative criteria is an important subject for the future research. Moreover, handling non-univocal solutions in ordinal regression is of great importance and the proposed framework has potentials to develop in this direction. A framework for analyzing robustness of the recommendations derived from a non-monotonic value function can be developed drawing on the Robust Ordinal Regression (ROR) framework (Greco et al. 2008) that seeks for necessary and possible relations amongst the set of alternatives based on all the compatible value functions. However, it must be noted that the set of compatible non-monotonic value functions should be considered separately at different levels of complexity.

The experimental evaluation of the expressiveness of an additive value function and the robustness of the derived recommendations, in chapter 4, indicates how these two qualities are affected by different characteristics of a decision problem. A special attention was paid to investigating the impact of both the number of characteristic points which define the shape of per-criterion marginal value functions and applying different procedures for their selection. With this regards, a new supervised technique for the smart selection of characteristic points based on the configuration of alternatives performances and supplied preferences was introduced. To our best knowledge, this study is the first that captures expressiveness and robustness together. It is also one of the very first studies using an extensive experimental analysis based on simulation in multiple criteria decision aiding literature.

Bibliography

Angilella, S., Bottero, M., Corrente, S., Ferretti, V., Greco, S., and Lami, I. M. (2016). Non additive robust ordinal regression for urban and territorial planning: an application for siting an urban waste landfill. Annals of Operations Research, 245(1-2):427–456.

Angilella, S., Corrente, S., Greco, S., and Słowiński, R. (2014). MUSA-INT: Multicriteria customer satisfaction analysis with interacting criteria. Omega, 42(1):189 – 200.

Azzalini, A. (1985). A class of distributions which includes the normal ones. Scandinavian Journal of Statistics, 12:171–178.

Babin, B. J., Hardesty, D. M., and Suter, T. A. (2003). Color and shopping intentions: the intervening effect of price fairness and perceived affect. Journal of business research, 56.7:541–551.

Beccacece, F., Borgonovo, E., Buzzard, G., Cillo, A., and Zionts, S. (2015). Elicitation of multiattribute value functions through high dimensional model representations: Monotonicity and interactions. European Journal of Operational Research, 246(2):517–527.

Ben-Akiva, M. E. and Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand, volume 9. MIT press.

Biba, M., Esposito, F., Ferilli, S., Di Mauro, N., and Basile, T. (2007). Unsupervised discretization using kernel density estimation. In Proceedings of the 20th International Joint Conference on Artifical Intelligence, IJCAI'07, pages 696–701, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Bisdorff, R., Dias, L. C., Meyer, P., Mousseau, V., and Pirlot, M. (2015). Evaluation and decision models with multiple criteria: case studies. Springer.

Bottomley, P. A. and Doyle, J. R. (2006). The interactive effects of colors and products on perceptions of brand logo appropriateness. Marketing Theory, 6.1:63–83.

Bous, G., Fortemps, P., Glineur, F., and Pirlot, M. (2010). ACUTA: A novel method for eliciting additive value functions on the basis of holistic preference statements. European Journal of Operational Research, 206(2):435–444.

Bouyssou, D. (1990). Building criteria: A prerequisite for mcda. In Readings in multiple criteria decision aid, pages 58–80. Springer.

Box, G. (1953). Non-normality and tests on variances. Biometrika, 40:318–335.

Branke, J., Corrente, S., Greco, S., Słowiński, R., and Zielniewicz, P. (2016). Using Choquet integral as preference model in interactive evolutionary multiobjective optimization. European Journal of Operational Research, 250(3):884–901.

Charnes, A. and Cooper, W. W. (1961). Management models and industrial applications of linear programming, volume 1. JSTOR.

Charnes, A. and Cooper, W. W. (1962). Programming with linear fractional functionals. Naval Research logistics quarterly, 9(3-4):181–186.

Chebat, J.-C. and Morrin, M. (2007). Colors and cultures: exploring the effects of mall décor on consumer perceptions. Journal of business Research, 60.3:189–196.

Ciomek, K., Kadziński, M., and Tervonen, T. (2016). Heuristics for prioritizing pair-wise elicitation questions with additive multi-attribute value models. Omega, (in press):–. http://dx.doi.org/10.1016/j.omega.2016.08.012.

Corrente, S., Doumpos, M., Greco, S., Sł owiński, R., and Zopounidis, C. (2015). Multiple criteria hierarchy process for sorting problems based on ordinal regression with additive value functions. Annals of Operations Research, pages 1–23.

Corrente, S. and Greco, S. (2012). Słowiński. R., Multiple criteria hierarchy process in robust ordinal regression, 53(3):660–674.

Corrente, S., Greco, S., Kadziński, M., and Słowiński, R. (2013). Robust ordinal regression in preference learning and ranking. Machine Learning, 93(2-3):381–422.

Corrente, S., Greco, S., Kadziński, M., and Słowiński, R. (2014). Robust ordinal regression. In Cochran, J., Cox, L., Keskinocak, P., Kharoufeh, J., and Smith, J., editors, Wiley Encyclopedia of Operations Research and Management Science, pages 1–10. John Wiley & Sons, Inc.

Corrente, S., Greco, S., Kadziński, M., and Słowiński, R. (2016a). Inducing probability distributions on the set of value functions by subjective stochastic ordinal regression. Knowledge-Based Systems, 112:26 – 36.

Corrente, S., Greco, S., Matarazzo, B., and Sł owiński, R. (2016b). Robust ordinal regression for decision under risk and uncertainty. Journal of Business Economics, 86(1-2):55–83.

de Almeida, A. T. (2007). Multicriteria decision model for outsourcing contracts selection based on utility function and ELECTRE method. Computers & Operations Research, 34(12):3569 – 3574.

de Almeida, A. T., de Almeida, J. A., Costa, A. P. C. S., and 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 Gemmis, M., Iaquinta, L., Lops, P., Musto, C., Narducci, F., and Semeraro, G. (2009). Preference learning in recommender systems. Preference Learning, 41.

Despotis, D. K. and Zopounidis, C. (1995). Building additive utilities in the presence of non-monotonic preferences. Advances in Multicriteria Analysis, 5:101–114.

Devaud, J., Groussaud, G., and Jacquet-Lagreze, E. (1980). Utadis: Une methode de construction de fonctions d'utilite additives rendant compte de jugements globaux. In European working group on MCDA, Bochum , Germany.

Devaut, J. M., Groussaud, G., and Jacquet-Lagreze, E. (1980). UTADIS: Une methode de construction de fonctions d'utilite additives rendant compte de jugements globaux. European Working Group on Multicriteria Decision Aid, Bochum.

Diakoulaki, D., Zopounidis, C., Mavrotas, G., and Doumpos, M. (1999). The use of a preference disaggregation method in energy analysis and policy making. Energy, 24(2):157–166.

Dias, L., Mousseau, V., Figueira, J., and Climaco, J. (2002). An aggregation/disaggregation approach to obtain robust conclusions with electre tri. European Journal of Operational Research, 138:332–348.

Doumpos, M. (2012). Learning non-monotonic additive value functions for multicriteria decision making. OR Spectrum, 34(1):89–106.

Doumpos, M., Xidonas, P., Xidonas, S., and Siskos, Y. (2015). Development of a robust multicriteria classification model for monitoring the postoperative behaviour of heart patients. Journal of Multi-Criteria Decision Analysis, 23(1-2):15–27.

Doumpos, M., Zanakis, S. H., and Zopounidis, C. (2001). Multicriteria preference disaggregation for classification problems with an application to global investment risk. Decision Sciences, 32(2):333–385.

Doumpos, M. and Zopounidis, C. (1997). A multicriteria decision aid methodology for the assessment of county risk. Investigaciones europeas de dirección y economía de la empresa, 3(3):13–34.

Doumpos, M. and Zopounidis, C. (1998). The use of the preference disaggregation analysis in the assessment of financial risks. Fuzzy Economic Review, 3(1):39–57.

Doumpos, M. and Zopounidis, C. (2002). Multicriteria decision aid classification methods. Springer Science & Business Media, 73:120–122.

Doumpos, M. and Zopounidis, C. (2004a). Developing sorting models using preference disaggregation analysis: An experimental investigation. European Journal of Operational Research, 154:585–598.

Doumpos, M. and Zopounidis, C. (2004b). A multicriteria classification approach based on pairwise comparisons. European Journal of Operational Research, 158(2):378–389.

Doumpos, M. and Zopounidis, C. (2013). Computational intelligence techniques for multi-

criteria decision aiding: An overview. Multicriteria Decision Aid and Artificial Intelligence, pages 1–23.

Eckhardt, A. and Kliegr, T. (2012). Preprocessing algorithm for handling non-monotone attributes in the uta method. Preference Learning: problems and applications in AI (PL-12) workshop.

Ehrgott, M. (2006). Multicriteria optimization. Springer Science & Business Media.

Einhorn, H. J. and Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgement and choice. Annual review of psychology, 32(1):53–88.

Falomir, Z., Museros, L., and Gonzalez-Abril, L. (2015). A model for colour naming and comparing based on conceptual neighbourhood. an application for comparing art compositions. Knowledge-Based Systems, 81:1–21.

Figueira, J., Ehrgott, M., and Greco, S. (2016). Multiple criteria decision analysis Second Edition. Springer.

Figueira, J., Greco, S., and Roy, B. (2009). ELECTRE methods with interaction between criteria: an extension of the concordance index. European Journal of Operational Research, 199:496–505.

Figueira, J., Mousseau, V., and Roy, B. (2005). Electre methods. In Multiple criteria decision analysis: State of the art surveys, pages 133–153. Springer.

Fishbein, M. (1976). A behavior theory approach to the relations between beliefs about an object and the attitude toward the object. In Mathematical Models in Marketing, pages 87–88. Springer.

Fishburn, P. C. (1970). Utility theory for decision making. Technical report, DTIC Document.

Fürnkranz, J. and Hüllermeier, E. (2010). Preference learning: An introduction. Springer.

Garber, L. L., Hyatt, E. M., and Starr, R. G. (2000). The effects of food color on perceived flavor. Journal of Marketing Theory and Practice, pages 59–72.

Ghaderi, M., Ruiz, F., and Agell, N. (2015). Understanding the impact of brand colour on brand image: a preference disaggregation approach. Pattern Recognition Letters, 67(1):11–18.

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

Ghaderi, M., Ruiz, F. J., and Agell, N. (2014). Understanding color trends by means of non-monotone utility functions. In CCIA, pages 107–115.

Greco, S., Ehrgott, M., and Figueira, J. (2016). Multiple Criteria Decision Analysis: State of the Art Surveys. International Series in Operations Research & Management Science, Vol. 233. Springer, Berlin.

Greco, S., Kadziński, M., Mousseau, V., and Słowiński, R. (2012). Robust ordinal regression for multiple criteria group decision problems: UTA$^{\text{GMS}}$-GROUP and UTADIS$^{\text{GMS}}$-GROUP. Decision Support Systems, 52(3):549–561.

Greco, S., Kadziński, M., and Słowiński, R. (2011a). Selection of a representative value function in robust multiple criteria sorting. Computers & Operations Research, 38(11):1620–1637.

Greco, S., Mousseau, V., and Słowiński, R. (2011b). Parsimonious preference models for robust ordinal regression. In in Proc. EWG-MCDA.

Greco, S., Mousseau, V., and Sł owiński, R. (2014). Robust ordinal regression for value functions handling interacting criteria. European Journal of Operational Research, 239(3):711–730.

Greco, S., Mousseau, V., and Słowiński, R. (2008). Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. European Journal of Operational Research, 191(2):415–435.

Green, P. E. and Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. Journal of Marketing research, pages 355–363.

Green, P. E. and Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. Journal of consumer research, 5(2):103–123.

Green, P. E. and Srinivasan, V. (1990). Conjoint analysis in marketing: new developments with implications for research and practice. The Journal of Marketing, pages 3–19.

Grigoroudis, E. and Siskos, Y. (2002). Preference disaggregation for measuring and analysing customer satisfaction: The musa method. European Journal of Operational Research, 143(1):148–170.

Grigoroudis, E. and Siskos, Y. (2009). Customer satisfaction evaluation: methods for measuring and implementing service quality, volume 139. Springer Science & Business Media.

Gustafsson, A., Herrmann, A., and Huber, F. (2013). Conjoint measurement: Methods and applications. Springer Science & Business Media.

Hayes, A. and Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. Behavior Research Methods, 39(4):709 – 722.

Hoegg, J. and Alba, J. W. (2007). Taste perception: more than meets the tongue. Journal of Consumer Research, 33.4:490–498.

Hwang, C.-L. and Yoon, K. (2012). Multiple attribute decision making: methods and applications a state-of-the-art survey, volume 186. Springer Science & Business Media.

Jacquet-Lagrèze, E. (1995). An application of the uta discriminant model for the evaluation of r&d projects. In Pardalos P. M., S. and Y., Z. a., editors, Advances in Multicriteria Analysis. Kluwer Academic Publishers, pages 203–211. Dordrecht.

Jacquet-Lagrèze, E. and Siskos, Y. (1982). Assessing a set of additive utility functions for multicriteria decision making: The uta method. European Journal of Operational Research, 10:151–164.

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

Kadziński, M., Ciomek, K., Rychł y, P., and Sł owiński, R. (2016). Post factum analysis for robust multiple criteria ranking and sorting. Journal of Global Optimization, 65(3):531–562.

Kadziński, M., Corrente, S., Greco, S., and Sł owiński, R. (2014). Preferential reducts and constructs in robust multiple criteria ranking and sorting. OR Spectrum, 36(4):1021–1053.

Kadziński, M., Ghaderi, M., Wąsikowski, J., and Agell, N. (2017). Expressiveness and robustness measures for the evaluation of an additive value function in multiple criteria preference disaggregation methods: an experimental analysis. Computers & Operations Research.

Kadziński, M., Greco, S., and Sł owiński, R. (2012a). Extreme ranking analysis in robust ordinal regression. Omega, 40(4):488–501.

Kadziński, M., Greco, S., and Sł owiński, R. (2012b). Selection of a representative value function in robust multiple criteria ranking and choice. European Journal of Operational Research, 217(3):541–553.

Kadziński, M., Greco, S., and Sł owiński, R. (2013). Ruta: a framework for assessing and selecting additive value functions on the basis of rank related requirements. Omega, 41(4):735–751.

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

Kadziński, M., Słowiński, R., and Greco, S. (2015). Multiple criteria ranking and choice with all compatible minimal cover sets of decision rules. Knowledge-Based Systems, 89:569 – 583.

Kadziński, M. and Tervonen, T. (2013). Robust multi-criteria ranking with additive value models and holistic pair-wise preference statements. European Journal of Operational Research, 228(1):169–180.

Kahneman, D. and Tversky, A. (1984). Choices, values, and frames. American psychologist, 39(4):341.

Kahneman, D., Tversky, A., et al. (1982). The psychology of preferences.

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

Kliegr, T. (2009). Uta-nm: Explaining stated preferences with additive non-monotonic utility functions. Preference Learning (PL-09) ECML/PKDD-09 workshop.

Koopmans, T. C. et al. (1951). Activity analysis of production and allocation. Number 13. Wiley New York.

Korhonen, P., Silvennoinen, K., Wallenius, J., and Oorni, A. (2012). Can a linear value function explain choices? an experimental study. European Journal of Operational Research, 219(2):360 – 367.

Krantz, D., D, L., Suppes, P., and Tversky, A. (1971). Foundations of Measurement, Vol. I: Additive and Polynomial Representations. New York Academic Press.

Labrecque, L. I. and Milne, G. R. (2012). Exciting red and competent blue: the importance of color in marketing. Journal of the Academy of Marketing Science, 40.5:711–727.

Labrecque, L. I., Patrick, V. M., and Milne, G. R. (2013). The marketers' prismatic palette: A review of color research and future directions. Psychology & Marketing, 30.2:187–202.

Lichtenstein, S. and Slovic, P. (2006). The construction of preference. Cambridge University Press.

Lightfoot, C. and Gerstman, R. (1998). Brands: the new wealth creators. 'Brand Packaging' in S. Hart and J. Murphy (eds),Basingstoke: Macmillan, pages 46–55.

Lohse, G. L. and Rosen, D. L. (2001). Signaling quality and credibility in yellow pages advertising: the influence of color and graphics on choice. Journal of advertising, 30.2:73–83.

Long, G. D. and Cliff, N. (1997). Confidence intervals for kendall's tau. British Journal of Mathematical and Statistical Psychology, 50:31–41.

Lovett, M., Peres, R., and Shachar, R. (2014). A data set of brands and their characteristics. Marketing Science, 33.4:609–617.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley Symp. Math. Stat. Probab., Univ. Calif. 1965/66, 1, 281-297 (1967).

Markowski, C. and Markowski, E. (1990). Conditions for the effectiveness of a preliminary test of variance. The American Statistician, 44:322–326.

Mousseau, V., Dias, L., Figueira, J., Gomes, C., and Clímaco, J. (2003). Resolving inconsistencies among constraints on the parameters of an MCDA model. European Journal of Operational Research, 147(1):72–93.

Özpeynirci, o., Özpeynirci, S., and Kaya, A. (2017). An interactive approach for multiple criteria selection problem. Computers & Operations Research, 78:154–162.

Pardalos, P. M., Siskos, Y., and Zopounidis, C. (2013). Advances in multicriteria analysis, volume 5. Springer Science & Business Media.

Patiniotakis, I., Apostolou, D., and Mentzas, G. (2011). Fuzzy utastar: A method for discovering utility functions from fuzzy data. Expert Systems with Applications, 38:15463–15474.

Pirlot, M., Schmitz, H., and Meyer, P. (2010). An empirical comparison of the expressiveness of the additive value function and the Choquet integral models for representing rankings. In Proc. of 25th Mini-EURO Conference "Uncertainty and Robustness in Planning and Decision Making" (URPDM 2010), Coimbra, Portugal.

Rao, V. R. (2014). Theory and design of conjoint studies (ratings based methods). In Applied Conjoint Analysis, pages 37–78. Springer.

Rao, V. R. et al. (2014). Applied conjoint analysis. Springer.

Rossi, P. E. and Allenby, G. M. (2003). Bayesian statistics and marketing. Marketing Science, 22(3):304–328.

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.

Sarabando, P. and Dias, L. (2010). Simple procedures of choice in multicriteria problems without precise information about the alternatives' values. Computers & Operations Research, 37(12):2239 – 2247.

Savage Leonard, J. (1954). The foundations of statistics. NY, John Wiley, pages 188–190.

Schindler, P. S. (1986). Color and contrast in magazine advertising. Psychology & Marketing, 3.2:69–78.

Simon, H. A. (1957). Models of man; social and rational.

Siskos, E., Askounis, D., and Psarras, J. (2014). Multicriteria decision support for global e-government evaluation. Omega, 46:51–63.

Siskos, J. (1983). Analyse de systèmes de décision multicritère en univers aléatoire. Foundations of Control Engineering, 8:193–212.

Siskos, Y. and Grigoroudis, E. (2010). New trends in aggregation-disaggregation approaches. In Zopounidis, C., Pardalos, P. M., Pardalos, P. M., and Hearn, D., editors, Handbook of Multicriteria Analysis, volume 103 of Applied Optimization, pages 189–214. Springer.

Siskos, Y., Grigoroudis, E., and Matsatsinis, N. F. (2005). Uta methods. In Multiple criteria decision analysis: State of the art surveys, pages 297–334. Springer, New York.

Siskos, Y., Grigoroudis, E., and Matsatsinis, N. F. (2016). Uta methods. In Multiple Criteria Decision Analysis, pages 315–362. Springer, New York.

Siskos, Y. and Yannacopoulos, D. (1985). Utastar: An ordinal regression method for building additive value functions. Investigacao Operacional, 5:39–53.

Slovic, P., Fischhoff, B., and Lichtenstein, S. (1977). Behavioral decision theory. Annual review of psychology, 28(1):1–39.

Słowiński, R., Greco, S., and Matarazzo, B. (2002). Axiomatization of utility, outranking and decision-rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle. Control and Cybernetics, 31(4):1005 – 1035.

Słowiński, R., Greco, S., and Mousseau, V. (2013). Inferring parsimonious preference models in robust ordinal regression. European Conference on Operational Research (EURO 2013), Rome, Italy.

Soylu, B. (2011). A multi-criteria sorting procedure with Tchebycheff utility function. Computers & Operations Research, 38(8):1091 – 1102.

Spliet, R. and Tervonen, T. (2014). Preference inference with general additive value models and holistic pair-wise statements. European Journal of Operational Research, 232(3):607–612.

Spronk, J., Steuer, R. E., and Zopounidis, C. (2016). Multicriteria decision aid/analysis in finance. In Multiple Criteria Decision Analysis, pages 1011–1065. Springer.

Spyridakos, A., Siskos, Y., Yannacopoulos, D., and Skouris, A. (2001). Multicriteria job evaluation for large organizations. European Journal of Operational Research, 130(2):375–387.

Theussl, S. and Hornik, K. (2015). Rglpk: R/GNU Linear Programming Kit Interface. R package version 0.6-1.

Train, K. (2003). Discrete choice methods with simulation. Cambridge university press.

Tsoukiàs, A. (2007). On the concept of decision aiding process: an operational perspective. Annals of Operations Research, 154(1):3–27.

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychological review, 79(4):281.

Tversky, A. and Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In Utility, probability, and human decision making, pages 141–162. Springer.

Tversky, A. and Kahneman, D. (1985). The framing of decisions and the psychology of choice. In Environmental Impact Assessment, Technology Assessment, and Risk Analysis, pages 107–129. Springer.

Vetschera, R. and de Almeida, A. T. (2012). A PROMETHEE-based approach to portfolio selection problems. Computers & Operations Research, 39(5):1010 – 1020.

Vetschera, R., Sarabando, P., and Dias, L. (2014). Levels of incomplete information in group decision models – a comprehensive simulation study. Computers & Operations Research, 51:160 – 171.

Von Neumann, J. and Morgenstern, O. (1944). Theory of games and economic behavior. Princeton university press.

Von Winterfeldt, D. and Edwards, W. (1993). Decision analysis and behavioral research.

Wakker, P. (1989). Additive Representations of Preferences: A new Foundation of Decision Analysis. Kluwer Academic Publishers, London.

Wu, Chia-Chien, Wick, F. A., and Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. Frontiers in psychology, 5.

Zaltner, P. M. (1975). John Dewey's aesthetic philosophy. John Benjamin Company, Amsterdam, The Netherlands.

Zheng, J., Takougang, S. M., Mousseau, V., and Pirlot, M. (2014). Learning criteria weights of an optimistic Electre Tri sorting rule. Computers & Operations Research, 49:28 – 40.

Zopounidis, C. (2000). Special issue on artificial intelligence and decision support with multiple criteria. Computers & Operations Research, 27(7–8):597 – 599.

Zopounidis, C. and Doumpos, M. (1999a). Business failure prediction using utadis multicriteria analysis. Journal of the operational Research Society, 50(11):1138–1148.

Zopounidis, C. and Doumpos, M. (1999b). A multicriteria decision aid methodology for sorting decision problems: The case of financial distress. Computational Economics, 14(3):197–218.

Zopounidis, C., Galariotis, E., Doumpos, M., Sarri, S., and Andriosopoulos, K. (2015). Multiple criteria decision aiding for finance: An updated bibliographic survey. European Journal of Operational Research, 274(2):339–348.

Zopounidis, C. and Pardalos, P. M. (2010). Handbook of multicriteria analysis, volume 103. Springer Science & Business Media.