Knowledge Discovery Techniques for Understanding Customer Behavior with Incomplete Data

Knowledge Discovery Techniques for Understanding Customer Behavior with Incomplete Data

by

Benoît Depaire

Proefschrift voorgelegd tot het behalen van de graad van Doctor in de Toegepaste Economische Wetenschappen, 28 september 2009.

in loving memory of

Linda Celotto
Victor Emile Depaire
Gertrude Bemelmans

Contents

List of Tables

List of Figures

Acknowledgements

More than ten years ago, in the summer of 1998, I decided I wanted to become a Master of Business Economics - Information Systems Engineering[1]. When I enrolled in 1998, I did not yet know that the period of five years foreseen for these studies were going to be prolonged with another six years of study. A couple months before getting my master's degree, I went to see my master thesis supervisor, Prof. Dr. Koen Vanhoof, about the job offer of "teaching assistant/Ph.D. student" which had become vacant. After successfully applying for this job, I had the luxury and the dilemma of having to choose between two jobs, one as a teaching assistent/Ph.D. student and the other as an employee of a major IT consultancy firm in Brussels. Instead of choosing for the job in Brussels, which could have made me a wealthier man in financial terms, I opted to become a Ph.D. student which hopefully made me an equally wealthy man in general.

September, 2003 marked the start of a new period in my academic career as a Ph.D. student under supervision of Prof. Dr. Koen Vanhoof en Prof. Dr. Geert Wets. I personally believe that they have done a wonderful job in supervising me which is hopefully proven by the quality of the work that lies in front of you, but I humbly let you, the reader, be my judge. However, as many have mentioned before, making a Ph.D. is a journey with many obstacles to overcome. Personally, my biggest obstacle was when usable research results kept out during the first two years as a Ph.D. student. My supervisors and I decided to stop working on the then research topic and reoriented on the topics in this thesis. Luckily, one does not have to travel this journey alone and I would like to take this opportunity to acknowledge my gratitude to all those people around me who have helped me to succeed finishing this Ph.D thesis.

Mentioning each one of them and their respective contribution would probably take another six years and will most likely fail in being exhaustive. Therefore, please forgive me for invoking my right to be lazy and for not embarking upon this even greater challenge. To those of you who have been a support to me along the way, thank you, I'm eternally indebted to all of you, you know who you are.

But I would like to explicitly show gratitude to some people who managed to support me in ways that have surpassed the levels a man can wish for. First of all, I would like to thank both my Ph.D. supervisors, Koen en Geert. Geert, thank you for offering me this opportunity, forcing me to become a more mature and professional person than I was six years ago and for your focus and emphasis on the need of

[1]Handelsingenieur in de Beleidsinformatica

publishing my work. If I may call myself an academic researcher, than you have certainly contributed to this.

Koen, as mentioned before, our paths go way back, first as my master thesis supervisor and afterwards as my Ph.D. supervisor. All these years under your supervision have been a great experience. Knowing what I know now and having seen plentiful of other examples, I realize that our collaboration has been of a rare kind. We both share a common interest for data mining and knowledge discovery techniques, focussing on the applications of these techniques without fleeing away from the technical details, which lead to many interesting discussion. You always made time for me, immediately. Whenever I gave you a piece of text I had written, I received your comments within days, often within hours. I never had to fear to run dry of ideas as you often called me into your office to discuss some new ideas or insights you had for my research. On the other hand, you always respected my freedom to select which ideas I wanted to elaborate on. As Seneca once said: "If a man does not know to what port he is steering, no wind is favourable to him". Koen thank you for offering me the map to my port, but letting me be the helmsman of my own ship. Your contribution in this achievement is undeniable.

Furthermore, I would like to thank the members of my jury, Prof. Dr. Tom Brijs, Prof. Dr. Józsefef Dombi, Prof. Dr. Etienne Kerre, Prof. Dr. Dirk Van den Poel and Prof. Dr. Da Ruan. Many thanks for the useful remarks which raised this thesis to a higher level. It is a true honor to have such respected members of the academic community judging the quality of my work and being the members of my Ph.D jury.

During the six years as a Ph.D. student I shared my office with many colleagues, but some of them have become much more than mere colleagues, they became the reason which made me joyful and look forward each morning to go to my office. I consider them as true friends.

Karolien, although it has been a long time already, I still remember us sharing an office in the F-corridor downstairs as if the echo's of our laughters and vivid discussions are still resounding against the rumbling walls. Know that those memories will always remain.

An, we were friends first, became colleagues later, and have shared so many events that many secrets disappeared over time. We have been students together, have lived and experienced Erasmus in London together, went on mountain bike and ski trips together and have shared 40 hours a week on average for the last five years. Despite our interminable discussions or perhaps because of them, I always truly enjoy our time together and hope that the future will bring more of the same.

Katrien, it has been a joy to share my office with you as I never had an office colleague with such rational reasoning. Although you are sometimes so possessed by the extremely small letters on your new computer screen that you create invisible walls around you in the office and seemingly ignore the presence of An and me, you are an invaluable asset in the joy of our current office and my partner in crime for teasing An.

Furthermore, I would like to thank my parents-in-law, Robert and Mieke, for the support and opportunities they offered to my wife and me. It certainly made this period in our lives more comfortable and easier.

Finally, at the risk of sounding ridiculous, I would like to thank those who are most precious to me, my family. My parents, my brother, my grandparents and my godfather and godmother, they all have contributed in important ways by offering me

the necessary environment to become the man I am today. They have always made me feel loved and save as a child, supported me in my actions and prevented me from harm. I still remember the evenings at my grandparents' place as a little child, with 'bompa en bomma' or with 'nonno and nonna', staying up late, watching television and eating ice lollies or apples and crawling close against one of my grandmothers. Or the less than ordinary gifts I received from Patrick which made me very excited and my mother very concerned, such as a set of boxing gloves for me and my brother or a chemistry box for which I was technically still too young. The visit of my godmother Julie each year with loads of gifts was another joyful occasion which I vividly remember and which luckily caused less troubles for my mother. I truly believe that all these joyful occasions and the feeling of being loved have provided me the necessary self-confidence to start and finish my Ph.D. studies. If it were not for them, you would not be reading this text.

Didier, we have played and fought a lot as little children as true brothers do, but I know I can count on you anytime. Therefore, thank you for being my brother.

'Mama en papa', thanks for everything and more. Both of you have always believed in me and supported me, even in things I didn't excel at but enjoyed doing. Obviously, the soccer genes were not split evenly. Both of you were always there for Linde and me, even now with the construction of our house, and provided me with the right values in life necessary to successfully finish this work. There are simply no means to ever pay back what I have received, but I hope that this achievement can make you both proud as this is also your achievement.

Last but not least, and probably most important, I would like to thank my wife, Linde. 'Schatje', thank you for enduring my tempers and providing me the most beautiful smile I've ever seen. The last six years haven't always been easy and you had to share me more than once with my Ph.D thesis. As you know better than anyone, I've had my bad patches along the way but you were always there to give me new courage. I'm glad that I could share this path with you and hope that I can offer you the same support along the way ahead. You have been my haven when I was restless, provided me with laughter when I was sad and took my burden when I was tired. For all of this, for making me feel alive every day again, I thank you more than I can express.

Benoît Depaire
September 28, 2009

Samenvatting

Geen enkele onderneming die actief is in een competitieve en klantgerichte markt zal het belang van klantentevredenheid onderschatten zonder hiervoor een prijs te moeten betalen. Wetenschappelijk onderzoek heeft al meermaals de positieve invloed van klantentevredenheid op verschillende aspecten van een bedrijf aangetoond. Tevreden klanten zorgen voor minder klachten en minder kosten die hiermee gepaard gaan. Ze genereren positieve en onbetaalbare mondreclame, zijn bereid meer te betalen voor eenzelfde product en vertonen een grotere mate van loyaliteit wat zich uit in het heraankopen van het product. Daarom hoeft het ook geen betoog dat bedrijven een competitief voordeel kunnen uitbouwen als ze erin slagen hun klanten beter te begrijpen en meer tevreden kunnen maken dan hun directe concurrenten. Menig bedrijf maakt dan ook geld en tijd vrij om de informatie over hun klanten te analyseren en zo te achterhalen 'wat' hun klanten tevreden stelt. Er ontstaat echter een probleem wanneer de analysetechnieken worden toegepast op onvolledige data wat kan leiden tot onvolledige of zelfs foute conclusies betreffende klantentevredenheid.

Onvolledige data tijdens het analyseren van klantentevredenheid komt vaker voor dan men op het eerste zicht verwacht. Vaak beschikken bedrijven slechts over twee soorten informatie als het op klantentevredenheid aankomt, i.e. 'hoe tevreden is de klant met het product' en 'hoe goed presteerde het product volgens de klant'. Zulke data veronderstelt dat klantentevredenheid enkel gestuurd wordt door het performantieniveau van het product, wat tegengesproken wordt door de huidige ideeën en theoriën omtrent het klantentevredenheidsproces. Volgens het 'Expectancy-Disconfirmation' (ED) paradigma, dat één van de meest dominante theorieën is in de onderzoeksliteratuur, heeft de productperformantie slechts een indirect effect op klantentevredenheid. Volgens deze theorie wordt binnen het klantentevredenheidsproces de performantie vergeleken met een referentieniveau wat in de meeste gevallen leidt tot een disconfirmatie van deze referentie. Volgens het ED paradigma is het dit disconfirmatie- en referentieniveau die samen het tevredenheids- of ontevredenheidsgevoel veroorzaken. Hoe positiever de disconfirmatie is, des te groter de klant zijn tevredenheid zal zijn, ceteris paribus. Hoe hoger het referentieniveau is, des te hoger de klantentevredenheid zal zijn, ceteris paribus. Als een onderneming enkel performantie- en tevredenheidsdata ter beschikking heeft voor het analyseren van klantentevredenheid, ontbreekt er dus belangrijke informatie die een rol heeft gespeeld in het tevreden stellen van de klant.

Maar zelfs als een onderneming tijdens het ondervragen van de klanten gepeild heeft naar het referentieniveau van de klant, kan de onderzoeker of marketeer nog steeds met onvolledige data achterblijven. Want hoewel de onderzoeksliteratuur de laatste twintig tot dertig jaar het ED paradigma uitgebreid heeft getest en

een algemene consensus bestaat over de validiteit van deze theorie, blijkt uit het literatuuronderzoek in deze thesis dat de onderzoekswereld het nooit eens is geraakt over wat een klant als referentieniveau hanteert. Terwijl de oorspronkelijke invulling gebeurde door middel van de verwachte productperformantie heeft later onderzoek aangetoond dat onder andere het prestatieniveau van andere producten ook als referentieniveau kunnen gebruikt worden. Dus zelfs als de onderzoeker bijvoorbeeld de verwachtingen van de klant ter beschikking heeft, dan bestaat nog steeds de mogelijkheid dat dit niet de juiste referentie is in het tevredenheidsproces van die klant. Net als in het vorige scenario blijft de onderzoeker achter met onvolledige data.

In het eerste deel van deze thesis werd een nieuwe analysetechniek ontworpen die toelaat om het klantentevredenheidsproces te modelleren met onvolledige data in overeenstemming met de richtlijnen van het ED paradigma. Deze techniek laat bedrijven toe om met enkel performantie- en tevredenheidsdata de klantentevredenheid te analyseren op een manier die overeenstemt met de theorie, i.e. het ED paradigma. Hiervoor werd het klantentevredenheidsproces eerst in drie stappen ontleed:

- de evaluatie op productattribuutniveau van de productperformantie ten opzichte van een referentieniveau wat leidt tot disconfirmatie op attribuutniveau,
- de aggregatie van de disconfirmaties op productattribuutniveau naar een algemene disconfirmatie,
- de interactie tussen de algemene disconfirmatie en de initiële klantentevredenheid die samen leiden tot een algemeen tevredenheidsgevoel.

In plaats van iedere stap te vertalen naar één specifieke wiskundige functie werd ten gunste van de modelleringsflexibiliteit per stap een set van wiskundige eigenschappen gedefinieerd waaraan een wiskundige functie moet voldoen om de desbetreffende stap op een correcte wijze te modelleren. Dit leidt tot een modelleringsraamwerk dat het LIED-raamwerk wordt genoemd en heeft als voordeel dat gebruikers zelf kunnen beslissen welke aspecten van de theorie omtrent klantentevredenheid zij willen integreren in het wiskundig model. Verder werd in deze thesis ook aangetoond dat onder zwakke voorwaarden elke aggregatiefunctie behorende tot de familie van gegenereerde functies een geldige implementatie oplevert van het LIED-raamwerk. Door dit gegeven heeft de onderzoeker toegang tot een uitgebreide set van functies die reeds uitvoerig zijn bestudeerd vanuit een wiskundig perspectief binnen het onderzoeksdomein van aggregatiefuncties en die rechtstreeks kunnen toegepast worden om klantentevredenheid te modelleren met onvolledige data volgens de principes van het ED paradigma. De kracht van het LIED-raamwerk ligt in het feit dat het een duidelijke en verstaanbare vertaling aanbiedt aan marketeers tussen de verschillende aspecten van de theorieën omtrent klantentevredenheid enerzijds en de wiskundige eigenschappen van de modelleringsfuncties anderzijds.

Naast de uitwerking van dit modelleringsraamwerk werd in het eerste deel van deze thesis ook een specifieke implementatie aangeboden en bestudeerd. Deze implementatie is gebaseerd op de evaluatieoperator van Dombi, een gegenereerde functie, en wordt de D-LIED implementatie genoemd. Deze implementatie heeft slechts één parameter waardoor het mogelijk wordt om voor iedere klant een apart model te schatten waardoor de onderzoeker in staat is het individueel verwachtingsniveau per klant te achterhalen. Het nadeel van deze implementatie is dat er slechts één verwachtingsniveau wordt gemodelleerd dat voor alle productattributen geldt. Dit nadeel werd in deze thesis omzeild door de hiërarchische structuur van de beschikbare data uit te buiten waardoor het mogelijk werd een verwachtingsniveau op productdimensieniveau te modelleren. Verder werden er verschillende experimenten uitgevoerd om de empirische validiteit van de schatter van het verwachtingsniveau te verifiëren. Al deze experimenten bevestigden dat de resultaten van de D-LIED implementie betrouwbaar zijn.

Vervolgens werd aan de hand van twee gevalstudies de meerwaarde van informatie over de verwachtingen van een klant, afkomstig uit het model, aangetoond. Zo bleek dat sommige productdimensies die op basis van de performantiedata als significant slecht geïdentificeerd waren, minder problematisch te zijn dan de performantiedata deed uitschijnen. Na analyse van de verwachtingen omtrent deze 'slecht presterende' dimensies bleek immers dat klanten niet veel verwachtten van deze dimensies en dat ze dus amper impact hadden op de klantentevredenheid. Anderzijds bleken sommige dimensies die op het eerste zicht zeer goed presteerde amper de verwachtingen te overtreffen.

De D-LIED implementatie werd in deze thesis ook gebruikt om een nieuw type van 'Importance-Performance Analysis' (IPA) uit te werken. Traditionele vormen van IPA gebruiken een enquête of regressieanalyse om de belangrijkheid van een productattribuut te meten. Recent onderzoek heeft echter aangetoond dat zulke aanpak gebreken heeft. Het voornaamste probleem is dat de productattribuutbelangrijkheid vaak wordt voorgesteld aan de hand van een puntschatter terwijl empirisch onderzoek suggereert dat de belangrijkheid van een productattribuut in het klantentevredenheidsproces verandert naarmate het performantieniveau van dat attribuut verandert. De D-LIED implementatie, die een wiskundige voorstelling is van het klantentevredenheidsproces, laat toe de impact van een verandering in het performantieniveau te simuleren. Hierdoor konden er twee nieuwe IPA types gemaakt worden. Het eerste IPA type evalueert en visualiseert het aandeel van een productattribuut in de algemene tevredenheidsscore. Het tweede IPA type visualiseert de impact op tevredenheid bij een stijging of daling van het huidige prestatieniveau van ieder productattribuut.

Tenslotte werd in deze thesis de schatting van het verwachtingsniveau, afgeleid met behulp van de D-LIED implementatie, gebruikt om de compatibiliteit tussen de verwachting van de klant en de performantie van het product te bepalen. De compatibiliteit betreft het absolute verschil tussen de verwachting en de performantie. Op basis van de 'reinforcement learning' theorie werden de hypotheses geformuleerd dat zowel het aanbevelen van een product door klanten bij derden als de neiging om het product opnieuw aan te kopen positief gerelateerd zijn met de compatibiliteit tussen performantie en verwachting. In deze thesis werden twee verschillende compatibiliteitsmaten gedefinieerd en empirisch onderzoek toonde aan dat beide positief gerelateerd zijn met klantenaanbeveling en klantenloyaliteit. Finaal werd aan de hand van een gevalstudie ook geïllustreerd hoe een bedrijf deze compatibiliteitsmaat kan gebruiken om de gevolgen van klantentevredenheid, zoals aanbeveling bij derden en klantenloyaliteit, beter te begrijpen.

Uit de verschillende experimenten en empirische resultaten in het eerste deel van deze thesis bleek dat het LIED-raamwerk en de D-LIED implementatie zeer krachtige technieken zijn om nieuwe kennis te extraheren uit 'onvolledige' klantentevredenheidsdata, i.e. wanneer het bedrijf enkel over performantie- en tevredenheidsdata beschikt. Het tweede deel van deze thesis verlaat het pad van klantentevredenheid en focust op een totaal andere vorm van onvolledige data in de analyse van klantengedrag. Dit deel van de thesis bestudeert het probleem van klantensegmentatie wanneer belangrijke en noodzakelijke observaties of variabelen ontbreken om de ware onderliggende klantensegmenten te detecteren. Een nieuwe fuzzy clustertechniek, PSO-CFC, werd ontwikkeld, dewelke de beperkingen van de data tracht te overstijgen door samen te werken met andere datasites zonder hierbij privacygevoelige data uit te wisselen. Deze clustertechniek is gebaseerd op 'collaborative fuzzy clustering' en gebruikt 'particle swarm optimization' om de mate van samenwerking tussen de verschillende datasites te bepalen. Ook werd aangetoond dat deze nieuwe PSO-CFC clustertechniek verschillende kenmerken deelt met andere gedistribueerde clustertechnieken en gepositioneerd kan worden binnen het domein van 'ubiquitous knowledge discovery'. Experimenten voor zowel situaties waar observaties ontbreken als voor situaties waar belangrijke variabelen ontbreken werden uitgevoerd en de empirische resultaten toonden aan dat PSO-CFC de klassieke lokale clustering overtrof in verschillende experimenten. Wanneer belangrijke variabelen ontbreken, kan PSO-CFC de clusters significant verbeteren wanneer deze elkaar overlappen. Wanneer belangrijke observaties ontbreken, overtreft PSO-CFC lokale clustertechnieken zolang de data willekeurig getrokken is uit iedere cluster. In het bijzonder als de clusters niet in gelijke mate vertegenwoordigd zijn in de data zal PSO-CFC betere clusterresultaten opleveren.

1 Introduction

Understanding customers is becoming increasingly important for companies operating in a consumer-oriented market. A thorough understanding of what customers expect from a product, why they are satisfied with one offering but not with another, and why some customers are not coming back for repurchase, is crucial to any company for being successful. The importance of understanding customers has been recognized by academic researchers in the marketing domain a long time ago. The past decades, several psychological theories have been proposed to explain consumer behavior, and as with any true science, the development of these theories has been verified empirically. A typical approach for developing a consumer behavior theory starts with the formulation of hypotheses which are suggested by the author's own beliefs and existing psychological theories. Once the author has formulated the hypotheses and decided which tests are necessary to verify them, the author knows which data are necessary and starts collecting them. Finally, the data are analyzed and the hypotheses are accepted or rejected.

While academic researchers focus on hypotheses and theory building, companies and their marketing manager are faced with more practical questions concerning consumer behavior such as: "Which aspects of our product have a large influence on the satisfaction of our customers and make them coming back for more?". To answer such questions, companies need to collect data, similar to academics, but their data collection is potentially of a lower quality due to any of the following reasons:

- A company suddenly decides it needs more insights into their customer's behavior and decides to set up a survey for collecting customer data. However, the survey is built and conducted without much thought given to which specific questions the company actually wants to see answered, to which analyses are needed to answer these questions and to which specific data is required for performing such analyses. As a result, the company ends up with general data about their customers which might not be perfectly suited for providing answers to their questions.

- Sometimes a company follows a more structured approach for setting up a survey and starts by analyzing the questions it wants to see answered. However, in such situations it is still possible that their marketing department lacks the necessary expertise and ignores existing consumer behavior theories when building the survey. This results in more specific data but might be insufficient to answer the predefined questions correctly.

- Finally, marketing practitioners are often forced to perform their analyses on secondary data, i.e. data collected in the past for other purposes. Because collecting data costs time and money, marketing practitioners will always try to answer the consumer-related questions by means of the data at hand. Only when these data fail to provide any acceptable answer, a new stage of data collection is considered.

In all three situations, a marketer is faced with incomplete data, i.e. the data that might not be specific enough, lack theoretical validity or might be missing some necessary variables. If the knowledge discovery (KD) techniques, used to extract new information from these data, ignore or are unaware of the data incompleteness, the marketer runs the risk of drawing incomplete, unreliable and possibly biased conclusions.

In this thesis, new KD techniques for understanding customers with incomplete data are presented. Each technique tackles a different type of incomplete data and focuses on a different aspect of customer understanding.

Knowledge Discovery techniques can be roughly divided into two groups, i.e. statistical modeling techniques and data mining techniques. Statistical techniques often start from a mathematical representation of a theoretical model which is believed to have generated the data. Next, the data is used to learn the parameters of the model and inferences are drawn from the model based on assumptions about the data. Obviously, incomplete data can make parameter estimation difficult, biased or impossible.

In contrast to statistical techniques, data mining techniques seldomly start from a mathematical representation of a domain-specific theory, nor do they make many to any assumptions. Data mining techniques typically extract knowledge directly from the data by means of computational power and heuristics. Over the last decades, the domain of data mining and machine learning has made a lot of progress generating a myriad of techniques ranging from *Naive Bayes Classifier, Decision Trees, K-mean Clustering* to more advanced techniques such as *Random Forests, Neural Networks* and *Fuzzy Clustering* to name a few. Most of these techniques have been developed under the umbrella of data mining and outside any particular application field. Since data mining techniques learn directly from data, they can only learn what is inside the data and incomplete data will most likely lead to incomplete conclusions.

Without making a further distinction between statistical modeling and data mining, it has become obvious that marketers could benefit from KD techniques which take data incompleteness into account. In contrast to many existing data mining techniques, such KD techniques will need to be application specific since each type of incompleteness requires a different approach. This thesis considers two scenarios of data incompleteness within the domain of customer behavior.

The first part of this thesis focuses on incomplete data in customer satisfaction. According to the Expectancy-Disconfirmation (ED) Paradigm, i.e. one of the more dominant theories in customer satisfaction, two constructs are directly responsible for customer satisfaction, i.e. the customer's expectation and the customer's perceived disconfirmation between the product's performance and the customer's expectation. Particularly with secondary data, marketers only have data about the customer's satisfaction and the product's performance. The fact that customer's expectation and disconfirmation are missing, makes the data incomplete and hinders direct application of the ED Paradigm. It is also said that in such situations, marketers are working with limited information, i.e. they only know how the customer perceives the product's performance but lack information about the customer's expectation and experienced disconfirmation. In this part of the thesis, the terms "incomplete data" and "limited information" are used interchangeably. The KD techniques in this part of the thesis study how the ED Paradigm can still be applied with incomplete data. This approach is called a theory driven KD technique since it starts from a theoretical model, i.e. the ED Paradigm, which is believed to have generated the data.

Chapter 2 provides a detailed discussion of the ED Paradigm. In this chapter, the author goes back to the days before the ED Paradigm and illustrates how the ED Paradigm came to be. To help the reader understand the dominant role of this theory, the existing literature on this topic is reviewed and various adaptations of the original model are discussed. The purpose of this chapter is to help the reader understand the theory on which the KD techniques in this part of the thesis are based.

Next, in **Chapter 3**, a general framework is created to model customer satisfaction according to the ED Paradigm when only performance and satisfaction data are available. This framework separates the customer satisfaction process into three steps and each step is represented by a mathematical function. Instead of already defining the functional form of these functions, the framework defines a set of properties which must be held by these functions. These properties are defined such that any function having them, implements aspects of typical customer behavior defined by the ED Paradigm. Therefore, the framework, which is called the LIED framework, does not allow a marketer to model customer satisfaction directly, but offers him the flexibility to use functions of any functional form. As long as these functions have the necessary properties, the ED Paradigm is modeled correctly from a theoretical point of view. This chapter clearly exposes the relationship between the properties and the aspects of customer satisfaction behavior, which allows the researcher or marketer to identify the implications of missing properties.

In **Chapter 4**, the data sets which are used throughout this thesis are discussed. These data sets are real-life data sets and had to be cleaned and preprocessed. The entire process of preparing the data for the analysis in this thesis can be found in this chapter.

In **Chapter 5**, an implementation of the LIED framework is presented. This implementation corresponds to a mathematical function which is known as Dombi's uninorm, an aggregation function belonging to the family of generated functions. This implementation has the advantage that a model can be estimated for each customer separately. This chapter concludes with two case studies illustrating how this implementation, which is termed the D-LIED implementation, can be applied to incomplete data providing new and useful information to marketers.

Chapter 6 contains a first application of the D-LIED implementation. This chapter focuses on the marketing tool of Importance-Performance Analysis (IPA) which provides companies insights on which product aspects need further improvement. However, as discussed in this chapter, the current IPA approach faces a couple of difficulties to determine the importance of the different product attributes. A new alternative IPA approach is presented which uses the D-LIED implementation. Finally, a case study shows the potential of this new approach.

In **Chapter 7**, the final chapter of the first part of this thesis, the author shows how the D-LIED implementation can be used to create a new construct, i.e. the performance-expectation compatibility. It is shown that this new construct can help companies to get a better understanding of their customer's intentions, such as the intention to repurchase and the intention of recommending the company to others.

The second part of this thesis focuses on customer segmentation. When marketers try to find homogenous clusters within their customer base, they can face two different types of data incompleteness. Firstly, it could be that the data set is missing important observations, e.g. if the data set only contains information about the current customers, it becomes difficult to identify groups of potential customers which are no customers yet. Secondly, the data can be missing important variables, e.g. if two different types of customers differ on a variable missing in the data, no cluster algorithm will be able to separate the two customer types. In this part of the thesis, a new clustering approach is suggested which overcomes these problems. Since the KD technique in this part does not assume any underlying theoretical marketing model but uses the information in the data to extract new knowledge, it is considered to be a information driven KD technique.

In **Chapter 8**, a fuzzy clustering algorithm is discussed which tries to overcome the limitations of incomplete data by collaborating with data sites from other companies. Typically, such approach is not viable because companies are not willing or might be prohibited by privacy laws to exchange their data. The clustering algorithm in this chapter tries to overcome this limitation by exchanging only a part of the local cluster solutions instead of raw data or the complete local cluster solutions such that data privacy is preserved.

Finally as the ED Paradigm dictates, in order to satisfy the reader of this thesis, it is important to set the expectations right. Therefore, the author would like to end this introduction by positioning the thesis within the field of knowledge discovery and customer satisfaction research. The main focus of this thesis lies on the development of knowledge discovery techniques for marketers studying their customers with incomplete data. The thesis should be positioned on the intersection of knowledge discovery and marketing theory. It has a strong emphasis on the modeling aspects of customer satisfaction, but differs from other KD research in a sense that the techniques developed in this thesis are not general but tailored to specific marketing problems. Some of these techniques might be used in different contexts, but this could require some modifications.

On the other hand, this thesis draws from existing customer satisfaction theories from the marketing domain, but does not have the intention to extend or develop new customer satisfaction theories, safe for chapter 7 which might provide new ideas for future marketing research. The role of marketing theory in thesis is mainly limited to provide a solid and theoretically valid foundation for new knowledge discovery techniques. The main contribution of this thesis is the introduction of new KD techniques aimed at marketers who want to understand their customers but are faced with incomplete data.

Part I

Theory Driven Knowledge Discovery

2 Explaining customer satisfaction: the expectancy disconfirmation paradigm

2.1 Introduction

"Customer satisfaction has come to represent an important cornerstone for customer-oriented business practices across a multitude of companies operating in diverse industries [122]". Customers who are dissatisfied have a tendency to complain [11] or even worse to produce negative word of mouth. These customers will less likely repurchase your products or services. On the other hand, making your customers satisfied could lead to repatronage, acceptance of other products in the same product line and favorable word-of-mouth publicity [22]. Furthermore, researchers have provided evidence that customer satisfaction is related to firm performance [5], willingness of consumers to pay more [63], consumer retention [89, 3], shareholder value [5] and can be considered a prerequisite for customer loyalty [36].

Given the great importance of customer satisfaction, the vast amount of academic literature spent on this topic does not come as a big surprise. When studying customer satisfaction, one can investigate two main questions, i.e. what makes customers satisfied and how or why do customers become satisfied. The first question, which is of importance to business people and marketers, asks which product dimensions, attributes or aspects contribute to customer satisfaction. Knowing this, allows companies to invest in important product dimensions and not waste money on less important product attributes.

Marketing researchers on the other hand are often less interested in the specific attributes which make a customer satisfied which is often product-dependent. Their main goal is to understand the customer satisfaction/dissatisfaction (CS/D) process with all its antecedents and consequents. Ultimately, understanding how and why a customer becomes satisfied will help the identification of important product attributes and their impact on satisfaction. During the last five decades, many researchers have studied various antecedents and consequents of customer satisfaction, together with various relations which were explained by a variety of different consumer theories. Recently, Szymanski and Henard [122] provided a meta-analysis of the empirical evidence gathered in the domain of customer satisfaction. Their article studies the findings of 50 empirical studies, which resulted in a collection of more than 500 correlations between satisfaction and satisfaction antecedents.

In their article, Szymanski and Henard investigated the relationship between satisfaction and five specific antecedents, i.e. expectation, disconfirmation, performance, affect and equity. The role of expectation can be twofold. Firstly, it creates a level of anticipation and is assumed to have a direct positive influence on the customer's satisfaction. Secondly, expectation can also act as a reference against which performance is compared, resulting into a positive or negative expectancy disconfirmation which is the second antecedent. It is assumed that positive disconfirmation has a positive influence on satisfaction, while negative disconfirmation has a detrimental influence on satisfaction. Since disconfirmation is the resultant of expectation and performance, some studies assumed that the effect of performance (and expectation) is subsumed by the disconfirmation effect. However, other researchers (e.g. [26]) have found that performance also has a direct positive influence on satisfaction, aside from its indirect effect through disconfirmation. Therefore, performance is considered as a separate antecedent by Szymanski and Henard. The fourth antecedent, i.e. affect, adds an affective component to the satisfaction process. It is assumed that emotions, elicited throughout the product consumption process, leave affective traces in the consumer's memory which are integrated during the satisfaction assessment. This suggests a positive relationship between affect and satisfaction. Finally, equity is a fairness judgement that consumers make in reference to what others receive. Equity theory suggests that consumers become satisfied as they perceive their equity ratio as proportionately greater than the ratio achieved by the referent person or group. Table 2.1 shows the five antecedents and their hypothesized relationship with customer satisfaction.

Table 2.1: Hypothesized relationships between customer satisfaction and its antecedents

Antecedent	Hypothesized relation with satisfaction
Expectation	Positively related
Disconfirmation	Positively related
Performance	Positively related
Affect	Positively related
Equity (ratio)	Positively related

Calculating the mean correlation for each of these antecedents with satisfaction, Szymanski and Henard found that all five antecedents showed a positive and statistically significant correlation with satisfaction. Among the five antecedents, disconfirmation and equity had the strongest relationship with satisfaction. In the remainder of this dissertation, only expectation, disconfirmation and performance are considered, which are the three antecedents of Oliver's Expectancy Disconfirmation (ED) paradigm [96]. Oliver's later customer satisfaction model [97], introduced in 1996, can be regarded as his original ED paradigm extended with concepts such as equity and affect. The choice to focus on the original ED model, ignoring affect and equity as possible antecedents, does not reject the potential influence of these antecedents on satisfaction. However, the goal of this part of the dissertation is to develop data mining tools for customer satisfaction research which are based on solid and well established marketing theories and which can be used directly by companies and marketers. Therefore it needs to be taken into account that most companies neither collect nor have the necessary data available to integrate equity or affect theory into the data mining models. Building such models would not only complicate the modeling and learning stage, but would render the final tool of less practical value. By focussing solely on the expectancy disconfirmation paradigm, some explanatory power is sacrificed, by putting affect and equity aside, for some computational parsimony and an increased practical applicability. Nonetheless, this approach still integrates three of the five antecedents found significant by Szymanski and Henard [122] and this model starts from the ED paradigm, one of the most dominant consumer satisfaction models over the last three decades.

In this chapter, the reader is introduced to the CS/D literature. Firstly, the research findings concerning satisfaction prior to the introduction of the ED paradigm [96] are examined, followed by an elaborated discussion on the ED paradigm and all related literature. Finally, a summary of all relevant CS/D research findings is provided as the starting point for the remainder of the first part of this thesis.

2.2 Before the expectancy disconfirmation paradigm

When Oliver [96] introduced the ED paradigm in 1980, the ideas it encompassed were not something researchers in the consumer psychology field were unfamiliar with. The impact of expectation and disconfirmation on customer's post-purchase behavior had already been studied by several authors. Oliver's main merit comes from the fact that he succeeded to combine the existing research results with theories from psychology into a solid framework linking various satisfaction antecedents and consequents together. Therefore, it is important that some of the influential research results preceding Oliver's article are reviewed. Note that in the early days of CS/D research, not all authors used the concept of satisfaction as the dependent variable. Often researchers used other concepts such as product evaluation or perceived product quality instead. Although these concepts are not necessarily the same as satisfaction, as was pointed out by LaTour and Peat [80], it is likely that they are highly correlated with satisfaction. Therefore, the empirical results of these studies can still provide interesting insights, but they should be interpreted with the necessary precaution.

One of the first articles trying to explain customer satisfaction, which is often cited by various authors, is the article by Cardozo [22]. In his work, he studied the effect of customer effort and negative disconfirmation, i.e. when customers receive less than they expected, on satisfaction. One can summarize Cardozo's hypothesized theory as follows:

Customers who put only little effort in choosing and buying a product will not feel strongly committed to their choice or the product itself. Therefore, if they experience that the product performs less than they expected, they will evaluate the product less favorably than customers who had expected this lower performance. While both type of customers would receive the same product, the first type will evaluate the product less because of the negative disconfirmation the customer experienced. However, if the customers, who experienced negative disconfirmation, did put a lot of effort in the buying process, the effect of negative disconfirmation will disappear because they become more committed to their choice.

Cardozo uses two different psychological models to predict the customer's product evaluation, depending on the amount of effort the customer has spent. The first theory is the contrast theory, for which Cardozo referred to Helson's article [61], which according to Cardozo holds for low-effort customers. This psychological theory assumes that customers magnify the negative discrepancy between the perceived product performance and their expectation, resulting in a less favorable evaluation. This theory proposes disconfirmation as the main influence of the customer's satisfaction outcome, i.e. the greater the negative disconfirmation, the less favorable the product evaluation.

The second theory is Festinger's theory of cognitive dissonance [47] which according to Cardozo holds for high-effort customers. This theory states that if customers have spent a great deal of effort in buying a product, the customers will feel committed to their choice and when facing an error of judgement, they will most likely experience mental dissonance. The customers will relieve this mental distress by raising the evaluation of the product towards their expectations. According to this psychological theory, expectation dominates the satisfaction process as the satisfaction outcome is drawn towards the expected performance level.

The results of a laboratory experiment largely confirmed Cardozo's theory. However, three remarks should be made. Firstly, Cardozo only studied the effect of negative disconfirmation and therefore the question remained if contrast theory also holds for positive disconfirmation. Secondly, Cardozo created negative disconfirmation by increasing the expected product performance. This implies that when he compares customers with no disconfirmation against customers with negative disconfirmation, he is also comparing customers with low and high expectation levels. This experimentally induced correlation between disconfirmation and expectation makes it difficult to isolate the effect of disconfirmation on the final product evaluation. Finally, several authors (e.g. [7, 100, 101]) argued that Cardozo made a methodological error in his experiment such that he cannot compare the product evaluations of the customers with negative disconfirmation against the product evaluations of the customers with no disconfirmation.

Five years later, in 1970, Cohen and Goldberg [28] published an article in which they continued the study of dissonance theory in post-purchase product evaluation. They argued that growing contradictory evidence existed with dissonance theory which led to disaffection with the theory. One of the main criticisms towards dissonance theory is that it assumes that customers do not learn from their mistakes. If customers keep raising the product evaluations towards their expectations to remove mental dissonance, they never come to realize that they are buying a low quality product. However, Cohen and Goldberg also mention that dissonance theory should not be completely ignored in consumer satisfaction research since there is a vast amount of evidence which confirms that people try to avoid and reduce cognitive dissonance. In their introduction they argue that dissonance theory can not act as a general theory which explains every satisfaction process, but does "offer a parsimonious explanation for many otherwise disconnected observation ([28], p.315)".

In their study, they investigated two different ways of post-purchase cognitive reevaluation. On the one hand, consumers might try to justify their decision, no matter the outcome, which comes down to the dissonance theory. On the other hand, customers might learn from the outcome of their purchase and accept the mistake of their decision if thee experience it as such. This is referred to as the outcome learning theory. Cohen and Goldberg set up an elaborated experiment with instant-coffee, both from several known national brands and an unknown test brand. The experiment contained four distinct stages, i.e. a decision stage, an immediate post-decision stage, a non-consumption stage and a post-consumption stage. The non-consumption stage refers to an inspection test where the customer can judge several product dimensions without actually consuming the product, i.e. the non-taste attributes such as the aroma and the appearance.

From their results, they could draw some interesting conclusions. Firstly, it appeared that dissonance theory played a role when evaluating the product on the non-taste aspect, i.e. during the non-consumption stage, but only for the national brand. They explained this by the fact that customers who chose one of the national brands have prior beliefs about the product, which makes those customers less susceptible to new information about the product's performance. However, during the post-consumption stage, the expectation effect disappeared in favor of a confirmation/disconfirmation effect, supporting the outcome learning theory.

Finally, it is worth mentioning that similar to Cardozo [22], Cohen and Goldberg only studied the effect of negative disconfirmation during the consumption stage. However, in contrast with Cardozo [22], Cohen and Goldberg did not change the customer's expectation in their experiment, but the product's performance by altering the taste of the test brand in a negative way. This avoided any experimentally induced correlation between disconfirmation and expectation.

Next in 1972, Olshavsky and Miller [100] wrote an article in which they questioned Cardozo's conclusion [22] which suggested that overstatement of the product's performance, causing a negative disconfirmation, leads to an unfavorable evaluation of the product. They believed that, in contrast with some of the results of both Cardozo [22] and Cohen and Goldberg [28], dissonance theory, which focuses on expectation rather than disconfirmation, is the proper theory to explain customer satisfaction. To test their theory, Olshavsky and Miller created an 2x2 factorial design experiment in which they varied the product's performance and the product's expectation from low to high, which resulted in 4 conditions: high expectation with high performance (HE-HP), high expectation with low performance (HE-LP), low expectation with high performance (LE-HP) and low expectation with high performance (LE-HP). For each condition, they measured the customer's overall product evaluation on a scale from 0 to 7. Table 2.2 shows the average ratings from their experiment.

Table 2.2: Average ratings on overall performance [Olshavsky and Miller [100], p.20]

	High expectation	Low expectation
High performance	$5.5(\sigma = 0.96)$	$4.5(\sigma = 1.18)$
Low performance	$3.7(\sigma = 1.60)$	$3.1(\sigma = 1.20)$

Based on the dissonance theory, Olshavsky and Miller formulated 2 hypotheses which were empirically verified. The first hypothesis stated that in case of negative disconfirmation, the product evaluation will be drawn towards the expectation, which is higher than the perceived performance, causing a more favorable product evaluation. To test this hypothesis, they compared the average product evaluation in the LP-LE case with the LP-HE case. The second hypothesis stated that in case of positive disconfirmation, the product evaluation will be drawn towards the expectation, which is lower than the perceived performance, causing a less favorable product evaluation. To test this hypothesis, they compared the average product evaluation in the HP-HE case with the HP-LE case. The results from Table 2.2, which is adopted from Olshavsky and Miller [100], confirm both hypotheses and consequently support the dissonance theory. They explained the contradicting results found by Cardozo [22] due to a methodological flaw: "[Cardozo's] results were based on product evaluations anchored in expectations produced by high- and low-expectation catalogs [100]".

Although their work provided evidence for the dissonance theory, which posits expectation as a main contributor to customer satisfaction, the results also seem to provide evidence for contrast theory, which focuses on disconfirmation. If one compares the situations HP-LE and LP-LE, one can see the effect of positive disconfirmation on customer satisfaction given a constant level of expectation. Table 2.2 shows that in this comparison, positive disconfirmation has a strong positive effect on satisfaction. This positive effect caused by positive disconfirmation contradicts assimilation theory and supports contrast theory. The same argumentation holds for the negative effect on customer satisfaction caused by negative disconfirmation, which is revealed when comparing the situations HP-HE and LP-HE. However, Olshavsky and Miller only focussed on dissonance theory and the effect attributable to expectation.

Until Anderson's article [7] in 1973, researchers always used either dissonance or contrast theory to explain customer satisfaction. Dissonance theory or assimilation theory assumes that disconfirmation causes a psychological discomfort which is resolved by assimilating perceived performance towards expected performance. Contrast theory is somehow the opposite and assumes that any disconfirmation is magnified during product evaluation.

In his work, Anderson added two more plausible psychological theories to explain customer satisfaction. The first is called the generalized negativity theory which assumes that any discrepancy between expectation and performance has a negative influence on the product evaluation or satisfaction. This theory is based on previous work by Carlsmith and Aronson [23]. The second additional theory is the assimilation-contrast theory which combines both dissonance and contrast theories. It assumes that customers have a zone of acceptance, wherein assimilation theory holds: i.e. if the expectancy disconfirmation is small enough, such that it falls in the zone of acceptance, customers will ignore the discrepancy and evaluate the product as if they received what they expected. Outside the zone of acceptance lies the zone of rejection in which the contrast theory holds: i.e. if the disconfirmation becomes too large, the customer will exaggerate the perceived discrepancy, both in a positive and a negative way. Figure 2.1 illustrates this theory.

In his article, Anderson discussed the results of an experiment in which he kept the objective performance of the product constant and changed the customer's expectation. He used this changing expectation as the independent variable and the product evaluation as his dependent variable. His results suggested that neither assimilation, neither contrast nor generalized negativity theory were supported. Only, the assimilation-contrast theory could not be rejected, from which Anderson concluded that this was the underlying theory in the satisfaction process. However, he did point out that his research concerned simple and easy to understand products, i.e. ballpoint pens, and argued that assimilation-contrast theory might not hold for more complex products since assimilation theory had been supported by previous work. This argument remained untested in his article.

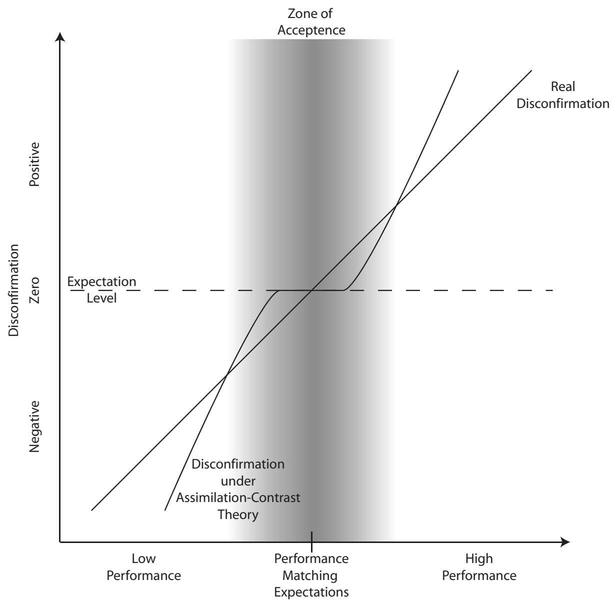


Figure 2.1: The zone of acceptance and its influence on disconfirmation.

Next, in 1976, an article by Olson and Dover [101] was published which introduced a precise definition of expectation. In their work, they claimed that previous research on consumer expectations contained some major theoretical and methodological problems. They argued that previous research implicitly assumed that customers' expectations could be manipulated through exposure of subjects to specific product related information. Only Anderson [7] actually performed a manipulation check to verify this assumption. Olson and Dover argued that previous research failed to identify alternative external sources for customers' expectations, such as advertisement, word-of-mouth, product observation and usage experience. Furthermore, none of the previous research recognized the possible longitudinal dimension of expectation, i.e. expectations are formed over time through several exposures to the product. Finally, they claimed that the expectancy manipulation was transparent to the subjects in some experiments and that none of the previous research specified the type of expectation they were measuring and manipulating.

According to Olson and Dover, the origin of all these problems is the absence of a precise and theoretical rich conceptual definition of expectation in previous research. In their work, they provided such definition and positioned expectation within the literature of cognitive structures and attitude formation and creation. They defined expectations as an individual belief element in a consumer's cognitive structure regarding the product. In other words, "an expectation is the perceived likelihood that a product possesses a certain characteristic or attribute, or will lead to a particular event or outcome ([101], p. 169)". In contrast with previous work, Olson and Dover did not consider expectation at the overall product level (e.g. a car), but at a product dimension levels (e.g. a car's fuel consumption). A customer can have different belief-expectation elements for a single product, i.e. one for each dimension.

They argued that these belief-expectation elements can explain the acquisition of a consumer's overall product attitude, which could be considered as a proxy for an overall product expectation construct.

In their work, Olson and Dover, performed an experiment with ground coffee in which they manipulated the customer's expectation of the dimension 'coffee bitterness', over a period of 12 days, both through ad-like communication and product usage. Their results showed that both assimilation-contrast theory and dissonance theory could explain the measured effects of disconfirmation. They pointed out that both theories predict the same pattern of results except for when expectations are very strongly disconfirmed.

Olson and Dover, concluded that product usage has a much stronger impact on belief-expectancy elements than information exposure, which has important implications for experiments trying to manipulate consumer expectations. Furthermore, they warned for possible spill-overs from manipulated expectation elements on non-manipulated expectation elements, which should be accounted for in future experiments. Finally, they also suggested two other theories which they considered useful for studying the CS/D process, i.e. Helson's [61] adaptation level theory and comparison level theory [124].

While Helson's adaptation level theory would later be used as the basis for Oliver's ED paradigm, the comparison level theory was picked up by LaTour and Peat [80]. Being unsatisfied with the existing theories explaining CS/D, LaTour and Peat provided an alternative conceptualization which was based on Thibaut and Kelley's [124] comparison level theory. In a purely theoretical article [80], they developed a model which considers the consumer's satisfaction to be determined by the discrepancy between the outcome and a standard of comparison known as the Comparison Level (CL). They argued that positive discrepancies would be satisfying, while negative discrepancies would be dissatisfying.

The discrepancy between the product performance and the comparison level should not be confused with the expectancy disconfirmation used by other authors in those days. Previous theories always used the customer's product expectation as the comparative referent which was often considered at a product level. Within LaTour and Peat's model, the comparative referent is more than just the expected performance. They defined the CL as a "function of experiences with similar products or services in the past, to a lesser extent the experiences of similar consumers, and to a still lesser extent outcomes promised by the manufacturer, retailer, and/or service provider [80]". Furthermore, LaTour and Peat also suggested that such CL existed for each product attribute and that the overall satisfaction would be an "additive function" of the discrepancies from the CL for each salient attribute.

It should be noted that within LaTour and Peat's comparison level adaptation, the CL has no direct effect on customer satisfaction, which stands in contrast with Oliver's ED paradigm. Also, LaTour and Peat's original article [80] did not contain any empirical evidence for their theory. They did however publish an article a year later [81], in which they empirically studied the three elements of the comparison level, i.e. the consumer's prior experiences, the situationally-produced expectations and the experiences of other consumers. From the empirical results they found that only prior experience had an effect on the overall satisfaction.

Meanwhile, in 1977, Oliver [95] published an article in which one can already recognize the foundations of his expectancy disconfirmation paradigm. He reviewed various articles which had been published on the topic of CS/D and concludes that only five studies in the late 1960's and 1970's were published in this domain. One set of studies focussed on the effect of disconfirmation, while the other studies considered the size of disconfirmation less important and focussed on the effect of the expectation level. However, all these studies only focussed on one of both effects.

Oliver argued that both constructs, i.e. expectation and disconfirmation, might simultaneously influence post-exposure product evaluation. For this idea, he referred to an analogue theory within the domain studying people's attitudes. He stated that expectation and disconfirmation can explain post-exposure evaluation "in much the same way that any revised attitude can be explained in terms of initial position plus degree and direction of change [95]". However, he also noted that if one wants to study disconfirmation and expectation simultaneously, one can get caught in a conceptual overdetermination. The three concepts of expectation, disconfirmation and product performance are all related and one can only manipulate two of them independently. In previous research, performance and expectation were always manipulated and disconfirmation was measured indirectly as the difference between expectation and performance. This implied an axiomatic negative correlation between expectation and disconfirmation, e.g. if one increased expectation to a higher level while keeping performance at a constant level, disconfirmation became negative. Obviously, such negative correlation between expectation and disconfirmation, makes it very difficult, if not impossible to measure the isolated effect of any of these constructs on customer satisfaction.

Furthermore, Oliver argues that the constructs expectation and disconfirmation do not have to be (negatively) correlated in real-life situations. He supports this argument by stating that expectation formation and disconfirmation experience occur in different moments in time. He even posits that disconfirmation experience is closer in time to the post exposure evaluation than expectation formation and consequently disconfirmation will have the greater effect on satisfaction.

Because of the danger of conceptual overdetermination in an experiment, Oliver investigated the relationships between disconfirmation, expectation and post-exposure evaluation in a field study. In his research, he measured both disconfirmation and expectation at an attribute level and aggregated level and post-exposure evaluation is measured by 'overall affect' and 'intention to buy'. The data revealed that expectation and disconfirmation are indeed uncorrelated. Only aggregated expectation and aggregated disconfirmation revealed a negative correlation, but this can be explained by a methodological flaw. The data also revealed that both expectation and disconfirmation have a positive relation with the post-exposure evaluation measures. For the measure 'overall affect', disconfirmation indeed had the greatest effect, while for 'intention to buy' the effects of disconfirmation and expectation were equal. Most likely, it were the results of this study which led to Oliver's next article in which he introduced the Expectation Disconfirmation Paradigm.

2.3 Oliver's expectancy disconfirmation paradigm

The expectancy disconfirmation paradigm

The discussion in the previous section gives a good picture of the kind of research results which were at hand in the late 1970's. One thing which strikes the attentive reader is probably the fragmented and sometimes contradicting nature of all these results. As for possible antecedents of satisfaction, one can detect two different lines of research. On the one hand there are researchers which focus on expectation, while others focus on disconfirmation. Closely related to this dichotomy are the different psychological theories which were proposed to explain the customer satisfaction process. While some researchers considered dissonance theory, others advocated contrast theory, while some combined both theories.

Based on all these results, the impression arises that both expectation and disconfirmation play a role in the satisfaction process and that both dissonance and contrast theory must explain a part of the relationships between antecedents, satisfaction and consequents. However, a coherent framework which connects these concepts and theories was still missing. Such a framework was introduced in 1980 by Oliver and is often referred to as the Expectancy Disconfirmation (ED) paradigm [96]. To date, this is still one of the most dominant customer satisfaction paradigms.

One of Oliver's main contributions is the idea that expectation and disconfirmation have an additive and simultaneous effect on satisfaction, whereas previous theories always focussed on only one of both constructs. Contrast theory for example contributed the effect on satisfaction fully to the magnitude and direction of the disconfirmation, while assimilation or dissonance theory hypothesized that satisfaction was positively related to expectation. Even in the assimilation-contrast theory, which combines both theories, it is assumed that, depending on the size of disconfirmation, either contrast (disconfirmation) or assimilation (expectation) theory is in play, but never both at the same time.

The use of such an additive interpretation is modeled well by Helson's [61] adaptation level theory, which was already being considered by researchers in the field of job satisfaction [68, 116] and suggested by Olson and Dover [101]. "[Helson's theory] posits that one perceives stimuli only in relation to an adapted standard. The standard is a function of perceptions of the stimulus itself, the context, and psychological and physiological characteristics of the organism. Once created, the 'adaptation level' serves to sustain subsequent evaluations in that positive and negative deviations will remain in the general vicinity of one's original position. Only large impacts on the adaptation level will change the final tone of the subject's evaluation. [96]"

Translated to the field of customer satisfaction, Oliver defined the following relationships between expectation, disconfirmation and satisfaction. Firstly, expectation acts as the adaptation level. It sets some kind of standard against which performance is evaluated and provides an initial level of satisfaction. Customers with high expectations will have a high initial level of satisfaction and vice versa. Next, after some period of consumption, the perceived performance will probably deviate from the adaptation level, which refers to the disconfirmation. According to Helson's adaptation level theory, the effect of this disconfirmation will be nullified if the performance remains in the general vicinity of the expected performance. Only if disconfirmation is large enough, it will have an effect on satisfaction. A positive disconfirmation will have a positive effect, while a negative disconfirmation will have a negative effect. Obviously, the disconfirmation effect in this theory is related to the concept of 'zone of acceptance'. Oliver mathematically expressed the relationship between expectation, disconfirmation and satisfaction as follows:

$$\text{satisfaction} = f(\text{expectations}, \text{disconfirmation}).$$

However, an additive interpretation of the expectation effect and the disconfirmation effect on satisfaction requires that both concepts, i.e. expectation and disconfirmation, are uncorrelated. This hypothesized absence of relation between expectation and disconfirmation has been tested and verified by Oliver in [95] and [96] and has also been confirmed by Szymanski and Henard's [122] meta-analysis.

The relationship between satisfaction, expectation and disconfirmation is only the first part of Oliver's ED paradigm and is illustrated in Figure 2.2 within the dashed box. Oliver continued his framework by integrating some satisfaction consequents into it. The first consequent he considers is the customer's attitude towards the product. The relationship between attitude and satisfaction is based on existing work by Howard and Sheth [67], Fishbein [48] and Helson's [61] adaptation theory again. Firstly, the pre-purchase attitude serves as the foundation for the post-purchase attitude. Subsequently, it also acts as an adaptation level for satisfaction. Oliver extends his framework even further and includes the customer's purchase intentions. Based on the work of Howard [66], Fishbein and Ajzen [49], Oliver hypothesizes that both satisfaction and post-purchase attitude influences post-purchase intentions, while pre-purchase intentions act as an adaptation level. Since this dissertation mainly focuses on satisfaction and its antecedents, we will limit our further discussion and review of Oliver's ED paradigm to the first part, i.e. the relationships between expectation, disconfirmation and satisfaction. From now on, when the ED paradigm is mentioned, it implicitly refers to the first part of it.

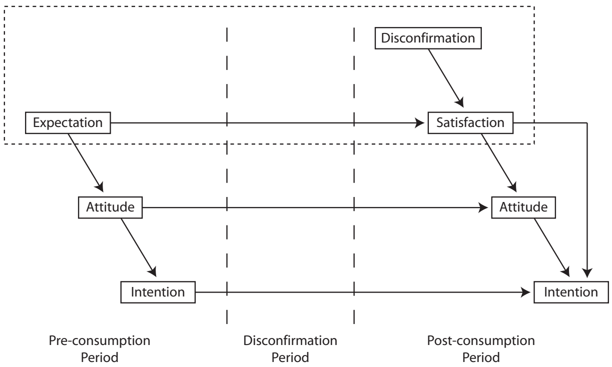


Figure 2.2: Oliver's expectancy disconfirmation paradigm [adapted from Oliver [96], p. 462].

From the results of the experiment in Oliver's [96] original article on the ED paradigm, three important conclusions can be drawn. Firstly, disconfirmation appeared to be uncorrelated to expectations which is crucial for considering satisfaction as an "additive function" of expectation and disconfirmation. Secondly, disconfirmation had the largest effect on satisfaction and was significant in all four experiments. Thirdly, the idea of an adaptation level influencing the satisfaction outcome was supported in three out of four experiments. Expectation acted as adaptation level in one experiment, while in two other experiments this role was fulfilled by pre-purchase attitude. Over the years various authors have provided support for the relationships between expectation, disconfirmation and satisfaction as hypothesized by the ED paradigm. Some of them are LaBarbera and Mazurski [79], Swan and Trawick [121], Bearden and Teel [11], Churchill and Surprenant [26] with their non-durable products experiment and Oliver and Desarbo [99]. As for which antecedent, expectation or disconfirmation, has the largest impact on satisfaction, the current amount of research suggests that in most cases the disconfirmation effect will be dominating the CS/D process. Szymanski and Henard [122]'s meta-analysis showed that the average sample-size adjusted correlation between satisfaction and disconfirmation is 0.37 and only 0.19 between satisfaction and expectation.

Over the years, many researchers tried to advance the ED paradigm by integrating new concepts such as complaint behavior [11], adding new antecedents such as performance [26, 127, 99] or redefining the interpretation of an antecedent such as expectation [134, 19, 127]. Over the years, the basic idea of the ED paradigm remained intact, albeit with some extensions and redefinitions of the antecedents. Next each antecedent and its role within the ED paradigm will be discussed separately.

Expectation

The role of expectation in the ED paradigm is dual. Firstly, it has a direct effect on satisfaction by setting the adaptation level and secondly it acts as a comparison referent for perceived product performance. The first role of expectation as a direct effect on satisfaction has been studied by various researchers with mixed results. Sometimes empirical evidence was found for a significant and positive effect (e.g. Anderson, Fornell and Lehmann [4]; Cadotte, Woodruff and Jenkins [19]; Churchill and Surprenant [26]; Oliver [96]; Tse and Wilton [127]) while other experiments found no significant effect (e.g. Oliver [96]; Churchill and Surprenant [26]; Anderson and Sullivan [6]). Some authors argued that the direct effect of expectation on satisfaction was influenced by other aspects. Churchill and Surprenant [26] did find a direct effect for their experiment with a house plant but not for their experiment with a video disc player. They argued that perhaps the ED paradigm only holds for non-durable products. As far as the author knows, this assumption has never been tested. Voss, Parasuraman and Grewal [133] explained the absence of a direct expectation effect on satisfaction by inconsistency between price and performance. If the price is fair in comparison to the delivered performance, i.e. low price for low performance and high price for high performance, satisfaction is directly influenced by the customer's expectation. However, Voss et al. argued that if companies ask an unfair price, i.e. a low price for a high performance or a high price for a low performance, expectation would not play a role in the CS/D process. They performed an elaborated experiment with hotel services and found empirical evidence for their hypothesis.

Although the overall picture of the direct expectation effect on satisfaction remains fuzzy, Szymanski's and Henard meta-analysis [122] supported a direct and positive effect as postulated by the ED paradigm. They found 8 studies investigating the expectation-satisfaction relationship which accounted for 17 measured correlations between expectation and satisfaction. Thirteen of these correlation were found significantly positive, while 1 correlation was positive but insignificant, 1 was significantly negative and 2 were negative but insignificant. Furthermore, the sample size adjusted average correlation between expectation and satisfaction was 0.19 and statistically significant.

Aside from its role as adaptation level, expectation also has an indirect effect on satisfaction by acting as a reference against which performance is compared, which results in disconfirmation. However, the definition or interpretation of the 'expectation as referent' concept has been much debated over the years. The first authors to formulate a precise definition for expectation were Olson and Dover [101]. They defined expectation as a belief concept or as "the perceived likelihood that a product possesses a certain characteristic or attribute, or will lead to a particular event or outcome". For example, a customer's expectation about a coffee's bitterness could be expressed as 'Customer X beliefs coffee Y is not bitter at all'. Note that Olson and Dover's definition of expectation positions expectation at the attribute/dimension level, rather than at the overall product level. Olson and Dover's definition of expectation was not only conceptually rich and precise, but also very appealing since it positioned consumer expectations squarely into the theoretically-rich literature dealing with cognitive structure and attitude formation and change.

Oliver [96] extended this definition by relating the concept of belief to an evaluation outcome. In his definition, expectation is the product between the belief of a specific outcome and the evaluation how good or bad this outcome is for the customer. This allows high expectations to be defined in two ways, i.e. the belief that desirable events will occur and belief that undesirable events will not occur. The rationale behind low expectations is analogue. This extended definition of expectation interprets the concept at the product attribute or dimension level. In order to end up with an overall product-level expectation, Oliver sums the various expectations with regard to different dimensions of the product. In short, Oliver defines the overall expectation as the sum of belief-evaluation products, which can be mathematically expressed as in Eq. 2.1. It should be noted that Olson and Dover [101] suggested that product attitude may be considered as a global expectation about the product which would explain why both expectation, in the form of sum of belief-evaluation products, and pre-purchase attitudes appeared as adaptation levels in Oliver's experiments.

$$\text{Overall Expectation} = \sum^{\text{dimensions}} \big(\text{Dimension evaluation} \times \text{Belief of occurence}\big) \qquad (2.1)$$

However, some authors have suggested that other types of expectations do exist and might be used as a comparative referent for product performance. For example, Miller [87] defined four types of expectations varying by level of desire. Expectation as it was originally interpreted by Olson and Dover and Oliver corresponds to Miller's expected performance level, which is also called the predicted level. The three other levels of performance identified by Miller were the ideal level, the deserved level and the minimum tolerable level. Miller denoted all four levels as expectation, but Woodruff et al. [134] pointed out that the ideal, deserved and minimum tolerable level are rather norms than expectations. These three definitions express what customer think they *should* receive rather than what a customers *expect* to receive.

Oliver [97] from his part, continued to use the terminology of Miller and called all four constructs expectation. He extended those four levels of expectation with work from Zeithaml et al. [103, 102] which resulted in 8 different levels of 'expectation' which are depicted in Figure 2.3. Oliver also identified two different zones of performance, i.e. the zone of indifference and the tolerance zone, which are depicted in the same figure. The former corresponds to the zone of acceptance which refers to a range of performance in which all performance levels are treated more or less equal

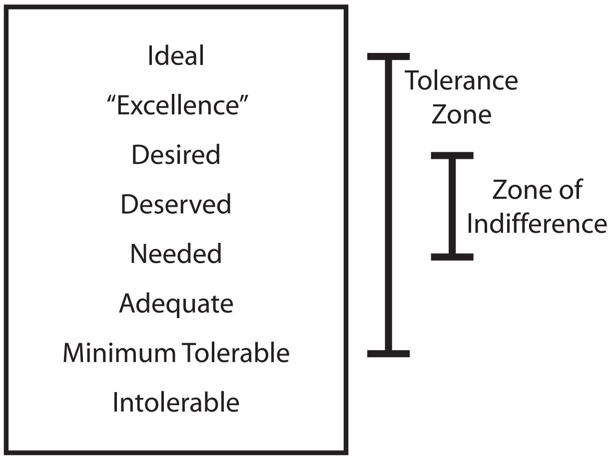


Figure 2.3: Expectations according to the level of desirability [adapted from Oliver [97], p.72].

to the expected performance level. The tolerance zone is a much broader range which represents all the performance levels which are tolerable to the consumer.

Cadotte, Woodruff and Jenkins [134, 19] suggested two other norms as a comparison referent in the ED paradigm. In contrast with expectation, their referents are based on knowledge of and experience with similar but other products from different brands. They called their type of referents 'experience based norms' and distinguished between product-based norms and best-brand norms. If for a specific product, the performance of a specific brand stands out from the rest on all product dimensions, they argue that customers use this performance level as a norm against which performance of other products is compared during the CS/D process. They also pointed out that if the focal product is from the superior brand, the norm used for comparison is equal to Oliver's concept of expectation.

If there is not a single brand which performs better than all the other brands for each product dimension, Cadotte et al. suggest that customers might average the performance of a group of similar brands, resulting in a product-based norm. From their empirical results, they found that the concepts expectation, brand-based norm and product-based norm are moderately correlated, revealing some overlap. On the other hand, they differ enough to consider them as three different concepts. They also tested all three potential comparative referents and found that the models with a product-based norm fitted the data best in two experiments and the model with the brand-based norm fitted the data best in the third experiment. Cadotte, Woodruff and Jenkins [19] concluded that expectation could not be ruled out as comparative referent, but empirical data suggested that other constructs such as experience-based norms are also viable candidates as comparative referent in the ED paradigm.

The idea of using 'experience-based norms' was not a new idea when it was suggested by Cadotte, Woodruff and Jenkins. As a matter of fact, LaTour and Peat [80] also introduced a comparative referent which was partially based on prior experiences. In contrast with the 'experience-based norms' of Cadotte et al., LaTour and Peat's comparative referent is a combination of three components. The first and most important component is the consumer's prior experiences with similar products, the second component is the customer's expectation about the product and the third and final component are the experiences of other consumers with the product. One could consider the comparison level of LaTour and Peat as a combination of customer's expectation and 'experience-based norms'.

While other authors debated on which referent was the right one, Spreng et al. [117] proposed a model with two referents, i.e. expectation and desires. The results of their experiment supported the hypothesis that customers compare the product performance against both an expected performance level and a desired performance level. Spreng et al. argued that it is possible that one referent might dominate the CS/D process. For example, desire as referent might dominate if it concerns a new product for which the customer has no expectations yet. On the other hand, expectation could be the dominating referent if it concerns a product which the customer has already bought many times before.

Disconfirmation

The second antecedent of customer satisfaction in the ED paradigm is disconfirmation, which is the discrepancy between the perceived performance and the customer's expectation. Within the literature, there are two alternative conceptualizations of the disconfirmation construct. The first, which is called subtractive disconfirmation or objective disconfirmation is modeled as the result of the difference between product performance and expectation [80]. Such type of disconfirmation can be easily applied to product attributes such as gas mileage of a car for which both expectation and performance can be measured objectively. As Oliver pointed out [97], subtractive disconfirmation can become more troublesome for attributes, such as driving comfort of a car, for which performance and related expectation cannot be measured with exact precision. In such cases, subtractive disconfirmation, being the difference between two imprecisely measured constructs, is imprecise too. Another problem with subtractive disconfirmation is that customers might evaluate the same discrepancy between expectation and performance differently. For example, one customer might consider a gas mileage disconfirmation of 1 liter per 100 kilometers as small, while another customer might find it unacceptable.

Some authors prefer subjective disconfirmation, which is measured directly on a bipolar 'better than expected / worse than expected' scale. Such bipolar scale gives rise to three different states, i.e. negative, zero and positive disconfirmation. Oliver [97] combined these three different states with the outcome of the three different components of disconfirmation, i.e. the event, its probability of occurrence, and its (un)desirability, to provide a precise definition of disconfirmation. Table 2.3 illustrates how the state of disconfirmation is related to these three components.

In general, subjective disconfirmation is preferred above subtractive disconfirmation for the reasons mentioned above. Also, Tse and Wilton [127] performed an experiment in which they compared the effect of both subtractive and subjective disconfirmation on satisfaction. Their results showed that subjective disconfirmation is superior to subtractive disconfirmation as conceptualization within the ED paradigm.

In the original version of Oliver's ED paradigm, disconfirmation is assumed to have a positive effect on satisfaction. Positive disconfirmation will increase satisfaction,

Table 2.3: Categories of disconfirmation and states of nature [Oliver [97], p.104]

State of Disconfirmation	Consumer's Experience
Positive	Low-probability desirable events occur, and/or high-probability undesirable events do not occur
Zero	Low- and high-probability events do or do not occur as expected.
Negative	High-probability desirable events do not occur, and/or low-probability undesirable events occur.

while negative disconfirmation will decrease satisfaction. In general this positive relation between disconfirmation and satisfaction seems to be valid. Szymanski and Henard found 137 measured correlations between disconfirmation and satisfaction among which 121 were significantly positive.

Anderson and Sullivan [6] suggested that the experienced level of disconfirmation is influenced by the level of difficulty to distinguish between high and low performance. Their experimental results supported the hypothesis that "when buyers find it easier to distinguish between high and low quality, they are more likely to experience disconfirmation [6]". Oliver [97] recognized the possible difference between the true disconfirmation and the experienced disconfirmation. He argued that the size of the disconfirmation effect on satisfaction depends on whether the perceived performance falls within the zone of acceptance. If disconfirmation is small such that the perceived performance falls inside the zone of acceptance, customers will ignore or not recognize the disconfirmation between expectation and performance and consequently, the effect of disconfirmation will be practically non existent. This situation can be explained by the assimilation theory and has been advocated by various other authors (e.g. Cardozo [22], Olshavsky and Miller [100] and Anderson [7] among others).

On the other hand, if disconfirmation is large enough such that the perceived performance falls outside the zone of acceptance, disconfirmation will have a positively related effect on satisfaction: more specifically, a negative disconfirmation will lower the satisfaction level, while a positive disconfirmation will increase the satisfaction. This corresponds well with the contrast theory. However, even if disconfirmation is large enough to have a contrasting effect, expectation still has an influence on satisfaction because it sets the adaptation level. Therefore, in contrast with previous research which advocated an assimilation-contrast theory, the ED paradigm assumes that both expectation and disconfirmation have an additive and simultaneous effect on satisfaction.

The zone of acceptance has already been illustrated in Figure 2.1 from the perspective of a subtractive disconfirmation, i.e. the X-axis measures the difference between expectation and performance. We now reconsider this illustration from the perspective of subjective disconfirmation in Figure 2.4. It becomes clear that the effect of disconfirmation on satisfaction is assumed to be positive but also non-linear.

In addition to this type of non-linearity, Mittal, Ross and Baldasare [90] studied an asymmetric effect of disconfirmation on satisfaction. They found that negative

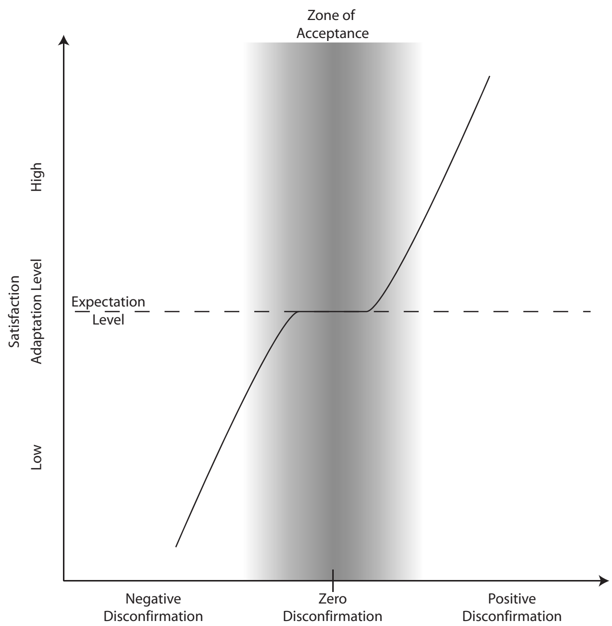


Figure 2.4: The effect of subjective disconfirmation on satisfaction.

disconfirmation had a stronger impact on satisfaction than positive disconfirmation. This was also found in an experiment by Anderson and Sullivan [6]. In general, these types of non-linearity are almost never considered in CS/D research. Most researchers focus on the correlation between disconfirmation and satisfaction, which implicitly assumes a linear relationship.

Performance

In the original ED paradigm, performance has an indirect effect on satisfaction. When compared against expectation it generates a level of disconfirmation which has a direct influence on satisfaction. In general, most authors never considered a direct effect of performance on satisfaction. Maybe this is caused by the fact that a direct effect of performance on satisfaction is of less interest to marketers because the level of performance is not under their control in contrast to customer's expectation and disconfirmation [26]. However, some authors did actually study a direct effect of performance on satisfaction additional to the direct effects of expectation and disconfirmation and found confirming results.

The first authors to study a direct effect of performance on satisfaction were Churchill and Surprenant [26]. Their original research question was whether it was necessary to include disconfirmation as an intervening variable affecting satisfaction or whether the effect of disconfirmation is adequately captured by expectation and perceived product performance. Their results were mixed. They performed their study on two product types, a durable good and a non-durable good. For the non-durable good, they found that disconfirmation not only had a significant impact on satisfaction, but it also had the largest impact on satisfaction. For the durable good, it showed that only performance had a significant impact on satisfaction, contradicting Oliver's ED paradigm. Although Churchill and Surprenant's results were not unanimously about the need for disconfirmation in a conceptualization of the CS/D process, it did provide the insight that performance could have a direct effect on satisfaction.

This remarkable result did not go unnoticed and was investigated in subsequent studies. Tse and Wilton [127], Oliver and Desarbo [99] and Anderson and Sullivan [6] also investigated the direct effect of perceived performance on satisfaction. They found significant positive direct relationships between performance and satisfaction. As Yi [145] pointed out, the accumulation of these results suggested a reexamination and extension of the original ED paradigm. This extension was formally introduced by Oliver in [97] and is illustrated in Figure 2.5. Besides the direct effect of performance, this model also integrates the difference between calculated (subtractive) and subjective performance. The model also indicates a possible relationship between expectation and performance, but does not specify the actual correlation between these two variables as they cannot be specified beyond the assumption that the relationship exists. According to Oliver [97], this expectation-performance relationship is idiosyncratic to the product or service being investigated.

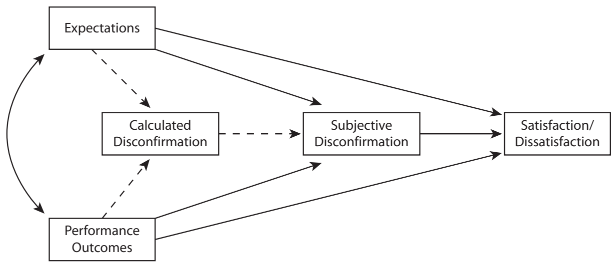


Figure 2.5: The complete expectancy disconfirmation paradigm with performance model. [Oliver [97], p.120]

Despite these findings supporting performance as a direct antecedent of CS/D, it should be noted that Spreng et al. [117] explicitly tested for a direct effect of performance on satisfaction but could not reject the null hypothesis that such direct effect was not present. They concluded that the direct effect of performance on satisfaction was completely mediated by the other model constructs. Spreng et al. also mentioned that "not all satisfaction researchers agree that an examination of the direct effect of performance is likely to be a fruitful theoretical approach [117]". Spreng et al. cite Oliver [98] who argued: "It says little, however, about the specific thought processes triggered by the product features. In particular, it fails to identify the mechanism by which performance is converted into a psychological reaction by the consumer."

As a conclusion, one could state that the opinions whether performance should be integrated as a direct effect in the CS/D process are rather mixed.

2.4 Summary

From the middle of the previous century on, CS/D received attention from researchers within the marketing and consumer psychology domain. From the early beginning two distinct antecedents of satisfaction were recognized, i.e. expectation and disconfirmation. One line of research focussed on expectation and explained the CS/D process by means of the assimilation theory, while another line of research suggested that disconfirmation caused (dis)satisfaction according to the rules of contrast theory.

In the beginning of the 1980's, Oliver reconciled both theories into the expectation disconfirmation paradigm. While previous research focussed on either expectation or disconfirmation, Oliver argued that both had a positive and additive effect on the customer's satisfaction experience. Firstly, expectation has a direct effect on satisfaction by setting the adaptation level. Secondly, the perceived performance is compared against expectation as a referent which causes disconfirmation. If disconfirmation is substantial, it will influence the final satisfaction outcome.

Over the years, the ED paradigm was tested and extended by various researchers. This caused expectation to lose its role as sole comparison referent in the CS/D process. Some authors found proof for customers using norms as referents, while some researchers have successfully operationalized the comparison referent by means of multiple referents. The direct effect of expectation on satisfaction, as assumed in the original ED paradigm, was also questioned. Although various researchers could not find the direct effect of expectation in their data, other empirical evidence showed that a direct effect exists in specific cases.

Disconfirmation and its role in the CS/D process has been less debated and almost all empirical evidence reveal disconfirmation as an important antecedent of satisfaction with a strong effect compared to the other antecedents. Various researchers argued that the disconfirmation process occurs at the attribute level [80, 97] and that the relationships between the antecedents and satisfaction are not necessarily linear and symmetric [90] as the amount of correlations in consumer satisfaction studies might suggest.

In general, the ED paradigm evolved from its original version in 1980 to an important, mature and valid CS/D model. Although other antecedents and models have been proposed, such as the ED model with performance or the equity model, the ED paradigm still remains an interesting framework to investigate the CS/D process.

3 A generic framework for modeling the ED paradigm with limited information

3.1 Introduction

The literature review in the previous chapter illustrated that over the past three decades, the ED Paradigm has become a theoretically and empirically strong model to analyze the CS/D process and to explain why and how customers become satisfied. If a company wants to analyze its customers according to the ED paradigm, it should measure constructs such as expectation, satisfaction and disconfirmation. However, many customer satisfaction surveys, carried out by marketing departments, only measure the product's performance and neither the disconfirmation experienced by the customers nor the customer's expectations.

As Oliver describes under the section titled 'Traditional Satisfaction Analysis' in [97], an all-too-familiar research scenario consists of asking consumers in a fairly direct manner to rate the importance and performance of various product attributes, together with an overall product evaluation. Next, marketers use these data to do descriptive analyses or to perform a linear regression analysis with the product attributes' performances as independent variables and the product satisfaction as dependent variable. The coefficients of such a regression are interpreted as the influence of a product attribute's performance on the overall satisfaction level. Consequently, in situations of such limited customer satisfaction data, the ED paradigm is ignored when analyzing the customer's satisfaction. This approach, which is e.g. used to derive the importance of an attribute in an Importance-Performance Analysis [83], completely ignores the expectancy disconfirmation paradigm and also lacks statistical reliability because it assumes a linear effect of performance on satisfaction, which is often not the case [33, 110].

Figure 3.1 illustrates the discrepancy between the theoretical model based on the 'Complete ED Paradigm with Performance' [97] and the frequently used 'Linear Regression' model when customer satisfaction data is limited to satisfaction and product performance. Obviously, the pragmatic approach does not correspond well with the existing theories.

Several reasons can be found why companies use this approach to learn about their customers' CS/D process. One could say that companies do not bother to look further than the intuitively obvious link between product performance and customer satisfaction. Furthermore, one could also argue that constructing a survey measuring performance and satisfaction is considerably easier, and consequently less expensive, than constructing a survey measuring satisfaction, performance, disconfirmation and

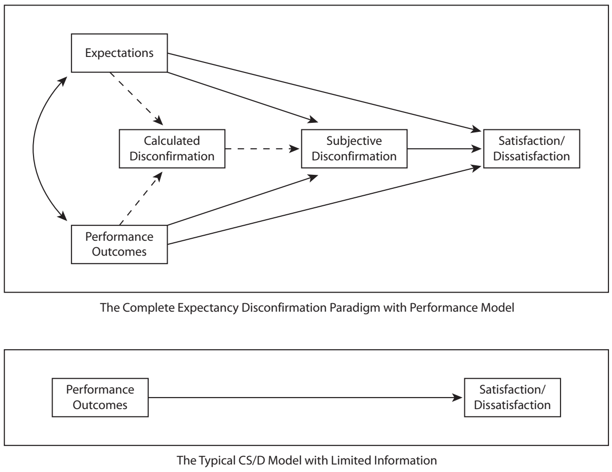


Figure 3.1: Comparison of the theoretical ED paradigm and the performance-based linear regression model.

expectation. The second approach implies measuring more constructs which makes the survey longer and possibly more tedious and expensive. Measuring perceived performance, expectation and disconfirmation at the same time can also be tricky since these three constructs are closely related. Furthermore, which type of referent is used by customers in their disconfirmation generating process is not always clear. Even if the construct which is used as a referent in the ED paradigm is known, it is not always possible to measure this prior to the customer's purchase. Many companies do not know their customers prior to the product purchase. Unfortunately, pre-purchase expectation can differ from post-purchase expectation since the latter is influenced by the consumption experience and therefore one should always try to measure pre-purchase expectation [97].

Despite the tendency of companies to use standard performance-satisfaction surveys instead of well constructed surveys which measure the various constructs of the ED paradigm, the author believes that the latter type of surveys are to be preferred. Measuring constructs such as expectation, disconfirmation, performance and satisfaction allows companies to model CS/D in correspondence with the theoretically valid ED paradigm. This will certainly increase the company's understanding of why their customers are satisfied or dissatisfied. However, the reality remains that many companies do not possess the right and complete data to fully model the ED paradigm, which introduces the goal of this part of the thesis:

How can companies with very limited information, i.e. only performance and satisfaction ratings, model their customer's satisfaction/dissatisfaction process in correspondence with the expectancy disconfirmation paradigm?

Although the limitation of the data can make it impossible to fully model the ED paradigm, it should not imply that the theoretical model has to be abandoned completely. Therefore, the goal of this section is to develop a framework which allows companies to model the CS/D process according to the principles of the ED paradigm and other established theoretical principles for customer satisfaction data which are limited to satisfaction measures and product performance measures.

Furthermore, the framework should be able to cope with performance data at a product attribute level, which is how it is typically measured. Oliver, who originally [96] examined the expectancy disconfirmation paradigm at the overall product level, already mentioned that the product evaluation ultimately takes place at the attribute level [97]. Also, Mittal et al. [90] supported this idea and stated that "studying satisfaction at the attribute level can help extend both conceptual and empirical understanding of the phenomenon".

Therefore, the framework to model the CS/D process with limited information, will need to aggregate various performance ratings, which only exist as external stimuli to the customer, into a single satisfaction rating. The aggregation itself represents the customer's interpretation of these stimuli which leads to (dis)satisfaction. Consequently, the framework will model the CS/D process by means of aggregation functions which are mathematical objects extensively studied over the past five decades within the lap of fuzzy set theory. Prior to the construction of a framework to model the CS/D process with limited information, the reader will be introduced to the domain of aggregation functions.

3.2 Aggregation functions

The origin of the aggregation functions research domain

The theory of aggregation functions (AF) as a separate research domain can be traced back to the origin of fuzzy set theory. Ever since, the mathematical study of AF has been systematically carried out under the joint umbrella of fuzzy set theory and non-classical decision theory [39]. Fuzzy set theory is a mathematical theory which originates from 1965 when it was introduced by Zadeh [146] and can be considered as an extension of the classical (crisp) set theory in mathematics.

Classical (crisp) sets are mathematical constructs which can be defined as "a collection of elements or objects $x_i \in X$ which can be finite, countable, or overcountable [147]". An object or element x_i either belongs to a set X or it does not. An object can not belong more or less to a set. If a crisp set does not contain too many elements, one can describe the set by enumerating its elements. For example, the crisp set A of 'natural numbers less than 4', can be written down as follows:

$$A = \{0, 1, 2, 3\}.$$

Fuzzy sets extend the notion of crisp sets by allowing elements to belong partially to a set. With which degree an element belongs to a fuzzy sets is called the membership degree and is expressed as a value within the interval $[0, 1]$. A fuzzy set is the extension of a crisp set in such a way that a crisp set can be regarded as a fuzzy set where each element belonging to the crisp set has a membership grade of 1 and all other elements have a membership grade of 0. Thus, the crisp set A of 'natural numbers less than 4' can be written down as a fuzzy set $\widetilde{A}$ as follows:

$$\widetilde{A} = \{\ldots, (-1; 0), (0; 1), (1; 1), (2; 1), (3; 1), (4; 0), (5; 0), \ldots\}.$$

The mathematical expression above represents a fuzzy set as a collection of tuples where the first part of each tuple is the element belonging to the fuzzy set and the second part of each tuple is the corresponding membership degree. Traditionally, elements of fuzzy sets with a membership grade equal to 0 are not explicitly written down. Of course, fuzzy set $\widetilde{A}$ is a rather artificial fuzzy set which does not illustrate the power of fuzzy sets. A better example would be the fuzzy set of 'natural numbers much smaller than 5'. Obviously, the definition of this set is rather vague because 'much smaller than' is not uniquely defined. A possible fuzzy set $\widetilde{B}$, representing this vague concept, could be:

$$\widetilde{B} = \{(0;1),(1;0.9),(2;0.5),(3;0.4),(4;0.1)\}.$$

Similar to classical crisp sets, fuzzy sets are accompanied with operations one can perform on the fuzzy sets. Zadeh [146] extended the well known intersection and union operations from classical set theory to fuzzy set theory. With fuzzy sets, the operations on sets are actually performed on the membership grades of the elements. For example, Zadeh defined the intersection of two sets as taking the minimum membership grade of each element and the union of two sets as taking the maximum membership grade of each element. The following example illustrates the union and intersection of two fuzzy sets according to Zadeh's original article [146].

$$\begin{aligned}
\widetilde{B} &= \{(0;1),(1;0.9),(2;0.5),(3;0.4),(4;0.1)\}\\
\widetilde{C} &= \{(0;0.7),(1;1),(2;0.7),(3;0.3)\}\\
\widetilde{B}\cup\widetilde{C} &= \{(0;1),(1;1),(2;0.7),(3;0.4),(4;0.1)\}\\
\widetilde{B}\cap\widetilde{C} &= \{(0;0.7),(1;0.9),(2;0.5),(3;0.3)\}
\end{aligned}$$

Using the minimum and maximum operator has the benefit that when applied to crisp sets, the same outcome is retrieved as when one applies the traditional intersection and union operator respectively. However, the minimum and maximum are not the only possible way to extend the crisp operators intersection and union to fuzzy set theory. Many other extensions have been proposed over the years and even new operators for fuzzy sets, which do not have an equivalent in crisp set theory, have been suggested (cf [44, 75, 147, 139]). All these operators have one thing in common: they combine the membership grades of an element in different fuzzy sets into a new membership grade. In other words, they aggregate the membership grades and are therefore called aggregation functions. The fact that the origin of aggregation functions is related to operators of fuzzy sets, combining membership grades, explains why aggregation functions are traditionally defined for the domain $[0,1]^n$ and maps to the codomain $[0,1]$. A proper definition of aggregation functions is the following [14] where $\mathbf{x} \leq \mathbf{y}$ implies that $x_i \leq y_i$ for all $i \in \{1,\ldots,n\}$:

Definition 1 (**Aggregation function**). An aggregation function is a function of $n > 1$ arguments that maps the (n-dimensional) unit cube onto the unit interval $f^n : [0,1]^n \to [0,1]$, with the properties

(i) $f^n(\underbrace{0,0,\ldots,0}_{\text{n-times}}) = 0$ and $f^n(\underbrace{1,1,\ldots,1}_{\text{n-times}}) = 1$,

(ii) $\mathbf{x} \leq \mathbf{y}$ implies $f^n(\mathbf{x}) \leq f^n(\mathbf{y})$ for all $\mathbf{x},\mathbf{y} \in [0,1]^n$.

Over the years, many families of aggregation functions have been developed and their mathematical properties have been studied. An introduction to the field of aggregation functions will be provided which focuses on the aggregation functions and the mathematical properties that are of interest to this thesis. A full review of aggregation function falls beyond the scope of this thesis and the interested reader is referred to several papers and specialized monographs [20, 77, 143, 14] for a thorough discussion. Note that in this thesis the superscript notation of an aggregation function refers to the arity of the aggregation function.

Important mathematical properties

The research on aggregation functions has always been heavily dominated by the study of mathematical properties. Although a thorough knowledge and understanding of the mathematical properties of aggregation functions is necessary for selecting the correct family of aggregation functions for a specific problem, a translation of these mathematical properties into understandable and meaningful interpretations of the aggregation behavior is equally important. As Beliakov mentions in [12], it is this clear and intuitive interpretation of aggregation functions which is sometimes lacking, in spite of the elaborated underlying mathematical theory. Therefore, the discussion of four mathematical properties, which are important for the further discussion in this thesis, will not solely focus on the mathematical definition, but will also try to describe the implications for the actual aggregation behavior.

The first mathematical property which is of importance for this work is associativity, which is defined as follows [20]:

Definition 2 (**Associativity**). An aggregation operator $f^{n+m} : [0,1]^{n+m} \to [0,1]$ is associative if

$$\forall n, m \in \mathbb{N}, \forall x_1, \ldots, x_n, y_1, \ldots, y_m \in [0,1] : \\ f^{n+m}(x_1, \ldots, x_n, y_1, \ldots, y_m) = f^2\left(f^n(x_1, \ldots, x_n), f^m(y_1, \ldots, y_m)\right).$$

Associativity can be a very powerful property because with this property it suffices to define the aggregation functions for two elements, i.e. the binary aggregation function. If more than two elements need to be aggregated, this can be easily done in an incremental way, i.e. each time a new element is aggregated with the preliminary aggregation result. Associativity is a very important property when modeling an aggregation process where input information arrives in steps and where the number of input elements can become astronomically large. On the other hand, associativity is a very strong and rather restrictive property [20] and recently some authors have argued that lack of associativity should not necessarily be considered a problem [108].

A second important mathematical property is symmetry or commutativity, which is defined as follows [14]:

Definition 3 (**Symmetry**). An aggregation function f^n is called symmetric, if its value does not depend on the permutation of the arguments, i.e.,

$$f^n(x_1, x_2, \ldots, x_n) = f^n(x_{P(1)}, x_{P(2)}, \ldots, x_{P(n)}),$$

for every $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and every permutation $P = (P(1), P(2), \ldots, P(n))$ of (1,2,...,n).

Symmetry can be interpreted as anonymity or equality of the input elements of the aggregation. This property is essential for explicitly modeling the irrelevance of the order of the input information in an aggregation process, such as the aggregation of anonymous votes into a final decision outcome. However, if the input elements have different semantical backgrounds such that reordering the input elements does not make sense, one should not strive for this property. As Rudas and Fodor [108] mentioned, this only results in an unnecessarily restrictive condition.

The third mathematical property, which is probably the most important in this thesis, is the neutral element, which is defined as follows [14]:

Definition 4 (**Neutral element**[1]). An aggregation function f^n has a neutral element $e \in [0, 1]$, if for every $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ with $x_i = e$, for some $1 \leq i \leq n$, and $n \geq 2$,

$$f^n(x_1, \ldots, x_{i-1}, e, x_{i+1}, \ldots, x_n) = f^{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

The neutral element of an aggregation function can be regarded as a specific value which has no impact on the aggregation outcome, i.e. adding more input elements with this value does not change the outcome of the aggregation. Originally, as defined above, a neutral value needs to represents a value which has no impact on the aggregation no matter which position it takes in the order of input elements. However, this is often a too restrictive condition in situations where each input element has a different semantical meaning and changing the order of the input elements does not make sense. Recently, the concept of neutral elements has been extended to neutral tuples which do take the position of input elements into account [13].

To properly define the concept of neutral tuples, first some notation needs to be introduced which comes directly from [13]. The vector with input elements is denoted as $\mathbf{x} \in [0, 1]^n$. If $\mathcal{I} = \{\mathcal{I}_1, \ldots, \mathcal{I}_m\} \subset \{1, \ldots, n\}$ is an index set with cardinality $|\mathcal{I}| = m > 0$, then $\mathbf{x}_\mathcal{I} = (x_{\mathcal{I}_1}, \ldots, x_{\mathcal{I}_m})$ will be used to denote the vector obtained from $\mathbf{x}$ by selecting the components whose indices are in $\mathcal{I}$ in the order $\mathcal{I}_1 < \ldots < \mathcal{I}_m$. In addition, if $\overline{\mathcal{I}} = \{\overline{\mathcal{I}}_1, \ldots, \overline{\mathcal{I}}_{n-m}\}$, with convention $\overline{\mathcal{I}}_1 < \ldots < \overline{\mathcal{I}}_{n-m}$, denotes the complement of $\mathcal{I}$ in $\{1, \ldots, n\}$, then $\mathbf{x}_{\overline{\mathcal{I}}}$ will denote the tuple $(x_{\overline{\mathcal{I}}_1}, \ldots, x_{\overline{\mathcal{I}}_{n-m}})$. For example, if $n = 5$, $\mathcal{I} = \{2, 4, 5\}$ then $\mathbf{x}_\mathcal{I} = (x_2, x_4, x_5)$ and $\overline{\mathcal{I}} = \{1, 3\}$ and $\mathbf{x}_{\overline{\mathcal{I}}} = (x_1, x_3)$. With this notation, it is possible to define a neutral tuple as follows [13]:

Definition 5 (**Neutral tuple**). Let f^n be an aggregation function and let $\mathcal{I} \subset \{1, \ldots, n\}$, $n > 1$, be an index set such that $0 < |\mathcal{I}| = m$. Then:

- A tuple $\epsilon \in [0, 1]^m$ is neutral for f^n with respect to $\mathcal{I}$ when

  $$f^n(\mathbf{x}) = f^{n-m}(\mathbf{x}_{\overline{\mathcal{I}}})$$

  holds for all $\mathbf{x} \in [0, 1]^n$ such that $\mathbf{x}_\mathcal{I} = \epsilon$

[1] In their work [14], Beliakov et al distinguish between a neutral element and a strong neutral element. We will not make this distinction and use the definition for a strong neutral element, which is the strongest version of both.

- The set made of all the tuples $\epsilon \in [0,1]^m$ which are neutral for f^n with respect to $\mathcal{I}$ will be denoted by $\varepsilon_m(f^n, \mathcal{I})$ and will be called the neutral set of f^n with respect to $\mathcal{I}$.

The definition of a neutral tuple allows a further refinement of the neutrality concept by focussing only on neutral tuples with a cardinality equal to 1. Such neutral tuples will be denoted as neutral singletons and are defined as follows:

Definition 6 (**Neutral singleton**)**.** Let f^n be an aggregation function and let $i \in \{1, \ldots, n\}$ be an index. Then:

- A singleton $\lambda^i \in [0,1]$ is neutral for f^n with respect to index i when

$$f^n(x_1, \ldots, x_{i-1}, \lambda^i, x_{i+1}, \ldots, x_n) = f^{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

  holds for all $\mathbf{x} \in [0,1]^n$

- The set made of all the singletons $\lambda^i \in [0,1]$ which are neutral for f^n with respect to index i will be denoted by $\Lambda^i(f^n)$ and will be called the neutral set of f^n with respect to index i.

If the neutral value represents the value which does not have an impact on the aggregation for any possible input position, then a neutral singleton λ^i represents a value which does not have an impact on the aggregation for input position i and $\Lambda^i(f^n)$ represents the set of values which do not have an impact on the aggregation for input position i. This allows to extend the principle of neutrality more easily to asymmetric functions.

Furthermore, the absorbing element or annihilator represents the opposite of the neutral value. The absorbing element is a value which completely determines the aggregation outcome regardless of the other input values. More specifically, if the aggregation contains one input element equal to the absorbing element, the outcome of the aggregation is equal to the absorbing element. One can interpret the absorbing element as an input value which can not be changed no matter which information is added to it. The proper definition of the absorbing element is as follows [14]:

Definition 7 (**Absorbing element**)**.** An aggregation function f^n has an absorbing element $a \in [0,1]$ if

$$f^n(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n) = a,$$

for every $\mathbf{x}$ such that $x_i = a$ with a in any position.

Classes of aggregation functions

In general, one can identify four general classes of aggregation function families, depending on how the outcome of the aggregation is positioned on the unit axis in relation to the position of the input elements. This is illustrated graphically in Figure 3.2 which shows the possible range of outcomes of the aggregation of two elements identified by the two black spots.

A disjunctive aggregation of two elements always provides an outcome which is at least as large as the largest input element. Therefore, a disjunctive aggregation function can be defined as [14]

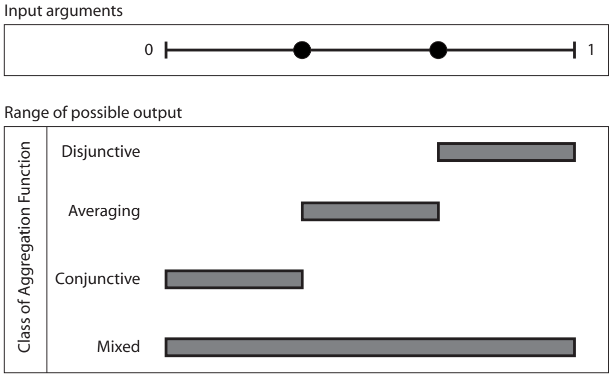


Figure 3.2: The four different classes of aggregation functions.

Definition 8 (**Disjunctive Aggregation Function**)**.** An aggregation function f^n has disjunctive behavior (or is disjunctive) if for every $\mathbf{x}$ it is bounded by

$$f^n(\mathbf{x}) \geq max(\mathbf{x}) = max(x_1, x_2, \ldots, x_n).$$

Disjunctive aggregation functions can be regarded as extensions of the crisp set union operator. Obviously, the maximum operator belongs to the class of disjunctive aggregation functions. Within the class of disjunctive aggregation functions, a large family of associative, symmetric aggregation functions with neutral element 0 exists which are called t-conorms. In general, disjunctive aggregation functions are used when upward reinforcement needs to be modeled, i.e. the input elements reinforce each other in a positive way such that the final outcome is greater than the largest input value.

On the other side of the spectrum, when negative reinforcement needs to be modeled, conjunctive aggregation functions are used. Negative reinforcement occurs when all the input elements are considered as “bad” and they have negative impact on each other causing the outcome to be less than the smallest input element. A conjunctive aggregation function can be defined as [14]

Definition 9 (**Conjunctive Aggregation Function**)**.** An aggregation function f^n has conjunctive behavior (or is conjunctive) if for every $\mathbf{x}$ it is bounded by

$$f^n(\mathbf{x}) \leq min(\mathbf{x}) = min(x_1, x_2, \ldots, x_n).$$

Conjunctive aggregation functions can be regarded as extensions of the crisp set intersection operator. Obviously, the minimum operator belongs to the class of conjunctive aggregation functions. Within the class of conjunctive aggregation functions, a large family of associative, symmetric aggregation functions with neutral element 1 exists which are called t-norms.

Next, there are some aggregation functions which result in a neither disjunctive, nor conjunctive outcome. These aggregation functions are called averaging aggregation functions and have no direct counterpart in classical crisp set theory. The definition of an averaging aggregation function is as follows [14]

Definition 10 (**Averaging Aggregation Function**). An aggregation function f is averaging if for every $\mathbf{x}$ it is bounded by

$$min(\mathbf{x}) \leq f^n(\mathbf{x}) \leq max(\mathbf{x}).$$

Averaging is possibly one of the most common ways to combine inputs. This type of aggregation function is mainly used to model compensatory behavior or to produce a value which can be considered representative for a set of input values. The most widely known averaging aggregation function and possibly most widely used one too is the arithmetic mean. Other well-known aggregation functions such as the weighted arithmetic mean, with weights $w_i \in [0,1]$ and $\sum w_i = 1$, or the geometric mean ($f^n(\mathbf{x}) = \sqrt[n]{x_1 x_2 \ldots x_n}$) also belong to the class of averaging aggregation functions. Besides these classical means, other types of averaging aggregation functions have been developed, such as the ordered weighted average (OWA) [137, 138], the Choquet integral [24, 55] and the Sugeno integral [120, 55].

Finally, the fourth class of aggregation functions, called mixed aggregation functions combines the aggregation behavior of all three other classes. This class of aggregation functions is defined as follows [14]:

Definition 11 (**Mixed Aggregation Function**). An aggregation function is mixed if it is neither conjunctive, nor disjunctive or averaging, i.e., it exhibits different types of behavior on different parts of the domain.

This class of aggregation functions is particularly useful when modeling aggregation of values on a bipolar scale. Bipolar scales have a neutral point such that all values below the neutral point can be considered negative, while all values above the neutral point can be considered positive. Commonly, negative values reinforce each other downwardly, positive values reinforce each other upwardly, and negative and positive values average each other. Later it will be shown that this type of aggregation behavior is useful when modeling the CS/D process at an attribute level. Therefore, focus is put on a particular family of aggregation behavior belonging to the class of mixed aggregation functions, i.e. uninorms [144, 38, 50, 141, 51]. Some other families of aggregation functions belonging to the class of mixed AF are nullnorms, T-S functions, symmetric sums and ST-OWA functions [14].

Uninorms, which are the result of the unification of the t-norm and the t-conorm, were formally introduced by Yager and Rybalov [144]. They defined a uninorm as follows:

Definition 12 (**Uninorm**). A uni-norm U is a function $U : [0,1] \times [0,1] \to [0,1]$ having the following properties:

(i) $U(x_1, x_2) = U(x_2, x_1)$ (Symmetry),

(ii) $x_1 \geq x_3$ and $x_2 \geq x_4 \Rightarrow U(x_1, x_2) \geq U(x_3, x_4)$ (Monotonicity),

(iii) $U(x_1, U(x_2, x_3)) = U(U(x_1, x_2), x_3)$ (Associativity),

(iv) There exists some element $e \in [0,1]$ called the neutral element such that for all $a \in [0,1], U(a, e) = a$.

The fourth property of this definition reflects the unifying nature of this aggregation function. T-norms, which are conjunctive aggregation functions, are, similar to uninorms, defined as symmetric, associative and monotone aggregation functions. However, contrary to uninorms, the neutral element of a t-norm is fixed to 1. On the other hand, t-conorms are symmetric, associative and monotone aggregation functions which have a neutral element equal to 0. Uninorms extend and unify both aggregation functions by allowing the neutral element to take any value within the [0,1] interval. At the same time, this results in a more complex aggregation behavior which is illustrated in Figure 3.3. This figure illustrates the important role of the neutral element e. Not only does it act as a null vote in the aggregation, it also defines the aggregation behavior. If both the input elements are smaller than e, the input elements will negatively reinforce each other during the aggregation. On the other hand, if both the input elements are larger than e, the input elements will positively reinforce each other. Finally, when the input elements are on either side of e, they will average each other out. In relation to the use of uninorms for values from a bipolar scale, this reveals that the neutral element can be somehow considered as the central element which defines positive and negative values. It should also be noted that the neutral element should not necessarily be located in the center of the scale as Figure 3.3 might imply.

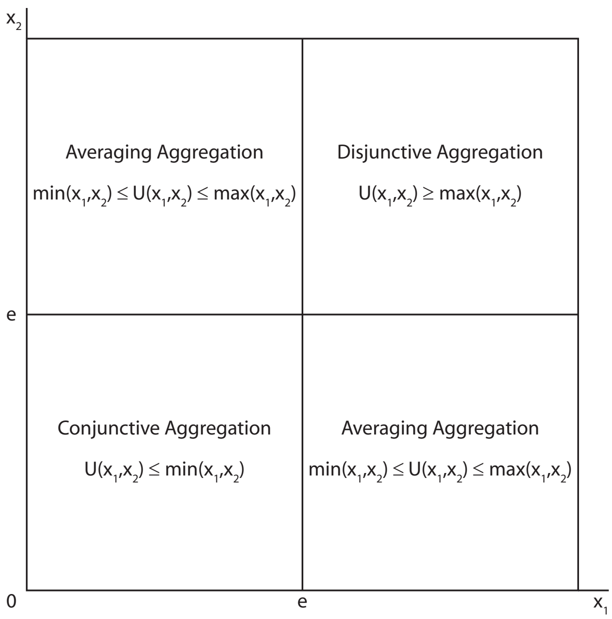


Figure 3.3: Aggregation behavior of a uninorm with neutral element e.

Among the family of uninorms, two types of uninorms can be discerned, i.e. conjunctive uninorms and disjunctive uninorms. This dichotomy is caused by the fact that for any uninorm $U(0,1) \in \{0,1\}$. If $U(0,1) = 0$, the uninorm is called conjunctive and has $a = 0$ as its absorbing element [14]. If $U(0,1) = 1$, the uninorm is called disjunctive and has $a = 1$ as its absorbing element [14]. Furthermore, uninorms are never continuous on the whole unit square $[0,1] \times [0,1]$ [14]. However, there are uninorms which are continuous everywhere except at the corners $(0,1)$ and $(1,0)$. These are called representable uninorms or generated uninorms. This particular class of uninorms was introduced by Dombi [38] long before Yager and Rybalov coined the term uninorm. A representable uninorm is built by means of univariate functions g, called generator functions or generators [14]:

Definition 13 (Representable uninorms). Let $g : [0,1] \to]-\infty, +\infty[$ be a strictly increasing bijection such that $g(e) = 0$ for some $e \in]0,1[$.

- The function given by

$$U(x_1, x_2) = \begin{cases} g^{-1}(g(x_1) + g(x_2)) & \text{if } (x_1, x_2) \in [0,1]^2 \backslash \{(0,1),(1,0)\}, \\ 0 & \text{otherwise} \end{cases}$$

  is a conjunctive uninorm with the neutral element e, known as a conjunctive representable uninorm.

- The function given by

$$U(x_1, x_2) = \begin{cases} g^{-1}(g(x_1) + g(x_2)) & \text{if } (x_1, x_2) \in [0,1]^2 \backslash \{(0,1),(1,0)\}, \\ 1 & \text{otherwise} \end{cases}$$

  is a disjunctive uninorm with the neutral element e, known as a disjunctive representable uninorm.

The function g is called an additive generator of the uninorm U. Note that since generator functions are unary, the superscript notation of the function's arity is not used.

A remarkable property of representable uninorms, which does not automatically hold for any uninorm, is that they are strictly increasing on $]0,1[^2$. This is caused by the fact that representable uninorms are built with strictly increasing generators.

Representable uninorms, which belong to the aggregation function family of uninorms can also be assigned to the family of generated functions. This family of aggregation functions, which encompasses aggregation functions of all classes, i.e. disjunctive, conjunctive, averaging and mixed, are all built by means of generator functions. Generated functions are defined as [14]:

Definition 14 (Generated function). Let $g_1, \ldots, g_n : [0,1] \to]-\infty, +\infty[$ be a family of continuous non-decreasing functions and let $h : \sum_{i=1}^{n} Ran(g_i) \to [0,1]$ be a continuous non-decreasing surjection. The function $f^n : [0,1]^n \to [0,1]$ given by

$$f^n(x_1, \ldots, x_n) = h\left(g_1(x_1) + \ldots + g_n(x_n)\right)$$

is called a generated function, and $\left(\{g_i\}_{i \in \{1,\ldots,n\}}, h\right)$ is called a generating system.

Generated functions are a very general family of aggregation functions. For example, the generating system

$$\left(\{g_i(x) = x\}_{\forall i \in \{1,\ldots,n\}}, h(x) = \frac{x}{n}\right)$$

results in the well known arithmetic mean, which is an averaging aggregation function. On the other hand, the generating system

$$\Big(\{g_1(x) = g_2(x) = \ldots = g_n(x) = g(x)\}, h(x) = g^{(-1)}(x)\Big)$$

with a strictly increasing bijection $g(x) : [0,1] \to]-\infty, +\infty[$ results in the mixed aggregation function known as a representable uninorm. A third and very interesting type of generated function are the continuous generated functions with a neutral element, which are generated by means of the generating system

$$\Big(\{g_1(x) = g_2(x) = \ldots = g_n(x) = g(x)\}, h(x) = g^{(-1)}(x)\Big)$$

with a continuous non-decreasing function $g : [0,1] \to]-\infty, +\infty[$ such that $g(e) = 0$, $g^{(-1)}$ as its pseudo-inverse, and $Ran(g) \subsetneq]-\infty, +\infty[$.

Pseudo-inverse functions are inverse functions which can also be defined for discontinuous functions. The exact definition is as follows:

Definition 15 (**Pseudo-inverse of a monotone function**). Let $g : [a,b] \to [c,d]$ be a monotone function, where $[a,b]$, $[c,d]$ are subintervals of the extended real line. Then the pseudo-inverse of g, $g^{(-1)} : [c,d] \to [a,b]$ is defined by

$$g^{(-1)}(t) = sup\{z \in [a,b] | \,(g(z) - t)\,(g(b) - g(a)) < 0\}.$$

3.3 A generic framework for the ED paradigm with limited information

Now that the basic concepts of aggregation functions have been introduced, the problems of modeling the CS/D process according to the ED paradigm with limited information can be reconsidered. As was mentioned in the introduction, companies which want to evaluate their customers' satisfaction with a specific product X, often apply the following scenario. Firstly, the company identifies different product attributes which are believed to have an influence on the overall satisfaction experience. Next, by means of a survey, the companies ask several customers to rate the performance of each product attribute and to give an overall satisfaction score. With the collected data, which contains rather limited information about the CS/D process, a linear regression of the performance ratings on the satisfaction score is performed. The regression coefficients provide an indication of the impact of a specific product attribute's performance on the overall satisfaction.

However, this approach has several problems which have been discussed before and the modeling mismatch between the linear regression approach and the ED paradigm is shown in Figure 3.1.

The best solution to these problems is to enrich the data with information about the customer's expectations and disconfirmation such that CS/D process can be modeled according to the ED paradigm. Most likely, this results in setting up a new and more comprehensive survey to gather the necessary data. However, for some companies this might not be a viable alternative because of cost constraints.

In such cases, a new framework is proposed as a second best alternative. This framework allows companies to use their performance data and satisfaction scores to model the CS/D framework according to the ED paradigm which consequently results in theoretical more sound inferences from the gathered data. The framework is called the LIED[2] framework, which stands for Limited Information Expectancy Disconfirmation framework.

Although, the data might seem too limited at first to model the ED paradigm, the LIED framework succeeds in doing so by considering the customer's expectation as an unknown, latent construct. This is illustrated in Figure 3.4. Considering expectation as a latent construct instead of measuring it directly has the benefit that the researcher does not need to decide which type of expectation to choose from to operationalize the customer's referent. The LIED framework simply assumes a latent construct which is used as comparative referent. Given this expectation, the (calculated) disconfirmation can be derived as the difference between the measured construct 'Performance' and the latent construct 'Expectations'. This way, four out of five constructs are recovered from the complete ED model. The only construct the LIED framework can not recover from the data is the subjective disconfirmation. Obviously, comparing Figure 3.4 with Figure 3.1 reveals that the LIED framework provides a much closer match with the theory than the linear regression approach.

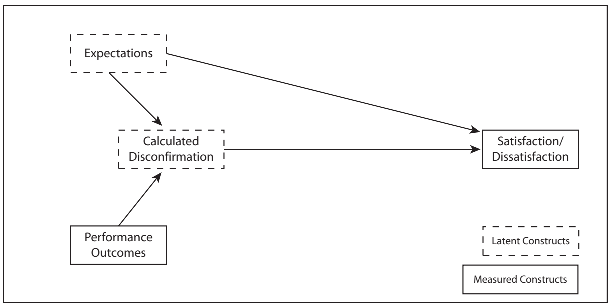


Figure 3.4: Modeling the ED paradigm with limited information.

Objective and subjective disconfirmation

Considering the ED paradigm in Figure 3.1 and focussing on the relationship between calculated disconfirmation and satisfaction, it can be seen that this effect is mediated through subjective disconfirmation. One could say that objective disconfirmation is first transformed into subjective disconfirmation according to a specific function f, and that subjective disconfirmation has an effect on satisfaction according to a specific function g. Following this notation, one can define the effect of objective disconfirmation on satisfaction as the function $h = g \circ f$. The fact that the LIED framework does not incorporate subjective disconfirmation as a separate construct, implies that the measured effect of objective disconfirmation on satisfaction must be equal to h and the resulting limitation of this framework is that it is impossible to disentangle the effects represented by f and g.

Oliver [97] argued that customers do not always evaluate performance at the objective level, especially when the product attribute is difficult to measure

[2]pronounce as "light"

objectively, such as for driving comfort. In such situations, our construct of disconfirmation is rather theoretical, but the function h still provides useful insights about the relationship between disconfirmation and satisfaction.

Finally, there is some critique in the existing literature against using calculated disconfirmation because it does not contain information about the valence of the disconfirmation: e.g. the fact that a car uses 1 liter of gas more than expected, gives no information whether customers consider this as a large or small disconfirmation. However, this critique is mainly directed at the measurement of disconfirmation and is not a real issue within the LIED framework. Another point of critique against objective disconfirmation is that for some product attributes it becomes nearly impossible to objectively calculate the disconfirmation because expectation can not be measured objectively with high precision. For example, while expected car mileage can be expressed on a high precision scale, it is much harder for customers to express their expected level of driving comfort with the same level of accuracy. However, since the LIED framework does not need to measure expectation, but derives it from the performance data, this critique is not a real problem either.

The LIED Framework

Based on the customer satisfaction theory discussed previously, expectation and disconfirmation are considered as the two direct antecedents of customer satisfaction. Note that although some research results suggested performance as a direct antecedent of satisfaction, the LIED framework only considers the indirect effect of performance, mediated through disconfirmation. This implies that the LIED framework remains theoretically close to the original formulation of the ED paradigm in [96], which is based on Helson's [61] adaptation theory. Consequently, customer satisfaction is represented as a function Φ_1 of the experienced disconfirmation $\mathbf{d} = \{d_1, \ldots, d_i, \ldots, d_n\}$ and the expectation level $\mathbf{e} = \{e_1, \ldots, e_i, \ldots, e_n\}$ for each attribute i.

$$\text{Satisfaction} = \Phi_1(\mathbf{d}, \mathbf{e}) \tag{3.1}$$

Because only performance data is available in the scenario of limited information, satisfaction is often modeled as a function Φ_2 of the perceived performance $\mathbf{p} = \{p_1, \ldots, p_i, \ldots, p_n\}$ for each attribute i.

$$\text{Satisfaction} = \Phi_2(\mathbf{p}) \tag{3.2}$$

Equations 3.1 and 3.2 illustrate how the performance based satisfaction modeling Φ_2 fails to correspond properly with the expectancy disconfirmation model Φ_1. To correct this mismatch, a specific fixed expectation level $\mathbf{e}$ is assumed which is integrated into model Φ_3 as a parameter. This allows the transformation of the performance levels into disconfirmation levels. In the following notation, expectation will always be explicitly mentioned in the function's signature, separated from the variables by a semicolon.

$$\begin{aligned}\text{Satisfaction} &= \Phi_3(p_1, \ldots, p_n; \mathbf{e}) \\ &= \Phi_3(p_1 - e_1, \ldots, p_n - e_n; \mathbf{e}) \\ &= \Phi_3(d_1, \ldots, d_n; \mathbf{e}) \\ &= \Phi_3(\mathbf{d}; \mathbf{e}) \qquad (3.3)\end{aligned}$$

This approach offers a better match with the expectancy disconfirmation paradigm (cf compare Eq. 3.1 and 3.3). The difference between the LIED modeling approach Φ_3 (Eq. 3.3) and the expectation disconfirmation model Φ_1 (Eq. 3.1), is that in the traditional ED modeling, the customer's expectation is measured directly by means of a survey and entered into the model as a separate variable. In the LIED approach, the expectation is unknown and has to be learned from the data. In other words, in the LIED model, expectation is integrated by means of parameters. However, if a single model is learned for all the customers, the expectation level derived from the model's parameters also holds for all the customers. Consequently, if customers are believed to have different expectation levels and the researcher wants to know the individual expectation levels, a separate model needs to be built for each customer.

The LIED modeling approach can be separated into three consecutive steps. First, the separate disconfirmation effect for each attribute is modeled. Next, all these effects are aggregated into an overall disconfirmation effect. Thirdly, satisfaction is modeled by means of this overall disconfirmation effect. Each step of the modeling framework will be separately discussed in detail below.

Modeling the attribute's disconfirmation effect on satisfaction

The first modeling step transforms the attribute's performance p_i into the attribute's disconfirmation effect δ_i on satisfaction, by means of a first-step function $\phi_i^1(p_i; e_i)$ which is defined as follows:

Definition 16 (First-step function). A first-step function $\phi_i^1(p_i; e_i)$ is a function with the following properties:

(i) $\forall x, y \in Dom(\phi_i^1) : x \leq y \Rightarrow \phi_i^1(x) \leq \phi_i^1(y)$,

(ii) $\exists [a, b] \subset Dom(\phi_i^1) : \forall p_i \in [a, b] \Leftrightarrow \phi_i^1(p_i; e_i) = 0$.

Note that the superscript notation of a step function refers to the current step in the LIED framework. The first step function is always unary. Furthermore, the first-step function has to be monotone to ensure that an increase in performance, which also implies an increase in disconfirmation given the fixed expectation level, does not result in a drop in satisfaction, ceteris paribus. The second part of the definition of a first-step function refers to the existence of a non-tempty interval of root elements, i.e. elements for which the function evaluates to zero. Note that because the first-step function is monotone by definition, no two disjunctive intervals of root elements exist.

Definition 17 (Root element). Element a is a root element of a unary function f, if and only if $f(a) = 0$.

The root elements of the first-step function fulfill a crucial role in the framework because they represent the performance levels which do not result in a disconfirmation effect. This is possible if the calculated disconfirmation, i.e. $p_i - e_i$, is zero or so small that it does not have an effect on satisfaction. In other words, this set of root elements contain performance levels which lead to disconfirmation levels to which customers are indifferent. Therefore, the following property can be formulated:

Property 1 (**Zone of indifference**). *The interval of root elements $[a, b]$ of the first-step function represents the zone of indifference.*

Obviously, one of the performance levels within the zone of indifference has to be the expected level of performance which results in a zero disconfirmation level. Therefore, if the interval of root elements is limited to one specific value, it must represent the customer's expectation. This is formulated in the following property:

Property 2 (**Expectation**). *If the interval of root elements of the first-step function is limited to one single element, then this element is equal to the customer's expectation e_i.*

The first-step function is defined as general as possible. All functions which are monotone and possess an interval of root elements are possible candidates for the first-step function in the LIED framework. However, to guide researchers in choosing a proper functional form for the first-step function, a few additional properties of the first-step function will be discussed which can help to increase the match between the framework and the existing CS/D theory.

Property 3 (**Asymmetric disconfirmation effect**). *Disconfirmation has an asymmetric effect on satisfaction iff $\exists d$ such that $(e_i + d), (e_i - d) \in Dom(\phi_i^1)$ and $\phi_i^1(e_i - d; e_i) \neq -\phi_i^1(e_i + d; e_i)$.*

Research by Mittal et al. [90] has provided evidence for an asymmetric effect of disconfirmation on satisfaction. Researchers who want to model such asymmetric disconfirmation effect can use property 3 to select an appropriate first-step function. This property is illustrated in Figure 3.5 which shows an inverse S-shaped function as the first-step function. This figure illustrates that attribute disconfirmation d_{i1} and d_{i2} have the same magnitude, but the effect on disconfirmation caused by d_{i1} is much larger than the effect on disconfirmation caused by d_{i2}.

Modeling the overall disconfirmation effect

The second modeling step in the LIED framework aggregates all the individual attribute disconfirmation effects δ_i into an overall disconfirmation effect δ_{total}, by means of the second-step function ϕ^2, who's arity is defined by the number of product attributes. Note again that the superscript notation does not refer to the function's arity but to the current step in the LIED framework. The second-step function ϕ^2 is defined as follows:

Definition 18 (**Second-step function**). A second-step function ϕ^2 is a function with $Dom(\phi^2) \supseteq Ran(\phi_1^1) \times \ldots \times Ran(\phi_i^1) \times \ldots \times Ran(\phi_n^1)$ with the following property

$$\forall \mathbf{x}, \mathbf{y} \in Dom(\phi^2) : \mathbf{x} \leq \mathbf{y} \Rightarrow \phi^2(\mathbf{x}) \leq \phi^2(\mathbf{y}).$$

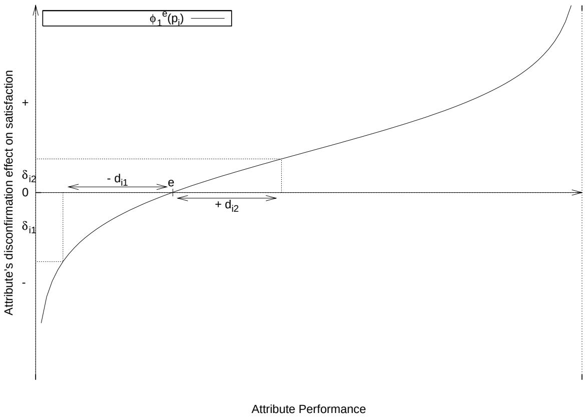


Figure 3.5: The asymmetric effect of disconfirmation on satisfaction.

The second step in the LIED framework actually models some kind of human aggregation behavior. It has been acknowledged that customers most likely experience disconfirmation at an product attribute or product dimension level, but to arrive at an overall satisfaction level, these disconfirmation effects need to be aggregated somehow. Therefore, the second step function is defined to be monotone. It would not make sense that the overall disconfirmation effect decreases when not a single attribute disconfirmation effect decreases.

The way people aggregate information has been extensively studied in domains such as information processing, psychology and attitude formation and decision making (cf [113, 43]). Within decision making, classic utility theory [132] offers an idealized mathematical model how people aggregate the valuations of an object's component characteristics into the utility of the object. To some extent, the first two steps of the LIED framework resemble the classic utility theory. Both models aggregate evaluations made at a more detailed level, i.e. product attribute disconfirmation or object component valuation, to arrive at a global level evaluation.

Within classic utility theory, mathematical tractability is obtained by assuming that the utilities of objects are compensatory [43]. Translated to the LIED framework, this would imply that a negatively valued disconfirmation is compensated by a positively valued disconfirmation. This results in the following property:

Property 4 (**Compensating behavior**). *A second step function has compensating behavior if for some non-empty part of its domain the following holds*

$$min(\mathbf{x}) \leq \phi^2(\mathbf{x}) \leq max(\mathbf{x}).$$

Although people aggregate information in a compensating way in real life [113], it has been criticized for being a too idealistic representation. Various researchers have suggested that some people will exhibit non-compensatory aggregation behavior in certain situations [113, 43, 54, 21]. Among the non-compensatory models, one can distinguish two types of aggregation behavior, i.e. conjunctive aggregation and disjunctive aggregation. This results in the following two properties:

Property 5 (**Conjunctive behavior**). *A second step function has conjunctive or downward reinforcement behavior if for some part of its domain the following holds*

$$\phi^2(\mathbf{x}) \leq min(\mathbf{x}).$$

Property 6 (**Disjunctive behavior**). *A second step function has disjunctive or upward reinforcement behavior if for some part of its domain the following holds*

$$max(\mathbf{x}) \leq \phi^2(\mathbf{x}).$$

Modeling customer satisfaction

The third modeling step transforms the overall disconfirmation effect on satisfaction into a final satisfaction response, by means of a third-step function ϕ^3 which is defined as

Definition 19 (**Third-step function**). A third-step function ϕ^3 is a function with $Dom(\phi^3) \supseteq Ran(\phi^2)$ with the following property

(i) $\forall x, y \in Dom(\phi^3) : x \leq y \Rightarrow \phi^3(x) \leq \phi^3(y)$,

(ii) $0 \in Dom(\phi^3)$.

Note that ϕ^3 is a unary function and the superscript refers to the third step in the LIED framework. The first part of the definition of a third-step function sets the monotonicity property of the third-step function. This ensures that satisfaction cannot decrease if the overall disconfirmation effect increases or remains the same, ceteris paribus. The second part defines that the element zero must be part of the third-step function's domain. This is crucial because the element zero has an important interpretation, it represents the situation where the customer does not experience any disconfirmation effect. Consequently, the satisfaction level when the overall disconfirmation effect is zero represents the initial satisfaction level or the adaptation level set by the customer's expectation (cf Oliver's ED Paradigm [96]). Therefore, the following property can be formulated:

Property 7 (**Adaptation Level**). *The value $\phi^3(0)$ corresponds to the initial satisfaction level or the adaptation level set by the customer's expectation level.*

It should be noted that the LIED framework does not mathematically impose a positive relationship between the customer's expectation $\mathbf{e}$, which was modeled as a parameter in the first-step function, and the adaptation level $\phi^3(0)$. Such mathematical restriction is not very practical because expectation is modeled as a parameter and not as a variable. Therefore, once a model is learned from the data, expectation is assumed to be fixed and equal for the entire population and the model cannot reveal how expectation influences satisfaction.

The only modeling situation when a direct positive relationship between $\mathbf{e}$ and $\phi^3(0)$ is of practical importance, is when multiple models are learned for various groups of customers. Such approach is necessary when the marketers expect customers to have widely varying expectation levels. Across these multiple models, the researcher might impose the restriction that the relationship between $\mathbf{e}$, as revealed by the first-step function, and the initial satisfaction level $\phi^3(0)$, as revealed by the third-step function, should be positive. Such restriction should however be considered with some circumspection. Firstly, defining a positive relationship between a vector $\mathbf{e}$ and a scalar $\phi^3(0)$ is not univocal. Secondly, although research results provided evidence for such a positive relationship between expectation and satisfaction, such a relationship was not always discovered by other researchers and the magnitude of the relationship is rather small compared to the effect of disconfirmation [122].

Finally, the effect of the overall disconfirmation on satisfaction is not necessarily linear. Mittal et al. [90] demonstrated that the overall satisfaction displays diminishing sensitivity to changes in the magnitude of performance for a given attribute. Therefore, it is not unlikely to assume that the satisfaction response also exhibits a diminishing sensitivity to the overall disconfirmation effect. This results in the following two properties:

Property 8 (**Positive diminishing returns**)**.** *The third-step function models diminishing returns for positive disconfirmation effects if the following holds:*

$$\exists a \in Dom(\phi^3) : x > a \Rightarrow \frac{d^2\phi^3(x)}{dx^2} < 0.$$

Property 9 (**Negative diminishing returns**)**.** *The third-step function models diminishing returns for negative disconfirmation effects if the following holds:*

$$\exists a \in Dom(\phi^3) : x < a \Rightarrow \frac{d^2\phi^3(x)}{dx^2} > 0.$$

The property of diminishing returns was separated in a positive and a negative part because it could be that customers only show diminishing returns for positive or negative disconfirmation.

The LIED framework continued

The framework models customer satisfaction in three consecutive steps, using attribute performance data p_i and assuming a specific expectation level $\mathbf{e}$. Each step consists of a specific type of function, which is defined in such a way that the framework is in accordance with the expectancy disconfirmation paradigm. Furthermore, for each step function, various properties were discussed which clarify the relationship between the mathematical function and the theories surrounding CS/D research. When the three step functions are combined, the entire framework can be written mathematically as follows.

$$\text{Satisfaction} = \Phi_3(p_1, \ldots, p_n; \mathbf{e}) = \phi^3\left(\phi^2\left(\phi_1^1(p_1; e_1), \ldots, \phi_n^1(p_n; e_n)\right)\right)$$

Researchers or marketers who want to use this framework to model customer satisfaction have to start by choosing a valid functional form for each step function. Each function can then be studied mathematically to verify which properties hold in order to fully understand the modeling characteristics. For example, by testing the sign of the second derivative of the third step function, it is possible to verify if diminishing returns are modeled or not (cf Properties 8 and 9).

The framework itself is very general and the restrictions imposed by definition on the functional form of the step functions are rather weak. This results in a myriad of possible functional forms for each step function and for the framework in general. In the next section, the relationship between the framework and a specific family of aggregation functions is discussed, which offers guidance to researchers and marketers in selecting appropriate functional forms for the three step functions.

3.4 Generated Functions

The LIED framework aggregates a set of attribute performance values into a single satisfaction value in three steps. Firstly, the input elements are transformed, then these transformations are aggregated and finally the aggregated value is transformed again. This aggregation procedure is almost exactly the same as the aggregation behavior of generated functions, which is a specific family of aggregation functions. Table 3.1 reveals the similarity between the framework and generated functions. It also shows that generated functions impose more restrictions on the function used for aggregation in the second step. Whereas the framework allows any monotone n-ary function, the family of generated functions exclusively uses the mathematical sum as aggregation function.

Table 3.1: Match between the modeling framework and generated functions

	Transformation	Aggregation	Transformation
LIED Framework	First-step function ϕ_i^1	Second-step function ϕ^2	Third-step function ϕ^3
Generated Function	Continuous non-decreasing function	$\sum$	Continuous non-decreasing surjection

Given the similarity between the framework and generated functions, the question arises whether all generated function are valid implementations of the framework. For this, all three steps need to be compared. The inner function $g_i(x_i)$ of a generated function $f^n(\mathbf{x}) = h(\sum_{i=1}^n g_i(x_i))$ is defined as a continuous monotone function mapping $[0,1] \to]-\infty,+\infty[$. In order to be a valid first-step function, $g_i(x_i)$ must be monotone and its image must contain the value 0. The condition of monotonicity holds by definition. Given the continuous nature of $g_i(x_i)$ it suffices that the image of $g_i(x_i)$ contains at least one positive value and one negative value, in order for 0 to be an image element of $g_i(x_i)$. Therefore, as long as the image of the inner function $g_i(x_i)$ contains the value 0, any valid inner function is a valid first-step function.

The framework defines the second-step function as any non-decreasing function with a domain which contains all possible vector elements produced in the first step of the framework. This definition gives the researcher or marketer a lot of freedom in choosing a valid second-step function. The definition of generated functions does not provide this freedom in the second stage of the aggregation but limits the implementation of the second stage to the mathematical sum. Considering Definition 18, the mathematical sum is a valid second step function since it is strict monotone increasing and its domain contains $\mathbb{R}$.

The third-step function is defined as a monotone function containing the value 0 in its domain. The outer function $h(x)$ of a generated function $f^n(\mathbf{x}) = h(\sum_{i=1}^{n} g_i(x_i))$ is monotone by definition and contains 0 in its domain if $0 \in \sum_{i=1}^{n} Ran(g_i)$. Therefore, any outer function $h(x)$ where $0 \in Dom(h)$ is a valid third-step function. From this follows:

Definition 20 (A valid generating system implementation). Any generated function built with the generating system $\left(\{g_i\}_{i\in\{1,\ldots,n\}}, h\right)$ is a valid framework implementation if the following holds:

(i) $\forall i \in \{1, \ldots, n\} : 0 \in Ran(g_i)$,

(ii) $0 \in Dom(h)$.

The large family of generated functions is now identified as valid implementations of the LIED framework under two specific conditions. Having a more restrictive nature than the framework, using generated functions to implement the framework also holds some implications. Firstly, generated functions limit the domain and the codomain to the $[0, 1]$ interval. This implies that satisfaction scores and attribute performance scores need to be measured in the $[0, 1]$ interval. Within the domain of CS/D, the $[0, 1]$ scale is not always used. Sometimes, researchers prefer the use of a bipolar scale such as $[-1, 1]$.

As mentioned by Beliakov et al. [14], the aggregation functions on two different closed intervals are isomorphic, i.e. any aggregation function on the scale $[a, b]$ can be obtained by a simple linear transformation from an aggregation function on $[0, 1]$. Thus the choice of scale is a question of interpretability, not of aggregation type. In the situation of modeling CS/D with the LIED framework, this implies that researchers must linearly transform their scale to the [0,1] scale and thee can transform the result back to the original scale for interpretation purposes.

Secondly, generated functions as implementations of the framework limit the second-step function to the mathematical sum. This results in a second step function with compensating, conjunctive and disjunctive behavior (cf Properties 4, 5 and 6). More specifically, the LIED framework implemented with generated functions has the following attractive property:

Property 10 (Mathematical sum as a compensatory, conjunctive and disjunctive second-step function). *If the mathematical sum is used as the second-step function, i.e. $\phi^2 = \sum$, the following aggregation behavior occurs:*

(i) A positive and a negative disconfirmation effect will compensate each other,

(ii) Two positive disconfirmation effects will positively reinforce each other,

(iii) Two negative disconfirmation effects will negatively reinforce each other.

In this thesis, the implementations of the new CS/D modeling framework will be limited to generated functions. As shown above, this provides a valid implementation given the two conditions mentioned in Definition 20. Albeit a bit more restrictive in its choice of the step functions than the original framework, generating functions still provide a lot of modeling flexibility. At the same time, it offers a solid mathematical structure based on extensive research in the field of aggregation functions which can help to identify the right implementation of the framework. Some of the mathematical properties of generated functions and their implications for the LIED framework are discussed next.

Continuity

A first important mathematical property of generated functions is continuity. This is a nice property to have because it makes the estimation of the model parameter easier, since many optimization techniques are based on the (partial) derivative of a function.

Generated functions such that $Dom(h) \neq]-\infty, +\infty[$ are always continuous on the whole unit cube. Generated functions for which $Dom(h) =]-\infty, +\infty[$ are also continuous except for the case where there exist $j, k \in \{1, \ldots, n\}$, $j \neq k$, such that $g_j(0) = -\infty$ and $g_k(0) = +\infty$.

From the modeling framework perspective, discontinuity arises when one of the first-step functions transforms an attribute performance p_i into $-\infty$ while another first-step function transforms the attribute performance p_j into $+\infty$. Given the fact that the mathematical sum is used as the second-step function, the overall disconfirmation effect is modeled as a sum with one term equal to $-\infty$ and another term equal to $+\infty$, which is mathematically undefined. In such cases the modeler has to decide whether $(-\infty) + (+\infty)$ results in $-\infty$, which results in a conjunctive generated function, or in $+\infty$, which results in a disjunctive generated function. To recognize any discontinuity in the modeling framework, it suffices to verify the image values of the various first-step functions for the attribute performance values 0 and 1. If 0 is transformed into $-\infty$ for some first-step functions and 1 is transformed into $+\infty$ for some other first-step functions, discontinuity is present in the modeling framework and needs to be handled appropriately.

In general, the easiest way to achieve continuity in the whole domain $[0, 1]$ is by using first-step functions for which $\phi_i^1(0) \neq -\infty$ and $\phi_i^1(1) \neq +\infty$.

Symmetry

A generated function is symmetric if and only if it can be generated by a generating system such that $g_1 = g_2 = \ldots = g_n = g$, where $g : [0, 1] \rightarrow]-\infty, +\infty[$ is any continuous non-decreasing function [78]. Using a symmetric generated function as the implementation of the framework implies that the attribute performances $(0.2, 0.4, 0.3)$ should result in the same satisfaction as the attribute performances $(0.4, 0.3, 0.2)$. In the CS/D modeling framework, this only makes sense if the three attributes are equally evaluated into a disconfirmation effect by a customer. If customers do not consider attributes equally important, it is possible that the customers are less sensitive to disconfirmation for the less important attributes, resulting in a smaller disconfirmation effect on satisfaction. Also, customers might hold different expectations for different attributes. Consequently, equal levels of performance will lead to different levels of disconfirmation and different disconfirmation effects. In both situations, the modeler will prefer an asymmetric generated function.

In general, if modelers use the same first-step function for each attribute, they should be aware that they are using a symmetric generated function and that they are making the assumption that the customers have the same expectation level for each attribute and that the disconfirmation process is the same for each attribute.

Neutrality

Some generated functions possess a neutral element, whereas others do not. Komornikova [78] provided the following definition of the class of generated functions possessing a neutral element:

Definition 21 (**Generated functions with neutral element**). A generated aggregation function f^n has a neutral element $e \in [0,1]$ if and only if for each $n > 1$, fn can be expressed as

$$f.(x_1, \ldots, x_n) = g^{(-1)}\left(g(x_1) + \ldots + g(x_n)\right)$$

where $g : [0,1] \rightarrow]-\infty, +\infty[$ is a continuous non-decreasing function such that $g(e) = 0$ with the pseudo-inverse $g^{(-1)}$.

From Definition 21 follows that the neutral element of a generated function corresponds to the root of the first-step function. According to Properties 1 and 2, this corresponds to the customer's expectation, if there is a single root element, or the customer's zone of indifference, if there is a set of root elements. Therefore, the neutral element of the generated function plays a very important role in the modeling framework, it reveals the referent(s) against which customers compare the perceived attribute performance(s).

Definition 21 also reveals that generated functions with neutral elements are always symmetric since they are constructed with the same first-step function for each attribute. However, mostly researchers or marketers want asymmetric generated functions to model CS/D unless they assume that the expectation for each attribute is equal.

Although asymmetric generated functions do not have neutral elements, they do have neutral singletons. The root element of a specific first-step function ϕ_i^1 which models the transformation from attribute performance p_i to disconfirmation effect δ_i for attribute i results by definition in a disconfirmation effect $\delta_i = 0$. Consequently, this attribute's performance has no influence on the outcome of the second step function and therefore no influence either on the final outcome of the model. By definition, the set of root elements of first-step function ϕ_i^1 is the neutral set $\Lambda^i(f,n)$ of the model f with n attributes with respect to attribute i.

Definition 22 (**Neutral singletons of a generated function**). Let f^n be a generated aggregation function with the generating system $\left(\{g_i\}_{i\in\{1,\ldots,n\}}, h\right)$. Then for each $i \in \{1, \ldots, n\}$, each root element x_0 of g_i is a neutral singleton λ^i for f^n with respect to i.

In general, when the generated function is symmetric, the neutral element represents the customer's expectation, which is equal for each product attribute. When the generated function is asymmetric, the neutral singletons for index i represent the customer's expectation for attribute i.

3.5 Implementing the LIED framework: an illustrative example

In this chapter, a mathematical framework has been developed to model customer satisfaction in conformity with the ED Paradigm. As mentioned before, this framework imposes only a few restrictions on the functional form of its equations and there are still many different ways to implement the framework. To provide guidance for implementing the framework, the previous section investigated the close relationship between the LIED framework and the well-known family of generated functions. This provides a strong mathematical foundation for selecting appropriate functions, but still leaves the researcher with a myriad of options. In the end, researchers still have to make decisions about the functional form of the different step functions. While making such decisions, it has to be taken into account that the selected functional form of the model equations have implications on both the model's theoretical assumptions and the model's estimability. This section illustrates how to select appropriate step functions considering the theoretical assumptions one wants to model. Assume the following situation:

"A company sells a very simple, low technological product which has to perform a simple single task. It is assumed that only two product attributes have an influence on the customers' satisfaction, i.e. the price and the performance of the product. There are many other competitors offering a similar product and it is assumed that customers are not really sensitive to the product's price as long as it is situated within an acceptable range. However, real high prices will lower satisfaction and real low prices have a positive effect on satisfaction. Since the product only performs a simple task and many alternatives exist, customers are likely to be more sensitive to bad performance than to good performance."

To provide a proper implementation of the LIED framework for the above situation, researchers must decide whether or not to exploit the relationship between the LIED framework and generated functions. If they decide to implement the LIED framework with a generated function, they must use continuous, monotone functions $\phi_i^1 : [0,1] \to]-\infty, +\infty[$ with $0 \in Ran(\phi_i)$ as first step functions, the mathematical sum as the second step function and a continuous monotone function $\phi^3 :]-\infty, +\infty[\to [0,1]$ with $0 \in Dom(\phi^3)$ as the third step function. Additionally, researchers who do not choose an implementation by means of a generated function need to define a proper second step function.

By choosing a generated function as the implementation of the LIED framework, three interesting properties of the model can be easily investigated, i.e. continuity, symmetry and neutral element. Continuity of the model which can be important for estimating the model can be easily achieved for a generated function by ensuring that $\phi_i^1(0) \neq -\infty$ and $\phi_i^1(1) \neq +\infty$. Symmetry of the model implies that both product attributes are evaluated against the same referent in the same manner, which is not the case in the above example. Therefore, symmetry is not required and ϕ_1^1 does not have to be the same as ϕ_2^1. If the generated function is no longer symmetric, it will not possess a neutral element. Instead, researchers must focus on the neutral singletons of the function as proxies for the customers expectation levels.

This first global analysis of the situation results in some basic guidelines for the LIED implementation:

- The first step functions $\phi_i^1 : [0,1] \to]-\infty,+\infty[$ must be continuous monotone with $0 \in Ran(\phi_i)$ and may differ for each product attribute
- $\phi_i^1(0) \neq -\infty$ and $\phi_i^1(1) \neq +\infty$ for both first step functions
- The second step function is implemented as a mathematical sum of two terms
- The third step function $\phi^3 :]-\infty,+\infty[\to [0,1]$ must be continuous monotone with $0 \in Dom(\phi^3)$

Irrespective of further modeling decisions, any implementation following the above will be continuous and the neutral singletons will represent the customers' expectations for each product attribute.

Next, both first and third step functions need to be analyzed and defined in detail. According to the above example, the first step function for the product performance attribute needs to be asymmetric, i.e. customers will be less sensitive to positive disconfirmation than to negative disconfirmation. A suitable functional form for this first step function is

$$\phi_1^1(x_1) = \begin{cases} a(x_1 - e_1) & x_1 \in [0, e_1] \\ b(x_1 - e_1) & x_1 \in]e_1, 1], \end{cases} \tag{3.4}$$

with $e_1 \in [0,1]$ and $a > b > 0$. Figure 3.6a is a plot of this first step functions with $a = 3$, $b = 1$ and $e_1 = 0.4$ and illustrates that this function is monotone continuous with $\phi_1^1(0) = -e_1 \neq -\infty$ and $\phi_1^1(1) = 1 - e_1 \neq +\infty$. Furthermore Property 3 on page 42 can be used to verify that ϕ_1^1 is asymmetric since $b < a$.

As for the first step function of the product price attribute, the information in the description of the example does not mention anything about asymmetry, but indicates the existence of a zone of indifference. According to Property 1 on page 42 this implies that the first step function needs to have an entire interval of root elements. An appropriate first step function for the product price attribute is:

$$\phi_2^1(x_2) = \begin{cases} c(x_2 - e_{2a}) & x_2 \in [0, e_{2a}] \\ 0 & x_2 \in]e_{2a}, e_{2b}[\\ c(x_2 - e_{2b}) & x_2 \in [e_{2b}, 1], \end{cases} \tag{3.5}$$

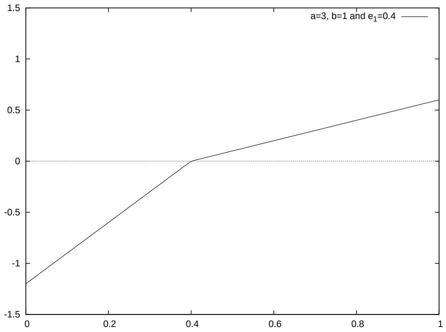
with $e_{2a} < e_{2b}$ and $c > 0$. Figure 3.6b which shows a plot of this first step function with parameter values $c = 3$, $e_{2a} = 0.4$ and $e_{2b} = 0.7$ clearly shows the zone of indifference $[e_{2a}, e_{2b}]$.

Finally, an appropriate third step function needs to be selected. The above example gives no specific information on how customers transform disconfirmation into satisfaction. Therefore, a linear transformation of disconfirmation into satisfaction was selected such that $0 \in Dom(\phi^3)$:

$$\phi^3(x) = Max(Min(1, dx + f), 0) \tag{3.6}$$

All three step functions are now clearly defined and the implementation of the full model can be formulated as follows:

(a) First step function $\phi_1^1(x_1)$ for product performance attribute (Eq. 3.4)



(b) First step function $\phi_2^1(x_2)$ for product price attribute (Eq. 3.5)

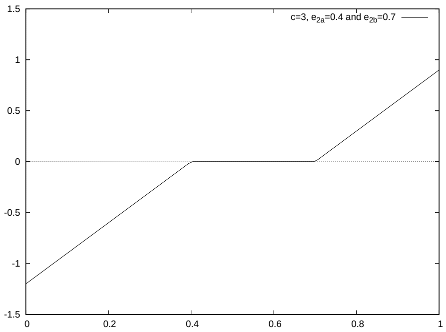


Figure 3.6: First step functions

$$\Phi_3(x_1,x_2) = \begin{cases} \phi^3(a(x_1-e_1)+c(x_2-e_{2a})) & x_1 \in [0,e_1], x_2 \in [0,e_{2a}] \\ \phi^3(a(x_1-e_1)) & x_1 \in [0,e_1], x_2 \in]e_{2a},e_{2b}[\\ \phi^3(a(x_1-e_1)+c(x_2-e_{2b})) & x_1 \in [0,e_1], x_2 \in [1,e_{2b}] \\ \phi^3(b(x_1-e_1)+c(x_2-e_{2a})) & x_1 \in]e_1,1], x_2 \in [0,e_{2a}] \\ \phi^3(b(x_1-e_1)) & x_1 \in]e_1,1], x_2 \in]e_{2a},e_{2b}[\\ \phi^3(b(x_1-e_1)+c(x_2-e_{2b})) & x_1 \in]e_1,1], x_2 \in [1,e_{2b}] \end{cases} \tag{3.7}$$

This elaborated example showed how the LIED framework and its connection with generated functions can be used to translate a specific consumer satisfaction setting into an appropriate model. However, the final model obviously does not excel in parsimonity and can become troublesome to estimate. The development of an appropriate estimation procedure for this specific models falls outside the scope of this thesis. In one of the following chapters, an other LIED implementation based on generated functions is introduced together with an appropriate estimation procedure.

3.6 Summary

In this chapter, the LIED framework has been introduced, which allows companies to model their customer's (dis)satisfaction process in a theoretically sound manner with only a limited amount of information needed. Many companies only gather performance data and satisfaction data from their customers, which makes it seemingly impossible to study the CS/D process according to the proven ED paradigm because no information about the customer's expectation is available. The LIED framework overcomes this problem by modeling expectation as a latent construct which must be learned from the data.

The LIED framework segments the CS/D process in three consecutive steps which are all modeled by means of a different step function. Each step of the framework has been studied from a mathematical point of view which allows a better understanding between the mathematical properties and the existing customer satisfaction theories.

To aid researchers and marketers to apply the framework to their existing data, the LIED framework was positioned within the theory of aggregation functions. It was shown that a large family of aggregation functions, i.e. generated functions, are valid implementations of the LIED framework given two additional imposed properties. By focussing on the family of generated functions, it becomes possible to study the implementation of the LIED framework in a structured manner. At the end of this chapter, we already gave some interesting interpretations of properties which are common for any implementation of the LIED framework by means of a generated function.

With the LIED framework, marketers who only posses performance and satisfaction data are offered a blueprint to model customer (dis)satisfaction in a theoretically more sound way than by means of a linear regression.

4 Data sets

4.1 Introduction

To test the value of a LIED framework implementation, it will need to be applied on some real-life data sets which reflect the situation for which the framework has been specifically designed as close as possible. To this end, three data sets have been selected which only contain satisfaction data, performance data and some socio-demographic information about the respondents. None of them really focussed on customer expectations or measured disconfirmation at a detailed level. These data sets represent typical situations where information is too limited to model the ED paradigm directly but for which the LIED framework can offer a solution.

All three data sets relate to a product or service from three different sectors, i.e. the energy market, the family amusement sector and the banking sector. Due to confidentiality restrictions, the original data had to be made anonymous and the data fields had to be rephrased. The extent to which the data can be discussed in detail is limited but this does not prevent a discussion of the general data structure.

Each data set has a hierarchical structure which is depicted in Figure 4.1. At the top level, customer satisfaction and satisfaction consequents, such as loyalty and recommendation, are measured. Next, several product or service dimensions were identified and customers were asked how well the product performs on each dimension. These performance scores are considered as dimension performances. Finally at the lowest level, various product attributes were distinguished for each dimension and each respondent gave a performance score for these product attributes. These scores will be denoted as the attribute performances. The LIED framework will typically use the dimension performances to model customer satisfaction instead of the attribute performances. Dimension performances are conceptually closer to customer satisfaction than attribute performances and the number of attribute performances easily exceeds 20 or more which can make proper estimation of all the model parameters very difficult. However, the attribute performances are very useful because it allows to verify the hierarchical structure. Also, it allows dimension performance to be treated as a proxy for the customer's satisfaction with the product dimension and to model dimension satisfaction with the attribute performance by means of the LIED framework.

4.2 Data cleaning

The implementation of the LIED framework to model the CS/D process should be considered as a single step of the entire KDD process. Prior to the actual discovery

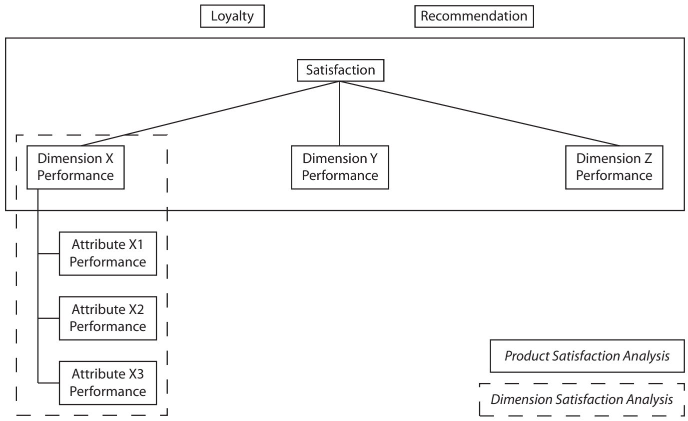


Figure 4.1: The hierarchical structure of the data sets.

of new knowledge within the data, several data preprocessing steps should be taken to prevent the discovery of meaningless and invalid patterns.

It is important to realize that from a methodological and academic point of view, the data owned by companies about their customers is seldom of high quality. Both the observational and secondary nature of the data can be identified as reasons for this phenomenon. With observational data, the researcher has only little control about the variance among the explanatory variables, as opposed to experimental data. Secondary data on the other hand implies that the data was not collected with the analysis in mind, as opposed to primary data. As a consequence, the data might violate several data assumptions since they were not taken into account during the data collection stage. In such situations, the necessary corrective actions must be taken. Luckily, many data mining analyses do not impose many assumptions on the data, in contrast to traditional statistical analysis. Also, due to the secondary nature of the data, not all necessary variables might have been measured, which is one of the reasons of existence of the LIED framework. This framework is particularly useful if customer satisfaction data lacks information about customer's expectations. Variables which are irrelevant for our research should be removed from the data set.

The data also has to be scanned for impossible or incorrect data entries. If the correct value for an erroneous data entry is obvious to retrieve, the correct value is imputed, otherwise, the data entry is reported as missing. Next, the problem of missing values in the data set should be tackled by means of an appropriate missing value analysis technique.

After the missing data are imputed and a rectangular data set is acquired, a factor analysis can be performed to reduce the number of variables. As mentioned above, the data sets contain a very large amount of attribute performances which can make the estimation of the LIED framework problematic given its non-linear nature. For this reason, it is more advisable to use product dimension performances to model the CS/D process, which are less in number and make the LIED framework more estimable. These dimension performances are conceptually closer to customer satisfaction and can be considered as a summary score of all attribute performances belonging to that dimension. However, dimension performances are measured by a single question which can be sensitive to measurement errors and there is no guarantee that the dimensions surveyed are distinct nor that the attributes are assigned to the correct dimension.

Both measurement errors and dimensionality of the data set can be reduced by means of factor analysis whose "fundamental intent is to determine the number and nature of latent variables or factors that account for the variation and covariation among a set of observed measures, commonly referred to as indicators" [18]. Within the data sets in this thesis, a factor analysis can look for structure among the attribute performances (indicators) such that highly correlated attribute performances are considered to measure the same dimension performance (construct). Firstly, this allows the construction of new dimension scores which are the average of the matching attribute performances, hereby reducing the risk of measurement errors. Secondly, factor analysis will also reveal a structure of product dimensions and corresponding product attributes which in the ideal case matches the original hierarchical structure in the data set.

All these data cleaning steps, the missing value analysis and the factor analysis were performed for each data set. Next, we will discuss the missing value analysis and factor analysis more in detail.

Missing value analysis

The most intuitive cure for missing data is the deletion of all cases with missing values. However, this rather common approach is far from ideal. Obviously, deleting cases could reduce the data set to an impractical sample size. But even worse, deleting cases with missing values could bias any conclusions about the population drawn from the collected sample [59, 76]. To understand why deleting cases with missing values should be avoided, the classification of missing values developed by Rubin [107] should first be introduced. Three types of missing data can occur, i.e. data which are not missing at random (NMAR), data which are missing at random (MAR) and data which are missing completely at random (MCAR).

The difference between these three types of missing data can be easily explained by means of an academic example. Assume a small data set with information about the respondents gender and annual income. Some respondents refused to answer the income question which leads to missing values. If there is no particular reason why people refused to answer the income question, the data is said to be completely missing at random (MCAR). In other words, there is no specific group of people which is more reluctant to answer the income question. From a statistical point of view, missing data is MCAR if the probability of missing values is independent from any other variable, both observed and unobserved. Only when missing data is MCAR, deletion of cases with missing entries does not result in biased conclusions. However, deletion of cases with MCAR data still reduces the sample size which might be problematic for further analysis.

When missing data is MAR, a specific group of people which are more reluctant to answer the income question does exist. However, the available data allows the researcher to detect this group of people. In the aforementioned example, the income data would be MAR if for example male respondents are much less likely to answer the income question than female respondents. Obviously, the probability of a missing income entry is dependent on an observed variable, i.e. gender, which is the statistical interpretation of MAR. If missing data is MAR, deleting records with missing entries not only reduces the sample size, but also results in biased conclusions. In the example, deleting records with missing values would mainly result in deletion of male respondents since their probability to answer the income question is lower. Consequently, if this reduced data set is used to estimate the gender ratio of the population, the amount of female individuals would clearly be overestimated.

Finally, missing data can also be NMAR which means that some specific group of respondents is more likely to generate missing values but the observed variables are not sufficient to determine this group. In the above example, missing income data would be NMAR if people with a university degree are more reluctant to answer the income question than people without a university degree. Assuming that people with a university degree have a higher annual income than people without a university degree, deletion of records with missing income data would provide an average population income which will be underestimated since mainly high income levels are missing. Also in case of NMAR, deletion of cases with missing values leads to biased conclusions.

Therefore, it is better to use specific imputation methods to tackle missing values than simply deleting cases with missing values. More specifically, multiple imputation methods and model-based imputation methods are preferable. These methods have the advantage of not lowering the sample size and being unbiased as long as the data is MAR or MCAR [59, 76]. The multiple imputation methods produce multiple imputed versions of the original data set. Each version contains different values for the missing entries, which represents the uncertainty of the imputed values. Since multiple imputation does not produce one single but multiple data sets, subsequent statistical analysis need to be performed for each imputed data set. The different results of each imputed data set needs to be combined into a single statistical result. This approach can be very daunting and unattractive for marketers and practitioners.

Therefore, another imputation method which is unbiased under MAR and MCAR is preferred, i.e. model-based imputation. This approach involves a maximum likelihood estimation of the underlying process which generates the missing values. To build this model, it uses all available data and every observed variable. In this dissertation, the expectation-maximization (EM) approach available in SPSS 16 was used. For more information about the implementation of the EM model-based imputation approach, the reader is referred to [69]. Model-based imputation has the advantage over multiple imputation that it produces a single imputed data set which is more attractive to work with, but has the disadvantage that it reduces the variance in the data set [111]. However, as pointed out by Scheffer [111], single imputation methods are acceptable if the level of missing data is less than 10%. Hair et al. even suggest that variables with levels of missing data around 20% to 30% can be remedied with EM imputation [59].

In this dissertation, the following approach to tackle missing data is suggested: Firstly, the level of missing data per variable is calculated. According to a rule of thumb [59], variables with more than 20% missing data are considered for deletion. Finally, it is verified if the missing data are MAR or MCAR by means of Little's MCAR test. A insignificant result of this test indicates that data are MCAR. If missing data are MAR and not MCAR, the researcher should keep in mind that the single imputation method might reduce the variation of the data when drawing conclusions from the sample. Irrespective of the type of missing data, the missing values are imputed with SPSS EM model-based imputation technique.

Exploratory factor analysis

After the missing value analysis, a factor analysis is performed. Within the field of factor analysis, a distinction must be made between confirmatory factor analysis and exploratory factor analysis. As the name suggests, the latter type of factor analysis has an exploratory nature and is rather information driven than hypothesis driven. With exploratory factor analysis, the researcher makes no a priori assumptions about the number of factors or the pattern of relationships between the common factors and the indicators. The main purpose of exploratory factor analysis is to determine the number of latent factors within the data and to reveal which variables are good indicators for the different factors. [18, 59]

Confirmatory factor analysis is often preceded by exploratory factor analysis and has a much more hypothesis-driven nature. Its main purpose is to test the validity of a hypothesized construct or the reliability of a measurement scale. In contrast to exploratory factor analysis, the research must have a firm a priori sense, based on past evidence and theory, of the number of factors that exist in the data, of which indicators are related to which factors, and so forth when performing a confirmatory factor analysis [18].

All three data sets have a hierarchical structure which already suggests a number of factors (product dimensions) and their corresponding indicators (product attributes). However, there is no theoretical evidence or underlying theory supporting the specific structure in each data set which makes the factor analysis rather exploratory and descriptive than confirmatory. Therefore, an exploratory factor analysis was conducted.

The methodology used to factor analyze the data corresponds to the steps outlined by Hair et al. [59]. The variables used as indicators in the factor analysis are the set of all product attribute performances. Firstly, the sample size is evaluated. As indicated by Hair et al. [59], the sample size should contain at least 100 cases and there should be at least 10 respondents per variable. Ideally, this respondent-variable ratio should be at least $20:1$.

Next, some measures of intercorrelation are calculated to verify if enough correlation exists among the indicators to justify a factor analysis. Firstly, the partial correlations are studied, which measure the correlation between two variables controlling for the effects of all the other variables. High partial correlations indicate a low intercorrelation among the variables, i.e. only little of the correlation between two variables is explained by the other variables. As a rule of thumb, partial correlations would be considered high when they exceed 0.7. Besides verifying for the absence of high partial correlations, the Bartlett test of sphericity is used. This test provides the statistical significance that the correlation matrix has significant correlations among at least some of the variables. Finally, the appropriateness of factor analysis is also verified by means of the measure of sampling adequacy (MSA) which ranges from 0 to 1. The measure can be interpreted as follows: 0.80 or above, meritorious; 0.70 or above, middling; 0.6 or above, mediocre; 0.50 or above, miserable; and below 0.50, unacceptable [59]. Both the overall MSA and the MSA per variable are examined.

Once the sample size is decided to be sufficiently large and enough intercorrelation exists among the indicators, a component analysis is performed. This is an exploratory factor analysis which considers the total variance among the indicators to derive latent factors. Three different criteria are used to determine the optimal number of factors. Firstly, the number of product dimensions in the original data set gives an indication of the number of factors. Secondly, the latent root criterion, which is the most commonly used technique to determine the number of factors, is used. With this criterion, only factors with a latent root or eigenvalue greater than 1 are considered. A factor with an eigenvalue less than 1 accounts for less variance in the data set than a single variable and should therefore be discarded. Thirdly, the scree test criterion is used which is a graphical method to determine the number of factors. A graph is created with the eigenvalue plotted for each factor in order of decreasing eigenvalues. The scree test criterion defines the first point at which the curve begins to straighten out as the number of factors to extract. It should be noted that all three criteria can reach a different conclusion on the number of factors and that each of them only give an indication. Ultimately, selecting the number of factors is the researcher's choice which should be based on the results of these three criteria, but also on the interpretation of the final factor model.

To interpret the factor model, the factor loadings, which are the correlations between each indicator and the factor, are studied. The initial factor solution typically results in a general factor with high loadings on almost every variable which accounts for the largest amount of variance in the data and several subsequent factors which are based on the residual amount of variance. This initial factor solution is often difficult to interpret and therefore a factor rotation is performed. A factor rotation actually rotates the axes of the factors which results in the same factor solution but with easier interpretations of the factors. Two types of factor rotations exist, orthogonal and oblique rotations. An orthogonal rotation respects the orthogonality of the reference axes of the factors and results in a factor solution where the factors are independent of each other. An oblique rotation lets go of this restriction and can result in factors which are correlated. Both the orthogonal rotation method Varimax and the oblique rotation method Oblimin, as implemented in SPSS 16, will be applied. Finally, the rotated solution which has the most meaningful interpretation is selected.

To interpret the rotated factor solution, the researcher must focus on the factor loadings which are practically and statistically significant. Table 4.1 is taken from [59] and gives guidelines about the level of factor loading necessary to be statistically significant given a specific sample size. Factor loadings greater than ± 0.50 are considered practical significant [59]. Variables which have no significant loadings for any factor are bad indicators and are considered for deletion. The same holds for indicators which have crossloadings, i.e. factor loadings which are significant on multiple factors. Prior to deletion of insignificant or crossloaded indicators, factor models with different rotations or different number of factors should be tested to see if these variables become significant or if the crossloadings disappear.

Finally, once a factor model is chosen and interpreted, each factor represents a product dimension. As mentioned before, in an ideal situation, the dimensions from the factor analysis should match the dimensions in the original structure. For each factor a new variable is added to the data set which represents a dimension performance. The scores on these new dimension performances are calculated by taking the average of all the attribute performance scores which are indicators of the specific factor. The dimension observed during the survey shall be denoted as the hierarchical product dimensions and the dimensions derived from the factor analysis as the factorized dimensions.

Each type of product dimensions serves a different goal. The hierarchical product dimensions are determined when the survey is designed and are based on expert-knowledge. The attributes related to a dimension are not a way of measuring the

Table 4.1: Guidelines for identifying significant factor loadings based on sample size [Hair et al. [59], p.128]

Factor Loading	Sample Size Needed for Significance[1]
.30	350
.35	250
.40	200
.45	150
.50	120
.55	100
.60	85
.65	70
.70	60
.75	50

[1] Significance is based on a .05 significance level (α), a power level of 80 percent, and standard errors assumed to be twice those of conventional correlation coefficients.

Source: Computations made with SOLO *Power Analysis*, BMDP Statistical Software, Inc., 1993.

same thing as the product dimension neither are they necessarily a good indicator of the product dimension score. The product attributes are assigned to a product dimension because they intuitively belong to that dimension and because they might explain that specific dimension performance. The focus of the hierarchical product dimensions and corresponding product attributes is on the hierarchical nature of the product. If one considers the dimension performance as a proxy for dimension satisfaction which can be modeled by means of the product attribute performances, the hierarchical dimensions are the preferable constructs to use.

On the other hand, the factorized product dimensions are constructed as multi-item measures of product dimension performances. The advantage of multi-item scales is that they allow measurement errors to cancel out against each other which increases the scale's reliability [106]. The factorized dimensions are not constructed during the survey design, but are derived from the observed product attributes. Not every observed product attribute necessarily relates to a factorized dimension. Those attributes which do belong to a factorized dimension can be interpreted as a way of (partially) measuring the product dimension performance. Factorized dimensions should sufficiently differ conceptually against each other and each factorized product dimension should be truly unidimensional. The goal of the factorized dimension performances is to provide valid and reliable measures which can be used to explain and predict customer satisfaction.

Therefore, once the factorized product dimensions are defined, some measures are calculated to asses the reliability. Reliability can be defined as the degree to which measures are free from error and yield consistent results [106] and can be measured by means of Cronbach's coefficient alpha [25, 106, 94]. As suggested by Churchill [25], coefficient alpha is used in conjunction with factor analysis. For each factor found by the factor analysis, coefficient alpha is calculated to measure the reliability of factorized dimension performance. As Nunnaly [94] mentioned, if α is very low, the test is either too short or the items have very little in common. Therefore, it is recommended to strive for α values greater than 0.7 which represent a modest level of reliability. However, if $\alpha < 0.8$ and there are more than 3 items, each item is considered for deletion to test if deletion might increase α.

4.3 Energy data set

The first data set concerns a satisfaction survey among customers of two companies active in the Belgian energy market. Table 4.2 shows the various variables of the data set which were retained for subsequent analysis. Besides a field indicating to which company a customer belongs, five other variables, i.e. ecology, choice commitment, gender, age and marital state, provide information about the type of customer. At the top level of the hierarchical structure, two other constructs were measured besides overall customer satisfaction, i.e. recommendation and intention to switch. The data set also contains information about the customers' overall level of disconfirmation and whether they received value for their money. In total the performance of 8 product domains were measured. Table 4.3 shows how many product attributes were measured for each product dimension. Note that for dimensions 7 and 8 no product attributes were measured.

Table A.1 in Appendix A.1 gives some univariate statistics for all the variables of the energy data set. The data set contains 1053 cases and except for the variable 'Value for money', all variables have less than 20% missing values. Since 'Value for money' has such a high amount of missing values and because this variable is of less importance for future analysis, it was decided to delete this variable. Little's MCAR test was significant at a 0.01 level and thus the missing data is MAR. This implies that deleting cases with missing values would probably introduce a bias in conclusions drawn from the cleaned data. Therefore, the SPSS EM model-based imputation technique is used. Once the missing data is imputed, the minima and maxima of each variable are verified. The imputation method might impute values which do not fall in the range of the original scales. Values lower than the minimum of the scale are recoded to the minimum and values higher than the maximum of the scale are recoded to the maximum of the scale. Table A.1 in Appendix A.1 also compares some univariate statistics, i.e. mean and standard deviations, between the original data set and the imputed data set. These results show that the imputation did not have a big impact on the overall univariate structure of the data.

Once the imputed data set is created, a factor analysis can be performed. The imputed data set contains 1053 cases which is far above the minimum sample size of 100 cases. With the 23 product attribute performances which will be used in the factor analysis, the respondent-variable ratio is approximately 45:1 which is well above the recommended value of 20 : 1. Therefore, the sample size is more than sufficient to perform the factor analysis. Besides a sufficiently large sample, factor analysis also requires a large amount of intercorrelations between the indicator variables. Table A.2 in Appendix A.1 shows the partial correlations between the variables. The diagonal

Table 4.2: Variables of energy data set

Variable	Values
Company	[1 – 2]
Overall Satisfaction	[1 – 10] [Unsatisfied – Satisfied]
Overall Disconfirmation	[1 – 5] [Very negative – Very positive]
Value for money	[1 – 5] [Very Low – Very High]
Recommendation	[1 – 5] [Certainly not – Certainly yes]
Intention to switch	[1 – 5] [Certainly not – Certainly yes]
Product Dimension 1: Availability	[1 – 5] [Very bad – Very good]
Product Dimension 2: Employees	[1 – 5] [Very bad – Very good]
Product Dimension 3: Problem Solving	[1 – 5] [Very bad – Very good]
Product Dimension 4: Information	[1 – 5] [Very bad – Very good]
Product Dimension 5: Service	[1 – 5] [Very bad – Very good]
Product Dimension 6: Invoices	[1 – 5] [Very bad – Very good]
Product Dimension 7: Price	[1 – 5] [Very bad – Very good]
Product Dimension 8: Product Quality	[1 – 5] [Very bad – Very good]
Ecology concerned customer	[1 – 5] [Absolutely not – Certainly yes]
Customer commitment	[1 –5] [Absolutely not – Very much]
Gender of customer	[Male – Female]
Age of customer	
Marital state of customer	[1 – 3]

of this table contains the MSA values per variable. The highest partial correlation is 0.439 between product attributes D4d and D4e which both belong to the 'Information' dimension, which is still considered small (< 0.7). The lowest MSA value per variable is 0.921 for the product attribute D6b of the 'Invoice' product dimension. This value is significantly larger than the threshold 0.8 which is already a meritorious value for a factor analysis. Also the Overall MSA measure with a value of 0.963 and the Bartlett's Test which is significant at 0.01 (cf Table 4.4) indicate a very high level of intercorrelations which makes this set of attribute performance a perfect candidate for factor analysis.

Next, the number of factors need to be determined. According to the hierarchical structure, 6 dimensions should be present within the attribute performance data, i.e. *availability*, *employees*, *problem solving*, *information*, *service* and *invoices*. Note that product dimensions *price* and *quality* do not have corresponding product attributes defined and are therefore not considered during the factor analysis. Table A.3 in Appendix A.1 gives the eigenvalue of each possible factor. According to the latent root criterion which states that any factor with an eigenvalue larger than 1 should be kept, 4 factors should be extracted from the data. The fifth factor still has an

Table 4.3: Product dimensions and product attributes of energy data set

Product Dimensions	Number of product attributes
D1: Availability	3 [D1a – D1c]
D2: Employees	4 [D2a – D2d]
D3: Problem Solving	5 [D3a – D3e]
D4: Information	5 [D4a – D4e]
D5: Service	3 [D5a – D5c]
D6: Invoices	3 [D6a – D6c]
D7: Price	0
D8: Product Quality	0

Table 4.4: Overall MSA and Bartlett's test of sphericity for energy data set

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.963
Bartlett's Test of Sphericity	Approx. Chi-Square	18297.820
	df	253.000
	Sig.	.000

eigenvalue of 0.911 which is close to 1 such that a five factor model could also be considered. Finally, the scree plot which is shown in Figure 4.2 shows two kinks in the slope of the graph, the first at two factors and the second at six factors. However, as can be read from Table A.3 in Appendix A.1, the two factor model only explains around 59% of all the variance in the data which is considered rather low. Therefore, only the six factor model is considered.

In summary, the four, five and six factor model are investigated. The four and five factor model were suggested by the latent root criteria, while the six factor model is supported by the scree plot and the hierarchical structure of the data. For each factor model, both the orthogonal Varimax and the oblique Oblimin rotation were performed. The tables with the factor loadings can be found in Appendix A.1. Factor loadings larger than 0.5, which are both practically and statistically significant are marked grey.

Both the Varimax and the Oblimin rotated version of the four factor model give more or less the same picture. They both show that dimensions Availability (D1), Information (D4) and Invoice (D6) are recognized as separate dimensions within the data set, while the other three dimensions are all captured by one factor. The third question (D4c) of the Information dimension appears to be a bad indicator for the Information dimension. The attributes of the hierarchical dimension Employees appear to be very closely related with the attributes of the hierarchical dimensions Availability (D1) and Problem Solving (D3).

The five factor models show a very similar picture of the structure, but suggest that at least two attributes of the Employees dimension should be considered as

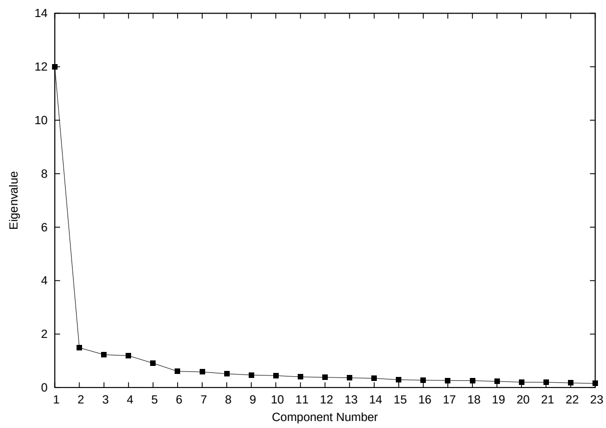


Figure 4.2: Screeplot energy data set.

a separate dimensions, i.e. D2b and D2c. The Varimax rotation illustrates that two other attributes D2a and D2d are more or less crossloaded on both the first and the third factor, although the loadings on the third factor are not practically significant (< 0.5). The only two hierarchical dimensions which are still not separated by the factor model are the Problem Solving and the Service Dimensions. Adding another factor to the model does not separate these two dimensions. The six factor model isolates a single Availability attribute from the other Availability attributes and suggests it as a separate factor.

Because a factor with only a single significant indicator is not desirable, the five factor model is preferred over the six factor model. Furthermore, because the five factor model explains more variance than the four factor model (73.18% versus 69.22%) and the five factor model has a structure which corresponds better with the original hierarchical data structure, the five factor model is also preferred over the four factor model. Furthermore, the Varimax rotated version of the five factor model was chosen because it only contains one variable without significant loadings compared to two insignificant variables in the Oblimin rotated factor model. The variable D4c will not be used in the calculation of the summated scores because it is not practically significant and crossloaded over two factors.

Based on the results of this five factor model, five new product dimensions are created. The score of a factorized dimension is calculated by averaging the scores of the attribute variables which loaded significantly on that dimension. Table 4.5 provides an overview of each factorized dimension and the variables used to calculate the summated score. Each factorized dimension was given a name based on the product attributes belonging to that specific dimension. Since all Cronbach's alphas were larger than 0.70, the summated scores can be considered as reliable scores for the factorized dimensions.

4.4 Family entertainment data set

The second data set comes from a large scale customer satisfaction survey which was performed across seven companies from the family entertainment sector. The original

Table 4.5: Factorized dimensions of energy data set

Factorized Product Dimension	Product Attributes	Cronbach's Alpha
Availability	D1a, D1b, D1c	0.78
Employees	D2b, D2c	0.74
Service	D2a, D2d, D3a, D3b, D3c, D3d, D3e, D5a, D5b, D5c	0.96
Information	D4a, D4b, D4d, D4e	0.87
Invoices	D6a, D6b, D6c	0.75

data set contains 70 fields and 4031 records. Each record represents a customer buying a package of products from one specific companies. Per customer, a unique identifier (ID) is retained along with information about the specific company from which the customer bought the product. The survey was not taken at a single moment in time but in six different waves, which allows companies to analyze evolutions in customer satisfaction. In addition the customers' value for money perception, their intentions to buy again and recommend the company to other people, performance on 9 different dimensions were also measured. Four of these dimensions represent the four possible products in a product package, i.e. *performance of products A*, *B*, *C* and *D*. The other dimensions are *the overall experience with the company*, *the company's accommodations*, *the company's personnel*, *the prices* and *company's communications*. For each dimension, a set of corresponding attribute performances were measured in addition to the dimension performances. Tables 4.6 and 4.7 provide an overview of all the variables present in the family entertainment data set and the existing hierarchical structure.

It stands out from analyzing the missing values in the data set that dimensions *product B*, *C* and *D* contained a lot of missing values, both on the dimension score as the corresponding attribute performance scores. A possible explanation is that not all customers bought a package containing all four products. Customers who reported missing values for the performance of a specific product because they did not experience that product are clearly different from customers who did experience the product but refused or forgot to answer. The satisfaction outcome of the first type of customer will not be influenced by the performance of the missing product dimension in contrast to the latter type of customer. Therefore, it was decided to remove all customers with missing values for an entire product dimension (A, B, C or D), i.e. where both dimension performance and the corresponding attribute performances are missing. These customers most likely did not buy the full package with all four products. This resulted in a reduction of the sample size from 4031 customer to 2122 customers, which is still a substantially large data set.

Table A.10 in Appendix A.2 shows that all variables have missing values levels below 20% except for variables D4g, D8c and D9a. Given their high levels of missing entries, it was decided to remove these variables from the data set. All three variables measure attribute performance from different product dimensions. By removing them, it is unlikely that the data set will lose crucial information since several other

Table 4.6: Variables of family entertainment data set

Variable	Values
ID	
Company	[1 – 7]
Wave	[1 – 6]
Overall Satisfaction	[1 – 10] [Low – High]
Value For Money	[1 – 10] [Low – High]
Recommendation	[1 – 10] [Very Unlikely – Very Likely]
Intentions to Repatronage	[1 – 10] [Very Unlikely – Very Likely]
Product Dimension 1: Overall Experience	[1 – 10] [Low – High]
Product Dimension 2: Product A Performance	[1 – 10] [Low – High]
Product Dimension 3: Product B Performance	[1 – 10] [Low – High]
Product Dimension 4: Accommodation Performance	[1 – 10] [Low – High]
Product Dimension 5: Product C Performance	[1 – 10] [Low – High]
Product Dimension 6: Product D Performance	[1 – 10] [Low – High]
Product Dimension 7: Personnel Performance	[1 – 10] [Low – High]
Product Dimension 8: Prices	[1 – 10] [Very Bad – Very Good]
Product Dimension 9: Communication	[1 – 10] [Very Bad – Very Good]

attribute measurements are available for each dimension. Little's MCAR test was significant at a 0.01 level implying that the data is not missing completely at random (MCAR). After the missing data were imputed by means of the SPSS EM model-based imputation technique, the minima and maxima of each variable were verified to see if they fell within the limits of the original scale. Values less than the minimum of the scale were recoded into the scale's minimum and values greater than the scale's maximum were recoded into the scale's maximum. The last two columns in Table A.10 in Appendix A.2 show the mean and standard deviations of the different variables for the imputed data. Inspection reveals that they are very similar to the means and standard deviations of the original data. The imputation did not have a significant effect on the univariate data structure.

After the missing value analysis, a data set with 67 variables and 2122 variables was retained. The imputed data set contains 51 product attributes across 9 product dimensions, which results in a 41 : 1 cases to variable ratio. Both the overall sample size as the cases to variable ratio are sufficient for a factor analysis. Next, both Bartlett's Test of Sphericity was performed and the MSA was estimated. The results are presented in Table 4.8 which indicate an extremely adequate level of intercorrelation between the variables for a factor analysis, i.e. the overall MSA is larger than 0.8 and Bartlett's Test is highly significant. In Table A.11 one can find the partial correlations between any pair of variables and the MSA for each variable

Table 4.7: Product dimensions and product attributes of family entertainment data set

Product Dimensions	Number of product attributes
D1: Overall Experience	9 [D1a – D1i]
D2: Product A Performance	7 [D2a – D2g]
D3: Product B Performance	4 [D3a – D3d]
D4: Accommodation Performance	8 [D4a – D4h]
D5: Product C Performance	6 [D5a – D5f]
D6: Product D Performance	3 [D6a – D6c]
D7: Personnel Performance	5 [D7a – D7e]
D8: Prices	7 [D8a – D8g]
D9: Communication	5 [D9a – D9e]

(cf the diagonal of Table A.11). The partial correlations are smaller than 0.45 and the variables MSA's are larger than 0.94 which supports the idea that there is enough intercorrelation in the data for a factor analysis to make sense.

Table 4.8: Overall MSA and Bartlett's test of sphericity for family entertainment data set

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.966
Bartlett's Test of Sphericity	Approx. Chi-Square	64000.688
	df	1275
	Sig.	.000

The original hierarchical structure of the data with nine product dimensions, i.e. *overall experience*, *product A*, *product B*, *accommodation*, *product C*, *product D*, *personnel*, *prices* and *communications*, suggests a nine component factor analysis. A factor analysis with nine dimensions is also supported by the latent root criterion. Table A.12 shows that the first nine components have an initial eigenvalue larger than 1. It also reveals that a nine factor model only captures around 64% of the variance within the data, which is still substantial. The third test to determine the number of components, i.e. the graphical scree plot test, is shown in Figure 4.3. The graph shows that the line of the eigenvalues starts to flatten out at ten components. Based on all three tests, both nine and ten factor models were considered. For each factor model, a Varimax rotated and an Oblimin rotated version were estimated.

In Tables A.13, A.14, A.15 and A.16, the factor loadings for each factor model can be found. All factor loadings which are considered practically and statistically significant, i.e. factor loadings greater than 0.5, are marked with a grey background. None of the models contain crossloadings. The two nine factor models correspond almost perfectly with the hierarchical structure of the data. The Varimax rotated

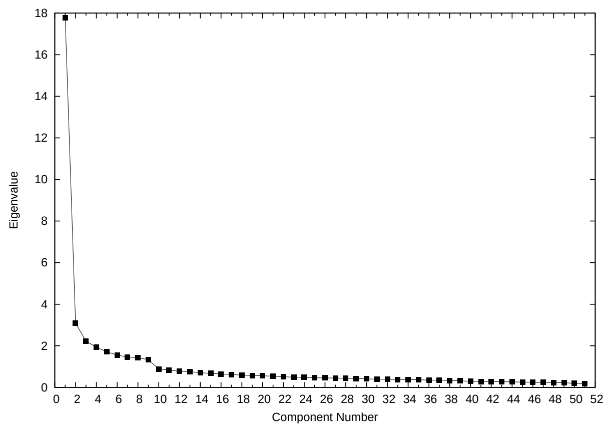


Figure 4.3: Screeplot family entertainment data set.

nine factor model only lacks a significant loading for product attribute D1h. Of all four factor models, this model gives the best results. The ten factor models actually identify attribute D1h as a separate factor, but still respect the original hierarchical structure. Such a factor with a single variable is an indication that D1h is not measuring the same dimension as the other variables of dimension D1. Therefore, the Varimax rotated nine factor model was selected and variable D1h was removed from the set of product attributes for the construction of a summated scale.

Based on the results of this nine factor model, nine factorized dimensions are created. The score of each factorized dimension is calculated by taking the average of the scores on the attribute variables which loaded significantly on the dimension. Table 4.9 provides an overview of each factorized dimension and the variables used to calculated the summated score. The factorized dimensions corresponded perfectly with the hierarchical dimensions and were given the same name. Since all Cronbach's alphas were significantly larger than 0.70, the summated scores can be considered as reliable scores for the factorized dimensions.

4.5 Banking product data set

The third data set contains satisfaction data gathered among 1101 customers of a company which sells products and services in the banking sector. Each customer has bought a durable product of the company together with a service contract. Each customer completed a survey of 82 questions. No socio-demographic data were collected about the customers. Table 4.10 and 4.11 provide an overview of all the variables in the data set.

Firstly, all cases with missing values for variables 'Satisfaction' and 'Recommendation' were removed because they will most likely appear as dependent variables in subsequent analyses. This reduced the sample size from 1101 to 1038. Several variables were removed because they had very high levels of missingness. Table A.17 in Appendix A.3 shows the levels of missingness for each variable. All variables with more than 20% missing entries were removed from the analysis. This caused the

Table 4.9: Factorized dimensions of family entertainment data set

Factorized Product Dimension	Product Attributes	Cronbach's Alpha
Overall Experience	D1a, D1b, D1c, D1d, D1e, D1f, D1g, D1i	0.90
Product A	D2a, D2b, D2c, D2d, D2e, D2f, D2g	0.82
Product B	D3a, D3b, D3c, D3d	0.85
Accommodation	D4a, D4b, D4c, D4d, D4e, D4f, D4h	0.82
Product C	D5a, D5b, D5c, D5d, D5e, D5f	0.91
Product D	D6a, D6b, D6c	0.88
Personnel	D7a, D7b, D7c, D7d, D7e	0.89
Prices	D8a, D8b, D8d, D8e, D8f, D8g	0.90
Communication	D9b, D9c, D9d, D9e	0.86

removal of dimensions 'Sales Service', 'Maintenance' and 'Administration' and their corresponding product attributes. Note that these variables were not deleted because they have no importance in the CS/D process, but because the available data is too limited to take them into account in subsequent analyses. Furthermore, two attribute performances of the dimension 'Invoices' and three product attribute performances of the dimension 'Communication' were also deleted. The latter two dimensions still contain respectively four and five attributes which is substantial enough to keep these dimensions in the data set. All these preprocessing steps reduced the data set to 1038 records and 46 variables, covering 6 product dimensions, i.e. *Image*, *Price*, *Quality*, *Product Performance*, *Invoices* and *Communications*, which have respectively 10, 6, 0, 9, 4 and 5 corresponding product attributes.

Little's MCAR Test was significant at a 0.01 level which implies that the remaining missing data is still not missing completely at random (MCAR). Therefore, the SPSS EM model-based imputation technique was applied and the minima and maxima of each variable were verified to see if they fell within the original scale's limits. Values outside the scale's limit were corrected in the same way as for the other two data sets. Table A.17 shows the mean and standard deviation of each variable after imputation. Comparing these values with the means and standard deviations of the original data shows that the univariate structure of the data remained intact after imputation.

With 1038 cases and 34 product attributes, the factor analysis has a large enough sample and a good respondent-variable ratio of 30 : 1. Both the overall MSA measure and Bartlett's Test of Sphericity (cf Table 4.12) indicate that the level of overall intercorrelation between the product attributes is high enough for a factor analysis.

Table 4.10: Variables of banking product data set

Variable	Values
ID	
Satisfaction	[1 – 10] [Very Low – Very High]
Disconfirmation	[1 – 5] [Very Negative – Very Positive]
Recommendation	[1 – 5] [Very Unlikely – Very Likely]
Usage	[1 – 3] [Low – High]
Value For Money	[1 – 5] [Very Low – Very High]
Dimension 1: Image	[1 – 5] [Very Bad – Very Good]
Dimension 2: Price	[1 – 5] [Very Bad – Very Good]
Dimension 3: Quality	[1 – 5] [Very Bad – Very Good]
Dimension 4: Product Performance	[1 – 5] [Very Bad – Very Good]
Dimension 5: Sales Service	[1 – 5] [Very Bad – Very Good]
Dimension 6: Maintenance	[1 – 5] [Very Bad – Very Good]
Dimension 7: Invoices	[1 – 5] [Very Bad – Very Good]
Dimension 8: Administration	[1 – 5] [Very Bad – Very Good]
Dimension 9: Communication	[1 – 5] [Very Bad – Very Good]

Table A.18 in Appendix A.3, which contains the partial correlations between the variables as well as the MSA measure for each variable on the diagonal, reveals a similar conclusion. Except for two variables, i.e. D7e and D9g, all MSA measures are at least mediocre with a value larger than 0.70. With MSA measures around 0.51 and 0.56, the results still comply with the rule of thumb in [59] which states that MSA values must exceed 0.50 for both the overall test and each individual variable. Furthermore, partial correlations are rather small with values less than 0.45 for all but one variable.

According to the original hierarchical structure of the product attributes which were entered in the factor analysis, five dimensions are expected to be found. The latent root criterion, which selects all factors with an initial eigenvalue larger than 1, suggests a 7 factor analysis (cf Table A.19 in Appendix A.3). The screeplot reveals two points where the curve flattens out, which is at four components and at eight components. Based on these three criteria, a four factor model, a five factor model a seven factor model and an eight factor model were considered. For each model, a Varimax and an Oblimin rotated version were built. The factor loadings for all these models can be found in Appendix A.3.

After a first analysis of the eight models, both four factor models and the Oblimin rotated five factor model were no longer considered as viable models because they had too many variables without significant factor loadings, i.e. respectively 10, 11 and 9 non significant variables. This high amount of non significant variables are an indication that the factor model does not fit the data. Also the eight factor

Table 4.11: Product dimensions and product attributes of banking product data set

Product Dimensions	Number of product attributes
D1: Image	10 [D1a – D1j]
D2: Price	6 [D2a – D2f]
D3: Quality	0
D4: Product Performance	9 [D4a – D4i]
D5: Sales Service	10 [D5a – D5j]
D6: Maintenance	15 [D6a – D6o]
D7: Invoices	6 [D7a – D7f]
D8: Administration	4 [D8a – D8d]
D9: Communication	8 [D9a – D9h]

Table 4.12: Overall MSA and Bartlett's test of sphericity for banking product data set

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.918
Bartlett's Test of Sphericity	Approx. Chi-Square	13317.636
	df	561.000
	Sig.	.000

models were removed from further analysis. These models seemed unappropriate because they contained various factors with only two or even one significant loading. Preferably, factors have three or more significantly loading variables in order to have a solid representation of a product dimension.

Of the three remaining models, the Varimax rotated 7 factor model appeared to be the best. Compared with the two other models, it has less insignificant variables and also the interpretation of the various factors made more sense for the Varimax rotated 7 factor model. Remarkable is the fact that none of the three models provided factors which perfectly matched the original hierarchical product dimensions. Even the five factor model, which has the same number of factors as there are hierarchical dimensions in the data, did not correspond perfectly. Apparently, some product attributes of both *Invoices* and *Communications* form a first factor, while the remaining attributes of both dimensions represent a second factor. Also in the seven factor models, it becomes clear that the product attributes belonging to the hierarchical dimensions *Invoices* and *Communications* do not result in two crisp factors. After analyzing the survey questions belonging to these product attributes, three new factorized dimensions can be found, i.e. 'Communication', 'Invoices' and 'Administrative Overload'. At the same time, the seven factor model shows that the hierarchical dimension 'Product Performance' actually consists of two separate dimensions, which are 'Product Performance' and 'Product Reliability'.

After deciding to continue with the Varimax rotated seven factor model, the non significant variables were removed one by one, starting with the variable with the

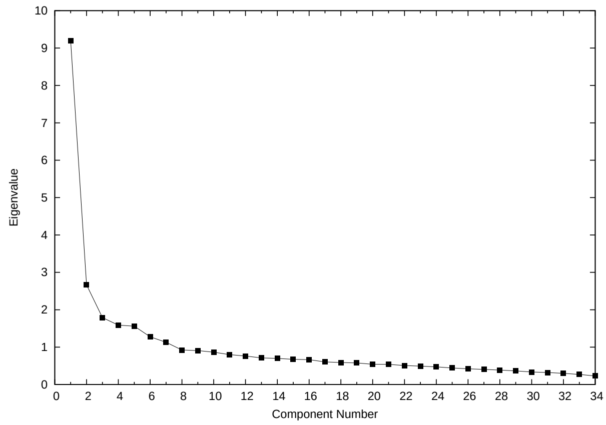


Figure 4.4: Screeplot banking product data set.

lowest highest loading, i.e. D4g whose highest loading was 0.405. Next, variable D2e was removed whose highest loading changed to 0.451, after deletion of D4g. Next, variable D1f was removed whose highest loading was still insignificant at a level of 0.487. Due to removal of these three variables, variable D4i also became insignificant with a maximal factor loading of 0.491. Therefore, this variable was also removed from the analysis. After this step, all variables had significant factor loadings on exactly one factor. The factor loadings of the final Varimax rotated factor model is shown in Table A.28 in Appendix A.3.

Based on this final seven factor model, seven factorized dimensions were created. The score of each factorized dimension is calculated by taking the average of the scores on the product attributes corresponding to the factorized dimension. Table 4.13 provides an overview of each dimension and the variables used to calculated the summated score. The factorized dimensions which corresponded perfectly with the hierarchical dimensions were given the same name. The other factorized dimensions were given a name based on the corresponding product attributes. Most Cronbach's alphas were significantly large except for dimensions 'Product Reliability' and 'Administrative Overload'.

The reliability of the dimension 'Product Reliability' could be increased by adding product attribute D4i again to the dimension. This attribute was removed during the factor analysis because it became insignificant after the removal of several other insignificant variables. By adding this variable back to the factorized dimension, the Cronbach's alpha increases up to 0.60. As for the dimension 'Administrative Overload', it was not possible to increase the measurement reliability. There were no candidate attributes to be added to this dimension and removing attributes from this dimension only decreased Cronbach's alpha. Because its reliability is rather low, this factorized dimension was removed. Furthermore, dimension 'Product Performance' with a mediocre Cronbach's alpha of 0.75 and 5 items makes a good candidate for improving the Cronbach's alpha by deleting a item. However, dropping a single item at a time did not improve alpha. The final set of factorized dimensions are given in Table 4.14.

Table 4.13: Factorized dimensions of banking product data set

Factorized Product Dimension	Product Attributes	Cronbach's Alpha
Image	D1a, D1b, D1c, D1d, D1e, D1g, D1h, D1i, D1j	0.86
Price	D2a, D2b, D2c, D2d, D2f	0.89
Product Performance	D4a, D4c, D4d, D4e, D4f	0.75
Product Reliability	D4b, D4h, D9h	0.53
Invoices	D7a, D7f	0.78
Communications	D9a, D9e, D9f	0.75
Administrative Overload	D7b, D7e, D9g	0.43

Table 4.14: Final set of factorized dimensions of banking product data set

Factorized Product Dimension	Product Attributes	Cronbach's Alpha
Image	D1a, D1b, D1c, D1d, D1e, D1g, D1h, D1i, D1j	0.86
Price	D2a, D2b, D2c, D2d, D2f	0.89
Product Performance	D4a, D4c, D4d, D4e, D4f	0.75
Product Reliability	D4b, D4h, D4i, D9h	0.60
Invoices	D7a, D7f	0.78
Communications	D9a, D9e, D9f	0.75

5 An implementation of the LIED framework: The Dombi's uninorm approach

5.1 Introduction

This chapter introduces a possible implementation of the LIED framework, which uses Dombi's aggregation operator as the underlying aggregation function. Dombi's aggregation function was introduced in 1982 [38] and belongs to the family of generated uninorms. Other authors [129, 130] have already used Dombi's aggregation function to model the CS/D process in the past, but it has never been presented as the implementation of a more general framework. Positioning this modeling approach in the broader context of the LIED framework has several advantages.

Firstly, because the LIED framework models customer satisfaction by the ED paradigm principles, it suffices to recognize Dombi's uninorm as a valid implementation of the LIED framework in order to know whether it is a valid candidate for CS/D modeling. Secondly, the LIED framework offers a structured three-step approach to identify and analyze assumptions Dombi's uninorm imposes on the modeling process. Finally, if some of these assumptions are unwanted, the structured framework helps making the necessary modifications to the aggregation function.

In the next section, Dombi's uninorm will be discussed in detail and some of its properties are introduced, while it is shown that Dombi's uninorm offers a valid implementation of the LIED framework. This implementation will be referred to as the D-LIED implementation. It will be shown that the D-LIED implementation imposes some additional assumptions which can be restrictive, but it also has the benefit of modeling expectations at the individual customer level. In the third section, the validity of the D-LIED implementation will be analyzed, both theoretically and empirically. Finally, the potential of modeling CS/D with the D-LIED implementation will be illustrated by some case studies in the final section.

5.2 Dombi's evaluation operator

Dombi's uninorm [38] belongs to the family of generated uninorms, but was published long before Yager and Rybalov [144] introduced the term uninorm. In his original work, Dombi started from the situation where an object is to be evaluated on all its properties. The evaluation of each property is done against a specific expectation level and the property evaluations have to be aggregated to form an overall object evaluation. Such situation shows strong analogies with multi-criteria decision problems and with the CS/D evaluation, which explains why other authors [129, 130] have previously used Dombi's evaluation operator to model customer satisfaction.

Dombi's binary evaluation operator started from an axiom system with five axioms. The first four axioms are:

Definition 23 (**Axiom system relating to Dombi's evaluation operator**)**.** The first four (out of five) axioms relating to Dombi's binary evaluation operator are:

- The evaluation operator is continuous except for the points $(0,1)$ and $(1,0)$,
- The evaluation operator is strictly monotone increasing in the interval $[0,1]$,
- The evaluation operator is associative,
- For the evaluation operator σ holds that $\sigma(0,0) = 0$ and $\sigma(1,1) = 1$.

The fifth axiom actually comprised a set of three mutually exclusive axioms which determined whether the evaluation operator was conjunctive, disjunctive or none of both (which was called aggregative by Dombi). A conjunctive function has 0 as the absorbing element, a disjunctive function has 1 as the absorbing element and the aggregative evaluation operator has no absorbing element and is undefined for the tuples $(0,1)$ and $(1,0)$. Later work on the family of representable uninorms, to which Dombi's evaluation operator belongs, makes only a distinction between conjunctive and disjunctive. In this thesis, when the evaluation function is mathematically undefined for $(0,1)$ or $(1,0)$, as was the case for Dombi's aggregative evaluation function, the correct behavior, i.e. conjunctive or disjunctive, is set by definition.

This axiom system led to a family of aggregation functions which are defined as follows:

Definition 24 (**Dombi's evaluation operator [38]**)**.** Dombi's evaluation operator are all functions which can be written in the form

$$\sigma(x_1, x_2) = f\left(f^{-1}(x_1) + f^{-1}(x_2)\right)$$

where $f(x)$ is continuous and strictly increasing.

Comparing the definition of Dombi's evaluation operator (cf. Definition 24) with the definition of a representable (or generated) uninorm (cf Definition 13 on page 37) shows that Dombi's evaluation operator is a representable uninorm if one defines the aggregation at the extreme points, i.e. $(0,1)$ and $(1,0)$. Because Dombi's evaluation operator is associative, the above definition can be easily extended to n-ary functions as follows:

Definition 25 (**Dombi's n-ary evaluation operator**)**.** Dombi's evaluation operator are all functions which can be written in the form:

$$\sigma(x_1, \ldots, x_n) = f\left(\sum_{i=1}^{n} f^{-1}(x_i)\right)$$

where $f(x)$ is continuous and strictly increasing.

In [38] three different generator functions $f(x)$ are presented, among which the following is the most popular:

$$f(x) = \frac{\alpha e^x}{1 + \alpha e^x}, \tag{5.1}$$

with the inverse function:

$$f^{-1}(x) = ln\left[\frac{x}{\alpha - \alpha x}\right]. \tag{5.2}$$

Constructing Dombi's uninorm by means of Definition 25 and the generator function in Eq. 5.1, results in the following aggregation function:

$$U(x_1, x_2, \ldots, x_n) = \frac{\prod_{i=1}^{n} x_i}{\prod_{i=1}^{n} x_i + \alpha^{n-1} \prod_{i=1}^{n}(1 - x_i)}. \tag{5.3}$$

For $\alpha = 1$, this results in another well known mixed aggregation function which is often referred to as the $3 - \prod$ function.

Dombi's uninorm also belongs to the family of generated functions because Dombi's uninorm is a representable uninorm (cf Definition 25, Definition 13 and Definition 14). Dombi's uninorm is also a valid LIED implementation because it is a generated function with $f^{-1}(x) = 0$ for $x = \frac{\alpha}{1+\alpha}$ and $0 \in Dom\,(f(x))$ (cf Definition 20). This allows using the properties of the LIED framework to analyze Dombi's uninorm as a modeling function for the CS/D process.

First, the general properties of the LIED framework when it is implemented by means of a generated function are considered. According to Property 10 it follows that Dombi's uninorm will show compensatory, conjunctive and disjunctive behavior. Dombi's uninorm will compensate a positive disconfirmation (i.e. more than the expected value) and a negative disconfirmation (i.e. less than the expected value), while two positive (negative) disconfirmations will be positively (negatively) reinforced. Furthermore, Dombi showed in his original work that his evaluation operator is symmetric. As has been mentioned in chapter 3 on page 48, this has some important implications for the model. Firstly, a LIED framework implemented by a symmetric generating function assumes that the expectation for each product attribute is equal and the attribute performance transformation into the attribute's disconfirmation effect is also the same for every attribute. This is a very restrictive assumption which might not always model the reality correctly, especially when the product attributes have different levels of importance in the CS/D process. However, using Dombi's uninorm to model the CS/D process also holds a big advantage over other generating functions which makes it worth considering. The strength of Dombi's uninorm lies in the fact that it only contains 1 parameter which needs to be estimated, i.e. α. This makes it possible to estimate the CS/D process for each customer separately, since only one record is needed to fit a model with a single parameter. Using only one case to estimate a single parameter model is of course more subject to noise and measurement errors than learning a model from multiple case. On the other hand, by using the factorized dimensions, measurement errors were already for a large part taken care of.

Learning a separate model for each customer allows the researcher to extract the individual expectation level for each customer. Given Definition 21, it follows that

Dombi's uninorm has a neutral element, which had also been proven by Dombi in his original article [38]. The neutral element e of Dombi's uninorm can be expressed as:

$$e = \frac{\alpha}{1+\alpha}. \tag{5.4}$$

According to Property 2 on page 42, the neutral value e of Dombi's uninorm corresponds to the customer's expectation. If one builds a separate model for each customer, the customer's expectation can be derived by means of Eq. 5.4 if parameter α is known. Let S represent the customer's satisfaction score and x_i the product attribute performance scores, then the following equation can be used to find α and subsequently e:

$$\alpha = \left[\frac{\prod_{i=1}^{n} x_i}{\prod_{i=1}^{n}(1-x_i)} \frac{1-S}{S}\right]^{\frac{1}{n-1}} \tag{5.5}$$

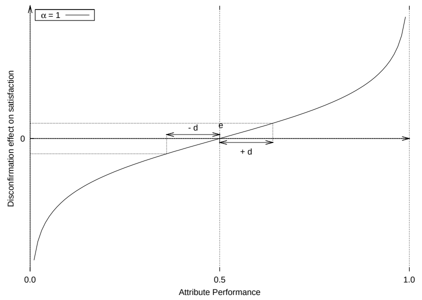
Not only can the LIED framework be used to study general modeling properties of Dombi's uninorm, it also allows to divide Dombi's uninorm into three parts, each corresponding to a specific step function. Each step function can also possess specific properties which allows researchers to increase their understanding of the CS/D modeling process. The first-step function of the D-LIED implementation is Eq. 5.2. The first two properties (cf Properties 1 and 2) of the first-step function relate to the customer's expectation level or zone of indifference, which have already been discussed above. The third property (cf Property 3) defines the symmetry of the disconfirmation effect on satisfaction. According to this property, Dombi's uninorm only models a symmetric disconfirmation effect when parameter $\alpha = 1$, which can be illustrated graphically.

Figure 5.1 shows three subfigures. Subfigure 5.1a shows the first-step function for a Dombi uninorm with $\alpha = 1$. This particular uninorm has a neutral element $e = 0.5$ (cf Eq. 5.4) around which it is perfectly symmetric. Subfigure 5.1a shows that a positive disconfirmation $+d$ has an effect on satisfaction equal in size as the effect caused by an equally large but negative disconfirmation $-d$. When $\alpha \neq 1$, this is no longer true as is being illustrated by Figures 5.1b and 5.1c. Changing the only parameter of the Dombi uninorm causes the first-step function to move vertically, which results in the loss of the disconfirmation effect's symmetry. More in particular, if $\alpha < 1$ then the neutral element e will be less than 0.5 and a negative disconfirmation has a larger effect on satisfaction than the effect of an equally large but positive disconfirmation. This can be explained as follows: a customer who is already having low expectations will be more sensitive to negative disconfirmation because negative disconfirmations are considered less likely to occur than positive disconfirmations given the low expectations.

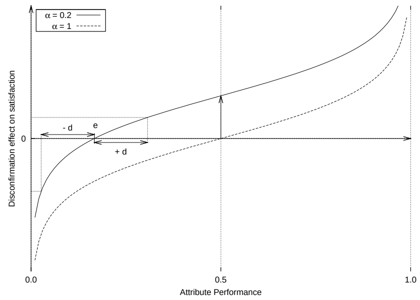
When $\alpha > 1$, the neutral element e will be larger than 0.5 and a positive disconfirmation will have a larger impact on satisfaction than a negative disconfirmation. This can be explained as follows: a customer with high expectations ($e > 0.5$) will be much more surprised if these expectations are exceeded than when the product fails to meet them. Therefore, the effect of exceeding expectations will be larger. However, one should not forget that when expectations are already high, it becomes increasingly more difficult for companies to exceed them.

The properties of the second step function, which is a mathematical sum in case of Dombi's uninorm, relate to the compensating, conjunctive or disjunctive behavior

(a) Symmetric disconfirmation effect when $\alpha = 1$



(b) Asymmetric disconfirmation effect when $\alpha < 1$



(c) Asymmetric disconfirmation effect when $\alpha > 1$

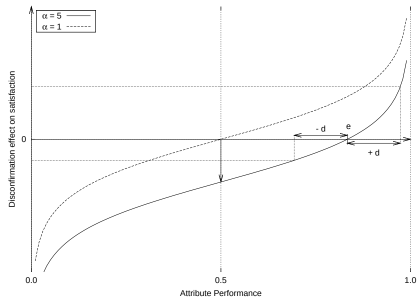


Figure 5.1: Asymmetric and symmetric disconfirmation effects modeled by Dombi's uninorm

of the model which have already been discussed previously when Property 10 was studied.

The third step function of the LIED framework has three more properties which might or might not hold for the D-LIED implementation. Firstly, Property 7 allows us to determine the customer's initial satisfaction level, also referred to as the adaptation level. In case of Dombi's uninorm, the adaptation level a is:

$$a = \frac{\alpha}{1+\alpha}, \tag{5.6}$$

which is exactly equal to the customer's expectation e.

Figure 5.2 illustrates the third step function for three different values of its parameter α. As alpha increases, the function is translated to the left and the initial satisfaction level, which is the intersection with the Y-axis, increases. Therefore, the customer's expectation and the customer's initial satisfaction level are positively correlated in the D-LIED implementation because $a = e$. Figure 5.2 also shows that Properties 8 and 9, implying diminishing returns of positive and negative disconfirmation effects, hold for Dombi's uninorm. These properties hold for any value of α since it is the only parameter in the model and it only influences the horizontal translation of the third step function. Therefore the shape of the third-step function, which incorporates diminishing return effects, remains intact. The fact

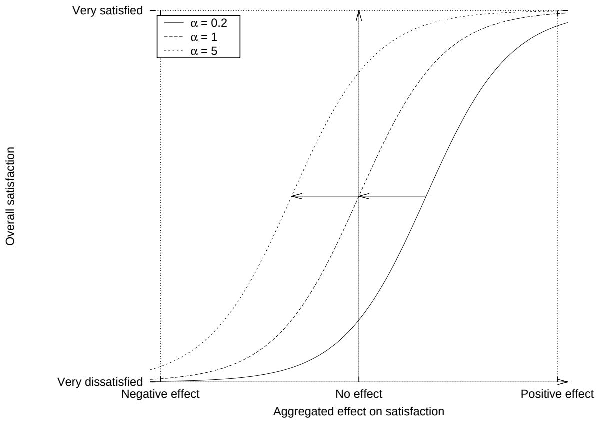


Figure 5.2: The third step function of Dombi's uninorm.

that Dombi's uninorm models diminishing return effects on satisfaction implies that as the overall disconfirmation effect grows larger (smaller), the additional final effect on satisfaction grows smaller.

This section illustrated the power of the LIED framework. Once it was proven that Dombi's uninorm is a valid LIED framework implementation, the framework served to analyze the modeling properties of Dombi's uninorm. It is now known that the D-LIED implementation takes the following into account when it models the CS/D process:

- It is assumed that each attribute's performance is compared against a specific reference level, which is captured by Dombi's uninorm neutral element and which is called expectation level because of historical reasons. Dombi's uninorm assumes that this reference level is global and equal for each attribute. On the other hand, it can be estimated for each customer separately and thus the model does not assume that all customers share the same expectation level. Dombi's uninorm trades unique expectation levels per attribute for unique expectation levels per customer.

- The D-LIED implementation assumes that the comparison between expectation and performance leads to disconfirmation which has an effect on satisfaction. When expectations are low, negative discrepancies between expectation and performance causes larger effects on satisfaction than equally sized positive discrepancies. When expectations are high, positive discrepancies have larger effects than equally sized negative discrepancies.

- Furthermore, two positive disconfirmation effects reinforce each other positively, two negative disconfirmation effects reinforce each other negatively and a positive and negative disconfirmation effect compensate each other.

- Finally, the final satisfaction outcome is the resultant of an initial satisfaction level which is equal to the expectation level. Therefore, when expectations are low, i.e. less than 0.5 on a $[0, 1]$ scale, the initial satisfaction level will also be low, i.e. less than 0.5 on a $[0, 1]$ scale and vice versa. Also, the final satisfaction score is diminishingly sensitive to increasing disconfirmation effects, both positively and negatively.

5.3 Validity

One of the key strengths of the D-LIED implementation is that it becomes possible to learn the global expectation level for each customer separately. However, analogous with what Peter [106] mentions in his article on reliability of marketing measures, an estimate which has not been demonstrated to have a high degree of validity cannot be considered a scientific estimate. Therefore, it is essential that the validity of the neutral element from the D-LIED implementation as an estimate of the customer's expectation level is verified. Following the ideas of Churchill [25] on validity of marketing constructs, two types of validity can be distinguished.

Firstly, the construct should "look right", which means that the way the construct is measured or the estimate is derived should make sense and give an impression of validity. This kind of validity is also referred to as *face* or *content validity*. Secondly, the estimate or measure should possess *construct validity*, which lies at the very heart of the scientific process and which means that the estimate is in fact measuring the underlying concept which the researcher wants to measure. Both validity concepts will be studied separately.

Theoretical validity

Face validity means that the estimate "looks right". For the D-LIED implementation, this implies that the expectation level "looks right", i.e. it corresponds to the existing CS/D theories. This type of validity has already been demonstrated throughout the previous sections of this thesis. Dombi's uninorm is proven to be a valid implementation of the LIED framework, which has been built based upon existing and dominating CS/D theories. Especially the ED paradigm has been used as the foundation for the LIED framework and consequently it can be said that Dombi's uninorm implements this paradigm. Furthermore, the analysis of the CS/D modeling properties of Dombi's uninorm in section 5.2 showed that the D-LIED implementation incorporates aspects as diminishing sensitivity of satisfaction to disconfirmation, asymmetric effects of disconfirmation on satisfaction and a non-linear relationship between performance and satisfaction. All these ideas have been suggested by other authors [90]. As a conclusion, the D-LIED implementation "looks right" according to the existing theories and therefore its estimate of the customer's expectation is theoretically valid.

Empirical validity

The construct validity of a measure can be determined in two ways [25]. Firstly, existing measures of the same construct can be used to verify to which degree they correlate with the new measure. Secondly, it can be verified if the new measure behaves as expected in relation to other constructs. In contrast with theoretical validity, conclusions about the construct validity are drawn from empirical results. Therefore, construct validity will also be referred to as empirical validity.

The option of measuring the correlation between the model-derived expectation level and some other direct measurements of the expectation level might be more difficult than appears at first sight. A symantical error is the cause of the misconception that such verification can be easily done. The expectation level derived from the D-LIED implementation represents the referent against which performance is compared, which is not necessarily the customer's expectation level. As have been mentioned in chapter 2, the original ED paradigm assumed that customers compared the perceived performance against their expectations. However, many authors have questioned expectation as the comparative referent and other referents have been suggested, such as e.g. experience-based norms [19]. To make it even more difficult, various authors have argued that expectation is not uniquely defined and various levels of expectations can be measured, e.g. the deserved expectation level or ideal expectation level. Some authors even proposed that customers use a combination of referents, which could be both expectations and norms. Finally, it was questioned whether all customers use the same type of referent. Maybe some customers use some kind of expectation level as their comparative referent, while others use a norm or a combination of both. Therefore, even if one succeeds in constructing a scale which produces reliable and valid measurements of the type of referent one wants to measure, one can still be measuring the wrong construct. Chapter 2 also mentioned that measuring expectations prior to product purchase can be difficult since it assumes possible identification of future customers while post-exposure expectation measures are possibly influenced by the product experience. Therefore, verifying the empirical validity of the D-LIED's estimate of customer expectation by comparing it with other measurements of the customer's comparative referent becomes near impossible and the second approach has to be used to prove empirical validity.

The second approach investigates if the construct or estimate behaves as it should in relation to other constructs. This approach studies if the D-LIED estimate of the customer's expectation level is related to concepts such as satisfaction, disconfirmation, performance in theoretically and empirically predictable ways. To this end, the empirical results of articles using Dombi's uninorm to model customer satisfaction shall be reviewed and new experiments on three different data sets were conducted.

Empirical evidence from previous research

In a first article by Vanhoof et al. [129], Dombi's uninorm was used to model customer satisfaction and the neutral value was assumed to be the customer's adequate expectation level, i.e. the expected performance level based on previous performances. This article contains two experiments which indicate that Dombi's neutral value has the expected behavior of a comparative referent. The first experiment studied the evolution of the expectation level for different levels of price/quality ratio. The authors argue that:

"Equity theory suggests that satisfaction with specific types of purchase transactions is determined by a consumer's comparison of the inputs and outcomes of companies on a quid pro quo basis. Customers expect service basics delivered at a level [commensurate] with the price they pay. A higher price results in a higher expectation level." [129]

They calculated the mean expectation level for different price/quality levels for two different data sets. The results of these analysis can be found in Tables 5.1 and 5.2. These tables show that Dombi's neutral element behaves as expectation, i.e. it increases as price/quality goes up.

Table 5.1: Dombi expectation level versus price/quality ratio [Vanhoof et al. [129], p.119]

Price/quality ratio	very low	low	equal	high	very high
Mean Dombi exp.	4.6	5.5	6.14	6.85	6.91
Number of cases	159	323	423	190	14

Table 5.2: Dombi expectation level versus price/quality ratio [Vanhoof et al. [129], p.119]

Price/quality ratio	very low	low	equal	high	very high
Mean Dombi exp.	4.77	5.73	6.38	7.39	8.07
Number of cases	18	58	165	138	43

In the same article, the authors also provided additional evidence that Dombi's neutral element behaves as a proxy for the expectation. They calculated the mean expectation for 9 different dimensions of the service provided by a particular company. They found that the dimensions *maintenance* and *products* had the highest expectation scores. This matches with the conclusions of Parasuraman's study [102] of sixteen focus group interviews with customers in six service sectors. Parasuraman stated that customers expect service companies to do what they are supposed to do. For the company in the study of Vanhoof et al. [129], the fundamental aspects were delivering a good product and providing high quality maintenance, which were also the aspects with the highest expectations.

In a second article by Vanhoof et al. [130], the authors used Dombi's uninorm to model customer satisfaction and derive the customer's expectation level. This allowed them to generate impact curves for the different product attributes. An impact curve of a specific attribute measures the contribution of each attribute performance level to the overall satisfaction score. These impact curves allow marketers to categorize attributes as excitement factors, performance factors or basic factors, which is a classification going back to Kano's three-factor theory [73].

A basic factor does not generate satisfaction when expectations are met or exceeded, they merely avoid dissatisfaction. Excitement factors on the other hand do not cause dissatisfaction when expectations are not met, but can surprise customers when expectations are met or exceeded and generated "delight". The performance factors are attributes which lead to satisfaction if performance exceeds expectations and to dissatisfaction when performance fails to meet expectations [83].

Vanhoof et al. applied their technique to a data set from a real company and confronted experts from the company with their results. The experts confirmed that the results were in congruence with previous findings. Although this only provided validity of the final results, i.e. the impact curves, this indirectly implies that the expectation values estimated from the Dombi uninorm model, which play a key role in generating the impact curves, must also have some level of validity.

These two articles already provide evidence which suggest that Dombi's neutral element is a valid estimate of the customer's expectation level. Before this body of evidence will be extended by various experiments on the three data sets used in this thesis, the application of the D-LIED implementation to model CS/D has to be discussed in detail.

Applying the D-LIED implementation

The D-LIED implementation is always applied at the individual customer level, which means that a new model is build for each customer. There are two levels at which the D-LIED implementation can be applied, i.e. the general level and the dimensional level.

At the general level, the model is built by learning the model's parameter α (Eq. 5.5) with the overall satisfaction score (S) and the dimension performances (x_i). Once α is learned for each customer, Eq. 5.4 can be used to learn the neutral element which represents the customer's general expectation level e. By subtracting the general expectation level e from the performance p_i of dimension i, the dimension disconfirmation D_i is found.

At the dimensional level, a D-LIED model is built separately for each dimension. For a specific dimension i, the model's parameter α is learned through Eq. 5.5 with the performance p_i for dimension i as the satisfaction level S and the performance p_{ij} of all attributes j belonging to dimension i as x_i. This approach uses the dimension performance p_i as a proxy for the dimension's satisfaction. Once α is learned for each customer and every dimension, Eq. 5.4 can be used to learn the neutral element which represents the customer's expectation level e_i for a specific dimension i. By subtracting the dimension expectation level e_i from the dimension performance p_i, the dimension disconfirmation d_i is found. Note that this dimension disconfirmation is calculated with dimension dependent expectations e_i, while the dimension disconfirmations from the general approach uses the general expectation level e. To stress this difference, the dimension disconfirmations at the general level is denoted by a capital letter D_i. If dimension disconfirmations are needed and the hierarchical data is available, the dimensional approach is preferred.

Table 5.3 provides an overview of the new constructs created with the D-LIED implementation. While the general approach can be applied to the original dimensions as well as the factorized dimensions, it makes little sense to apply the dimensional approach to the factorized dimensions because the factorized dimensions are not measured directly but are the average of the corresponding attribute performances. To distinguish the general approach applied to the factorized dimensions from its application to the original dimensions, the superscript F is used, i.e. p_i^F represents the performance on factorized dimension i, e^F represents the general expectation level learned with p_i^F and D_i^F represents the factorized dimension disconfirmation.

The D-LIED implementation uses Eq. 5.2 as a first step function which has $[0, 1]$ as its domain. Therefore, the data has to be transformed to the $[0, 1]$ interval. As mentioned in chapter 3, no matter which scale is originally used to measure the

Table 5.3: New constructs created with Dombi's uninorm implementation of the LIED framework

Level	New Constructs	Formula	Input
General	general expectation e	Eq. 5.4 and 5.5	dimension perf. p_i general sat. S
	dimension disc. D_i	$D_i = p_i - e$	dimension perf. p_i general exp. e
Dimensional	dimension expectation e_i	Eq. 5.4 and 5.5	dimension perf. p_i attribute perf. p_{ij}
	dimension disc. d_i	$d_i = p_i - e_i$	dimension perf. p_i dimension exp. e_i

variables, they can always be linearly transformed to the $[0, 1]$ scale without affecting the results. However, one has to be careful with selecting a proper transformation to the $[0, 1]$ domain. Dombi's uninorm considers the elements 0 and 1 as the absolute extreme values on the $[0, 1]$ scale. No value larger than 1 or smaller than 0 exists. The value 0 is so small, that no matter which other values (different from 1) are aggregated with 0, the outcome is always 0. At the same time, the value 1 is so large, that no matter which other values (different from 0) are aggregated with 1, the outcome is always 1. These values act as absorbent values. An aggregation with both 0 and 1 values has 0 or 1 as outcome, depending on whether Dombi's uninorm is respectively conjunctive or disjunctive. The values 0 and 1 can be compared with the values $-\infty$ and $+\infty$. On the other hand, the maximum value which can be given on a survey for an attribute performance is most likely not interpreted by the customer as the absolute maximal performance level possible. It most likely just represents a very high performance level. Therefore, this survey score should not be transformed to the value 1. The same holds for the lowest value a respondent can give for the performance of a product attribute. This score most likely represents an extremely low performance level, but not necessarily the lowest performance level possible. Therefore, the lowest score possible on the survey should not be transformed to the value 0.

To determine to which values the maximum and minimum scores on a survey should be mapped is not a trivial task. The problem is that the data does not contain information to help finding the correct transformation. Within this thesis, the following rule of thumb was used which appeared to produce good results: If the question on the survey was e.g. measured on a scale from $[1-5]$, the scale was divided by 6, which is the maximum scale value plus 1.

New evidence: analysis 1

To validate if the new constructs learned by the D-LIED implementation behave appropriately, information is needed on how they relate to each other and to other satisfaction related constructs in reality. Szymanski and Henard [122] performed an extensive meta-analysis of the satisfaction-related correlations found in satisfaction research from the last three decades. Table 5.4 repeats the results found by Szymanski and Henard and compares them with the results gathered from the three different data sets. A distinction is made between the results from the general approach and the results of the dimensional approach.

Table 5.4: Satisfaction-related correlations: Szymanski and Henard's meta analysis [122]

	Positive Correlations[a]	Negative Correlations[a]	Nonsignificant Correlations
Perf.-sat.	136	17	6
Perf.-disc.	22	0	1
Exp.-sat.	13	1	3
Exp.-perf.	22	1	2
Exp.-disc.	6	6	11
Disc.-sat.	121	1	15

[a] Statistically significant at alpha ≤ 0.05

Table 5.5: Satisfaction-related correlations: general approach

	Positive Correlations[a]	Negative Correlations[a]	Nonsignificant Correlations
Perf.-sat.	20	0	0
Perf.-disc.	20	0	0
Exp.-sat.	3	0	0
Exp.-perf.	20	0	0
Exp.-disc.	10	7	3
Disc.-sat.	15	0	5

[a] Statistically significant at alpha ≤ 0.05

Table 5.4 contains the correlations found by Szymanski and Henard between the constructs performance, satisfaction, disconfirmation and expectation. The same correlations were calculated for all three data sets following the general approach based on factorized dimension scores, i.e. using satisfaction, general expectation e^F, factorized dimension performances p_i^F and factorized dimension disconfirmation

Table 5.6: Satisfaction-related correlations: dimensional approach

	Positive Correlations[a]	Negative Correlations[a]	Nonsignificant Correlations
Perf.-sat.	20	0	0
Perf.-disc.	20	0	0
Exp.-sat.	20	0	0
Exp.-perf.	19	0	1
Exp.-disc.	0	20	0
Disc.-sat.	8	0	12

[a] Statistically significant at alpha ≤ 0.05

scores D_i^F. Table 5.5 provides an overview of the signs of the correlation. Table 5.6 contains the correlations derived from the dimensional approach, i.e. using dimension performance p_i as satisfaction, dimension expectations e_i, attribute performances p_{ij} and dimension disconfirmation scores d_i.

These tables show that for both approaches, all dimension performances were significantly correlated with satisfaction in a positive way. This corresponds with the results of Szymanski and Henard who found that the majority of performance-satisfaction correlations were positively statistically significant. Note that neither satisfaction nor dimension performance are constructs derived from the D-LIED implementation. They are directly measured or constructed by means of a factor analysis. Therefore, the fact that the Performance-Satisfaction correlations match the results found by Szymanski and Henard merely prove that the three data sets have a similar structure as the ones used in the meta analysis, in terms of satisfaction and performance. This finding is important for the validity and interpretation of the other correlations.

The performance-disconfirmation correlation is exclusively positive in a statistically significant way for both the general approach and the dimensional approach. Almost the exact same pattern was found by Szymanski and Henard. Thus, the factorized disconfirmation and the regular disconfirmation scores, which are both derived with the D-LIED model, relate to product performance in the same way as disconfirmation measures from other studies.

The correlation between expectation and satisfaction is predominantly positive and statistically significant within the literature review done by Szymanski and Henard. Almost the exact same pattern was found between the expectations derived from the D-LIED implementation and the measured satisfaction level, both for the general approach and the dimensional approach. These results provide evidence that general expectations derived from the factorized dimensions as well as dimension expectations derived from attribute performances behave in the same way as directly measured expectations.

The same holds for the relationship between the D-LIED derived expectations and the directly measured dimension performance scores. These correlations are predominantly positive and statistically significant in Szymanski and Henard's meta-analysis as well as in the dimensional and general analysis of the three data sets in this thesis. This further strengthens the validation of the way that expectation behaves in relation to other satisfaction-related constructs.

The only correlation which does not exactly match with the results found in the meta-analysis, is the correlation between the two model-derived constructs, i.e. expectation and disconfirmation. In the meta-analysis, most of these correlations are non significant and if they are significant, they can be both positive or negative. The dimensional approach only produces statistically significant correlations which are negative. This is a direct result of the way disconfirmation is calculated in the dimensional approach, i.e. $d_i = p_i - e_i$. For the general approach, the results have a better match with the meta-analysis. Although these factorized disconfirmations are also calculated as the difference between performance and expectation, they produce both positive and negative correlations which are statistically significant.

The correlation between disconfirmation and satisfaction is the last correlation from the meta-analysis to compare with. In the meta-analysis, disconfirmation and satisfaction are predominantly positive and statistically significant. The same holds for the correlations between factorized disconfirmation and general satisfaction in the general approach. The dimensional approach has trouble producing significant correlations between disconfirmation and satisfaction, but those which are statistically significant are also exclusively positive.

The constructs derived from the D-LIED implementation correlate with other satisfaction-related constructs in a valid and realistic way, which adds to the empirical validation of Dombi's neutral value as a proxy for the customer's expectation. The only relation which should be treated with some precaution is the one between expectation and disconfirmation in case of the dimensional approach. The researcher should keep in mind that this correlation is merely a reflection of the way disconfirmation is calculated. It should be noted that although the LIED framework operates under the assumption that the ED paradigm is a valid theory for explaining customer satisfaction, the correlations found in this study suggest that the D-LIED implementation produces valid results in general since Szymanski and Henard's meta-analysis concerns customer satisfaction theory in general and not the ED paradigm in particular.

New evidence: analysis 2

The next experiment is inspired by results from Oliver [95], which are also published in [97]. Oliver argued that both expectation and disconfirmation have separate but additional effects on customer satisfaction and he uses Table 5.7 to illustrate this idea. This table shows that as one keeps expectation constant, overall satisfaction will increase as disconfirmation moves from negative disconfirmation to positive disconfirmation. It also shows that if the disconfirmation level is kept constant, the satisfaction increases as the expectation level increases. These results correspond with the basic idea behind the ED paradigm and similar results are expected from empirically valid estimates of expectation and disconfirmation.

Therefore, similar tables were created by means of the D-LIED implementation for all three data sets. The general approach was used to derive the general expectation level e^F for each customer and the disconfirmation score D_i^F for each dimension. These disconfirmation scores and expectation scores are continuous and have to be transformed into categorical attributes to reproduce Table 5.7. As for the expectation

Table 5.7: Tabular results from Oliver (1977)[1]

Respondent Expectation Level	Negative Disconfirmation	Confirmation	Positive Disconfirmation
Low	2.97	3.88	5.06
High	3.87	5.06	5.93

[1] A 1-to-7 scale; cell means are not equal; total $n = 243$.

variable, all values below the median are transformed into low expectation and all values above the median into high expectation. The disconfirmation variables were transformed such that all values within the interval $[-\frac{\sigma}{2}, +\frac{\sigma}{2}]$ are coded as confirmation, all values less than $-\frac{\sigma}{2}$ are coded as negative disconfirmation and all values more than $\frac{\sigma}{2}$ are coded as positive disconfirmation, with σ representing the standard deviation.

Tables B.1, B.2 and B.3 in appendix B reproduce Table 5.7, showing the mean satisfaction score for customers with a specific expectation level (high or low) and a specific disconfirmation level (negative, none, positive). Each product dimension has its own disconfirmation scores which results in a table per product dimension. Additionally, for each data set the average disconfirmation [1] was calculated which was categorized in the same way as the other dimension disconfirmations. The average disconfirmation score can be considered as a proxy for the disconfirmation at product level, which is the level at which Oliver measured disconfirmation for construction of Table 5.7. The mean satisfaction scores for customers with a specific expectation level and a specific product disconfirmation level are shown in Table 5.8.

To verify the statistical significance of the differences in mean satisfaction between different cells, a two-way ANOVA analysis was performed. For full details of this statistical approach, the reader is referred to statistical books which deal with ANOVA analysis such as [93]. A two-way ANOVA analysis has two important assumptions, i.e. normality of the error terms and constancy of the error variance, which were violated for almost every dimension.

Various transformations were tried to eliminate the violations of these two assumptions, such as the square root transformation, the logarithmic transformation and the reciprocal transformation. Also the Box-Cox procedure was used in order to find a suitable transformation. However, no transformation was found which successfully transformed the error terms to normally distributed error terms with constant variance. An alternative would be to use a non-parametric version of the two-way ANOVA analysis, which can be established by running a normal two-way ANOVA on ranked data [30]. However, Seaman et al. [112] warned that such transformations might be inappropriate in certain situations.

Instead of working with a non-parametric two-way ANOVA technique, it was preferred to transform the data such that a one-way ANOVA analysis becomes possible. The transformation combines the expectation variable with two categories and the disconfirmation variable with three categories into a single variable with six categories, which are:

[1] The average was taken over all dimensions, not over all customers.

Table 5.8: Mean satisfaction for different levels of expectation and product disconfirmation.

(a) Energy data set

		Disconfirmation		
		-	0	+
Expectation	Low	1.58	$3.49^{\{a\}}$	$4.36^{\{a\}}$
	High	$3.24^{\{b\}}$	$4.42^{\{a,b\}}$	$5.28^{\{a,b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

Satisfaction is expressed on a 1-to-7 scale

(b) Family entertainment data set

		Disconfirmation		
		-	0	+
Expectation	Low	6.49	$7.81^{\{a\}}$	$8.68^{\{a\}}$
	High	$8.09^{\{b\}}$	$9.23^{\{a,b\}}$	$9.74^{\{a,b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

Satisfaction is expressed on a 1-to-7 scale

(c) Banking product data set

		Disconfirmation		
		-	0	+
Expectation	Low	1.37	$3.09^{\{a\}}$	$4.26^{\{a\}}$
	High	$2.60^{\{b\}}$	$4.09^{\{a,b\}}$	$4.77^{\{a,b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

Satisfaction is expressed on a 1-to-7 scale

- low expectation with negative disconfirmation (LEN),
- low expectation with confirmation (LEC),
- low expectation with positive disconfirmation (LEP),
- high expectation with negative disconfirmation (HEN),
- high expectation with confirmation (HEC),
- high expectation with positive disconfirmation (HEP).

This approach allows the comparison of the different cell means with any of the various non-parametric one-way ANOVA techniques which can be used if the assumption are violated. The downside of this approach is that differences between factor level means, e.g. mean satisfaction for low expectation vs mean satisfaction for high expectation, and interaction effects can no longer be statistically tested. However, the main purpose of this analysis is to compare the cell means rather than the factor level means or interaction effects.

Before the one-way ANOVA analysis was performed, its assumptions were tested. The normality of the residuals were tested by means of a one-sample Kolmogorov-Smirnov Test, both asymptotically and with a Monte Carlo experiment based on 10000 tests. The constancy of the error variances was tested with Levene's Test of equality of error variances. If both assumptions were met, an F-test is performed to test if the cell means differ in a statistically significant way and the Tuckey multiple comparison procedure was used as a post-hoc test to verify the difference between any combination of two cell means. If one of both assumptions were violated and no transformation provided any help, the non-parametric Kruskal-Wallis test was used instead of the F-test and the Games-Howell test was used for the post-hoc analysis of all combinations of two cell means. The Games-Howell test is a non-parametric alternative for the Tuckey procedure which was recommended above other non-parametric tests by Toothaker [125] if the number of cases per cell are 6 or more.

If the F-test or the Kruskal-Wallis test indicated a significant difference between the cell means, the following cell differences were tested with the Tuckey test or the Games-Howell test.

- $\Delta_1 = LEC - LEN$
- $\Delta_2 = LEP - LEC$
- $\Delta_3 = HEC - HEN$
- $\Delta_4 = HEP - HEC$
- $\Delta_5 = HEN - LEN$
- $\Delta_6 = HEC - LEC$
- $\Delta_7 = HEP - LEP$

The results of the one-way ANOVA analyses are integrated in Tables B.1, B.2 and B.3 in Appendix B and in Table 5.8. If the mean satisfaction of a cell differs statistically significant from the mean satisfaction of the cell with the same expectation level but a lower disconfirmation level, it receives superscript *a*. If the mean satisfaction of a cell differs statistically significant from the mean satisfaction of the cell with the same disconfirmation level but a lower expectation level, it receives superscript *b*.

According to the ED paradigm and confirmed by the results found by Oliver [95], the tested cell differences should be significantly greater than zero. In total, 156 of the 161 comparisons have the correct sign, and according to the results of the various ANOVA analysis, 106 of these comparisons were statistically significant (cf Tables B.1, B.2, B.3 in Appendix B and 5.8). This implies that the D-LIED framework reproduces Table 5.7 such that 97% of the cell comparisons have the correct sign and 66% of these comparisons are statistically significant. The D-LIED framework models CS/D such that both an increase in disconfirmation, ceteris paribus, or an increase in expectation, ceteris paribus, results in a statistically significant increase in satisfaction, as can be expected from the ED paradigm. The fact that the D-LIED framework produces results which behave as expected with regard to other constructs such as satisfaction provides additional empirical validity.

The results in Table 5.8 show that all comparisons with the average disconfirmation were statistically significant. This strengthens the believe that the average disconfirmation could be used as a proxy for product disconfirmation. Finally, note that none of the 6 comparisons which had an incorrect sign were statistically significant.

5.4 Case studies

The body of evidence suggesting that the D-LIED implementation produces valid approximations for customer expectation and customer disconfirmation from limited information is substantially large, both theoretically and empirically, justifying the use of the D-LIED implementation on customer satisfaction data. However, it should never be forgotten that the D-LIED implementation is an implementation of the LIED framework which is based on the ED paradigm. Therefore, the D-LIED implementation as well as any other LIED framework implementation implicitly assumes that the underlying CS/D process is captured by the ED paradigm. If a researchers or marketers are convinced that an other paradigm should be used to explain the CS/D process of their customers or product, the LIED framework should not be used.

In this section, two case studies will illustrate how new information can be extracted from the data with the D-LIED framework and how a company can benefit from it. Each case study assumes that the ED paradigm is the correct paradigm to model CS/D.

Case study 1

The first case study illustrates how univariate analyses of customer expectations can shed a new light on the performance of the company. This case study is applied on the Banking Product Data Set and uses the original hierarchical dimensions instead of the factorized dimensions because dimension expectations need to be calculated. Table 5.9 gives an overview of the limited information which is available before applying the D-LIED framework.

Table 5.9: Limited information available in banking product data set

Variable	Scale
Satisfaction	[1-10]
Image Performance	[1-5]
Price Performance	[1-5]
Quality Performance	[1-5]
Product Performance	[1-5]
Invoices Performance	[1-5]
Communications Performance	[1-5]

With this information at hand, the company can perform a simple descriptive and graphical analysis of their product's performance by averaging the dimension performances for all customers and plotting them as in Figure 5.3. Based on this plot for the Banking Product Data Set, the company arrives at the following conclusions:

- The *Price* dimension has a low performance rating, i.e. less than 2.5 on a scale from 1 to 5.
- Of all dimensions, *Price* performs much worse than the other dimensions and should receive full attention in the future.
- *Product Performance* and *Product Quality* are the best performing dimensions.
- The dimensions *Product's Image*, *Invoices* and *Communication* are performing above average.

This is where a traditional descriptive analysis would stop, but by applying the D-LIED implementation, the hierarchical structure of the data can be exploited and the data can be enriched with the customer expectations for each dimension. For each dimension, expectation e_i was estimated by means of Eq. 5.4 and 5.5 with the dimension performance p_i as a proxy for the dimension satisfaction and the attribute performances p_{ij} as input. This was possible for all dimensions except for *Quality*, which did not have any corresponding product attributes p_{ij}. The best possible estimate of the customer's expectation for this dimension is the general expectation level e which can be estimated with the D-LIED implementation by using the general satisfaction score and the various dimension performances p_i as input. Next, disconfirmation d_i was calculated for each dimension by subtracting the dimension expectation e_i from the dimension performance p_i. Table 5.10 shows the average performance, disconfirmation and expectation for all customers. This information is used to create Figure 5.4 which shows that some dimensions are performing much better than expected, while others are creating negative disconfirmation.

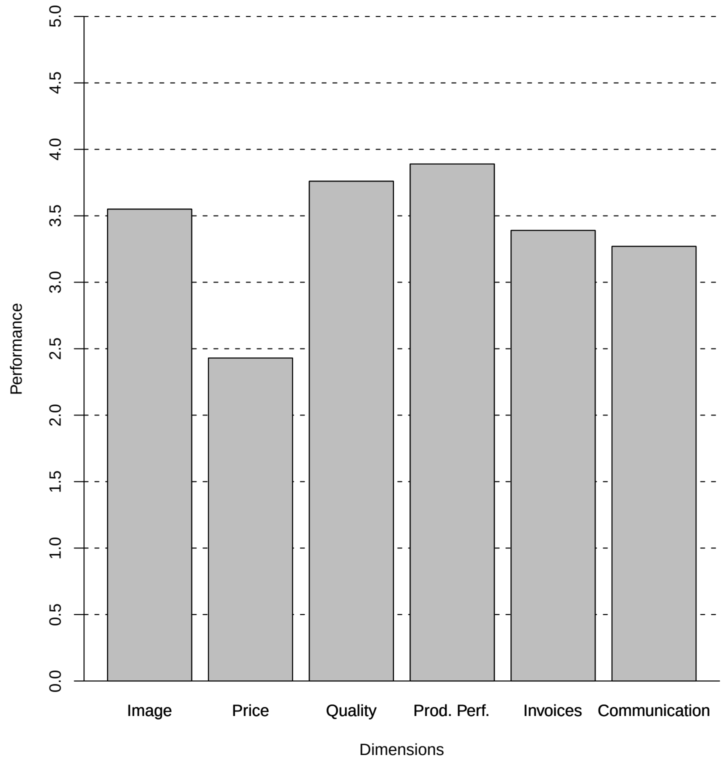


Figure 5.3: Banking product dimension performances.

The averages shown in Table 5.10 are sample means, which give only information about the surveyed sample of customers. The results in Table 5.10 and Figure 5.4 tell a manager for example that the customers they surveyed expected a better price on average. Most often, companies are not extremely interested in the surveyed sample, but much more in the entire population of customers. Although the sample mean provides an unbiased estimate of the population mean, additional information about the accuracy of the estimate is required to make inferences about the population. In this case study, the sample mean accuracy is expressed by means of 95% confidence intervals (CI).

According to [114], the single-sample t-test should be used to construct such confidence intervals since the true population variance is unknown. The $1-\alpha$ confidence interval can be constructed as follows:

$$CI_{1-\alpha} = \overline{X} \pm (t_{(\frac{\alpha}{2};n)})(\frac{\tilde{s}}{\sqrt{n}}) \tag{5.7}$$

where $\overline{X}$ represent the sample mean, $\tilde{s} = \sqrt{\frac{\sum(X-\overline{X})^2}{n-1}}$ represents an unbiased estimate of the population standard deviation, n represents the sample size and $t_{(\frac{\alpha}{2};n)}$

Table 5.10: Mean dimension performance, mean dimension expectation and mean dimension disconfirmation for banking product data set

Dimension	Performance	Expectation	Disconfirmation
Image	3.55	3.50	0.05
Price	2.43	2.52	-0.09
Quality	3.76	3.11	0.65
Product Performance	3.89	3.81	0.08
Invoices	3.39	2.78	0.61
Communications	3.27	3.34	-0.08

Performance and Disconfirmation are measured on a 1-to-5 scale

is the critical two-tailed value in the t distribution for $df = n - 1$, below which a proportion equal to $[1 - (\frac{\alpha}{2})]$ of the cases falls.

According to Sheskin [114], the t-test assumes that the underlying distribution of X is normal which was tested for each dimension disconfirmation d_i with a one-sample Kolmogorov-Smirnov Test, both asymptotically and with a Monte Carlo experiment based on 10000 tests. None of the dimension disconfirmation scores appeared to be normally distributed. Fortunately, for large samples the t-test is rescued by the central limit theorem which states that if the sample size n goes to infinity, the mean sampling distribution becomes normally distributed [88]. According to Cohen [29], any sample size larger than 30 can be considered as large. With 1038 records in the Banking Product Data Set, it is safe to use the t-test. Table 5.11 repeats the dimension disconfirmation, together with the 95% confidence intervals.

Table 5.11: Mean dimension disconfirmation with 95% CI for banking product data set

Dimension	Mean Disconfirmation	95% CI
Image Performance	0.05	$[0.00, 0.10]$
Price Performance	-0.09	$[-0.15, -0.04]$
Quality Performance	0.65	$[0.60, 0.70]$
Product Performance	0.08	$[0.03, 0.13]$
Invoices Performance	0.61	$[0.53, 0.69]$
Communications Performance	-0.08	$[-0.15, 0.00]$

Disconfirmation is measured on a 1-to-5 scale

Figure 5.4, which plots both dimension performance and dimension expectation, and Table 5.11 provide more insight into the company's performance and allow the companies to draw the following additional conclusions:

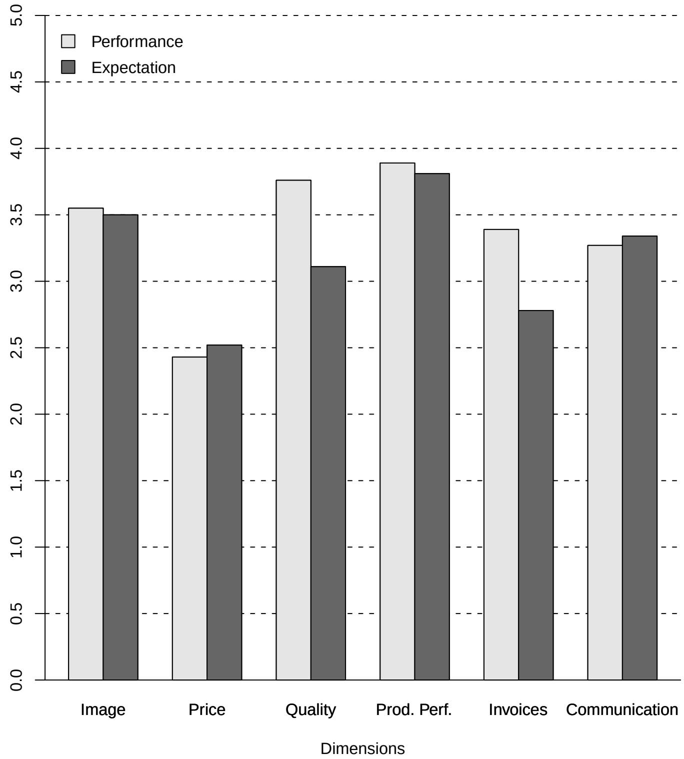


Figure 5.4: Dimension performances with expectations.

- The performance of *Price* is less dramatic than appeared to be during the descriptive analysis of product performance. *Price* is not performing very high and is performing slightly worse than expected, but the disconfirmation is limited because the customers are not expecting very much from this dimension. This sounds reasonable because the company has a quasi-monopolistic position in the market, which results in price-taking customers. These results reflect the fact that customers think that the price is not performing good, but they accept this situation because they cannot change anything.

- *Communication*, which was regarded as a good performing dimension seems to perform less than expected, causing negative disconfirmation. However, Table 5.11 indicates that the disconfirmation could also be 0.

- *Product Performance* appeared to be the overall best performing dimension in the descriptive performance analysis, but dimensions *Quality* and *Invoices* succeed much better in exceeding the customer's expectation. It would be wrong to decrease effort on *Product Performance* since it performs only slightly better than expected. If the company has to decrease focus on one of these dimensions, they should probably choose the *Invoices* dimension. Even if performance would drop for this dimension, the company would still be performing better than expected.

Comparing these conclusions with the previous ones illustrates that the D-LIED implementation adds new and important insights to the descriptive performance analysis.

Case study 2

The previous case study illustrated how a traditional analysis of descriptive performance statistics could be extended with a descriptive analysis of expectation and disconfirmation. These univariate analyses are a good starting point for studying the customer satisfaction process, but provide only limited insights in the interpretation of the satisfaction data. Researchers wishing to go beyond the univariate analysis of descriptive satisfaction statistics, typically calculate the statistical impact of the individual dimension performance scores on the overall satisfaction score with a regression analysis [97]. By studying the regression coefficients, the CS/D *key drivers* can be identified.

Oliver discussed that analyzing the CS/D process by means of only performance and satisfaction scores can be both correct and incorrect. Regressing satisfaction on the dimension performances is a first step in gaining insights about the levels and potential causes of satisfaction. Aided with expert interpretation of the results and follow-up validation, *nothing* may be wrong with this approach. However, even if the regression results identify problematic product dimensions, it will fail to explain why particular dimensions are a problem (or a benefit) for the consumer. That is, the traditional regression approach ignores the psychological process which customers use to interpret the experienced levels of performance. Fortunately for the numerous firms which conduct surveys in the traditional manner, consumers do bring their psychological interpretations to performance ratings. At the same time, traditional regression analysis is unable to reveal these interpretations [97].

In this case study, we will illustrate how expectations and disconfirmations, extracted from the data by means of the D-LIED implementation, can extend the traditional regression analysis and provide insights in the psychological process underlying the customer's interpretation of the experienced levels of performance. The case study uses the Energy Data Set which contains customer satisfaction data for two competitors. Within this case study, two different regression models will be estimated for each company.

The first regression analysis uses the satisfaction score as the dependent variable and the eight dimension performance scores as the explanatory variables. The dimensions are: *availability* (p_1), *employees* (p_2), *problem solving* (p_3), *information* (p_4), *service* (p_5), *invoices* (p_6), *price* (p_7), *product quality* (p_8). The first regression model (cf Eq. 5.8) has 9 regression coefficients. Coefficient α_0 is the intercept term which can be interpreted as the basic level of satisfaction, i.e. the satisfaction when all the performance scores are equal to zero. In theory this is the minimum level of satisfaction, unless some product dimensions have a negative influence on product satisfaction, which would not make sense. Coefficients β_1 to β_8 measure the influence of a unit increase in dimension performance on satisfaction. These coefficients allow the researcher to identify dimensions which have a statistically significant influence on satisfaction and to distinguish between less influential dimensions and real key satisfaction drivers. Note that the hierarchical dimension scores are used instead of the factorized dimension scores. Although the latter type of dimension scores are multi-item scores which are less sensitive to measurement errors, the hierarchial dimension scores were used for a correct comparison with the next regression model which incorporates dimension expectations learned at the dimensional level by the D-LIED implementation. Also keep in mind that the performance and satisfaction scores have been transformed to the $[0, 1]$ interval by dividing the satisfaction score by 11 and the performance scores by 6. This results in satisfaction scores between 0.09 and 0.91 and performance scores between 0.17 and 0.83. This regression model is referred to as the performance regression model.

$$\text{Satisfaction} = \alpha_0 + \sum_{i=1}^{8} \beta_i p_i \tag{5.8}$$

The second regression model, shown by Eq. 5.9, divides each performance score in an expectation score and a disconfirmation score. The expectations are learned at the dimensional level except for dimensions *price* and *product quality*, which do not have corresponding product attribute performances. For the latter two dimensions, the average expectation of the other dimensions was used as a best estimate of the dimension expectation. Using expectation learned at the general level as an estimate for the *price* and *product quality* expectation, as is done in the previous case study, is not possible. Learning expectation at the general level is a function of general satisfaction and dimension performance. Using this very construct in a regression with general satisfaction as the dependent variable leads to a circular path of modeling. The disconfirmations in this second regression model are the dimension expectations subtracted from the dimension performances. The final regression model does not incorporate the expectation levels for dimensions *price* and *product quality* because this would lead to perfect collinearity between the expectation variables.

$$\text{Satisfaction} = \alpha_0 + \sum_{i=1}^{8} \beta_i d_i + \sum_{i=1}^{6} \gamma_i e_i \tag{5.9}$$

The second regression model contains 15 regression coefficients. Coefficients β_1 to β_8 measure the impact on customer satisfaction of a unit increase in dimension disconfirmation, while coefficients γ_1 to γ_6 measure the impact of a unit increase in dimension expectation. The difference with the previous regression is that dimension performance is divided in an expectation element and a disconfirmation element which reveals the psychological interpretation of the customers. While the performance regression might identify a dimension as having a large impact on satisfaction, it cannot tell if companies should try to keep expectations equal and create positive disconfirmation or if companies should try to increase the customer's expectations and keep disconfirmation equal. With this type of regression, such conclusions are possible. Only the expectation variables for dimensions *price* and *product quality* are missing. Instead, the intercept term α_0 can be interpreted as the average impact of the expectations for these two dimensions on the customer's satisfaction. This regression model is referred to as the disconfirmation regression model.

Traditionally, these regression models are conducted as ordinary least squares (OLS) multiple linear regression analyses. The benefit of a linear regression model is its parsimony and its ease of interpreting the coefficients, which can be important during a presentation to people unfamiliar with statistics. On the other hand, a linear model also assumes that the impact of performance, disconfirmation or expectation is independent of the current level of performance, disconfirmation or expectation. Some authors [90] question the validity of this assumption but in this case study, it was chosen to build a linear model to achieve an easy interpretation of the results while keeping in mind that the model might only approximate the possible non-linear structure of the underlying process.

A multiple linear regression also relies on a set of assumptions to estimate unbiased coefficients and efficient and unbiased confidence intervals. One of these assumptions is the fact that the error terms are normally distributed with a mean equal to zero and a constant variance σ^2, which is referred to as the normality of the error terms. A second and important assumption is that the variance of the error terms is equal for any value of the explanatory variables, which is referred to as the equal variance or homoscedasticity assumption. [93, 58]

Both assumptions were tested for all four regression analysis. Testing the normality of the error terms is done by checking the normality of the residuals by means of a quantile comparison plot. The quantile comparison plots in Figure 5.5 plot the studentized residuals against their expected value under a t-distribution with $(n - k - 2)$ degrees of freedom [2]. The plots also show a 95 percent pointwise confidence envelope for the studentized residuals, using a parametric version of the bootstrap. The method used is from Atkinson [9, 53]. If the residuals follow a normal distribution, the studentized residuals should align with the diagonal through the origin. These plots reveal that the residuals are heavily skewed to the left for all four regression models, suggesting non-normal error terms.

To test the homoscedasticity assumption, the Breusch-Pagan test [17] was employed which calculates a test statistic X^2_{BP} which follows approximately the chi-quare distribution with one degree of freedom under the assumption of homoscedasticity [93, 53]. Table 5.12 shows the results of the Breusch-Pagan test for all four regression model and the results are highly significant, strongly suggesting heteroscedastic error terms.

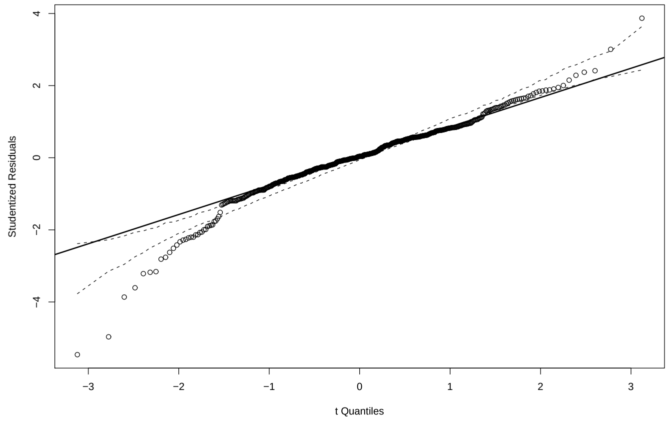
Table 5.12: Results of the Breusch-Pagan test for the energy data set

	Performance Reg.		Disconfirmation Reg.	
Company	1	2	1	2
X^2_{BP}	30.03	49.70	29.01	44.71
p-value	< 0.01	< 0.01	< 0.01	< 0.01

These results indicate that two important assumptions of the OLS linear regression model are violated which can lead to unreliable regression coefficients. Furthermore, Figure 5.5 reveals that several observations in all four models have residuals which

[2] n refers to the number of cases, k refers to the number of parameters.

(a) Performance regression model: company 1



(b) Performance regression model: company 2

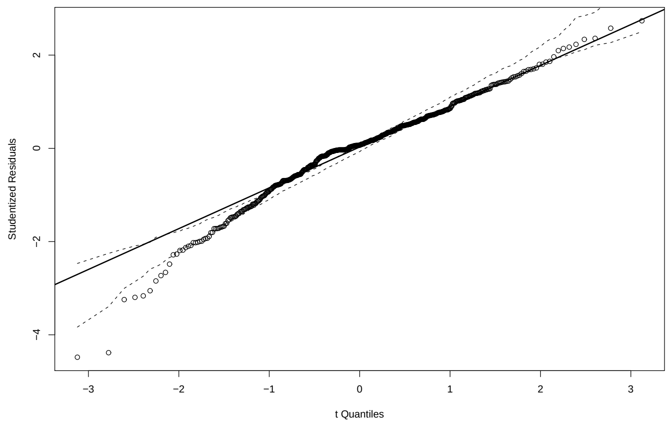


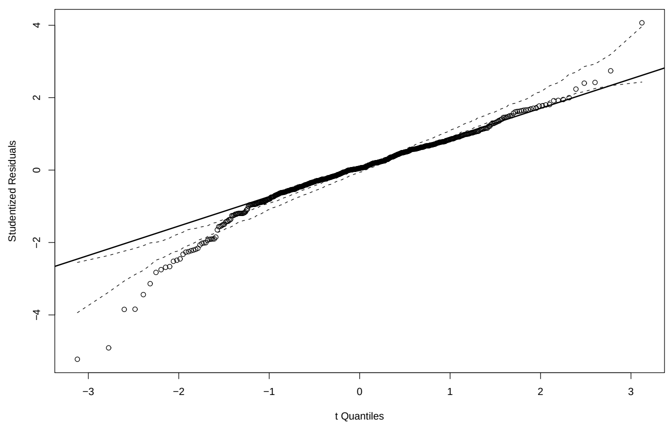
Figure 5.5: Quantile-comparison plots of the studentized residuals for the energy data set

are much lower than expected. These observations can be considered as outliers and can wreak havoc with least squares estimates.

All these results suggest that the data is not appropriate for conducting multiple linear regression. One could try various transformations of the dependent and independent variables to resolve the issues of the non-normal and heteroscedastic residuals, but even if a proper transformation was found, the ease of interpretation of the regression coefficients would be lost. To limit the impact of outliers, one could omit these cases from the data set, but this would substantially decrease sample size and might omit important information from the data set.

Instead, a robust regression technique with non-parametric bootstrapping for the construction of confidence intervals was used. The iteratively reweighed least squares

(c) Disconfirmation regression model: company 1



(d) Disconfirmation regression model: company 2

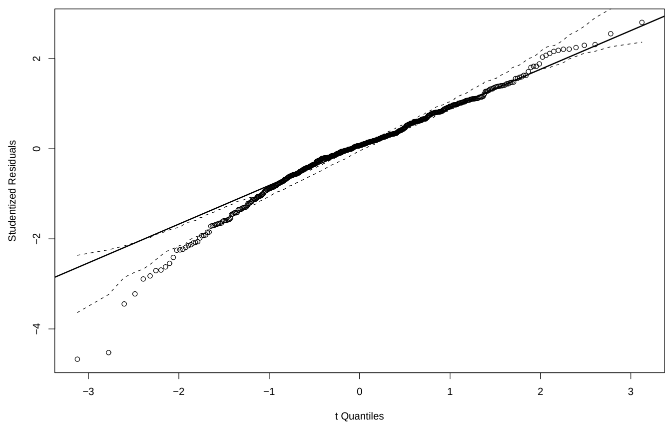


Figure 5.5: Quantile-comparison plots of the studentized residuals for the energy data set [continued]

(IRLS) robust regression technique was selected, which uses the weighted least squares procedures to dampen the influence of outlying observations [93]. The basic idea behind this approach is that observations with large residuals get small weights and observations with small residuals receive large weights. The weights are determined by the Huber weight function which is defined as follows:

$$w = \begin{cases} 1 & |u| \leq 1.345 \\ \frac{1.345}{|u|} & |u| > 1.345 \end{cases} \tag{5.10}$$

with u denoting the scaled residual. On Figure 5.6, which shows the Huber weight function, it can be seen that any observation with a scaled residual between -1.345

and 1.345 receives a maximum weight of 1, while observations with scaled residuals larger than 1.345 or smaller than −1.345 receive lower weights. The value of 1.345 is called the tuning constant and is chosen such that the IRLS robust procedure is 95 percent efficient for data generated by the normal error regression model [93].

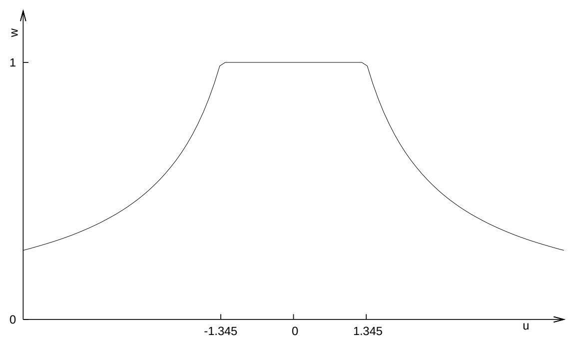


Figure 5.6: Huber weight function.

The advantage of the robust regression method is that it is less sensitive to outliers, but the drawback of this approach is that the evaluation of the precision of the estimated regression coefficients is more complex. To calculate 95 percent confidence intervals, a non-parametric bootstrap technique must be used. Bootstrapping, which was developed by Efron [41], is based on creating new bootstrap samples S_b^* by sampling with replacement from the original sample S until the new sample size is equal to the original sample size. This process is repeated a number of times which results in a set of B bootstrap samples, denoted as $\{S_1^*, \ldots, S_b^*, \ldots, S_B^*\}$. The key idea is that all these bootstrap samples can be considered as samples from the unknown population (or at least they look like the unknown population).

Next, the regression coefficients are estimated for each bootstrap sample, which are denoted as $\widehat{\beta}_{ib}^*$, while the regression coefficients estimates for the original sample is denoted as $\widehat{\beta}_i$. According to the bootstrap theory, the sampling distribution of the regression coefficient estimate $\widehat{\beta}_i$ around the true regression coefficient β_i, which is needed to construct confidence intervals, is analogous to the distribution of $\widehat{\beta}_{ib}^*$ around $\widehat{\beta}_i$. Thus, by making B very large, the bootstrap approach can construct a detailed empirical distribution of $\widehat{\beta}_{ib}^*$ around $\widehat{\beta}_i$ which serves as an estimate of the true sampling distribution of $\widehat{\beta}_{ib}^*$ [52]. This estimate of the sampling distribution can be used to construct confidence intervals around $\widehat{\beta}_i$. Various approaches to construct bootstrap confidence intervals exist, such as the *normal-theory interval*, the *bootstrap percentile interval* and the *bias-corrected, accelerated percentile intervals* (BC_a). The first approach assumes a normally distributed bootstrap sampling distribution of $\widehat{\beta}_{ib}^*$, while the latter two use the quantiles of the bootstrap sampling distribution to establish the end points of a confidence interval non-parametrically. Since the BC_a confidence intervals are an adjustment of the percentile confidence intervals, the BC_a

approach was used to construct 95% CI. In this case study, the confidence intervals are based on 1000 bootstrap samples.[3]

By using a robust regression technique with bootstrapped confidence intervals, various problems encountered by OLS multiple regressions are avoided. Firstly, the regression technique is robust and consequently less sensitive to outliers. Secondly, non-normality and heteroscedastic regression coefficients pose no problem for nonparametric bootstrapping to construct 95 percent confidence intervals. For a more detailed discussion on robust regression and bootstrapping, the interested reader can refer to [93, 52, 42].

Table 5.13: Coefficients of the performance regression model and the disconfirmation regression model for company 1 in the energy data set

	Perf. Regression		Disc. Regression	
	Coef.	95% CI	Coef.	95% CI
α_0 (Intercept)	0.35	$[0.28, 0.41]$	0.35	$[0.28, 0.41]$
β_1 (Availability)	0.03	$[-0.02, 0.1]$	0.01	$[-0.07, 0.08]$
β_2 (Employees)	0.02	$[-0.13, 0.14]$	-0.03	$[-0.16, 0.09]$
β_3 (Problem Solving)	0.22	$[0.12, 0.34]$	0.10	$[-0.02, 0.25]$
β_4 (Information)	-0.02	$[-0.12, 0.07]$	-0.02	$[-0.11, 0.08]$
β_5 (Service)	0.12	$[0.00, 0.24]$	0.09	$[-0.02, 0.22]$
β_6 (Invoices)	-0.02	$[-0.09, 0.07]$	-0.03	$[-0.11, 0.05]$
β_7 (Price)	0.19	$[0.12, 0.26]$	0.20	$[0.13, 0.26]$
β_8 (Product Quality)	0.07	$[-0.01, 0.15]$	0.05	$[-0.03, 0.14]$
γ_1 (Availability)			0.11	$[0.03, 0.20]$
γ_2 (Employees)			0.07	$[-0.07, 0.21]$
γ_3 (Problem Solving)			0.30	$[0.18, 0.42]$
γ_4 (Information)			-0.07	$[-0.19, 0.04]$
γ_5 (Service)			0.17	$[0.04, 0.31]$
γ_6 (Invoices)			0.03	$[-0.06, 0.11]$

Coefficients significant at 5% are marked in grey

The coefficient estimates of both the performance regression model and the disconfirmation regression model for company 1 are presented in Table 5.13. In order to interpret these results, it is important to have some background information about this company. Company 1 used to have a monopolistic position before the energy market was liberalized and is still the biggest player on the market. Many customers have established a long time relationship with this company which might explain why so few people switched companies after the liberalization of the energy market.

[3] The same reasoning holds for the other regression coefficients α_0 and γ_i.

In general, the results of both regression models support Oliver's conjecture that performance-based approaches can identify the key satisfaction drivers but fail to reveal the psychological process underneath. Table 5.13 illustrates that the performance regression model identifies dimensions *Problem Solving*, *Service* and *Price* as key drivers. Since companies in this market are not really selling a tangible product, *Service* and *Problem solving* could be considered as their core business activities. The fact that the core business activities are identified as the key drivers of satisfaction is in line with the findings of Parasuraman et al. [102]. In addition to these three key drivers, the disconfirmation model succeeds in identifying one additional key driver, i.e. the *Availability* dimension. The fact that the disconfirmation model finds the same key drivers as the performance model furthermore validates the results of the disconfirmation model. All the significant coefficients are positive as can be expected from the theoretical relationship between satisfaction on one hand and performance, disconfirmation and expectation on the other hand.

Up to this point, the disconfirmation model merely confirms the conclusions of the performance model. However, in contrast to the performance model, the disconfirmation model does offer an explanation of the CS/D process. According to the results in Table 5.13, most key drivers, except *Price*, have an influence on satisfaction only through expectation. For example, customers with a positive disconfirmation on the dimension *Problem Solving* will not have a higher satisfaction score on average than customers with a neutral disconfirmation level. On the other hand, customers with a higher expectation on the dimension *Problem Solving* will also have a higher satisfaction level. Customers are only sensitive to disconfirmation for the dimension *Price*. These results demand some explanation.

Firstly, non-significant coefficients do not imply the absence of a real effect on satisfaction. It only implies that there is not enough statistical evidence to reject the null hypothesis of no effect. The 95 percent confidence intervals of β_3 (*Problem Solving*) and β_5 (*Price*) for the disconfirmation model fall almost completely into the positive range, suggesting a positive effect on satisfaction. On the other hand, even if there is a positive effect on satisfaction, it will be relatively small.

Why disconfirmation has almost no effect on satisfaction can be explained by the fact that company 1 used to be the only player on the market and almost all its customers have a long history with this company, making them less sensitive to disconfirmation (almost insensitive). Oliver [96] suggested that expectation acts as an adaptation level based on previous experience. Given the long history of experiences with this company, it is reasonable to assume that customers have built a very strong adaptation level which discards almost any disconfirmation experienced. Customers are only sensitive to disconfirmation on the dimension *Price*, which can be an effect of the liberalization of the market, making customers more price-aware and weakening their adaptation level for this dimension.

Company 2 is a new player which entered the energy market directly after the liberalization. This company has more customers who consciously selected their company, which explains why both regressions identified more key drivers of satisfaction compared with company 1. Similar to company 1, the core business activities, i.e. *Problem Solving* and *Service*, are identified as key drivers by the performance model, as could be expected from the results of [102]. Additionally, the performance model also identified dimensions *Invoices*, *Price* and *Product Quality* as key drivers. Similar to company 1, the disconfirmation model identified one additional

Table 5.14: Coefficients of the performance regression model and the disconfirmation regression model for company 2 in the energy data set

	Perf. Regression		Disc. Regression	
	Coef.	95% CI	Coef.	95% CI
α_0 (Intercept)	0.34	$[0.23, 0.42]$	0.35	$[0.26, 0.44]$
β_1 (Availability)	-0.07	$[-0.17, 0.03]$	-0.03	$[-0.12, 0.09]$
β_2 (Employees)	0.00	$[-0.15, 0.15]$	0.02	$[-0.14, 0.17]$
β_3 (Problem Solving)	0.27	$[0.13, 0.41]$	0.21	$[0.08, 0.34]$
β_4 (Information)	0.10	$[-0.01, 0.21]$	0.15	$[0.03, 0.27]$
β_5 (Service)	0.19	$[0.04, 0.35]$	0.14	$[0.01, 0.29]$
β_6 (Invoices)	0.18	$[0.10, 0.27]$	0.13	$[0.05, 0.23]$
β_7 (Price)	0.10	$[0.03, 0.19]$	0.13	$[0.05, 0.21]$
β_8 (Product Quality)	-0.18	$[-0.29, -0.05]$	-0.15	$[-0.26, -0.03]$
γ_1 (Availability)			-0.11	$[-0.22, 0.01]$
γ_2 (Employees)			-0.16	$[-0.35, 0.01]$
γ_3 (Problem Solving)			0.35	$[0.16, 0.51]$
γ_4 (Information)			0.01	$[-0.10, 0.13]$
γ_5 (Service)			0.25	$[0.10, 0.43]$
γ_6 (Invoices)			0.23	$[0.14, 0.33]$

Coefficients significant at 5% are marked in grey

key driver, i.e. *Information.* The latter key driver was not a key driver for company 1, which can be explained by the fact that customers of company 1 have a long history with company 1 which makes quality of information less important than for customers of a new company.

All coefficients were positive as was expected, except for dimension *Product Quality.* The performance model suggests that if product quality increases, customer's satisfaction will drop. The disconfirmation model suggests that if disconfirmation of product quality performance becomes more positive, the satisfaction will drop. Both conclusions are contradicting with the existing CS/D theory and can not be explained logically. Since both models find this result, company 2 should perform further analysis and gather data more to discover what is going on.

The disconfirmation model reveals that customers with higher expectations on dimensions *Availability*, *Problem Solving* and *Service* are also more satisfied. Additionally, the disconfirmation model also reveals that customers of company 2 are much more sensitive to disconfirmation than customers of company 1, which can be explained by the fact that customers of company 2 do not have the same long history with their company as customers of company 1.

The results of the analysis on both companies clearly illustrate that performance regression only provides part of the story behind customer satisfaction which could lead to incorrect decisions. For example, the performance model identified *Problem Solving* and *Service* as key drivers, which could lead to an increase of effort on these two dimensions. However, the disconfirmation model showed that any additional positive disconfirmation created on these dimensions has almost no effect on satisfaction. Apparently, the customer's expectation of both dimensions have a much larger impact on the customer's satisfaction and companies could perhaps focus on raising the expectation levels. If company 1 wants to increase satisfaction by putting more effort in one of the dimensions, it should try to improve the performance of the *Price* dimension, which is much more sensitive to an increase in positive disconfirmation.

Furthermore, the disconfirmation model also revealed an interesting difference between the customers of both companies, which remained hidden with the performance regression model: while the performance model already showed that company 2 has more key drivers than company 1, it failed to show that customers of company 2 are much more sensitive to disconfirmation than customers of company 1.

5.5 Summary

In this chapter, the D-LIED implementation of the LIED framework was introduced, which is based on Dombi's evaluation operator. It was shown that this aggregation function provides a valid implementation of the LIED framework and that the neutral element of the aggregation function acts as an estimate for the customer's expectation level. It was also illustrated that the D-LIED implementation can be applied on two different levels in case of hierarchical data, i.e. the general and the dimension level. At the general level, the D-LIED implementation provides an overall expectation score and at the dimension level, the D-LIED implementation extracts an expectation score per dimension. With the customer's expectation, one can estimate the customer's disconfirmation by subtracting the expectation from the performance score.

Next, both the theoretical and empirical validity of expectation and disconfirmation, extracted with the D-LIED framework, were evaluated. The empirical validity encompassed both existing and new research. The results of these validation analyses showed that the expectation and disconfirmation learned from the data with the D-LIED implementation could be used in customer satisfaction analysis. Two different case studies showed how the new information can be imported directly into traditional satisfaction analysis. These case studies gave a first impression of what is possible with the D-LIED implementation and the next two chapters will continue to illustrate the potential of this implementation.

6 Application 1: importance-performance analysis

6.1 Introduction

Importance-Performance Analysis (IPA) is a graphical analysis of the product attributes (or dimensions) to provide insights on necessary and corrective business decisions. IPA calculates the importance and average performance for each product attribute (or dimension) and plots each attribute on a two-dimensional grid, which is illustrated by Figure 6.1.

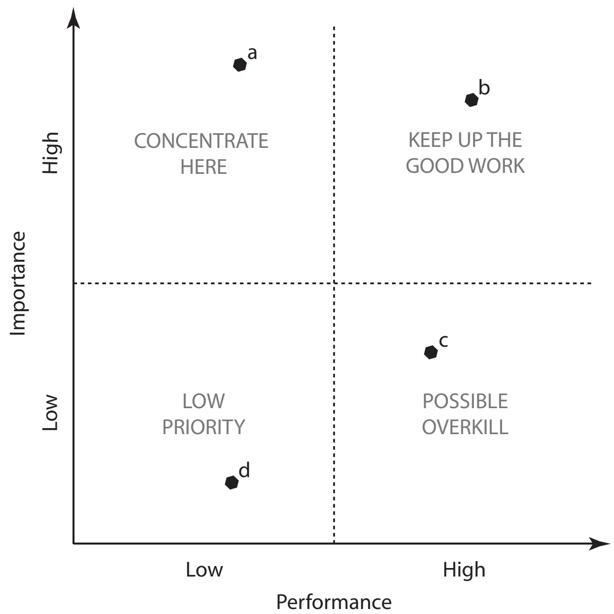


Figure 6.1: The importance-performance analysis.

The X-axis represents the attribute's performance and is divided in a *Low Performance* part and a *High Performance* part. The Y-axis represents the attribute's importance and is also divided in a *Low Importance* part and a *High Importance* part. In total, the IPA plot is divided in four quadrants. Figure 6.1 shows four different product attributes, each in a different quadrant. Product attribute a is situated in the upper left quadrant which is called the *Concentrate Here* quadrant.

Product attributes in this quadrant are highly important but are performing bad and need immediate attention. Product attribute b, which is situated in the *Keep up the Good Work* quadrant represents a product attribute of high importance which is performing well. Product attribute c is situated in the lower right quadrant which is called the *Possible Overkill* quadrant because the attributes in this quadrant are performing above average but have low importance. The fourth quadrant, which contains attribute d, is called the *Low Priority* quadrant because it contains attributes which are not performing well but are not very important either.

The IPA technique, originally introduced by Martilla and James [82], is particularly easy to interpret which makes it an attractive manager's tool for developing marketing strategies and a useful technique for evaluating the elements of a marketing program. Due to its strengths and managerial attractiveness, IPA has been adopted in other domains besides marketing strategy development. Slack uses IPA as an important part of operations strategy formulation [115], but IPA has also been used as an analysis tool in the field of customer satisfaction management [84, 83]. Sampson et al. [110] used IPA as part of a service quality improvement program of a school's district Food Services Department. Without question, IPA has become a frequently applied analytical technique in various domains.

Recently, research publications in the field of customer satisfaction in general and Importance-Performance Analysis in particular, revealed some hazardous pitfalls and complications when measuring the product attribute's importance . In this chapter, these problems and their consequences are discussed for the two most commonly used techniques and a new approach, based on the D-LIED implementation, is developed. The next section will discuss the evolution of the IPA technique as well as the recent developments which question the measurement and interpretation of the importance figures. The third section will elaborate on the two most commonly used techniques for measuring attribute importance, while the fourth section introduces a new method to measure importance figures. Finally, the possible problems of the traditional technique are tested empirically and a case study illustrates the application of the D-LIED approach.

6.2 Evolution of importance-performance analysis

In their original work [82], Martilla and James emphasized that the main contribution of IPA does not lie in the absolute positioning of the attributes on the grid, but in the prioritization of the attributes, which is based on the relative positioning. In their work, they implicitly assumed that importance and performance are independent.

Over the years, different authors have suggested several extensions of the basic IPA model, which pertained to conceptualization and measurement of attribute performance [110], but the performance-importance independence assumption remained intact. One of the first authors to question this assumption was Nigel Slack [115], who suggested that attribute prioritization was a function of attribute performance interpreted in light of attribute importance. Sampson et al. [110] continued this line of reasoning and argued that a causal relationship had to exist between attribute importance and attribute performance. They assumed that attribute importance was negatively correlated with attribute performance, i.e. attribute importance increases when attribute performance drops.

A similar result was found by Mittal et al. [90], who investigated the relationship between attribute performance and their impact on overall satisfaction. They showed that a negative attribute performance had a greater impact on overall satisfaction, than a positive attribute performance. Although their research pertains to the field of customer satisfaction, it can be linked to Importance-Performance Analysis, because the impact on overall satisfaction is often used as a proxy for the relative attribute importance [84].

Both Mittal et al. and Sampson et al. assumed a negative correlation between attribute performance and attribute importance. Their findings largely confirmed this hypothesis but some anomalies in their results indicated that there was still a part that remained unexplained. Sampson et al. noticed that the Performance-Importance Response Function was not linearly descending, but had a convex pattern, curving back up at the highest attribute performance scores. Mittal et al. discovered that one attribute showed a completely opposite pattern than expected. This attribute became more important as performance increased.

Matzler et al. addresses these peculiarities by assuming a causal relationship between attribute performance and attribute importance, without specifying the sign of correlation [83]. They used Kano's three-factor theory [73] and Oliver's ED paradigm [96] to explain the fact that not all attributes show the same sign of correlation between attribute performance and importance.

According to Kano, three different types of attributes can be discerned: basic factors, performance factors and excitement factors. Basic factors cause dissatisfaction if not fulfilled but do not lead to customer satisfaction if fulfilled. These, attributes show a negative correlation between attribute performance and attribute impact. Performance factors lead to satisfaction if performance is high and to dissatisfaction if performance is low. These attributes have the same importance when performance is low as when performance is high. Finally, excitement factors increase satisfaction if performance is high, but do not cause dissatisfaction if they are not fulfilled. These attributes show a positive correlation between attribute performance and attribute impact. [83]

Evidence is accumulating that the original assumption of performance-importance independence is incorrect, which calls the applicability of current IPA analysis into question [83]. Particularly, the determination of an attribute's importance can be problematic and unreliable. The next section discusses the two most commonly approaches to determine attribute importance in detail.

6.3 Traditional approaches to measure attribute importance

Direct measurement

While attribute performance is always measured directly by means of a customer survey, the measurement of attribute importance scores is not always trivial [110]. One of the most straightforward methods to measure attribute importance is by surveying the customers. A variety of measurement scales exist for attribute importance: e.g. direct rating, constant-sum scale, anchored scale and partial ranking among others. According to a study by Griffin and Hauser [57], there are no significant differences between these different measurement approaches. Their results were also confirmed by Matzler et al. [84].

One of the advantages of measuring attribute importance directly by means of a survey is the level of information obtained. The importance is measured at the level of a single respondent, providing information at the micro level. The data can always be aggregated if information is needed at a higher level, such as the average attribute importance for the entire population. The direct measurement approach also seems to offer extensive control over the validity of the measures, because the researcher can choose the appropriate measurement scale and question formulations. However, both Sampson et al. [110] and Matzler et al. [84] argue that the existing direct measurement scales for attribute importance are less valid than they appear to be.

Sampson et al. claim that the survey question 'How important is this attribute?' will be interpreted as 'How important is this attribute at the current level of attribute performance?'. They postulate that attribute importance depends on attribute performance level, which reflects in the answers of the respondents. This assumption is supported by their results, which show a negative correlation between attribute importance and attribute performance.

Also Matzler et al. question the validity of the directly measured importance scores, but do not agree with Sampson et al. on the source of the problem. According to Matzler et al., the respondent's self-stated importance does not reflect the true performance-importance relationship because these measures do not express the importance of the current performance level on customer satisfaction. Instead, customers interpret importance as the relative importance of an attribute compared to other attributes, irrespective of the current performance level. Consequently, basic factors will be rated as most important, performance factors as less important and excitement factors as the least important. Their empirical results do show a substantial difference between the directly measured importance scores and the statistically derived importance scores.

Furthermore, when collecting attribute importance scores by direct measurement, it is important to make sure that the importance related survey answers are not contaminated by the performance related survey answers. Martilla and James warn about this pitfall in their original work [82] and suggest to separate the importance measures and the performance measures to minimize compounding and order effects.

Finally, direct measurement of attribute importance can also pose a practical problem. Adding questions of attribute importance to a survey which also contains the measurements of attribute performance makes the survey more elaborated which could lead to a more costly survey or a survey which takes too much time from the respondents. Secondly, it is possible that the IPA analysis has to be performed with secondary data, i.e. data originally collected for other research, which does not contain attribute importance data. In such situations, the researcher is forced to use other approaches for measuring attribute importance.

Regression based measurements

Several articles on IPA and customer satisfaction analysis, describe the use of linear regression, with satisfaction as the dependent variable and attribute performances as the independent variables, to determine the attribute importance scores [90, 83, 37]. The regression coefficients measure the impact of the attributes on the overall satisfaction, which can be used as proxies for the attribute importance.

This indirect approach of measuring attribute importance has several advantages. Firstly, it allows researchers to use secondary data and to derive the attribute importance scores from customer satisfaction data. Furthermore, linear regression is a well-known technique among researchers and is easy to apply. Linear regression also allows researchers to build confidence intervals and to test statistical significance to increase the validity of their results.

On the other hand, linear regression comes with a large set of assumptions which should always be tested. One of these assumptions is that the model is specified correctly. As shown in the previous chapter, a regression model with satisfaction and attribute performance fails to capture the ED paradigm which is considered to be a valid approximation of the CS/D process. Even if the performance regression is assumed to hold for the available data, various other problems might occur with the linear regression model.

The performance regression model assumes that the attribute's impact on customer satisfaction can be represented as a point-estimate which is contradicted by results from Matzler et al. [83], Mittal et al. [90] and Sampson et al. [110]. Even if confidence intervals were to be used in an IPA, the performance regression provides a single importance estimate per attribute. This implies that the importance of an attribute is independent of the attribute's performance level which is not true according to several authors [110, 84]. Therefore, the coefficients only capture a fraction of the true impact on satisfaction and the remaining part hides inside the error term, making it heteroscedastic. The danger of heteroscedasticity in a performance regression was already mentioned by Sampson et al. [110], but they never tested it for their data.

The regression can be adapted to capture the dependence between attribute importance and attribute performance by creating a set of dummy variables for every attribute performance variable. Each set comprises one dummy variable per attribute performance level, except for the reference performance level. These dummy variables measure the impact of a specific attribute performance level on satisfaction. If the original attribute performance variable is continuous, this approach requires discretization first.

A side effect of this approach is the explosion of the number of explanatory variables. For example, if the original regression uses 4 attributes, all measured on a categorical scale from one to ten, the adapted regression will have 36 explanatory variables. This explosion of explanatory variables can cause a new problem, i.e. near micronumerosity. This phenomenon is closely related to near perfect multicollinearity and has the same symptoms. This problem generally arises when the number of observations barely exceeds the number of parameters to be estimated [58]. It is also possible to recode every explanatory variable into less dummy variables, but this makes interpretation of the regression coefficients more difficult and less useful.

Furthermore, past research also assumed that no interaction effects between attribute performance variables existed. This implies that the impact on satisfaction of one attribute is independent from the performance level of other attributes. For example, the impact of a performance increase of 10% on product quality has the same impact on satisfaction when all other attributes perform very high or when all other attributes perform extremely low. This is a rather unrealistic assumption and can be resolved by adding interaction effects to the regression. However, the researcher still has to make assumptions about the type of interaction and decide which attributes interact.

A final limitation of this technique is the fact that only information at the level of the entire population can be derived. The calculated attribute importance scores measure the attribute impact on satisfaction on average for the entire sample, in contrast with directly measured importance, which identifies the impact on satisfaction at the respondent's level.

6.4 Attribute importance measure derived with D-LIED implementation

The attribute importance derivation

The two traditional approaches to measure attribute importance, which are discussed in the previous section, are widely adopted in IPA research, but have their limitations. The direct measurement approach might suffer from incorrect interpretation of the questions and is impossible to use with secondary data. The indirect regression approach starts off wrong by applying a performance-based model which ignores the dominant ED paradigm and might suffer from violations of OLS regression assumptions making the results unreliable. This section proposes a new approach which starts from the D-LIED implementation, allowing a much better modeling of the CS/D process. This D-LIED approach only requires satisfaction and performance data which makes it perfectly suitable for secondary customer satisfaction data sets with limited information.

This new approach starts by applying the D-LIED implementation at the general level to learn the general expectation e_k for each customer k (cf Eq. 5.5 and 5.4). Next, Eq. 5.5 can be used to find parameter α_k for each customer k, which can be plugged into the general modeling function of the D-LIED implementation (cf. Eq. 5.3). This provides a unique customer satisfaction modeling function for each customer, i.e.

$$U_k(x_{1k}, x_{2k}, \ldots, x_{nk}) = \frac{\prod_{i=1}^{n} x_{ik}}{\prod_{i=1}^{n} x_{ik} + \alpha_k^{n-1} \prod_{i=1}^{n}(1 - x_{ik})}. \tag{6.1}$$

With a unique CS/D modeling function for each customer, it becomes possible to evaluate the customer's satisfaction level in a hypothetical situation, which differs from the experienced one. The first interesting hypothetical situation is where the performance of attribute x_i is equal to the customer's expectation. This situation can be modeled as follows:

$$S_{ik}^{e} = U_k(x_{1k}, \ldots, x_{(i-1)k}, e_k, x_{(i+1)k}, \ldots, x_{nk}), \tag{6.2}$$

where S_{ik}^{e} represents the satisfaction level experienced by the customer if attribute x_i performed as expected (e_k). Since e_k is the neutral element of the uninorm in Eq. 6.1 and given Definition 4 of a neutral element, S_{ik}^{e} can also be considered as the satisfaction level the customer would experience if there was no attribute x_i in the CS/D process.

A second hypothetical situation which is interesting to simulate is where the performance of attribute x_i increases with a value δ. This situation can be modeled as follows:

$$S_{ik}^{\delta} = U_k(x_{1k}, \ldots, x_{(i-1)k}, x_{ik} + \delta, x_{(i+1)k}, \ldots, x_{nk}), \tag{6.3}$$

where S_{ik}^{δ} measures the satisfaction of customer k if the performance of attribute x_i increases with δ.

Both S_{ik}^{e} and S_{ik}^{δ} can be used to measure the impact of attribute x_i on the customer's satisfaction by subtracting it from the experienced satisfaction S_k. A more general way of expressing the impact of attribute x_i is as follows:

$$\text{Impact}(x_{ik}) = S_k - S'_{ik}, \tag{6.4}$$

with S'_{ik} representing the hypothetical satisfaction level.

If S_{ik}^{e} is used as S'_{ik}, then Impact(x_{ik}) represents the impact on satisfaction for customer k, caused by the experienced disconfirmation for attribute x_i. Note that because of the monotonicity property of Dombi's uninorm, the impact of x_{ik} on satisfaction is positive if the attribute performs better than expected and negative if the attribute performs worse than expected. These impacts can be used as proxies for the importance of the attributes. The greater the impact of an attribute on satisfaction, the more important it becomes.

However, these impact scores are at the customer level and need to be aggregated to the population level for an IPA analysis. The easiest approach would be by taking the average of all impact scores for a specific attribute x_i for the entire sample. However, not every customer experiences the same performance level for attribute x_i and taking the average would ignore the dependency between attribute performance and attribute importance. Therefore, the average is taken conditional on the performance level z of attribute x_i. The disconfirmation impact scores $I_{(x_i=z)}^{e}$ are calculated as follows:

$$I_{(x_i=z)}^{e} = \frac{\sum_{k=1}^{N_{iz}} (\text{Impact}(x_{ik}) \mid x_i = z)}{N_{iz}}, \tag{6.5}$$

with N_{iz} representing the number of customers which have $x_i = z$ and Impact(x_{ik}) calculated as follows:

$$\text{Impact}(x_{ik}) = S_k - S_{ik}^{e}. \tag{6.6}$$

These disconfirmation impact scores $I_{(x_i=z)}^{e}$ measure the average effect on satisfaction if attribute x_i would go from its current performance level z to the expected performance level e and are useful to measure the importance of the product attributes in the current situation.

If S_{ik}^{δ} is used as S'_{ik}, then Impact(x_{ik}) represents the impact on satisfaction for customer k, caused by a performance increase of δ for attribute x_i. Similar to the disconfirmation impact scores, these impact scores are also aggregated to the population level by a conditional average. The marginal impact scores $I_{(x_i=z)}^{\delta}$ are calculated as follows:

$$I_{(x_i=z)}^{\delta} = \frac{\sum_{k=1}^{N_{iz}} (\text{Impact}(x_{ik}) \mid x_i = z)}{N_{iz}}, \tag{6.7}$$

with N_{iz} representing the number of customers which have $x_i = z$ and Impact(x_{ik}) calculated as follows:

$$\text{Impact}(x_{ik}) = S_k - S_{ik}^{\delta}. \tag{6.8}$$

The marginal impact scores are particularly useful in what-if analysis where managers want to verify the possible outcome of efforts to increase the performance of a particular product attribute. A case study in a following section will illustrate how both disconfirmation and marginal impact scores can be used in the context of an Importance-Performance Analysis.

Strengths and limitations of D-LIED derived attribute importance

For deriving the attribute importance, the D-LIED approach holds several advantages over the two traditional approaches. Compared with the direct measurement approach, the D-LIED approach can be used on secondary data and because it is an indirect approach it does not suffer from incorrect interpretation of survey questions. Compared with the regression approach, the D-LIED approach has the benefit of providing importance scores at the respondent level. But the most important difference with the regression approach is that the D-LIED approach is based on the ED paradigm. In contrast with the regression approach which only focuses on attribute performance, the D-LIED approach focuses both on customer's expectation and disconfirmation, providing a better theoretical fit.

The importance score of a specific attribute x_i obtained with the regression approach measures how satisfaction changes as the performance of x_i changes, which corresponds to the first partial derivative of the regression function with respect to x_i. Since the regression function is linear, the first partial derivative is the regression coefficient,

$$\frac{\partial(\alpha_0 + \sum_i^n \beta_i x_i)}{\partial x_i} = \beta_i, \tag{6.9}$$

which implies two unrealistic assumptions. Firstly, because the derivative is independent from x_i, the regression approach assumes that performance and importance are independent. Secondly, because the derivative is independent from any other attribute x_j, importance is also independent from the performance of other attributes.

The D-LIED approach does not make such assumptions in most situations, as can be illustrated by Figure 6.2 which plots the partial derivative of the D-LIED modeling function U for a product with two attributes x_1 and x_2 and an expectation level of 0.5. Note that the derivative of U corresponds to $I^{\delta}_{(x_i=z)}$ with δ infinitesimally small.

Three different situations need to be distinguished to fully understand the relationship between the marginal impact of attribute x_1 on one hand and the performance of x_1 and x_2 on the other hand: attribute x_2 performs worse than expected, attribute x_2 performs exactly as expected and attribute x_2 performs better than expected. All three situations are plotted in Figure 6.2 as separate functions. Firstly, the marginal impacts are always positive in each scenario which implies that an increase in performance of x_1 always results in an increase in satisfaction. This is a logical behavior and is directly related to the fact that U is monotone increasing. Next, Figure 6.2 shows that the marginal impact of x_1 is only constant when attribute x_2 performs as expected. Thus, when the other attribute causes no disconfirmation and the only disconfirmation effect comes from attribute x_1, each equal increase in disconfirmation leads to an equal increase in satisfaction.

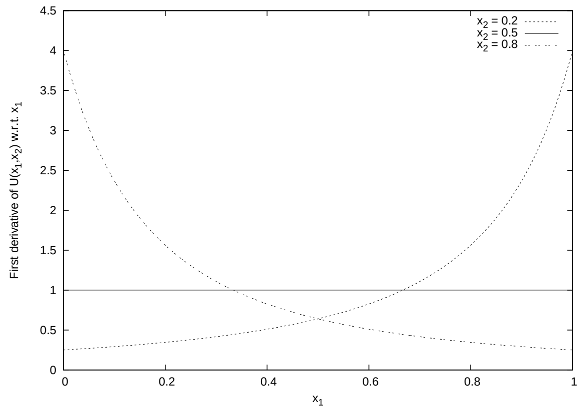


Figure 6.2: The first derivative of $U(x_1, x_2) = \frac{x_1 x_2}{x_1 x_2 + (1-x_1)(1-x_2)}$ with respect to x_1 for different values of x_2.

However, in both other scenarios, the marginal impact changes as the performance of x_1 changes, which implies a dependency between the attribute's importance and its performance. When attribute x_2 performs worse than expected, attribute x_1 needs to perform better than expected before it has a large marginal impact on satisfaction. This reflects the fact that low performance of x_2 causes such a negative disconfirmation effect, which first must be compensated before the performance of x_1 really stands out and has a large marginal impact on satisfaction. On the other hand, when attribute x_2 performs better than expected, the impact of x_1 is already large for low performance on x_1 and gradually fades out when performance on x_1 increases. This implies that if one attribute already performs very good and creates a high positive disconfirmation, any other low performing attribute stands out and even the smallest improvement will be clearly noticed and has a large impact on satisfaction. Finally, the fact that the impact of x_1, for a fixed performance level of x_1 differs between the three scenario's illustrates that the attribute importance is also dependent on the performance level of the other attributes.

The above discussion about the dependency between an attribute's importance and the performance of that and other attributes, has focussed on a product with only two attributes. Note that the conclusions above can be easily extended to products with more attributes by means of the associativity property of the D-LIED modeling function. This is shown by the following equation.

$$U(x_1, x_2, \ldots, x_n) = U(x_1, U(x_2, \ldots, x_n)) = U(x_1, x_2') \qquad (6.10)$$

Since $x_2' = U(x_2, \ldots, x_n)$ represents a satisfaction score, while x_2 represents a performance score, it is not easy to provide a clear interpretation for the interaction between the importance of x_1 and the performance of $x_1, \ldots, x_n$. Such an interpretation is much easier to derive if the D-LIED modeling function is studied in detail:

$$U(x_1, x_2, \ldots, x_n) = f\left(\sum_{i=1}^{n} f^{-1}(x_i)\right) = f\left(f^{-1}(x_1) + \sum_{i=2}^{n} f^{-1}(x_i)\right) \tag{6.11}$$

Based on the three-step interpretation of the LIED framework, $\sum_{i=2}^{n} f^{-1}(x_i)$ can be interpreted as the total disconfirmation effect caused by all attributes except x_1. This allows the formulation of the following assumptions which are made by the D-LIED framework regarding the importance of x_1:

- The marginal impact of x_1 is positive and constant if the total disconfirmation effect of all other attributes is equal to zero.
- The marginal impact of x_1 is positive and large for small performance levels, while decreasing as the performance of x_1 increases, if the total disconfirmation effect of all other attributes is positive.
- The marginal impact of x_1 is positive and large for high performance levels, while decreasing as the performance of x_1 decreases, if the total disconfirmation effect of all other attributes is negative.

These assumptions have obviously more face validity than the assumptions made by the regression approach.

Finally, the D-LIED approach also has a limitation compared with the regression approach. Neither the marginal impact scores, nor the disconfirmation impact scores can be considered as decompositions of the overall satisfaction score. That is, one can not reconstruct the customer's satisfaction by summing the impact scores of each attribute multiplied by the attribute's performance level. With the regression approach, the sum of the regression coefficients multiplied by the attribute performance levels results in the satisfaction score by definition.[1]

6.5 Empirical study

Purpose

In the previous section on regression based importance measures, two specific modeling assumptions, frequently made in IPA analysis, were discussed. Firstly, the regression approach assumes that attribute importance can be captured by a single point-estimate. Sampson et al. argued that this was incorrect and causes heteroscedastic error terms in the regression, making the regression coefficients less reliable. Secondly, the standard regression approach does not incorporate interaction terms, suggesting that the importance of attribute x_1 is independent of the performance on other attributes. The purpose of this empirical study is to investigate both assumptions.

In this empirical study, the regression based approach will be applied to the Family Entertainment Data Set. This data set contains customer satisfaction and product

[1] Actually, there is a difference between the measured satisfaction and the reconstructed satisfaction, which is captured by the residual.

performance data for seven different companies. For each company, a regression model is built with *Satisfaction* as the dependent variable and the 9 different factorized dimensions as the explanatory variables. Note that this study uses the factorized product dimensions instead of the original product dimensions because there is no need for expectation at the dimension level and factorized dimension scores are less susceptible to measurement errors.

In the first analysis, the homoscedasticity of the regression's error terms are studied to verify the assumption of independence between an attribute's importance and its performance level. As Sampson et al. pointed out, a heteroscedastic error term is probably caused by violation of this assumption. The second analysis adds various interaction terms to the regression function and verifies if there are significant interaction terms and if the model's fit increases significantly when the interaction effects are added.

In Importance–Performance Analysis and other customer satisfaction research, the regression model often assumes that no interaction effects exist between the attributes' performances. However, [62] suggests that the importance of an attribute depends on the performance of another attribute.

> "When Japanese companies entered the color television market, they changed the way in which products won orders from predominantly price to product quality (in terms of conformance) and reliability in service. ...By the early 1980s, manufacturers that lost orders raised product quality so that they were again qualified to be in the market. As a result, the most important order-winning criterion in this market has reverted back to price."

The purpose of this analysis is to test the validity of the 'no interaction' assumption. The goal of this study is not to identify all possible interaction effects. The only type of interaction effects tested are two-factor interaction effects of type $\frac{X_i}{X_j}$. To fully understand the way this interaction effect influences the impact of an attribute X_i on satisfaction, assume the following regression function:

$$Y = \alpha_0 + \beta_1 X_1 + \beta_2 X_2 + \gamma_{12}\frac{X_1}{X_2}. \tag{6.12}$$

The impact of attribute X_1 on Y can be measured by the partial derivative of Y with respect to X_1, which is

$$\frac{\partial Y}{\partial X_1} = \beta_1 + \gamma_{12}\frac{1}{X_2}. \tag{6.13}$$

This illustrates that if $\gamma_{ij} > 0$, the impact of product attribute X_i on satisfaction Y will decrease as the performance of X_j increases. If $\gamma_{ij} < 0$, the impact of product attribute X_i on satisfaction Y will increase as the performance of X_j increases. If significant interaction terms are found, the assumption of independence between an attribute's importance and the performance of other attributes is violated.

Note that both violations, i.e. heteroscedastic error terms and model misspecification, make the regression terms unreliable and possibly biased.

Homoscedasticity of the regression approach

"The Gaussian, standard, or classical linear regression model (CLRM), which is the cornerstone of most econometric theory, makes 10 assumptions" [58]. One of these assumptions is the homoscedasticity or equal variance of the disturbance term u_i, which implies that no relationship exists between the variance of u_i and the regressors X_i. Mathematically, this can be formulated as follows:

$$\mathbf{var}(u_i \mid X_i) = \sigma^2 \tag{6.14}$$

However, according to Sampson et al. [110], this basic assumption is violated when the customer satisfaction is regressed on the attribute performances, which is a common approach in IPA and customer satisfaction research. Sampson et al. refer to the dependency between impact on overall satisfaction and attribute performance, as the cause of the heteroscedastic disturbance term, which mathematically can be formulated as follows:

$$\mathbf{var}(u_i \mid X_i) = \sigma_i^2 \tag{6.15}$$

To verify the presence of heteroscedasticity in a performance-satisfaction regression, 7 regressions models were built for the Family Entertainment Data Set, i.e. one for each company. As mentioned before, each regression used satisfaction as the dependent variable Y and the factorized dimension performances as explanatory variables X_1 to X_9. The regression coefficients and the regression fit are presented in Table 6.1.

These results show that the adjusted R^2, which measures the amount of variance in customer satisfaction explained by the variation in attribute performances and corrected for the number of coefficients, falls between 28% and 40%. Furthermore, almost all companies have customers whose satisfaction is significantly influenced by the performance of *Product A* and the *Overall experience*. Product dimension *Price* has a significant impact on satisfaction for 4 out of 7 companies. *Accommodation* and *Product B* seem to have a negative impact on satisfaction which makes no sense, i.e. an increase of performance cannot cause a decrease of satisfaction. Note that such results cannot be modeled by the D-LIED implementation because a valid LIED implementation uses a monotone increasing modeling function.

Next, to investigate whether heteroscedasticity is in play, the formal Goldfeld-Quandt (GQ) heteroscedasticity test was applied. "This popular method is applicable if one assumes that the heteroscedastic variance, σ_i^2, is positively related to one of the explanatory variables in the regression model"[58]. However, it is also possible that the variance of the residuals u_i decrease as the independent variable X_i increase. Therefore, the GQ approach was adapted to test for both type of relationships.

Traditionally, the GQ starts by ranking the observations in increasing order, according to the values of X_i. Secondly the data is split into two sets, omitting the middle n observations. Thirdly, an OLS regression is fitted to both data set 1 (smallest X_i values) and data set 2 (largest X_i values). If no heteroscedasticity is present, the residual sum of squares (RSS) of both regressions should not differ significantly. The RSS is the sum of the squared residuals for each observation, with the residual being the difference between the observed Y and the Y predicted by the regression model. Therefore, the following ratio is computed:

Table 6.1: Coefficients of the performance regression on family entertainment data set

	Company						
	1	2	3	4	5	6	7
(Intercept)	0.19	0.17	0.20	0.32	0.07	0.27	0.21
Overall Experience	0.37	0.30	0.49	0.30	0.57	0.31	0.79
Product A	0.33	0.26	0.10	0.34	0.32	0.28	0.18
Product B	-0.08	-0.05	-0.03	0.01	-0.17	-0.03	0.03
Accommodation	-0.07	0.15	0.03	-0.22	-0.03	0.06	-0.20
Product C	0.06	0.12	0.05	-0.01	-0.13	-0.01	0.18
Product D	0.06	-0.04	-0.04	0.00	0.09	-0.04	-0.07
Personnel	0.03	-0.05	0.11	0.05	0.11	0.00	-0.11
Prices	0.04	0.03	0.09	0.14	0.12	0.11	0.05
Communication	0.03	0.06	-0.05	0.01	0.05	-0.01	-0.09
Adj. R^2	0.34	0.28	0.33	0.29	0.40	0.32	0.39
number of cases	295	239	314	255	270	266	483

Coefficients significant at 5% are marked in grey

$$\lambda = \frac{RSS_2 \backslash df}{RSS_1 \backslash df} \tag{6.16}$$

If the disturbance terms are homoscedastic, it can be shown that λ follows the F distribution[2]. To test for a negative relationship between u_i and X_i, it suffices to rank the observations in decreasing order or to calculate λ differently, as follows:

$$\lambda = \frac{RSS_1 \backslash df}{RSS_2 \backslash df} \tag{6.17}$$

Table 6.2 shows the GQ test statistic calculated with Eq. 6.16, with the middle 60 observations deleted. If the variance of the residuals are positively correlated with the independent variable X_i, the test statistic should be greater than 1. If the test statistic is less than 1, a negative correlation is assumed and Eq. 6.17 is used to determine the statistical significance. All regressions suffer from heteroscedasticity according to the GQ tests. Companies 2, 5, 6 and 7 have many product attributes causing heteroscedasticity, while companies 1, 2 and 3 have only a few product attributes which cause heteroscedasticity. From an other point of view, product dimensions *Overall experience*, *Product A*, *Product B*, *Product C* and *Communication* cause heteroscedasticity in at least 4 regressions (out of 7). This suggest that the impact of these dimensions are not independent from their performance.

[2]Actually, this approach also assumes that the disturbance terms u_i are normally distributed.

Table 6.2: Goldfeld-Quandt test statistics for the family entertainment data set

	Company						
	1	2	3	4	5	6	7
Overall Experience	1.50	2.02	1.16	1.02	0.52	0.59	0.39
Product A	1.06	2.27	0.85	0.67	0.89	0.46	0.51
Product B	0.84	1.99	0.68	0.90	0.73	0.49	0.77
Accommodation	1.19	1.20	0.85	0.71	0.72	0.91	0.63
Product C	1.53	1.27	1.57	0.99	0.64	0.79	0.51
Product D	1.21	1.71	0.81	0.98	0.59	0.60	0.82
Personnel	1.41	2.54	0.89	0.75	1.15	0.64	0.80
Prices	0.87	1.93	0.91	0.77	0.76	0.83	0.54
Communication	1.09	1.53	1.18	0.87	0.63	0.67	0.72

Coefficients significant at 5% are marked in grey

Note that the GQ test studies the homoscedastic nature of u_i in relation to each independent variable X_i separately, generating a test statistic per independent variable. In the previous chapter, the Breusch-Pagan test [93] was used to test for heteroscedasticity which produced only a single test statistic per model. The benefit of the GQ test is that it allows to identify the problematic product attributes. However, for sake of completeness, the Breusch-Pagan test was also conducted on all seven regression models.

Table 6.3: Breusch-Pagan test statistics for the family entertainment data set

	Company						
	1	2	3	4	5	6	7
B-P Test Statistic	3.81	3.69	2.69	4.47	13.61	19.51	67.02
p-value	0.05	0.05	0.10	0.03	0.00	0.00	0.00

The results from this test largely confirm the results of the GQ-tests, suggesting moderate to heavy heteroscedasticity in performance regressions, making the regression coefficients unreliable.

Interaction effects between product attributes

To test for interaction between product attributes with regard to the impact on satisfaction, 14 regression models were built for the Family Entertainment Data Set. For each company a full model with interaction effects and a reduced model with only main effects were estimated. Each model regresses the customer's satisfaction against the product attribute performances. The reduced model can be written mathematically as follows:

$$Y = \alpha_0 + \sum_{i=1}^{9} \beta_i X_i \tag{6.18}$$

The reduced model contains 9 main effects which correspond to 9 (factorized) product dimensions. Table 6.4 illustrates the correspondence between the variables and the product dimensions.

Table 6.4: Variables and corresponding product dimensions for family entertainment data set

Main effect	Product Dimension
X_1	Overall Experience
X_2	Product A
X_3	Product B
X_4	Accommodation
X_5	Product C
X_6	Product D
X_7	Personnel
X_8	Prices
X_9	Communication

The full model contains the main effects and various interaction effects. If all interaction effects are added to the model, the full model can be written as follows:

$$Y = \alpha_0 + \sum_{i=1}^{9} \beta_i X_i + \sum_{i=1}^{9} \sum_{\substack{j=1 \\ j \neq i}}^{9} \gamma_{ij} \frac{X_i}{X_j} \tag{6.19}$$

Because of the high number of interaction effects[3] and the possible multicollinearity among them, the AIC stepwise variable selection method [131] was used to add (or delete) interaction effects to (from) the reduced model. The AIC stepwise approach starts off by calculating the Akaike Information Criterium (AIC) [2] of a base model with the following formula where k denotes the number of parameters:

$$\text{AIC} = -2\text{maximized log-likelihood} + 2k \tag{6.20}$$

Next, various extended models are built and the AIC for each of them is calculated. Each extended model adds a new interaction effect to the base model or removes a recently added interaction effect from the base model. The extended model with the lowest AIC is compared with the base model. If the AIC of the base model is lower, the stepwise procedure finalizes. If the selected extended model has a lower AIC than the base model, the extended model becomes the new base model and the procedure is repeated. In this study, the first base model corresponds with the reduced model, represented by Eq. 6.18.

[3]There are 72 two factor interaction effects in total!

Table 6.5: Regression coefficients: company 1 from family entertainment data set

		Reduced Model	Extended Model
Intercept	α_0	0.19	1.35
Main effects	β_1	0.37	0.35
	β_2	0.33	-1.98
	β_3	-0.08	0.71
	β_4	-0.07	0.28
	β_5	0.06	-0.11
	β_6	0.06	-0.83
	β_7	0.03	1.62
	β_8	0.04	-0.01
	β_9	0.03	0.73
Interaction effects	γ_{36}		-0.59
	γ_{29}		0.39
	γ_{75}		-1.46
	γ_{45}		1.33
	γ_{74}		0.42
	γ_{42}		-1.27
Adjusted R^2		0.34	0.39

Coefficients significant at 5% are marked in grey

Table 6.5 presents the regression coefficients of the reduced and the extended model for company 1. The reduced model only has two significant main effects, i.e. the performance of dimensions *Overall Experience* and *Product A*. This is remarkable because the survey to collect the data was created with the family entertainment sector in mind and only product dimensions which were expected to be significant were measured. The extended model gives a different picture which corresponds better with a priori expectations. The extended model has 6 significant main effects (out of 9), and 5 significant interaction effects. Also note that the amount of variation in satisfaction explained by the various effects in the model, after correction for the number of model parameters, increases from 34% to 39%. These results strongly indicate that interaction effects are present in the customer satisfaction process for company 1.

The same analysis was conducted for the other 6 companies. The regression coefficients of the other reduced and extended models, together with the adjusted R^2, can be found in Appendix C. Table 6.6 provides a summary of the results for

Table 6.6: Extended model for all seven companies from the family entertainment data set - summarized results

Company	♯ Interaction effects*	Δ Adjusted R^2	F-test statistic	Pr(>F)
1	5	5%	5.26	< 0.01
2	5	6%	3.97	< 0.01
3	9	10%	5.23	< 0.01
4	3	6%	5.43	< 0.01
5	6	11%	9.33	< 0.01
6	6	7%	6.18	< 0.01
7	16	10%	5.93	< 0.01

* statistically significant at 5%

all 7 companies. These results show that all extended regression models contained several statistically significant interaction effects and that the explanation power, measured with adjusted R^2, substantially increased for all regression models when these interaction effects were added. Furthermore, the reduced model with only main effects is a restricted version of the extended regression model, with the restrictions that the coefficients of the interaction effects are all equal to zero. This allows the use of an F test to test the null hypothesis that the interaction effect coefficients are indeed equal to zero. The F-test statistic is calculated as follows [58]:

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)}, \tag{6.21}$$

with RSS_R denoting the residual sum of squares of the restricted model, RSS_{UR} denoting the residual sum of squares of the unrestricted model, m denoting the number of linear restrictions, k representing the number of parameters in the unrestricted regression and n denoting the number of observations. The null hypothesis was rejected for all seven companies, indicating that at least one of the regression coefficients of the interaction effects is statistically significant from zero.

6.6 Case study: the D-LIED IPA

The purpose of this case study is to illustrate how the D-LIED approach can be used to perform an IPA analysis. This analysis was done for each company of the Family Entertainment Data Set, but only the results of company 4 will be discussed in detail. The IPA plots for the other companies can be found in Appendix D.

The data contains one satisfaction score and 9 factorized dimension scores for each customer. For an IPA analysis, a performance and an importance score is needed for each dimension. The performance score $\overline{x}_i$ of a dimension i is calculated as the average factorized dimension performance p_i^F:

$$\overline{x}_i = \frac{\sum_{i=1}^{n} p_i^F}{n} \tag{6.22}$$

The importance score of a dimension i can be measured in two different ways with the D-LIED approach, i.e. with disconfirmation impact scores $I^e_{(x_i=z)}$ or with marginal impact scores $I^\delta_{(x_i=z)}$. Both impact scores lead to different interpretations of the IPA plot and have different purposes. First, an IPA is made based on the disconfirmation impact scores $I^e_{(x_i=z)}$. Firstly, the data is recoded to the $[0, 1]$ interval by dividing it by 11, next the D-LIED implementation is applied and S^e_{ik} and the corresponding $\text{Impact}(x_{ik})$ are computed for each customer. The next step would be to take the conditional average as follows:

$$I^e_{(x_i=\overline{x}_i)} = \frac{\sum_{k=1}^{N_{iz}} \left(\text{Impact}(x_{ik}) \mid x_i = \overline{x}_i\right)}{N_{i\overline{x}_i}}. \tag{6.23}$$

However, since the factorized dimensions x_i are quasi-continuous, it is possible that no observation is found for which $x_i = \overline{x}_i$. Even when such observations are found, the number of observations are probably low which makes the average less reliable. Therefore, the 25% percentiles of x_i were computed first and it was checked in which percentile $\overline{x}_i$ fell[4]. Next, the average was computed for all the observations from the 25% percentile which contained $\overline{x}_i$. This resulted in a disconfirmation impact score for each product dimension which was then multiplied by 11 to retrieve the original scale. The absolute value of these impact scores are used as the importance scores for each dimension. The absolute value was taken because the disconfirmation impact score can be negative and large negative impacts are equally important as large positive impacts. To mark the difference between negative and positive impacts, a different marker is used in the IPA plot, which is shown in Figure 6.3.

The X-axis of this plot was positioned at the mean customer's expectation. The Y-axis was positioned at the mean impact score. Other ways of positioning the axes exist, such as putting it half way between the maximum impact (performance) score and the minimum impact (performance) score or by positioning it such that the number of attributes on one side of the axis equals the number of attributes on the other side of the axis. However, the main point of an IPA analysis is the relative positioning of the attributes, rather than the absolute positioning [82].

The importance dimension of the IPA plot in Figure 6.3 represents the impact of the attributes on satisfaction caused by the disconfirmation currently experienced by the customers. Therefore, this grid allows managers to analyze the current situation. For company 4, this IPA shows that four attributes perform less than expected and have a negative impact on satisfaction, i.e. *Communication*, *Price*, *Product C* and *Product D*. However, only *Price* ends up in the "Concentrate Here" quadrant, while the other three dimensions have only a small impact on satisfaction and belong to the "Low Priority" quadrant. Furthermore, there are two dimensions within the "Keep up the good work" quadrant, i.e. *Personnel* and *Accommodation*. These two dimensions are the best performing dimensions and have a large contribution to the customer satisfaction outcome. The final three attributes, i.e. *Overall experience*, *Product B* and *Product A* are not really performing much worse than *Personnel* and *Accommodation*, but have a much lower impact on satisfaction which is why they belong to the "Possible Overkill" quadrant.

This IPA analysis provides a manager meaningful insights into the current situation. For example, for company 4 it shows that attention should be given to

[4]This approach would not be necessary for the observed dimensions which are categorical.

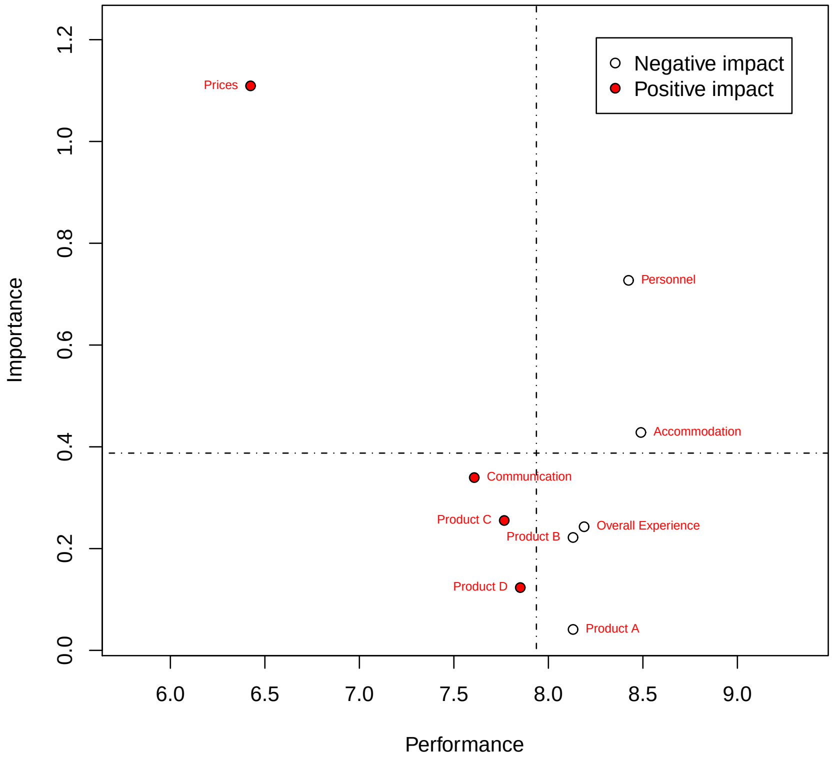


Figure 6.3: D-LIED IPA based on disconfirmation impact.

the performance of the *Price* dimension and an effort should be made to keep the performance of *Personnel* and *Accommodation* at the current level.

However, managers will be equally interested in what the future brings if they try to increase a specific attribute's performance. This kind of information can be retrieved by using the marginal impact scores $I^{\delta}_{(x_i=z)}$ as the importance score. The marginal impact scores measure the impact on satisfaction when performance is increased by δ. In this study, δ was set to 1. The importance scores for each dimension are calculated completely analogous as the disconfirmation impact scores, except no absolute value was taken because marginal impact scores are always positive. Figure 6.4 shows the IPA plot based on marginal impact scores. The axes were positioned similar to the previous IPA plot.

This IPA analysis provides new and important information. Firstly, increasing the performance of dimension *Price*, which appeared to be a priority in the previous IPA analysis, has only low impact on satisfaction. This raises several questions. Is it worthwhile putting effort in the performance of this dimension given the limited impact, even if it currently has a rather large negative influence on satisfaction? And,

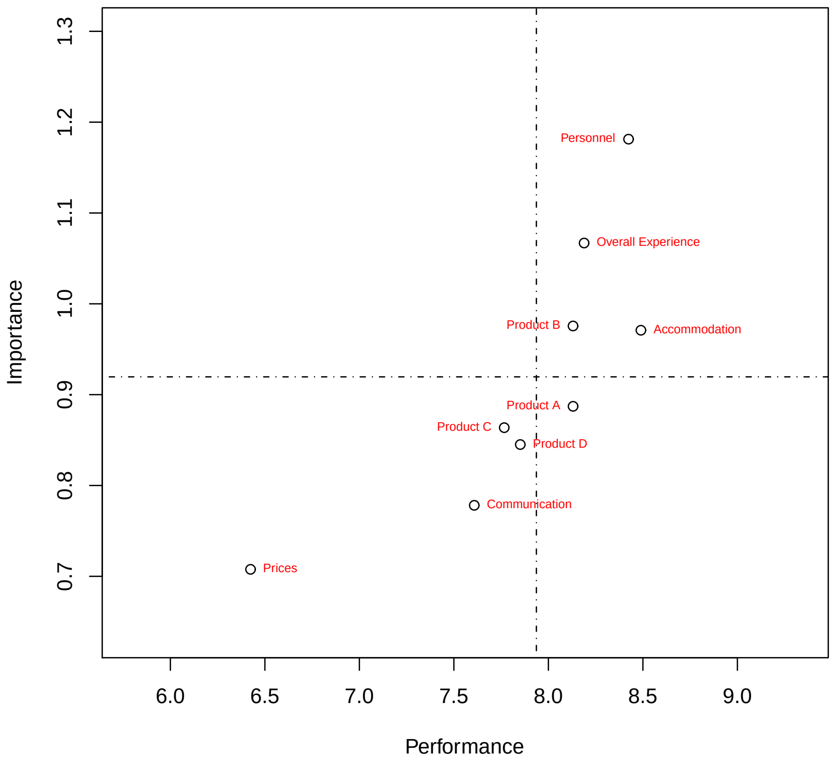


Figure 6.4: D-LIED IPA based on marginal impact.

should the performance of dimension *Price* perhaps be increased with more than 1? The first question requires a managerial decision, but the second question can be answered by constructing another marginal IPA analysis with a larger δ.

Secondly, this marginal IPA analysis also shows that increasing the performance of *Accommodation*, which belonged to the "Keep up the good work" quadrant has much less impact than increasing the performance of *Personnel* which also belonged to the "Keep up the good work" quadrant or even *Overall Experience* which belonged to the "Possible Overkill" quadrant. Apparently, *Overall Experience* is currently not having a large impact on satisfaction despite its high performance level, which puts it in the "Possible Overkill" quadrant in the previous IPA, but the marginal IPA analysis reveals that it might be worthwhile to increase its performance level by 1. This result is supported by recent research which states that an attribute's impact is non-linear and performance dependent [90].

However, the final decision of which attributes the company is going to focus on, does not only depend on the results of the IPA analysis. For example, increasing the performance of the *Price* dimension, as suggested by the first IPA analysis, might be impossible because of financial reasons. Nevertheless, for developing proper marketing strategies, both an IPA analysis which provides insights about the current situation and an IPA approach which can simulate changes and takes the nonlinearity aspect into account are recommended. The D-LIED IPA approach has the advantage of offering both and allowing to simulate any change in the parameters, i.e. performance and expectation, to measure the effect.

6.7 Summary

Importance-Performance Analysis is an important managerial marketing tool. It plots each product attribute on a two dimensional grid, visualizing the attribute's performance and importance. While consensus exists about measuring the performance of a product attribute, the importance dimension is much more problematic. In this chapter two common approaches are discussed to determine the importance of a product attribute, i.e. the direct measurement approach and the regression based approach. The direct measurement approach uses a survey to ask respondents about the attribute's importance but has the danger of measuring the wrong construct because respondents might interpret importance questions incorrectly, as has been argued by different authors [110, 84]. The regression based approach also faces several problems. Firstly, the regression is likely to be heteroscedastic because it assumes independence between attribute importance and attribute performance which is unlikely to be true according to existing literature [110, 90]. Secondly, the regression approach most likely suffers from model misspecification because it does not integrate interaction effects between product attributes. In this chapter, elaborated tests were performed to evaluate how strongly the assumptions of homoscedasticity and correct model specification are violated and both analysis confirmed the assumption violations. These violations make the coefficient estimates, which act as the dimension's importance, unreliable and possibly biased. Future research on the nature of and solutions for these problems will be necessary.

Furthermore, an alternative approach to determine the attribute's importance has been presented which is based on the D-LIED implementation. As has been shown throughout this chapter, this D-LIED approach has several benefits. Firstly, impact measures are calculated at the level of a single respondent in contrast to the regression based approach which only provides population level impact measures. Even when these impact measures are averaged, in order to obtain population level scores, the population level impact scores are still conditional to the performance score, which is in line with the assumption of performance-importance dependency. Secondly, the D-LIED approach implicitly takes interaction effects into account and focuses on disconfirmation instead of attribute performance, allowing a close match with existing customer satisfaction theory.

Finally, two different ways of utilizing the D-LIED IPA for managerial ends was presented by means of a case study. The first way, which is based on the disconfirmation impact scores, allows an analysis of the current situation, identifying the attributes with a large positive or negative impact on the current satisfaction level. The second way, which is based on the marginal impact scores, allows the manager to perform what-if simulations, determining the impact on customer satisfaction in case of a change in the attribute performance level or a change in the expectation level.

To conclude, the use of the D-LIED implementation within the context of IPA possesses a great deal of potential as a novel technique in the field of customer satisfaction research and Importance-Performance Analysis.

7 Application 2: expectation-performance compatibility

7.1 Introduction

During the last four decades, customer (dis)satisfaction has taken an important role in marketing research, both from an academic as from a managerial point of view [130, 122]. Although this has not always been the case, customer (dis)satisfaction is now widely recognized as an important cornerstone for customer-orientated companies, irrespective of the industry they operate in [129, 122]. Research results have shown that customer (dis)satisfaction has an influence on several important aspects of a competitive business, such as repurchase intention [122], consumer retention [89, 3], firm performance [4], customer complaining behavior [122], negative word of mouth behavior [122] and eventually on shareholder value [5].

Several theoretical models have tried to explain the human behavior in a customer satisfaction context. One of the more dominant models which tried to do so is Oliver's expectancy disconfirmation paradigm. This paradigm has been introduced and discussed in chapter 2 of this dissertation. Basically, it identifies two direct causes of satisfaction, i.e. customer expectation and experienced disconfirmation, which is the difference between the experienced performance and the expected performance. Expectation has a direct influence on customer (dis)satisfaction: "Consumers are thought to assimilate satisfaction levels to expectation levels in order to avoid dissonance that would arise when expectations and satisfaction levels diverge. This assimilation effect results in satisfaction judgments being high (low) when expectations are high (low)" [122]. On the other hand, disconfirmation is believed to have a contrasting effect on the overall (dis)satisfaction level. If consumers perceive a great positive (negative) disconfirmation, their satisfaction level will rise (drop).

While most research publications discuss the ED paradigm at the aggregate (product) level which assumes a single performance, expectation and disconfirmation score for the entire product or service, Oliver [96] already argued in a 1980 that "disconfirmation takes place at the individual attribute level" [97]. This implies that the consumer has expectations at the attribute level and perceives a specific performance and disconfirmation for each product or service attribute (dimension). The consumer will aggregate all these expectations, performance perceptions and subjective disconfirmation scores into an overall satisfaction score. This aggregation has been captured by the LIED framework and can be applied to satisfaction-performance data with the D-LIED implementation, which has been extensively discussed in the previous chapters.

This chapter will proceed one step further in the process of customer (dis)satisfaction and studies the customer's intentions, such as the customer's loyalty and the customer's recommendation intentions. More specifically, the role of the customer's expectation in the process of the customer's intention will be investigated. In the next section, the hypothesis is formulated that the degree of compatibility between the customer's expectation and the perceived product performance has a direct influence on the customer's intentions and it will be shown how the D-LIED implementation can be used to estimate the expectation-performance compatibility. Next, the results of an empirical study will be presented which tries to assess the validity of hypothesis and which compares the different measures of compatibility. Finally, an overall conclusion will be given.

7.2 Expectation-performance compatibility

The idea of considering the expectation-performance compatibility as a construct in the customer's intention process stems from the participatory learning paradigm introduced by Yager [140, 32, 142]. The basic premise of this paradigm is that learning takes place in the framework of what is already learned and observations too conflicting with the learner's beliefs are discarded. Central in the formulation of the participatory learning system is the compatibility measure. A high compatibility between the learner's beliefs and the new information has a positive effect on the learning process.

Analogously, the customer's intentions are also formed in the framework of what has already been learned about the product. Furthermore, the product experiences result in the customer's expectation, which will act as a comparison reference for subsequent experiences. Therefore, it is believed that a high compatibility between the customer's expectation and the perceived performance must have a positive effect on the customer's intentions. Note that incompatibility implies unexpected product experiences, both negative and positive. These unexpected product experiences can make the customer uncertain about the product's capabilities, which has a negative influence on the customer's loyalty and recommendation intentions. This leads to the following two hypotheses:

Hypothesis 1. The expectation-performance compatibility is positively correlated to the customer's intended loyalty.

Hypothesis 2. The expectation-performance compatibility is positively correlated to the customer's recommendation intention.

It is important to stress that these hypotheses imply that both positive and negative incompatibility are negatively related to the customer's intention. Incompatibility, caused by a much better product performance than expected, is also assumed to have a negative effect on the customer's repurchase and recommendation intention because it creates uncertainty about the product's performance.

In order to test these hypotheses, a formula is needed to express the expectation-performance compatibility at the product or service level. This product-level compatibility measure is based on the incompatibility between the product's performance and the customer's expectation at the product attribute level. The incompatibility at the attribute level can be expressed as the difference between the perceived performance and the expected performance. The perceived performance is typically measured by means of a survey. In this chapter, it is assumed that product attribute performance is measured on a $[0,1]$ scale, with 0 denoting the lowest performance level and 1 denoting the highest performance level. Thus, if the original survey measures performance and satisfaction on different scales, they first need to be transformed to the $[0,1]$ interval. Next, customer expectation is derived for each customer by means of the D-LIED implementation. If x_{ik} represents the performance of attribute i as it was experienced by customer k and S_k represents the satisfaction level of customer k, then Eq. 5.5 can be used to find the D-LIED parameter α_k for each customer as follows:

$$\alpha_k = \left[\frac{\prod_{i=1}^{n} x_{ik}}{\prod_{i=1}^{n}(1 - x_{ik})} \quad \frac{1 - S_k}{S_k} \right]^{\frac{1}{n-1}} \tag{7.1}$$

Once α_k is known, Eq. 5.4 can be used to find the expectation level e_k for each customer as follows:

$$e_k = \frac{\alpha_k}{1 + \alpha_k} \tag{7.2}$$

Next, attribute's i incompatibility for customer k can be expressed as follows :

$$I_{ik} = x_{ik} - e_k \tag{7.3}$$

Note that the domain of I_{ik} is the $[-1,1]$ interval and that both -1 and 1 reflect complete incompatibility while 0 reflects perfect compatibility. Next, the various attribute incompatibility scores need to be aggregated into a compatibility score C_k at the product level such that the following properties hold.

Property 11.

a) $0 \leq C_k \leq 1$

b) $\forall i : I_{ik} = 0 \Rightarrow C = 1$

c) $\forall i : I_{ik} \in \{-1, 1\} \Rightarrow C = 0$

Property 11 a) sets boundaries to the compatibility measure which makes it easier to interpret and compare the final compatibility outcome. Property 11 b) states that if the performance of each product attribute is compatible with the expectation, the product level compatibility measure should be maximized. Property 11 c) implies that if each product attribute is completely incompatible with the expected performance, the product level compatibility should be minimized.

The first aggregation function which has Property 11 and which is used to define the product's compatibility with the customer's expectation is the same as the compatibility measure in Yager's participatory learning paradigm [140]. This compatibility measure will be denoted as C_k^1 and is defined as follows:

$$C_k^1 = 1 - \frac{1}{n} \sum_{i=1}^{n} |I_{ik}| \tag{7.4}$$

A second possible aggregation function which has the necessary properties for being a compatibility measure is also based on the original compatibility measure of Yager's participatory learning paradigm but takes the square of the attribute incompatibilities instead of the absolute value. This compatibility measure is denoted as C_k^2 and can be expressed as follows:

$$C_k^2 = 1 - \frac{1}{n}\sum_{i=1}^{n} I_{ik}^2 \tag{7.5}$$

It can be easily verified that both compatibility measures possesses all the required properties. Next, the proposed hypotheses will be tested by means of an empirical study.

7.3 Empirical study

Data

This empirical study uses the Energy Data Set which comprises customer satisfaction data for two Belgian companies. The survey measured the overall satisfaction level and the performance of various product attributes and product dimensions. A factor analysis was performed to construct multi-item dimension performance scores which are less sensitive to measurement errors than the directly measured dimension performances. The details of the factor analysis can be found in Chapter 4.

The factorized dimension scores are expressed on a scale from 1 [very low] to 5 [very high] and the overall satisfaction was measured on a scale from 1 [extremely dissatisfied] to 10 [extremely satisfied]. Additional to performance and satisfaction scores, the data set also contains information about the customers' intentions, such as their intention to switch companies (disloyalty) and their intention to recommend the company to a friend. Intention to switch was measured on a scale from 1 [No intention to switch at all] to 5 [Strong intention to switch]. The recommendation score was also measured on a scale from 1 [No intention to recommend] to 5 [Strong intention to recommend].

Next, the D-LIED implementation was used to calculate the expectation score e_k and both compatibility scores, C_k^1 and C_k^2 for each customer. The expectation score is transformed back to the same $[1, 5]$ scale as the performances by multiplying it by 6.

Table 7.1 provides descriptive statistics about the data for company 1, while Table 7.2 shows the descriptive statistics for company 2. In total, the data set contains 1053 records (525 observations for company 1, 528 observations for company 2).

Results

The purpose of the empirical study is to assess the validity of the proposed hypotheses. Therefore, the Pearson correlations between both compatibility measures on the one hand and loyalty and recommendation on the other hand were calculated. Both hypotheses are compared against the null hypothesis which states that no correlation exists between the compatibility measures and the customer's intentions. The results are presented in Table 7.3 for company 1 and Table 7.4 for company 2.

These results reveal that all correlations are statistically significant, which rejects the null hypotheses in favor of the alternative hypotheses. These results

Table 7.1: Descriptive statistics: company 1

	Minimum	Maximum	Mean	Std. Dev.
Satisfaction	1	10	7.90	1.45
Intention to Switch	1	5	3.63	1.07
Intention to Recommend	1	5	2.29	1.26
Availability	1	5	3.51	1.02
Employees	1	5	4.14	0.75
Service	1	5	3.80	0.83
Information	1	5	3.39	1.00
Invoices	1.33	5	4.22	0.75
Expectation	1.23	5.36	3.69	0.80
Compatibility C^1	0.72	1	0.92	0.05
Compatibility C^2	0.92	1	0.99	0.01

Table 7.2: Descriptive statistics: company 2

	Minimum	Maximum	Mean	Std. Dev.
Satisfaction	1	10	7.01	2.10
Intention to Switch	1	5	3.09	1.26
Intention to Recommend	1	5	2.71	1.38
Availability	1	5	3.81	0.95
Employees	1	5	4.22	0.80
Service	1	5	3.55	1.06
Information	1	5	3.06	1.15
Invoices	1	5	3.85	1.08
Expectation	1.17	5.54	3.68	0.89
Compatibility C^1	0.66	1	0.90	0.05
Compatibility C^2	0.88	1	0.98	0.02

confirm the assumed relationship between performance-expectation compatibility and the customer's intention. A positive correlation exists between the product's performance-expectation compatibility and the customer's intention to recommend the product. Furthermore, a negative correlation can be found between the product's performance-expectation compatibility and the customer's intention to switch to another company. The latter conclusion is analogue to saying that compatibility is positively correlated with the customer's loyalty. These results reveal some new and interesting insights into the customer's post-purchase intentions process.

Table 7.3: Company 1: Pearson correlations

	C^1	C^2
Intention to Switch	-0.21^a	-0.19^a
Intention to Recommend	0.29^a	0.26^a

[a] Statistically significant at 1%

Table 7.4: Company 2: Pearson correlations

	C^1	C^2
Intention to Switch	-0.39^a	-0.33^a
Intention to Recommend	0.47^a	0.44^a

[a] Statistically significant at 1%

Furthermore, the results in Tables 7.3 and 7.4 seem to suggest that compatibility measure C^1 has a stronger correlation with the customer's intentions than compatibility measure C^2. To verify if the correlation r_{S,C^1} between *Intention to Switch* and compatibility measure C^1 is truly larger than the correlation r_{S,C^2} between *Intention to Switch* and compatibility measure C^2, further statistical analysis of the difference between r_{S,C^1} and r_{S,C^2} is needed. The same holds for the difference between r_{R,C^1}, i.e. the correlation between *Intention to Recommend* and C^1, and r_{R,C^2}, i.e. the correlation between *Intention to Recommend* and C^2.

The test statistic for evaluating the null hypothesis that the correlations between X and Z differs statistically significantly from the correlation between Y and Z is computed with Eq. 7.6. In this equation, n refers to the sample size and r_{XY} refers to the correlation between X and Y. For a more detailed discussion of this test statistic, the reader can refer to [118, 27, 114].

$$t=(r_{YZ}-r_{XZ})\sqrt{\frac{(n-1)(1+r_{XY})}{2\left[\frac{(n-1)}{(n-3)}\right]\left[1-r_{YZ}^2-r_{XZ}^2-r_{XY}^2+2r_{YZ}r_{XZ}r_{XY}\right]+\left[\frac{r_{YZ}+r_{XZ}}{2}\right]^2\left[1-r_{XY}\right]^3}} \tag{7.6}$$

Table 7.5 shows the results of the comparison of the correlations for both companies. The reported p-values show that each hypothesis is statistically rejected at a significance level of 5%, except for the difference between r_{S,C^1} and r_{S,C^2}. However, for both companies, the test whether C^1 has a stronger correlation with the customer's intentions than C^2 actually consists of two separate hypothesis which are tested on the same sample[1]. Therefore, Bonferroni's correction of the statistical significance level needs to be applied [93, 31]. This correction implies that each of the h hypotheses in a set of hypotheses need to be tested at an adjusted significance level $\alpha' = \alpha/h$ to achieve a significance level α for testing the set of hypothesis. This implies that the hypotheses in Table 7.5 need to be tested against a 2.5% significance level to

[1] Note that there are 2 separate sets of two hypothesis because the samples for both companies are different

Table 7.5: Statistical test to verify which compatibility measure has the strongest correlation with the customer's intentions

	H_0	p-value (2-tailed)
Company 1		
	$r_{S,C^1} - r_{S,C^2} = 0$	0.13
	$r_{R,C^1} - r_{R,C^2} = 0$	0.01
Company 2		
	$r_{S,C^1} - r_{S,C^2} = 0$	0.00
	$r_{R,C^1} - r_{R,C^2} = 0$	0.02

achieve 5% significance for the test that C^1 is stronger correlated with the customer's intentions than C^2. Even at this adjusted significance level, the conclusions remain. For company 2 C^1 has a statistically significant stronger correlation with the customer's intentions than C^2. For company 1, the null hypothesis that C^1 is equally strong correlated with the customer's intentions as C^2 cannot be rejected, but $|r_{R,C^2}|$ is statistically smaller than $|r_{R,C^1}|$. All these results suggest that C^1 should be preferred over C^2 for predicting and explaining customer's intentions. The reason why C^1 has a stronger correlation with the customer's intentions than C^2 might be caused by the fact that incompatibilities are diminished in C^2 by taking the square of these incompatibilities.

Table 7.6: Mean comparison of compatibility measures and customer intentions between both companies

	Mean		Mann-Whitney Test
	Company 1	Company 2	p-value (2-tailed)
Switch	2.29	2.71	0.00
Recommend	3.63	3.09	0.00
C^1	0.92	0.90	0.00
C^2	0.99	0.98	0.00

Next, the study focuses on the differences between the companies by computing and studying the mean *Intention to Switch*, the mean *Intention to Recommend*, the mean compatibility measure C^1 and the mean compatibility measure C^2, which are presented in Table 7.6. These results show that customers of company 1 are statistically significantly more loyal and more likely to recommend their company than customers of company 2. Furthermore, Table 7.6 also shows that company 1 succeeds better in meeting the customer's expectations than company 2. Although the difference between the average compatibility of both companies is rather small, it is still statistically significant at the 1% level. The statistical significance of the difference between two means was calculated with the Mann-Whitney Test. The classical t-test for two independent samples (cf [114]) was unreliable because the assumption of equal variance between the two samples was heavily violated, which was tested with Levene's test for equality of variances (cf [93]). Also the normality assumption, was violated according to the Kolmogorov-Smirnov test. Instead of relying on the robustness properties of the t-test, the non-parametric Mann-Whitney Test (cf. [114]) was applied which does not assume normality or equal variance. The p-values of the Mann-Whitney Test were both derived asymptotically and with a Monte-Carlo experiment with 10000 samples. Both approaches resulted in the same p-values.

The results of Table 7.6 and the positive correlation found between performance-expectation compatibility and the customer's intentions suggest that the difference in customers' intentions between both companies can be partially explained by company 1 having a better ability to meet customer expectations. This difference makes sense because the first company is a company which is in the market for decades while company 2 is the newcomer to the market.

Finally, to illustrate how the compatibility measure helps a company understand the customer's intentions, a regression model with *Intention to Recommend* as the dependent variable and *Satisfaction*, *average performance* and *compatibility* as explanatory variables is built for company 2. C^1 is used as *compatibility* measure because the previous results indicated that this measure has the highest correlation with the customer's intentions. *Average performance* is a new variable which represents the mean performance of the five factorized dimensions.

Note that the compatibility measure is a function of expectation and dimension performance and expectation is a function of Satisfaction and dimension performance. Therefore, the compatibility measure is actually a function of Satisfaction and dimension performance. Even though it has been shown that the compatibility measure is related to the customer's intentions, the question remains whether the compatibility measure merely summarizes the information from Satisfaction and dimension performances or if it truly extracts new information hidden in the relationship between these constructs. By adding all three constructs to the regression function, i.e. *Satisfaction*, *average performance* and *compatibility*, it can be tested if *compatibility* has a significant effect on the customer's intention which has not been captured by *Satisfaction* or *average performance*.

Intentions to Recommend is measured on a discrete five-point scale which is a categorical scale rather than an interval scale. Therefore, a multinomial logistic regression was conducted instead of a multiple linear regression model. A multinomial logistic regression is much more appropriate for regressions with a categorical dependent variable than a linear regression. The customer's *Intentions to Recommend* was recoded to a variable with three categories, i.e. "no intention to recommend" (values 1 and 2 of the original variable), "undecided" (value 3 of the original variable) and "intention to recommend" (values 4 and 5 of the original variable). Figure 7.1 shows a histogram of the new *Recommendation* variable.

A multiple linear regression would try to predict the customer's *Intention to Recommend* by means of the customer's *Satisfaction*, the *average performance* and the *compatibility*. However, this approach would most likely predict values for the dependent variable which are impossible in reality, i.e. all values except for $\{1, 2, 3\}$. Instead, a multinomial logit model does not try to predict the customer's *Intention*

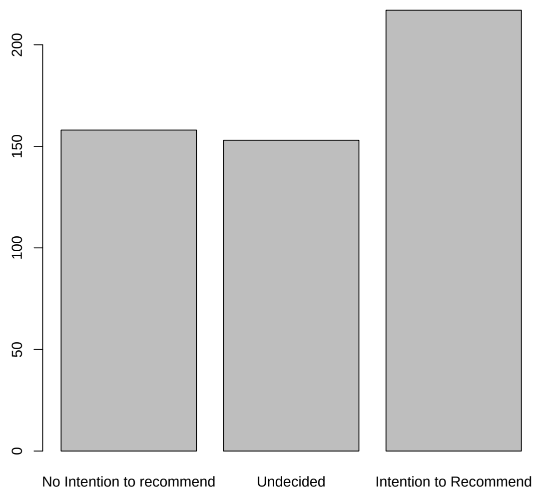


Figure 7.1: Distribution of customer's intention to recommend.

to Recommend, but the logit of the independent variable. The logit of a variable is defined as the natural logarithm of the ratio of the probability that $Y = i$ to the probability that $Y = r$, with r representing the reference category [85]. In this analysis, category 3 ("Intention to Recommend") was selected as the reference category. The regression model can be expressed as the set of following two equations, with Y denoting *Intentions to Recommend*, X_1 denoting *Satisfaction*, X_2 denoting the *average performance* and X_3 denoting the *compatibility* C^1:

$$ln\left(\frac{P(Y=1)}{P(Y=3)}\right) = \beta_{01} + \beta_{11}X_1 + \beta_{21}X_2 + \beta_{31}X_3 \tag{7.7}$$

$$ln\left(\frac{P(Y=2)}{P(Y=3)}\right) = \beta_{02} + \beta_{12}X_1 + \beta_{22}X_2 + \beta_{32}X_3 \tag{7.8}$$

One of the benefits of the multinomial logistic regression model over the linear regression model is that the multinomial model does not make normality or homoscedasticity assumptions about the error terms. On the other hand, the multinomial model does make the assumption that no strong multicollinearity between the independent variables exists and that the model is linear in the logit [85]. Both assumptions need to be tested before the regression model can be analyzed.

First, the assumption of multicollinearity was tested. Table 7.7 shows the variance inflation factors (VIF) for each variable. The VIF of a variable shows how the variance of the estimator is inflated by the presence of multicollinearity. As a rule of thumb, Gujarati [58] suggests that a VIF of a variable exceeding 10 indicates a high level of multicollinearity. Table 7.7 shows that multicollinearity is not a problematic issue in this regression model.

Table 7.7: Collinearity diagnostics (variance inflation factors)

X_1	X_2	X_3
1.55	1.76	1.43

Next, the assumption of linearity in the logit is tested. This assumption implies that the logistic regression model has a linear form and the change in logit(Y) for a one-unit change in X is equal to the logistic regression coefficient [85]. Menard [85] suggests to use the Box-Tidwell transformation described by Hosmer and Lemeshow (cf [65],p. 90), which involves adding a term of the form $(X)ln(X)$ to the model. Statistically significant coefficients for this variable suggest that the relationship between logit(Y) and X is nonlinear.

Table 7.8: Linearity in the logit test

	χ^2	df	p-value
$X_1 ln(X_1)$	2.23	2	0.33
$X_2 ln(X_2)$	3.15	2	0.21
$X_3 ln(X_3)$	0.18	2	0.91

Table 7.8 shows the statistical significance of the coefficients for the Box-Tidwell transformations. The statistical significance is computed by means of a χ^2 statistic which is the difference in -2 log-likelihood between the model with the Box-Tidwell transformed variable and the model without the Box-Tidwell transformed variable. The results of Table 7.8 show that all independent variables are linearly related to the logit of *Intentions to Recommend*.

Table 7.9: Goodness of model fit

	McFadden R^2
Model with C^1	0.22
Model without C^1	0.20

With none of the assumptions violated, the final model can be analyzed in detail. Table 7.9 shows McFadden's R^2, which is a goodness of fit measure. McFadden's R^2 is the multinomial regression variant of the R^2 in a multiple linear regression and indicates how much the inclusion of the independent variables in the model reduces the variation [85]. The results in Table 7.9 show that the model with *Satisfaction*, *average performance* and *compatibility* as independent variables explain 22% of the variation in the customer's *Intention to Recommend*. McFadden's R^2 was also computed for the same regression model but without *compatibility* as an explanatory variable. Apparently, without *compatibility*, the model could only explain 20% of the variation in the customer's *Intention to Recommend*, which is a first

Table 7.10: Likelihood ratio tests

	χ^2	df	p-value
Intercept	59.61	2	0.00
Satisfaction	57.43	2	0.00
Average Performance	24.34	2	0.00
Compatibility (C^1)	16.36	2	0.00

indication that *compatibility* contains information not captured by *Satisfaction* or *average performance*.

Next, Table 7.10 shows the statistical significance of each independent variable by means of a χ^2 statistic which is the difference in -2 log-likelihood between the model with the independent variable and the model without the independent variable. These results show that all independent variables are statistically significant including the *compatibility* measure which reconfirms that compatibility contains information not captured by *Satisfaction* or *average performance*.

Table 7.11: Parameter estimates

	β	β^*	e^β	p-value
Intention to Recommend = "No intention"				
Intercept ($\beta_0 1$)	19.52	-	-	0.00
Satisfaction ($\beta_1 1$)	-0.66	-0.52	0.52	0.00
Average Performance ($\beta_2 1$)	-1.06	-0.32	0.35	0.00
Compatibility ($\beta_3 1$)	-12.38	-0.25	$4e^{-6}$	0.00
Intention to Recommend = "Undecided"				
Intercept ($\beta_0 2$)	12.88	-	-	0.00
Satisfaction ($\beta_1 2$)	-0.34	-0.42	0.71	0.00
Average Performance ($\beta_2 2$)	-0.81	-0.39	0.45	0.00
Compatibility ($\beta_3 2$)	-8.21	-0.26	$3e^{-4}$	0.00

β^* is the standardized regression coefficient.

Finally, Table 7.11 shows the estimates of the regression coefficients which allow the interpretation of the relationship between the independent variables and the dependent variable. The final column of this table shows that all variables are statistically significant at 1%. Furthermore, the table provides the estimates of the regression coefficients β as well as the estimates for the standardized regression coefficients β^* and the odds ratio's e^β. The odds ratios provide a much easier interpretation than the original regression coefficients. For example, an odds ratio

$e_{11}^{\beta} = 1.2$ would imply that if attribute X_1 increases with one unit, the odds $\frac{P(Y=1)}{P(Y=3)}$ would increase with factor 1.2. This allows the formulation of the following conclusions for company 2.

- As the customer's *Satisfaction* increases with a single unit, e.g. from 4 to 5, the odds of being indecisive about recommendation against having the intention to recommend drops with 39% and the odds of having no intention to recommend against the intention to recommend drops with 48%.
- As the *average performance* increases with a single unit, e.g. from 4 to 5, the odds of being indecisive about recommendation against having the intention to recommend drops with 55% and the odds of having no intention to recommend against the intention to recommend drops with 65%.
- As the *compatibility* increases with a 0.01, e.g. from 0.90 to 0.91, the odds of being indecisive about recommendation against having the intention to recommend drops with 8% and the odds of having no intention to recommend against the intention to recommend drops with 12%.[2]

Note that compatibility is measured on a $[0, 1]$ scale which makes the impact on the odds by an increase of 1 meaningless, which is why the impact of 0.01 was studied instead. This indicates that comparing the odds ratios between variables which are measured on different scales is prone to lead to incorrect conclusions. For example, from the discussion above it might appear that the *average performance* has a larger impact on the customer's *Intention to Recommend* than the customer's *Satisfaction* which is positive news for a manager since the average performance is easier to control than the customer's satisfaction. However, a unit increase of *Satisfaction*, which is measured on a [1,10] scale is less difficult than a unit increase of *average performance* which is measured on a [1,5] scale. Therefore, to compare the relative impact of the different independent variables, one needs to refer to the standardized regression coefficients. The standardized regression coefficients in Table 7.11, which are calculated by the procedure in [85] (p. 53), reveal that the strength of the relationship between *Satisfaction* and *Intention to Recommend* is the strongest and the strength of the relationship between *compatibility* and *Intention to Recommend* is the weakest.

From all these results, one can conclude for company 2 that compatibility is positively correlated with the customers' intention to recommend the company to their friends. Although average performance and customer satisfaction is more strongly related to the customer's intentions of recommendation, it would be incorrect to ignore the importance of performance-expectation compatibility.

7.4 Summary

In this chapter, expectation-performance compatibility was defined as a new construct in the process of customer's intention behavior. Two different approaches were suggested to compute the expectation-performance compatibility by means of the D-LIED implementation, i.e C^1 and C^2. Empirical results proved that these

[2]The impact of an increase of 0.01 in X on the odds is equal to $e^{0.01\beta}$.

compatibility measures are positively correlated with the customer's intentions, which implies that both positive or negative incompatibility between the customer's expectation and the perceived performance has a negative effect on customer loyalty and recommendation.

Furthermore, the empirical results suggested that C^1 should be preferred in customer intention analysis. To illustrate the use of the compatibility measure in such customer intention analysis, a multinomial logistic regression model was built. The results of this regression model indicated that the compatibility measure succeeds in extracting new information from the performance and satisfaction data and helps managers to better understand the customer's intention process. Finally, it should be noted that future research might benefit from adding external causes to explain customer intentions, such as drastic changes in the customer's personal situation (e.g. moving to an other country) which might be relevant to the customer's purchase behavior.

Part II

Information Driven Knowledge Discovery

8 PSO driven collaborative clustering: a clustering algorithm for ubiquitous environments

8.1 Introduction

All previous chapters in this thesis were based on the Expectancy-Disconfirmation Paradigm. This chapter leaves that path and considers a problem where a marketer is faced with another type of incomplete data and where the ED Paradigm is of no use. In contrast with the first part of this thesis, the data mining technique introduced in this chapter will not start from an existing marketing theory, model or paradigm but will extract knowledge from information within incomplete data by letting the data speak for itself. The benefit of a information driven knowledge discovery technique is that it can be much more flexible and powerful because it is not constrained by assumptions imposed by theory. On the other hand, data mining techniques which do not start from a well-accepted theory to extract knowledge from data is more susceptible to learn noise. Therefore, with techniques such as presented in this chapter, one must always carefully study the face validity of the extracted knowledge. Much more than with theory driven techniques, expert knowledge is required to distinguish results which reflect true knowledge and results which are sampling artifacts or noise.

The marketing problem faced in this chapter is one of customer segmentation. Traditionally, marketing research has employed clustering analysis to find meaningful segments within a customer base which allows companies to target their customers in a more efficient way [135]. Cluster analysis is a unsupervised data mining technique which groups objects in clusters such that the distance between objects of the same cluster is minimized and the distance between objects from different clusters is maximized [59]. Not only have clustering algorithms been used in marketing, but they are also popular in other domains such as engineering, computer sciences, medical sciences, earth sciences, social sciences and traffic safety research [135, 34, 8, 45, 56, 60, 71, 92, 109]. Many clustering algorithm assume that researchers have the necessary data available at their computer on which they can run the clustering algorithm.

However, sometimes companies might believe that their customer data is too limited to perform a meaningful clustering analysis. The following two motivating examples describe such situations where data is incomplete. These two examples illustrate the type of applications for which the clustering algorithm in this chapter is designed.

Motivating example 1.

A company holds information on a set of potential customers which it wishes to segment. This allows them to identify new opportunities and act appropriately. At the same time, other companies hold other information on the same set of potential customers and have similar needs to identify different segments. Due to privacy, security or business reasons, these companies are unwilling or prohibited to exchange their data. This prevents them from consolidating their data and performing analysis on the enriched data set. Yet an overall discovery of common patterns through some collaboration mechanisms enforced over the companies could be highly profitable in contrast to a confined discovery of local knowledge structures (clusters). In some sense, these companies could be regarded as members of a peer-to-peer ubiquitous environment where data and computing power are distributed. They might benefit from a KDUbiq clustering algorithm which allows them to segment the customers by using local data and findings coming from other companies without violating privacy, security or business constraints.

Motivating example 2.

Two companies retain the same type of information about their customers. Both companies have a different customer base and are not necessarily active in the same market. They want to segment their customers to identify customer stereotypes and they expect the stereotypes of the first company to show some overlap with the stereotypes of the second company. To increase the validation of specific customer stereotypes found by cluster analysis, both companies would prefer to enrich their own data with the data from the other company. However, privacy, security and business restrictions prevent them from exchanging their data. Yet again, an overall discovery of common patterns through some collaboration mechanisms enforced over the companies could be highly profitable in contrast to a confined discovery of local knowledge structures (clusters). The companies can be considered as members of a peer-to-peer ubiquitous environment where data and computing power are distributed and where both could benefit from an adequate KDUbiq clustering algorithm.

Both examples illustrate an environment where data is distributed such that the data at each site can be considered as incomplete. This chapter, introduces a new clustering algorithm which is a combination of particle swarm optimization (PSO) and collaborative fuzzy clustering (CFC) and which is particularly suited for this type of problems. This algorithm, denoted as PSO-CFC, is a modified version of the technique introduced by Falcón et al. [46] and can be interpreted as a Ubiquitous Knowledge Discovery (KDUbiq) clustering technique. KDUbiq is a fairly new research domain which is shortly introduced in the next section. The third section of this chapter continues by discussing other relevant literature on related existing data mining techniques from domains such as distributed clustering. The fourth section is devoted to a discussion of the original CFC algorithm, introduced by Pedrycz [104, 105] while the fifth section introduces PSO and shows how it cooperates with CFC in this KDUbiq environment. The sixth section contains the results of elaborated experiments which test the added value of the PSO-CFC clustering technique compared to local fuzzy clustering (i.e./ without collaboration between data sites). Finally, the limitations of the current technique and the directions for future research will be discussed before the final conclusions are drawn.

8.2 Ubiquitous knowledge discovery

Computing environments and technologies are increasingly evolving towards mobile, finely distributed, interacting and dynamic environments containing massive amounts of heterogeneous, spatially and temporally distributed data sources. Examples are peer-to-peer systems, grid systems and wireless sensor networks. These ubiquitous computing environments pose new challenges to the field of knowledge discovery and data mining which remain unsolved by traditional data mining techniques. The research discipline fostering the inception of innovative data mining methodologies which are capable to handle these new challenges has been termed Ubiquitous knowledge discovery (KDUbiq).

KDUbiq is a very wide research area with features setting it aside from traditional data mining and distributed data mining. KDUbiq algorithms typically operate in environments with distributed computing power and distributed data sources. KDUbiq algorithms must be capable of communication with different data sources and computing sites and should also be usable in environments too large to set up a master computer which collects and consolidates the different site results. Therefore, a KDUbiq environment typically consists of several computing devices performing local data mining on site using limited information at hand while communicating with other sites. Sometimes computing and power resources are limited and require resource-aware KDUbiq techniques. Sometimes, privacy and security restrictions hold which prevent raw data to be communicated or to be gathered centrally. Sometimes, KDUbiq algorithms have to deal with data streams and must be able to process the data in real-time. Not all of these features have to be present at the same time for an environment to be ubiquitous and neither does a KDUbiq algorithm necessarily have to deal with all these challenges. It all depends on the application. A sensor network will e.g. need algorithms which are very much resource-aware and limit communication to the absolute minimum while grid systems have much less resource constraints. On the other hand, KDUbiq algorithms that work on a peer-to-peer network are more likely to have to deal with privacy issues than algorithms working in a sensor network. For a more elaborated discussion of the different KDUbiq characteristics, the interested reader may refer to the KDUbiq Blueprint [1] which will be available as a Springer book by Fall 2008. Next, two motivating examples are given to illustrate the type of applications our KDUbiq algorithm can deal with and to identify the KDUbiq challenges faced in such applications.

If the motivating examples from the previous section are revised, it becomes clear that they can be considered as ubiquitous environments with distributed data sources and distributed computing power where privacy issues prevent the exchange of raw data. Such environments require a clustering technique which can run locally, but which can find similar results as if all data were consolidated at one site, without violating the privacy restrictions. Other KDUbiq features, such as resource limitations and real-time data mining are of much less importance in these types of applications and it should be noted that PSO-CFC is not designed to consider these challenges.

8.3 Relevant work

Not only can PSO-CFC be considered a KDUbiq clustering algorithm, it shares various properties with other clustering algorithms which are often studied under the umbrella of distributed or parallel computing. While distributed and parallel computing are closely related, there are some important differences to be made. Parallel clustering starts with a centralized data set and tries to increase clustering performance. It distributes the data among different clients and all clients cooperate intensively to simulate a global clustering. Examples of parallel clustering are the parallel version of DBSCAN in [136] and the parallel version of k-means in [35]. Distributed clustering on the other hand assumes that the data is not centrally available and the different clients work independently from each other. Each client learns a local cluster model which is sent to a central node where the local models are combined to a global cluster solution. Examples of distributed clustering are the Collective Hierarchical Clustering algorithm [72], the distributed k-windows algorithm [123] and the Density Based Distributed Clustering [70] among others. From this perspective, PSO-CFC belongs to the domain of distributed clustering.

Within the field of distributed clustering, it is important to discern the way the data is distributed. Horizontally partitioned data means that each party has the same set of records/instances but a different set of features/attributes. In [72], Johnson and Kargupta introduce a distributed clustering algorithm which tackles this type of distributed data. Vertically partitioned data refers to data distributed among various parties such that each party contains a joint set of features/attributes but a disjoint set of records/instances. Some algorithms which tackle this type of data are presented in [35, 123].

Both among vertically and horizontally distributed clustering techniques, two different motivations can be discovered behind the various methodologies aimed at amalgamating information from multiple clustering analyses such as multi-cluster combiners [10], consensus clustering [91] or the well-known cluster ensembles [119, 126]. A first reason for this fusion of knowledge structures is to assess the robustness of a specific clustering algorithm to the variability of sampled data, which is a pivotal issue in noisy and uncertainty-permeated environments. The representative approach in this category is consensus clustering [91], i.e./ a rather simple resampling technique executing the same clustering method multiple times over a set of perturbed instances of the same original data set. Finally, it attains an agreement or consensus among the multiple runs.A second motivation behind these approaches stems from the need to combine the output of manifold unsupervised learning algorithms, following the footsteps of well-settled meta-classifier techniques, as a way of getting a coherent picture of the underlying data dynamics which exists across the different data sites. From this perspective, PSO-CFC belongs to the second group of distributed clustering techniques which try to discover the global structure across various data sites.

However, most of the algorithms mentioned so far were not specifically designed with the concept of privacy preservation in mind. Recently, some authors have focussed their attention on privacy preserving distributed clustering algorithms, which have resulted in several publications [119, 86, 128].

The privacy-preserving distributed clustering approach “cluster ensembles” [119] has proven to be an efficient and general framework for mapping a set of hard partitions into an optimal, combined partition, even when the number of clusters in each partition may vary. Assuming a unique, global collection of patterns (possibly distributed over several feature spaces), data confidentiality at the local level is enforced by exchanging only the crisp labels corresponding to the cluster assignments of each data object. Given the high computational complexity of the ensuing “median partition" problem as posed in [119], the need for heuristic methods to derive approximate solutions of an acceptable quality becomes relevant to the topic. This is less of an issue with PSO-CFC, where good solutions to the clustering guided by the augmented objective function can be obtained with relatively modest computational effort.

A second type of privacy-preserving distributed clustering is presented in [86] and is concerned with distributed object collections which are described by the same group of features, i.e./ vertical clustering. The proposed framework is general enough to embrace unsupervised and semi-supervised scenarios and supports a broad range of data types and learning algorithms. However, it suffers from several drawbacks compared to collaborative fuzzy clustering, among them the assumption about the existence of an underlying probabilistic distribution of the information dwelling at each local repository. Such distributions are to be learned (e.g. Gaussian with full-variance, von Mises-Fisher, etc.) and their condensed parameters are sent to a central location for further aggregation into a unified model. However, the quantification of the privacy requirements or the weights associated with the local models makes room for subjective criteria which lead to undesirable effects at the global level.

A third privacy-preserving distributed clustering technique is the fine-grained distributed version of the very popular K-Means algorithm with full privacy preservation put forward in [128] which has a strong analogy with collaborative fuzzy clustering. It targets the horizontal mode (i.e./ feature-scattered clustering) in such a way that an elaborate degree of isolation in the communication phase among the data sites has been envisioned. As part of the secure multiparty computation approach enforced to prevent exchange of valuable data, not even cluster prototypes or distances between patterns to clusters are disclosed, which turns the algorithm more complicated and awkward. A problem with the algorithm in [128] is the way it computes the closest cluster for each data point, which is believed to be the sum of the partial distances found at the local level. While this is true for the most common distance measures (e.g. Euclidean, Manhattan, etc.), this feature severely restricts the shape of the clusters to be found and thus narrows the scheme's applicability to non-traditional scenarios.

A major difference between PSO-CFC and the other privacy-preserving distributed clustering techniques is that the former assigns fuzzy cluster labels to observations. With fuzzy cluster labels an observation can belong to several clusters with different degrees of membership. Such approach provides a more powerful representation of the reality where crisp memberships are more the exception than the rule.

8.4 Collaborative fuzzy clustering

PSO-CFC is based on Collaborative Fuzzy Clustering (CFC) which was introduced in 2002 by Pedrycz [104] as a novel clustering algorithm intended to reveal the overall structure of distributed data which at the same time respects any restrictions preventing data sharing. As discussed in the previous section, this approach exhibits both similarities and differences with other distributed clustering techniques (cf [105]).

The collaborative clustering scheme consists of two major steps. First, a local clustering analysis is performed at each individual data site separately. Next, the local findings are exchanged and an augmented clustering algorithm is applied at each data site. This augmented clustering algorithm takes the local data as well as the results of the other sites into account. The second step is repeated until some termination criterium is met.

Typically, two types of collaborative clustering can be distinguished, i.e./ the horizontal mode and the vertical mode. The horizontal mode assumes that each data site holds information on the same set of objects but described in different feature spaces, as is the case in the first motivating example. The vertical mode assumes that each data site holds information on different objects described in the same feature space. This is the case in the second motivating example. Next, the two versions of CFC will be explained in detail with the motivating examples in mind.

Horizontal CFC

There are P companies and each company $[ii]$ measured $A[ii]$ variables $x_{a[ii]}$ for the same set of N customers. The companies agreed to segment their customers in C clusters. The first step of the collaborative scheme performs a local fuzzy C-means cluster analysis (FCM) at each company $[ii]$ separately using only the local data. The generic version of the FCM method was proposed by Dunn [40] and Bezdek [15] in the 1980s, but has undergone significant changes over the years. The reader may refer to Hoppner et al. [64] for a comprehensive reference on this topic.

FCM identifies C cluster centers and assigns each record k (i.e./ a customer in our case) with a specific membership degree u_{ik} to cluster i. The membership degrees u_{ik} for $i = 1, \cdots, C$ are constrained to sum to 1. The FCM analysis tries to minimize the objective function

$$Q[ii] = \sum_{k=1}^{N} \sum_{i=1}^{C} u_{ik}^2[ii] d_{ik}^2[ii] \tag{8.1}$$

where d_{ik} denotes the distance between case k and cluster center i and where $[ii]$ refers to one of the companies at which the local analysis is performed.

The local analysis provides each company with a $C \times A[ii]$ cluster prototype matrix containing the cluster centers and a $N \times C$ partition matrix containing the membership degrees of each case k to each cluster i. Note that the size of the cluster prototypes differ across data sites because each site $[ii]$ has a different feature set size $A[ii]$.

In the second stage of CFC, the companies have to exchange their local results, without violating the privacy restrictions. In horizontal CFC the communication between data sites is realized by exchanging the partition matrices. The partition matrices are comparable between data sites in horizontal CFC because they relate to the same customers and clusters. At the same time, no private information about the customers is exchanged and without the prototype matrices, which are not exchanged, it is impossible to retrieve the original data. By realizing the communication at the level of granular information, collaborative clustering succeeds in complying to any privacy, security or business constraints.

Once the companies receive the partition matrices, the true collaborative FCM can be applied, which minimizes an augmented objective function (cf. Eq. 8.2). This function integrates the information from the other companies with the local data and uses collaboration links $\alpha[ii, jj]$ to control the extent of collaboration between two companies $[ii]$ and $[jj]$. The set of all collaboration links is called the collaboration matrix.

$$Q^*[ii] = Q[ii] + \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii, jj] \sum_{k=1}^{N} \sum_{i=1}^{c} (u_{ik}[ii] - u_{ik}[jj])^2 d_{ik}[ii] \tag{8.2}$$

The augmented objective function adds a second term to the goal function which quantifies the differences between the partition matrices of every data site. This forces the CFC algorithm to search for similar clustering results across data sites. During this collaborative stage, the entries of the partition matrix and prototype matrix will be recomputed. Next, the new partition matrices will be exchanged and each company will minimize the augmented objective function again. This is repeated until some termination criterion is reached, which relies on the changes of the partition matrices obtained in successive iterations of the clustering method. Algorithm 1 displays the breakdown of the horizontal collaborative clustering scheme.

Algorithm 1 The horizontal collaborative clustering scheme

for each data location $[ii]$ **do**
    Perform standard FCM clustering, minimizing objective function $Q[ii]$
end for
repeat
    Exchange the current partition matrices between the data locations
    for each data location $[ii]$ **do**
        Run the collaborative FCM clustering, minimizing $Q^*[ii]$
    end for
until some termination criterion is reached

Vertical CFC

In the vertical version of CFC, there are P companies and each company $[ii]$ has a different set of $N[ii]$ customers which are all measured by the same set of A features $\{x_1, \ldots, x_a, \ldots, x_A\}$. The companies agreed to segment their customers in C clusters. The first step of the collaborative scheme performs a local fuzzy C-means cluster analysis (FCM) for each company $[ii]$ using only the local data. This step is exactly the same as for the horizontal CFC variant. The local analysis provides each company with a $C \times A$ cluster prototype matrix containing the cluster centers and a $N[ii] \times C$ partition matrix containing the membership degrees of each case k to each cluster i. Note that this time the size of the partition matrices differ across data sites because each site $[ii]$ has a different set of $N[ii]$ customers.

In the vertical CFC, the prototype matrices are exchanged instead of the partition matrices. Since the feature sets are equal across data sites, the cluster prototypes can be compared among data sites which allows the measurement of the difference between the local cluster solutions. Only prototype matrices are exchanged which preserves the privacy of the data.

Once the companies receive the prototype matrices, the true collaborative FCM can be applied, which minimizes an augmented objective function (cf. Eq. 8.3). As in the horizontal version, a new term is added to the goal function which quantifies the inequality between the cluster solutions at different data sites. The second term of the augmented function compares the local membership degrees $u_{ik}[ii]$ with the membership degree $u_{ik}[jj]$. This is the membership a case k would have if cluster center i was positioned at the location of cluster center i from data site $[jj]$. The vertical version also uses collaboration links $\alpha[ii,jj]$ to control the extent of collaboration between two companies $[ii]$ and $[jj]$.

$$Q^*[ii] = Q[ii] + \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii,jj] \sum_{k=1}^{N} \sum_{i=1}^{c} (u_{ik}[ii] - u_{ik}[jj])^2 d_{ik}[ii] \tag{8.3}$$

During the collaborative stage, the entries of the partition matrix and prototype matrix will be recomputed and the new prototype matrices will be exchanged. Next, each company will minimize their augmented objective function again and this is repeated until some termination criterion is reached. Algorithm 2 displays the breakdown of the vertical collaborative clustering scheme.

Algorithm 2 The vertical collaborative clustering scheme

for each data location $[ii]$ **do**
    Perform standard FCM clustering, minimizing objective function $Q[ii]$
end for
repeat
    Exchange the current prototype matrices between the data locations
    for each data location $[ii]$ **do**
        Run the collaborative FCM clustering, minimizing $Q^*[ii]$
    end for
until some termination criterion is reached

8.5 Optimizing the collaboration matrix

One of the parameters of CFC is the collaboration matrix which can be set by using expert knowledge. Company experts should set high collaboration links with companies they want to cooperate strongly with and low collaboration links with companies whose data is not believed to be very compatible. Choosing the right collaboration links can be a difficult task which could lead to unbalanced results if chosen incorrectly. There is no guarantee that collaboration will yield meaningful results no matter how strong the connection between two companies might be.

Falcón et al. [46] developed a technique to optimize the collaboration matrix during the clustering analysis by applying the evolutionary optimization technique of Particle Swarm Optimization (PSO). Their objective was to a maximize the level of collaboration which differs from the objective in this chapter. The goal of the PSO-CFC approach is to mimic the situation where all companies would consolidate their data sets and perform a single global cluster analysis, which is impossible due to privacy constraints. Therefore, the PSO-CFC approach will need a PSO objective function which focuses on finding similar cluster solutions across companies. This section presents a modification of the PSO objective function to suit the needs of this KDUbiq environment.

It should be noted that the KDUbiq approach only makes sense if companies consider their own data as incomplete and want to improve the quality of their local analysis by collaborating with other companies which they believe have compatible data. If the data from other companies are not compatible or relevant or companies do not want to dilute the knowledge structures found by local analysis in favor of the knowledge structures revealed through a global clustering approach, companies should stick with their local analysis.

Particle Swarm Optimization (PSO) is an evolutionary optimization technique developed by Kennedy and Eberhart [74], inspired by the swarming behavior of bird flocks and fish schools. The optimization algorithm initializes Z particles x_z, each representing a possible solution to the optimization problem. Next, the particles start to fly throughout the solution space and at each time interval t, the fitness of the solution is evaluated by means of a fitness function. During their flight, each particle remembers its own best position p_z. The direction of a particle in the solution space is influenced by the particle's current location $x_z(t)$, the particle's current velocity $v_z(t)$, the particle's own best position p_z and the global best position p_g among all particles. The particle's new position $x_z(t+1)$ is calculated by Eq. 8.4 and Eq. 8.5:

$$v_z(t+1) = wv_z(t) + c_1 r_1 (p_z - x_z(t)) + c_2 r_2 (p_g - x_z(t)), \tag{8.4}$$

$$x_z(t+1) = x_z(t) + v_z(t+1), \tag{8.5}$$

where w is the inertia weight and c_1, c_2 are the acceleration constants drawing the particle toward the local and global best locations, respectively. The stochastic components of the PSO meta-heuristic are given by r_1 and r_2, which represent two uniformly distributed random numbers. All particles keep moving in the solution space until some criterion is met. The global best position in the end is the solution to the optimization problem. For a broader insight about this widespread optimization technique, refer to [16].

In the PSO-CFC approach, a single particle will represent an entire collaboration matrix and the flight of the particles represents the search for a collaboration matrix which optimizes the similarity of the cluster solutions across data locations. To achieve such optimization, an appropriate fitness function needs to be developed which represents the average dissimilarity between cluster solutions across data sites. The goal of the PSO algorithm is to minimize this function.

The fitness function can be defined in three steps. The first step, which is different for the horizontal and vertical variant, measures the dissimilarity between cluster i from data site $[ii]$ and cluster j from data site $[jj]$. In the horizontal variant, cluster dissimilarity must be based on the partition matrices which is the only information available about the clusters from the other data sites. In horizontal PSO-CFC, a cluster $C_i[ii]$ is redefined as a set of membership degrees $\{u_{1i}[ii], \cdots, u_{Ni}[ii]\}$ and the dissimilarity between cluster i from data site $[ii]$ and cluster j from data site $[jj]$ are measured as follows:

$$d(C_i[ii], C_j[jj]) = \frac{1}{N} \sum_{k=1}^{N} |u_{ik}[ii] - u_{jk}[jj]|. \tag{8.6}$$

This dissimilarity measure becomes zero, which is the lower bound, when all patterns belong to both clusters with equal membership degree. On the other hand, it will become 1, which is the upper bound, when both clusters are crisp and do not have any pattern in common.

In the vertical variant, dissimilarity between two clusters must be based on the prototype matrices which is the only information known from the cluster solutions at the other data sites. In vertical PSO-CFC, the dissimilarity between cluster i from data site $[ii]$ and cluster j from data site $[jj]$ is measured by calculating the Euclidean distance between the two cluster centers $v_i[ii]$ and $v_j[jj]$ (cf Eq. 8.7 where $v_{ia}[ii]$ represents the a^{th} feature of cluster center $v_i[ii]$).

$$d(C_i[ii], C_j[jj]) = \sqrt{\sum_{a=1}^{A} (v_{ia}[ii] - v_{ja}[jj])^2} \tag{8.7}$$

This distance measure has a lower bound equal to 0 which is reached when both cluster centers are at the exact same location in the feature space. No upper bound exists.

The second step is to measure the average dissimilarity between the clusters of data site $[ii]$ and data site $[jj]$. This requires a proper mapping between the clusters of data site $[ii]$ and the clusters of data site $[jj]$. Instead of developing an algorithm to perform this non-trivial mapping, our approach uses a simple heuristic which appears to work very well. This heuristic maps a cluster i from data site $[ii]$ to the least dissimilar cluster j from data site $[jj]$. Dissimilarity is measured with Eq. 8.6 for horizontal PSO-CFC and Eq. 8.7 for vertical PSO-CFC. The average dissimilarity between two data site $[ii]$ and $[jj]$ is calculated by means of Eq. 8.8. Note that this measure equals 0 when both cluster solutions are identical.

$$D[ii, jj] = \frac{1}{c} \sum_{i=1}^{c} \mathop{\mathbf{Min}}_{j=1}^{c} \left[d(C_i[ii], C_j[jj]) \right] \tag{8.8}$$

The third and final step of constructing the fitness function represents the level of dissimilarity present between all data sites. The PSO fitness function, which will be termed ρ, measures the average dissimilarity of all data sites pairs. With P data sites, there are $\frac{P(P-1)}{2}$ data site combinations, which results in the following fitness function:

$$\rho = \frac{2}{P(P-1)} \sum_{ii=1}^{P} \sum_{jj>i}^{P} D[ii, jj] \tag{8.9}$$

The PSO-CFC algorithm uses this fitness function to determine the optimal set of collaboration links. In the KDUbiq clustering setting, this implies that aside from data locations, which are called data nodes, a computing location is needed which performs the PSO algorithm. This location will act as the coordination node. It should be noted that the coordination node can be the same physical location as a particular data node, but this is not necessary. Algorithm 3 shows how the collaborative clustering scheme and the particle swarm optimization are integrated to automate the determination of the collaboration links.

Algorithm 3 The horizontal collaborative clustering scheme

Initialize Z particles x_z *(coordination node)*
repeat
  for each particle x_z **do**
    Perform Alg. 1 or 2 with collaboration matrix x_z *(data nodes)*
    Send the partition matrices to the coordination node
    Calculate the fitness function ρ *(coordination node)*
    Update p_z *(coordination node)*
  end for
  Update p_g *(coordination node)*
  for each particle x_z **do**
    Calculate the new position $x_z(t+1)$ *(coordination node)*
    Send $x_z(t+1)$ to the data nodes
  end for
until some termination criterion is reached *(coordination node)*
Send the optimal collaboration links to the data nodes
Perform Alg. 1 or 2 with the optimal collaboration matrix *(data nodes)*

8.6 Experiments

Methodology

Each analysis compares three different clustering approaches, i.e./ the global clustering (GC) approach, the local clustering (LC) approach and the collaborative clustering (CC) approach, and requires a separate data set per data site for the CC and LC approach and a consolidated data set for the GC approach. The LC approach performs a standard FCM clustering for each data set separately. This represents the situation where companies only have access to their own data and are not willing to collaborate with other companies. The GC approach represents the other extreme where no privacy constraints hold and where companies consolidate their data sets to achieve a single data set of higher quality, i.e./ a larger feature space or more observations. This approach is tested by performing a FCM cluster analysis on the consolidated data set. The CC approach represents the KDUbiq environment described in the motivating examples where privacy constraints prevent companies from sharing their data. However, in contrast to the LC approach, where companies work as isolated sites, the CC approach uses a collaboration mechanism to approximate the results of the GC approach. The CC approach is tested by performing a PSO-CFC clustering analysis across all data sites with the following parameters: 20 particles, 100 iterations, $c_1 = c_2 = 2.0$ and the inertia weight dynamically varied from 1.4 to 0.4. The purpose of each analysis is to evaluate the quality of the clustering results from all three approaches. In this chapter, cluster quality is defined as the number of correctly assigned cases. Analyses are performed on both artificial data sets and real-life data sets.

Artificial data sets have the benefit of keeping full control over the structure of the data, which allows the researcher to isolate the effect of a specific data structure property on cluster quality. It also has the benefit that the true cluster memberships are known which allows an exact evaluation of the cluster quality. Once the artificial data structure is designed, a sample can be drawn for each data site and the consolidated data set is created by joining all local data sets. Next, the three approaches are applied to the data sets and the number of incorrectly assigned observations are counted for each approach. Cluster assignment is based on the cluster with the highest cluster membership.

Assume that the cluster quality of the GC approach has to be compared against the cluster quality of the CC approach. This comparison is done by subtracting ϵ_{cc}, i.e./ the number of errors of the CC approach, from ϵ_{gc}, i.e./ the number of errors of the GC approach, which results in the new variable $d = \epsilon_{cc} - \epsilon_{gc}$. Note that one experiment leads to a single observation of d which cannot lead to reliable conclusions. Therefore, for each analysis, 30 experiments are performed. Each experiment z draws new data from the artificial data distributions and d_z is calculated. This results in a sample of cluster quality comparisons $\{d_1, \ldots, d_z, \ldots, d_{30}\}$. Ultimately, inferences about the population mean μ_d need to be drawn which indicates if CC performs better than GC on average or not. To draw conclusions about μ_d, an estimator of this population parameter is needed, which is typically the sample average $\overline{d} = \frac{\sum_{z=1}^{30} d_z}{30}$. However, the true population mean μ_d might differ to some extent from its estimator $\overline{d}$. Therefore, confidence intervals need to be constructed to know how reliable $\overline{d}$ is as an estimator of μ_d. To build such confidence intervals, the sampling distribution of $\overline{d}$ around μ_d has to be known.

If the underlying variable d is normally distributed, the t-test could be used to evaluate the sample mean and to construct confidence intervals [114, 29]. However, the underlying distribution of d is unknown and the sample size is not very large in the experiments in this chapter ($z = 30$) which makes the t-test more sensitive to violations of the normality assumption. Instead of relying on the robustness of the t-test, it was opted to use the non-parametrical bootstrap technique to construct confidence intervals.

Nonparametric bootstrap [42] is a recently fashionable way for statistical inference for quantities for which theoretical and/or even asymptotic results are hard to derive. The basic idea behind bootstrapping is that new samples S_b^* are created by resampling with replacement from the original sample $S = \{d_1, \ldots, d_z, \ldots, d_{30}\}$ until the new sample size is equal to the original sample size. This process is repeated a number of times which results in a set of B resamples, denoted as $\{S_1^*, \ldots, S_b^*, \ldots, S_B^*\}$. The key idea is that all these resamples can be considered as samples from the unknown population (or at least they look like the unknown population).

If the sample average $\overline{d}$ based on sample S_b^* is denoted as $\overline{d_b^*}$, then the distribution of $\overline{d_b^*}$ around $\overline{d}$ is analogous to the sampling distribution of $\overline{d}$ around the population mean μ_d [52]. Since B, i.e./ the number of resamples, can be made very large, a very detailed empirical distribution of $\overline{d_b^*}$ can be acquired which provides a detailed estimate of the sampling distribution of $\overline{d}$. This estimate of the sampling distribution can be used to construct confidence intervals around $\overline{d}$. Various approaches to construct bootstrap confidence intervals exist, such as the *normal-theory interval*, the *bootstrap percentile interval* and the *bias-corrected, accelerated percentile intervals* (BC_a). According to Fox [52], the latter are preferable and for a 95% BC_a confidence interval, the number of bootstrap samples should be on the order of 1000 or more. For more technical details about the construction of bootstrap confidence intervals and the use of bootstrapping to evaluate the results of machine learning algorithms, the interested reader should refer to [52, 29].

In summary, the following methodology is used for each analysis:

a) An artificial data distribution is designed.

b) For each of 30 experiments, samples from the artificial data distribution are drawn for each data site and the consolidated data set is created.

c) For each experiment, the measurements of interest are constructed and the average for the thirty experiments is calculated.

d) Bootstrapping is used to generate BCa confidence intervals, based on 10000 bootstrap samples.

Artificial data sets give researchers full control over the structure of the data which allows them to assign changes in cluster performances to specific causes, similar to a controlled laboratory experiment. On the other hand, artificial data sets might not always reflect reality which decreases the practical value of the results and conclusions. Therefore, some analysis were conducted on a real-life marketing data set. A first problem with real data is that the true cluster memberships of the customers are unknown. In the experiments in this chapter, the cluster assignments of the GC approach were used as the estimates for the true cluster memberships. The quality of the CC and LC approach is measured by counting the number of customers they assign differently than the GC approach.

A second problem is that there is only one data set and the underlying distribution is unknown which prevents the generation of new samples. Therefore, the bootstrap principle was used and new data sets were generated by sampling from the original data set with replacement until the size of the new data sets are equal to the original data set size. According to the bootstrap principle, these resamples can be considered as samples from the unknown population (or at least they look like the unknown population).

The subsequent steps of calculating the average quality difference between CC and LC and the construction of confidence intervals is completely analogous to the methodology followed for the artificial data sets.

Horizontal clustering: empirical results

Artificial data sets.

First, three different analysis were performed on artificial data to evaluate the relative quality of the LC and CC approach in different situations. Each analysis uses different data structures, which represent increasing clustering task complexity. First the data structure of each analysis will be discussed, prior to the results.

All three analyses concern two data sites. Each data site has a data set from three clusters, i.e./ C_1, C_2 and C_3, with 200 observations in each cluster. The difference between the data sites are the features used to describe the observations. The consolidated data set is the combination of the local data sets and has four features, i.e./ X_1, X_2, X_3 and X_4.

The first analysis represents the easiest clustering task. It uses cases drawn from three different multivariate normal distribution, i.e./ one per cluster, with mean vector

μ_1, μ_2 and μ_3 for respectively clusters C_1, C_2 and C_3 and with the same covariance matrix Σ for all three clusters.

$$\mu_1 = \begin{bmatrix} 3 & 1 & 1 & 2 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} 1 & 3 & 3 & 2 \end{bmatrix}$$

$$\mu_3 = \begin{bmatrix} 2 & 3 & 1 & 2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}$$

Figure 8.1 shows the 2-dimensional scatterplots for a random data set generated from the artificial data structure discussed above. These plots show that features X_2 and X_3 are sufficient to separate the three clusters while feature X_4 does not help the identification of the three clusters at all. In the first analysis, data site A has access to features X_1, X_2 and X_3 and data site B has features X_3 and X_4. Figure 8.1d reveals that data site A should have no problem assigning the cases to the correct cluster, while Figure 8.1f shows that it is impossible for site B to separate clusters C_3 and C_1. The first analysis simulated a KDUbiq environment where one company has all necessary features to solve a perfectly separable clustering problem, while the second company lacks important features.

The situation simulated in the second analysis differs from the first analysis because in the second analysis no single company has all necessary features to solve the clustering problem. However, the clusters can still be separated for the consolidated data set. The data sets generated for the second analysis use the same distribution as the first analysis, but this time data site A has only access to features X_1 and X_2, while data site B has access to features X_3 and X_4. Figure 8.1a shows that data site A can no longer separate the three clusters as there is some overlap between clusters C_2 and C_3, while Figure 8.1f shows that data site B still has problems separating clusters C_3 and C_1.

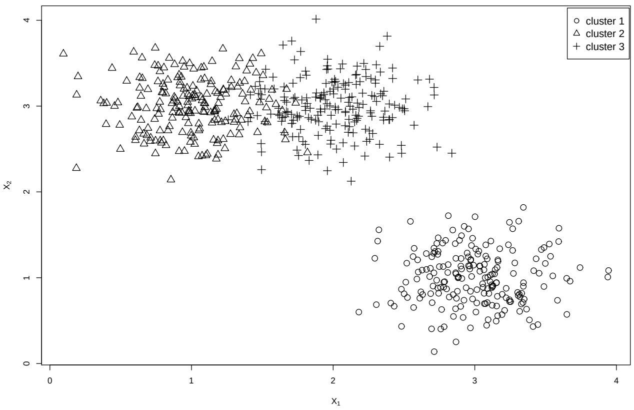
The third analysis represents the same situation as analysis 2, but now for a problem which is not perfectly separable, even if all features are available. This is a more realistic situation. The data generated for the third analysis still uses three multivariate normal distributions with the same covariance matrix as above, but with the following mean vectors for respectively clusters C_1, C_2 and C_3:

$$\mu_1 = \begin{bmatrix} 3 & 1.5 & 1 & 2 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} 1 & 2 & 1.5 & 2 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 2 & 2 & 1 & 2 \end{bmatrix}$$

Analogue to analysis 2, data site A received features X_1 and X_2, while data site B has access to features X_3 and X_4.

For all three analysis, the methodology described in subsection 8.6 was used and the results are presented in Table 8.1. The ϵ statistics refer to the number of incorrectly assigned cases in a certain data site by a specific approach. For example, ϵ_{lc}^{AB} are the number of incorrectly assigned cases for data site A and B by the LC approach. The ϵ statistics for the GC approach are always calculated for the

(a) X_1 vs X_2



(b) X_1 vs X_3

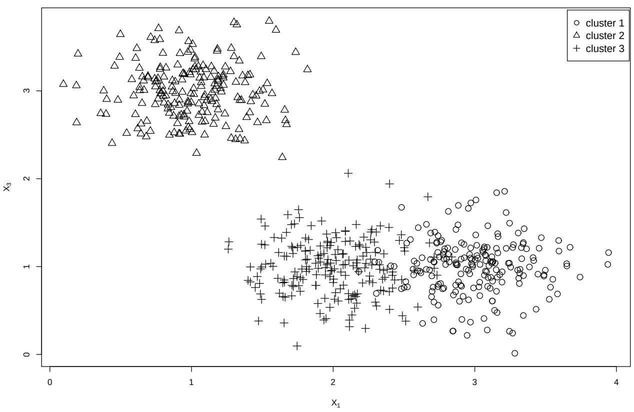
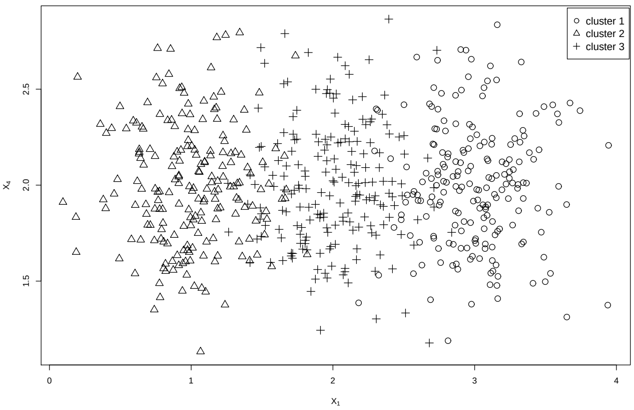


Figure 8.1: Scatterplot of full artificial data set

consolidated data set and the superscript denoting the data sites is left out of the notation.

Analyzing the results of the first analysis, the LC approach succeeded in assigning the cases of the first data site to the correct clusters, while it had great troubles assigning the cases of the second data site. These results were expected given the data structure of both sites. The CC approach did succeed to overcome the clustering problem of the second data site. The information exchanged during the collaboration phase of the CC approach successfully lowered the number of incorrectly assigned cases at data site B to 14 cases out of 600, compared with 202 cases out of 600 for the LC approach. On the other hand, it is remarkable that the CC approach did make assignment errors for data site A, which should be perfectly separable. Further

(c) X_1 vs X_4



(d) X_2 vs X_3

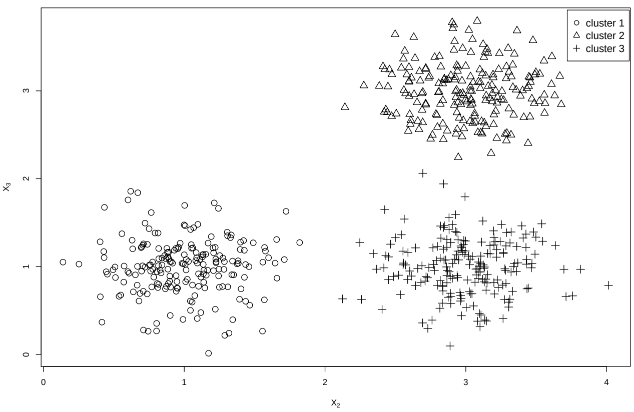
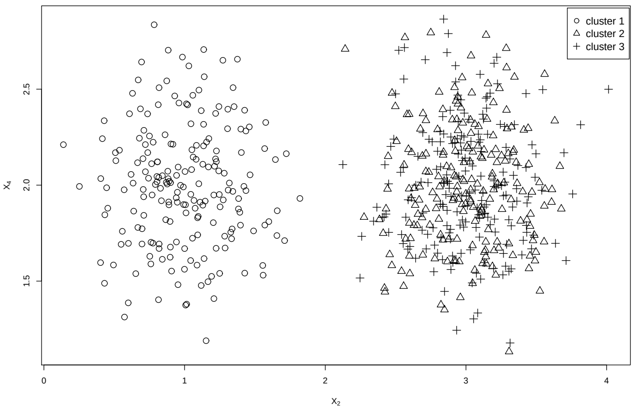


Figure 8.1: Scatterplot of full artificial data set [continued]

analysis of the results revealed that the PSO-CFC algorithm got stuck in a local optimum during three of the thirty experiments which caused these errors. Initially, the PSO-CFC algorithm had even more problems getting stuck in local optima. By changing the default number of particles to 40, the number of local optima problems could be reduced, but could not be eliminated completely. These adjusted settings were used for all horizontal PSO-CFC experiments. Comparing the CC and LC approach, the results are very promising. When one data site contains all information to segment the customers correctly, other data sites clearly benefit by applying the CC approach.

The second analysis changed the environment such that no single data site has all information available to solve a separable cluster problem. This is reflected in the

(e) X_2 vs X_4



(f) X_3 vs X_4

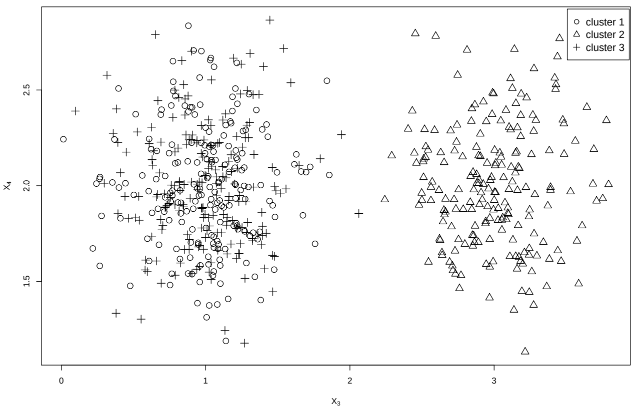


Figure 8.1: Scatterplot of full artificial data set [continued]

results. While the LC approach did not make assignment errors for data site A in the previous analysis, it now makes 20 assignment errors on average. Note that these errors are caused by the fact that data site A is no longer perfectly separable into three clusters. The number of errors made by the LC approach for data site B remains the same as in the first analysis, which was expected since it concerns the same data structure. Also in the second analysis, the CC approach succeeded to cluster the observations of data site B much better than the LC approach, i.e./ 26.4 incorrect cluster assignments versus 199.6 incorrect cluster assignments. However, the number of errors made on data site B by the CC approach is higher than in the first analysis, although the data structure of data site B did not change and the problem is still perfectly separable given all features. Furthermore, the number of errors made by the

Table 8.1: Cluster quality comparisons between the GC, LC and CC approach in a horizontal clustering environment on artificial data

		Analysis		
Statistic		1	2	3
ϵ_{LC}^{A}	mean	0.2	21.7	84.6
	95% CI	[0.1,0.4]	[20.4,22.9]	[79.9,90.7]
ϵ_{LC}^{B}	mean	202.1	199.6	307.7
	95% CI	[200.9,203.3]	[197.8,201.3]	[302.7,314.8]
ϵ_{CC}^{A}	mean	14.27	26.4	108.4
	95% CI	[3.6,38.83]	[22.0,33.8]	[94.9,129.0]
ϵ_{CC}^{B}	mean	14.3	26.4	108.4
	95% CI	[3.6,38.9]	[22.0,33.8]	[94.9,129.0]
$\epsilon_{LC}^{AB} - \epsilon_{CC}^{AB}$	mean	201.8	221.0	220.1
	95% CI	[200.5,203.0]	[218.5,223.5]	[213.3,227.9]
$\epsilon_{CC}^{AB} - 2\epsilon_{GC}$	mean	28.0	52.4	44.6
	95% CI	[6.7,77.0]	[43.6,67.1]	[17.87,85.87]
$\epsilon_{LC}^{AB} - 2\epsilon_{GC}$	mean	173.7	168.6	175.5
	95% CI	[124.0,195.0]	[152.1,177.8]	[132.8,203.1]

CC approach on data site A and B together is more or less equal to the difference in number of errors between the CC and the GC approach. This indicates that the GC approach still successfully separates the clusters in the second analysis. Therefore, it can be concluded that when not all necessary features are present at a single data site, the CC approach will make more errors, but still significantly decreases the number of errors made compared to the LC approach, i.e./ 221 errors less on average for both data sites.

The third analysis reflects the most realistic situation where observations are not perfectly separable given all features. Because of this, the GC approach now also makes assignment errors (86.1 on average, with [83.7,88.6] as 95% CI). Not only does GC make more errors than in the previous two analysis, also the LC and CC approach experience more difficulties clustering the data. However, comparing CC and LC, the results show that the CC approach makes 220.1 errors less on average for both data sites than the LC approach. Thus, even when data is no longer perfectly separable, the CC approach improves the clustering results compared to the LC approach.

Customer satisfaction data set.

Besides the artificial data set, experiments were also conducted on a real data set to see how the PSO-CFC approach behaves in a real world setting. This data set is a sample of 666 observations from the Family Entertainment Data Set. From this data set, the dimension performances for *Product A*, *Product B*, *Product C* and *Product D* were used as a proxy for the dimension's satisfaction. Additionally, the corresponding attribute performances were taken into account during the cluster analysis. Table 8.2 shows the number of attributes for each product. Although all products were sold by the same company, the data could also reflect 4 companies selling a single product to the same customer population. In the remainder of this chapter, the latter situation will be assumed.

Table 8.2: Attribute dimensions.

Attribute dimension	Number of attributes
Product A	7
Product B	4
Product C	6
Product D	3

If no privacy or security issues would exist and all four companies were willing to exchange private customer information, they could consolidate all their customer data and use this to segment the customer population into different groups. This would be the GC approach. However, companies often do not want to share private customer information or privacy constraints forbid to do so. Therefore, the common situation is that companies only use their own limited data to perform a customer segmentation, which is the LC approach. In general, the global clustering approach is assumed to provide better results since the clustering algorithm has access to more information about the customers. In this chapter, also the CC approach is analyzed, i.e./ a third approach trying to approximate the GC results without violating privacy restrictions.

The purpose of this analysis is to analyze the differences between the clusters found by all three approaches. Given the context of customer satisfaction and the fact that all attributes measure performance or satisfaction on a "low-to-high" scale, a 2-cluster model was considered. The cluster assignments of the GC approach were used as an estimate of the true cluster memberships. The methodology described in section 8.6 was used to calculate the average number of incorrectly assigned cases for both the CC and the LC approach, together with their bootstrap confidence intervals. The results are shown in Table 8.3

Table 8.3: Cluster quality comparisons between the LC and CC approach in a horizontal clustering environment on real data

Statistic		LC	CC
$\epsilon_{\cdot}^{A}$	mean	223.9	185.5
	95% CI	[203.5,242.3]	[170.6, 199.0]
$\epsilon_{\cdot}^{B}$	mean	224.1	185.5
	95% CI	[208.2,238.2]	[170.6, 199.0]
$\epsilon_{\cdot}^{C}$	mean	214.8	185.5
	95% CI	[202.0,226.0]	[170.6, 199.0]
$\epsilon_{\cdot}^{D}$	mean	246.1	185.5
	95% CI	[233.2,257.9]	[170.6, 199.0]

These results are very promising. Under the assumption that the cluster assignments of the GC approach are good approximations of the true cluster assignments, the CC approach outperforms the LC approach on all four data sites. Summed over all four data sites, the CC approach makes 166.7 errors less for 2664 cluster assignments, with a 95% confidence interval of $[152.8, 181.4]$. These results indicate that companies could achieve better cluster compositions by using the CC approach instead of the LC approach. Also note that no local optima problems were experienced for the real data set.

Not only does the collaboration aspect of the CC approach has an impact on the cluster assignments, it also influences the cluster centers. Figure 8.2 show the profiles for all the cluster centers found by the three approaches for each data site. These figures are based on the original data. A profile shows the values of a cluster center for each feature and gives an idea about the distance between the cluster centers. All four profiles show that each cluster solution contains two clusters which can be identified as a high satisfaction/performance group and a medium satisfaction/performance group of customers. For company A, it can be seen that the cluster centers are more separated in the LC approach than in the GC approach. The results also show that the CC approach provides cluster centers which approximate the GC approach solution much better. This pattern can be found for all four companies. This implies that if companies would share their private information, they would find more balanced customer clusters due to the additional customer information. These results confirm that the CC approach can approximate the GC solution without revealing private customer information.

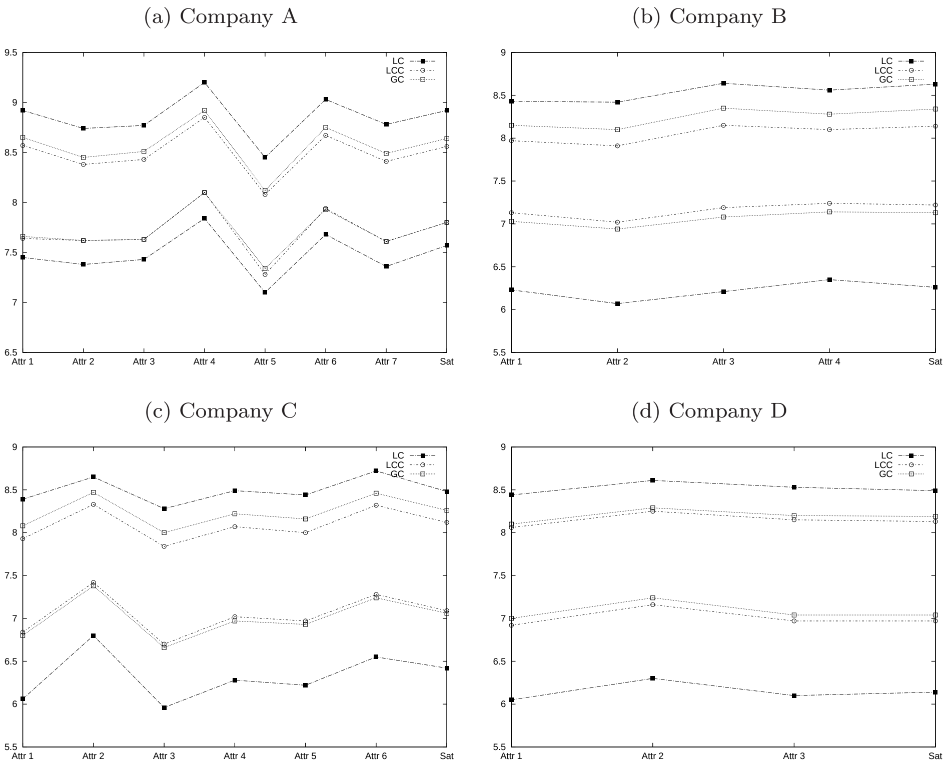


Figure 8.2: Cluster profiles

Vertical clustering: empirical results

Artificial data sets.

To assess the performance of the vertical variant of PSO-CFC, four different analysis were conducted on artificial data sets. Each analysis represents a different task, which ranges from easy to difficult. Each analysis assumes two data sites with their own local data sets containing 300 cases for the first three analysis and 450 cases for the last analysis. Both data sites have their data described by two features, X_1 and X_2. The consolidated data set, which is used for the GC approach, always consists of the combination of both local data sets. The data sets in the first analysis contain data from two clusters, while the other analysis use data containing three clusters.

The first analysis represents a very easy clustering problem for both data sites. Both sites draw their data from two multivariate normal distributions, i.e./ one per cluster, with the following mean vectors μ_1, μ_2 for respectively clusters C_1, C_2 and with the same covariance matrix Σ for both clusters:

$$\mu_1 = \begin{bmatrix} 4 & 4 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} 0 & 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix}.$$

The data sets in analysis 1 represent a randomly drawn data sample, where both clusters are equally represented and which is perfectly separable.

The second analysis tries to make the clustering problem more challenging by adding a third cluster and by changing the data distribution of the clusters such that the clusters do overlap which makes a perfect cluster assignment no longer possible. This analysis uses the following Σ as a covariance matrix and μ_1, μ_2, μ_3 as mean vectors for respectively C_1, C_2 and C_3:

$$\mu_1 = \begin{bmatrix} 2 & 3 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} 2.5 & 2 \end{bmatrix} \qquad \mu_3 = \begin{bmatrix} 3 & 3 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}.$$

The second analysis represents a situation where the data is drawn completely at random and where both clusters are equally present in the data, but for a problem which is no longer perfectly separable.

The third analysis makes the situation more realistic and more complex by changing the distribution of the clusters within the data. The data are still sampled at random from the same data distributions as in analysis 2, but this time cluster C_3 is oversampled and C_2 is undersampled in data site A, while C_2 is oversampled and C_3 is undersampled in data site B. The exact sampling distributions of the clusters are shown in Table 8.4. This analysis represents a situation where data is not perfectly separable and where some clusters might be more present within the data than others. This is a very common situation in real-life data sets. For example, a company which mainly sells to one type of customer will most likely have a majority of customers of this type in its data set. However, it should be noted that conditional on the cluster, the data are still a random sample.

Table 8.4: Cluster distribution for third analysis

	C_1	C_2	C_3
Site A	100	50	150
Site B	100	150	50

The fourth analysis goes one step further by creating so called biased samples for both data sites. The data for data site A and B are still drawn from the same data distributions as in the previous two analysis, but no longer at random, except for the observations of cluster C_3. Cluster C_1 was split in two parts based on the test expression $X_2 - X_1 > 1$. Observations passing this test, belong to the first part of C_1, those failing the test belong to the second part. Cluster C_2 was split with the test expression $X_2 > 2$. Table 8.5 shows the exact sampling distribution for the fourth analysis. Note that the clusters are not equally represented locally, neither are they drawn at random.

The results of all four analysis are presented in Table 8.6. The results of this first analysis show that the starting problem was indeed very easy, since none of the approaches made significant cluster assignment errors. The second analysis, which represented a problem which was no longer perfectly separable was a bigger challenge.

Table 8.5: Cluster distribution for fourth analysis

	C_1		C_2		C_3
	$X_2 - X_1 > 1$	$X_2 - X_1 \leq 1$	$X_2 > 2$	$X_2 \leq 2$	
Site A	120	30	60	90	150
Site B	30	120	90	60	150

The results show that both the CC and LC approach make some cluster assignment errors, i.e./ around 56 cases out of 600. However, no statistically significant difference can be found between the error rate of the CC and LC approach. It should be noted that the LC approach still performs as good as the GC approach, which suggests that CC can only improve on LC in a vertical PSO-CFC setting if LC performs worse than GC. This confirms our idea that the CC approach mimics the GC approach.

The results from the third analysis, show that the error rate of the LC approach increases compared to the second approach. This must be caused by the fact that the clusters are no longer equally present within the local data sites. Apparently, unequal representations of clusters can make it more difficult to find the true clusters within the data when the clusters overlap. However, the CC approach does not seem to suffer from this new cluster challenge. The error rate of the CC approach on both data sites remain equal to the previous analysis. A direct comparison of the cluster quality of the CC approach versus the GC approach, reveals that the CC approach still performs equally good as the GC approach. Compared with the LC approach, the results show that the CC approach results in significantly less cluster assignment errors.

The results of the fourth analysis, were rather surprising. Apparently, the LC approach outperforms both the GC and CC approach. The biased sampling produced data sets which were easier to cluster individually, than collectively. This suggests that the CC approach (and the GC approach!) should only be used under the assumption that the observations are sampled at random conditional on the cluster.

8.7 Remarks and future research

Some general remarks about the limitations of the current study should be made. Firstly, the horizontal PSO-CFC algorithm appears to get stuck in local optima for some analysis on the artificial data sets. Although this did not occur on the real-life data sets and the overall conclusions on the artificial data sets are still favoring the CC approach, future research should investigate this. The analysis should be repeated for other real-life data sets to verify if this problem is perhaps caused by the structure of the artificial data sets. On the other hand, it might be promising to search for an improvement of the current algorithm to prevent this situation.

Secondly, the current version of the algorithm uses a heuristic (cf Eq. 8.8) to perform the cluster mapping across data sites, which does not offer a guarantee of a perfect mapping in every situation. However, the results show no problems with the heuristic. Empirical results gave the author the impression that the CFC algorithm

Table 8.6: Cluster quality comparisons between the GC, LC and CC approach in a vertical clustering environment on artificial data

		Analysis			
Statistic		1	2	3	4
ϵ^{A}_{LC}	mean	0.1	56.46	69.3	61.9
	95% CI	[0.0,0.2]	[54.4,58.7]	[66.6,72.1]	[59.0,64.6]
ϵ^{B}_{LC}	mean	0.1	57.8	67.9	87.9
	95% CI	[0.0,0.2]	[55.8,59.9]	[65.6,70.9]	[83.5;93.0]
ϵ^{A}_{CC}	mean	0.1	56.7	57.6	68.4
	95% CI	[0.0,0.2]	[54.7,58.7]	[55.9,59.1]	[65.9,70.7]
ϵ^{B}_{CC}	mean	0.1	56.6	56.7	91.4
	95% CI	[0.0,0.2]	[54.1,58.9]	[54.2,58.9]	[88.5,94.8]
$\epsilon^{AB}_{LC} - \epsilon^{AB}_{CC}$	mean	0.1	1.2	22.9	-10.0
	95% CI	[0,0]	[-0.5,2.6]	[20.4,25.5]	[-12.9,-6.8]
$\epsilon^{AB}_{CC} - \epsilon_{GC}$	mean	0.1	2.0	23.5	-4.9
	95% CI	[-0.2,0.1]	[0.4,3.3]	[20.8,25.9]	[-7.8,-1.1]
$\epsilon^{AB}_{LC} - \epsilon_{GC}$	mean	0.1	0.8	0.6	5.1
	95% CI	[-0.2,0.1]	[0.0,1.9]	[-1.03,2.33]	[3.4,6.9]

automatically creates a partially correct mapping because incorrect mappings are penalized by the second term of the augmented objective function of CFC. However, if the algorithm starts with an incorrect mapping, the CFC algorithm needs various runs only to create a correct mapping, before it can do the real collaboration. The author has the impression from empirical experiments that this mapping phase takes longer as the number of data sites and clusters increases. Currently, an idea of integrating a binary integer programming problem to solve the mapping problem such that the CFC algorithm can directly start with the collaboration phase is being studied. Future research must show if this approach can increase the speed of the algorithm.

The speed aspect can be important if the current algorithm has to become resource-aware. In this chapter, the PSO-CFC algorithm has been introduced as a KDUbiq algorithm, but has not considered the computational complexity of the algorithm because the algorithm was developed with a KDUbiq situation in mind where computing power is not a scarce resource, such as in P2P networks or grid computing. The current code has not been developed with resource restrictions in mind as can occur in other KDUbiq environments such as sensor networks. It could be interesting for future research to create a sensor-network variant of the PSO-CFC algorithm. Currently, the complexity of the PSO-CFC algorithm is mainly dependent on the complexity of the CFC algorithm, since the PSO part only repeats the CFC algorithm for each particle a number of iterations long. Empirical experiments suggested that the algorithm is rather sensitive to the number of data sites and the number of clusters. However, exact complexity analysis have not been performed in this study. As for the communication requirements, which is also important in sensor networks, the vertical version is much more interesting than the horizontal version since it only passes cluster prototypes. These are much smaller than partition matrices, which are passed by the horizontal version. There is still a lot of ground to cover on the resource-awareness aspect of the algorithm which creates some interesting paths for future research.

8.8 Summary

In this chapter, the authors presented a PSO driven collaborative clustering algorithm to the KDUbiq community. This technique can address some typical issues in KDUbiq research, such as privacy constraints and distributed computing. Previous research of PSO-CFC in non-KDUbiq environments and the new empirical results in this study demonstrate the quality of this collaborative clustering approach.

As for the horizontal variant of PSO-CFC, the analysis on artificial data sets showed that the CC approach always outperforms the LC approach. Even if not a single data site has all the necessary features to separate the clusters or when a perfect separation of the clusters is not possible for the consolidated data set, the CC approach makes less assignment errors and approximates the GC result more closely. The latter was also confirmed by the analysis on the real-life data set.

As for the vertical variant of PSO-CFC, the analysis on artificial data sets showed that CC performs equally well as LC and GC when the data is sampled completely at random and the clusters are equally represented in the different data sets. When the data is drawn at random conditional on the clusters, but the clusters are no longer equally represented in the data sets, the CC approach significantly outperforms the LC approach and performs equally well as the GC approach. Only when the sampling is biased, it is not certain that CC performs equally well as LC. The analyses showed that a biased sampling led to better clustering results for LC. However, one has to be careful to generalize this conclusion since the experiment only implemented one specific type of bias.

Overall, PSO-CFC appears to be an interesting clustering approach for companies which have only access to incomplete customer data and which are willing to collaborate with their partners to find better customer segments.

9 Conclusions and future research

At the beginning of this thesis it was suggested that companies trying to understand their customers might be applying knowledge discovery techniques to incomplete data which can lead to incomplete conclusions. A possible scenario of incomplete data occurs when companies try to model customer satisfaction with only product performance data and customer satisfaction data. Such models implicitly assume that product performance is the only and main construct influencing customer satisfaction which is in contradiction with most of the current beliefs and theories regarding the CS/D process. The Expectancy-Disconfirmation Paradigm, i.e. one of the most dominant models, considers performance as an indirect cause of customer satisfaction, which is compared against a customer's expectation level and results in an experience of disconfirmation. According to the ED Paradigm, customer satisfaction is directly influenced by this experience of disconfirmation and the customer's expectation level. The literature review on the ED Paradigm illustrated the important position it takes in the marketing literature. However, it also showed that the research community never reached a consensus on the expectation construct. Several definitions of expectation were formulated in the literature and some authors argued that it were not expectations but norms set by other companies which acted as the comparative referent in the CS/D process.

Given the lack of consensus on how to measure the comparative referent, two scenarios of incomplete data can be identified under the assumption that the ED Paradigm is the appropriate CS/D model. Firstly, companies do not have information about the customer's expectation and/or disconfirmation levels and are modeling customer satisfaction directly with product performance, hereby completely ignoring existing marketing theory. Secondly, companies did measure the customer's expectation level, but have measured the wrong construct. The goal of the first part of this thesis was to model CS/D according to the principles of the ED Paradigm with incomplete data, i.e. containing only product performance and customer satisfaction data, and to offer several KD techniques to marketers allowing them to draw richer conclusions from their incomplete data.

Firstly, the CS/D process was separated into three parts, i.e. the evaluation of the product's performance against the customer's expectation at an attribute level, the aggregation of the attribute disconfirmation effects to a product level disconfirmation effect and the interaction between the customer's initial satisfaction level and the product level disconfirmation effect causing the final satisfaction outcome. For each step, mathematical properties were defined to enforce a correct modeling representation of the the ED Paradigm and other consumer satisfaction related research. This resulted in a very flexible modeling framework, called the LIED framework, which encompasses a wide range of mathematical functions to model the ED Paradigm with incomplete data. Furthermore, it is shown that within the mathematical research domain of aggregation functions, a family of functions, i.e. the generated functions, can be identified as valid implementations of the LIED framework. This results in a large set of functions, which have already been extensively studied from a mathematical point of view in the domain of aggregation functions, which can be directly used to model CS/D according to the principles of the ED Paradigm when companies only have access to product performance data and customer satisfaction data.

The match between the LIED framework and the family of generated functions offers great opportunities to marketers. Because the LIED framework offers a clear translation of CS/D properties to mathematical properties, marketers can identify the mathematical properties required for modeling CS/D as they believe it should be modeled. Next, the vast amount of literature on aggregation functions allows them to identify an appropriate function which possess all the necessary mathematical properties. Furthermore, the LIED framework can also be used to modify existing functions such that they provide a better match with the theory.

In this thesis, the D-LIED implementation of the LIED framework was studied, which is based on Dombi's evaluation operator, i.e. a representable uninorm. This implementation has only one parameter which makes it possible to learn the model for each customer separately resulting in an estimate of the customer's individual expectation level. The drawback of this advantage is that the customer's expectation is estimated at a general product level instead of a product attribute level. This disadvantage was circumvented by exploiting the hierarchical structure in the available data which made it possible to learn expectation at the product dimension level. Furthermore, several experiments were conducted to evaluate the empirical validity of the expectation estimate and they all confirmed that results from the D-LIED implementation made practical sense.

The first two case studies proved that having information about the customer's expectation enriched the conclusions drawn from the performance data. For example, product dimensions which were identified as performing much worse than all the other dimensions suddenly appeared to be much less problematic because customers did not expect a great deal from these dimensions. On the other hand, dimensions which were performing very good appeared to be only slightly exceeding the customer's expectations.

The D-LIED model also allowed the author to create a new type of Importance-Performance Analysis. Traditional IPA analysis uses a survey or regression analysis to measure the product dimension importance, which calls for several precautions. The main problem is that importance is often represented as a point estimate while theory and empirical evidence suggests that the importance of a product dimension changes as the performance changes. Because the D-LIED implementation is a mathematical representation of the customer's satisfaction process, the impact of changes in dimension performance can be simulated. This allowed the creation of two new types of IPA analysis, i.e. an IPA analysis to assess the shares of the current dimension performance levels in the overall satisfaction score and an IPA analysis to calculate the impact on satisfaction of an increase or decrease in dimension performance.

Finally, the expectation estimate of the D-LIED model could also be used to measure the compatibility between the customer's expectation and the product's performance. Based on the theory of reinforcement learning, it was hypothesized that customer's intentions, such as intentions to recommend the company or to repurchase, are positively related to the performance-expectation compatibility. Two different compatibility measures were defined and empirical results confirmed that both were positively related to the performance-expectation compatibility. Finally, a case study illustrated how companies can use this new construct to advance their understanding of customer satisfaction consequences.

The LIED framework, together with the D-LIED implementation, appears to be a very powerful approach to extract new knowledge from product performance data. The first part of the thesis showed that the model can be used in various ways, each way enriching the conclusions drawn from the data and the understanding of the customer's satisfaction process. At the same time, the flexibility and generic nature of the LIED framework offers many opportunities for future research. It might be very interesting to search for other implementations of the LIED framework which allows the marketer to estimate the customer's expectation at a product dimension level without the need of hierarchical data. Such implementation will automatically have more than one parameter which makes the estimation more difficult. Firstly, most generated functions are highly non-linear which require appropriate estimation techniques to prevent them from getting stuck in local optima. Secondly, expectation and the CS/D process is often assumed to differ between customers, which require different models to be learned for different customer groups. However, the researcher does not know which customers have a similar level of expectation and CS/D process. A possible inspiration to tackle this problem might come from the theory of finite mixture modeling which faces similar challenges. Finally, it might also be worthwhile to further investigate the results found regarding performance-expectation compatibility from a theoretical marketing point of view. Currently, the idea for the relationship between compatibility and customer's intentions comes from information processing theory, but the evidence suggests that it might be transferred to the marketing domain.

The second part of this thesis focuses on a totally different type of incomplete data. In this part, the problem of customer segmentation is studied in situations where important observations or variables are missing from the data which are necessary to find the true underlying customer clusters. The author studies a new fuzzy clustering approach, PSO-CFC, which tries to overcome the limitations of the data by collaborating with other data sites without exchanging privacy-protected data. This clustering approach is based on collaborative fuzzy clustering and uses particle swarm optimization to determine the level of collaboration between the different data sites. It is also shown that the PSO-CFC clustering technique shares several aspects with other distributed clustering techniques and can be positioned within the domain of ubiquitous knowledge discovery. Experiments for both situations where observations are missing and where important variables are missing were conducted and the empirical evidence illustrated that the PSO-CFC approach outperforms a classical local clustering analysis on several occasions. When important variables are missing, the results show that PSO-CFC significantly improves the clusters when they are not perfectly separable. When important observations are missing, the results show that PSO-CFC outperforms local clustering as long as the data is sampled at random from each cluster. Particularly when the clusters are not equally represented in the data, PSO-CFC provides better clustering results.

This new clustering approach also offers some interesting new directions for future research. One particular interesting idea is to analyze the final collaboration matrix to learn more about the different data sets. Data sites which contain almost the same information might generate different collaboration matrices than data sites which really complement each other and have almost no overlap in information. If such information can be retrieved from the collaboration matrix, PSO-CFC might even be able to detect concept drifts when applied on different data sets which represent the same observations at different moments in time. Overall, PSO-CFC offers a new and interesting clustering approach when several companies are willing to collaborate but when privacy restrictions prevent them from exchanging raw data and future research might even further extend the application possibilities of this new technique.

Appendices

A Factor analysis: additional results

A.1 Energy data set

Table A.1: Univariate statistics

	Original data			Missing		Imputed data	
	N	Mean	Std. Dev.	Count	Perc.	Mean	Std. Dev.
Company	1053			0	.0		
Satisfaction	986	7.46	1.904	67	6.4	7.45	7.46
Disconfirmation	950	2.71	.681	103	9.8	2.71	2.71
ValueForMoney	816	3.28	.809	237	22.5		
Recommendation	961	3.36	1.229	92	8.7	3.36	3.36
IntentionToSwitch	1010	2.50	1.356	43	4.1	2.50	2.50
ChoiceCommitment	1037	4.18	.997	16	1.5	4.18	4.18
D1a	975	4.04	1.099	78	7.4	4.05	4.04
D1b	1005	3.42	1.290	48	4.6	3.42	3.42
D1c	990	3.50	1.285	63	6.0	3.51	3.50
D1	1013	3.63	1.111	40	3.8	3.64	3.63
D2a	982	3.82	1.087	71	6.7	3.81	3.82
D2b	1005	4.20	.895	48	4.6	4.20	4.20
D2c	1008	4.17	.871	45	4.3	4.16	4.17
D2d	937	3.80	1.031	116	11.0	3.78	3.80
D2	999	3.94	.866	54	5.1	3.93	3.94
D3a	938	3.73	1.084	115	10.9	3.72	3.73
D3b	949	3.58	1.237	104	9.9	3.58	3.58
D3c	982	3.70	1.170	71	6.7	3.70	3.70
D3d	882	3.46	1.273	171	16.2	3.48	3.46
D3e	923	3.74	1.121	130	12.3	3.74	3.74
D3	945	3.61	1.094	108	10.3	3.62	3.61
D4a	952	3.41	1.268	101	9.6	3.40	3.41

Continued on next page

continued from previous page

	Original data			Missing		Imputed data	
	N	Mean	Std. Dev.	Count	Perc.	Mean	Std. Dev.
D4b	942	3.51	1.254	111	10.5	3.50	3.51
D4c	997	3.82	1.123	56	5.3	3.82	3.82
D4d	947	3.03	1.334	106	10.1	3.04	3.03
D4e	952	2.94	1.419	101	9.6	2.96	2.94
D4	1003	3.54	1.046	50	4.7	3.54	3.54
D5a	995	3.72	1.172	58	5.5	3.72	3.72
D5b	933	3.72	1.168	120	11.4	3.71	3.72
D5c	967	3.52	1.235	86	8.2	3.52	3.52
D5	999	3.73	1.013	54	5.1	3.72	3.73
D6a	1001	3.93	1.202	52	4.9	3.92	3.93
D6b	997	4.25	1.084	56	5.3	4.24	4.25
D6c	956	3.99	1.220	97	9.2	3.94	3.99
D6	995	3.87	1.115	58	5.5	3.86	3.87
D7	924	3.13	1.162	129	12.3	3.14	3.13
D8	1026	4.34	.735	27	2.6	4.34	4.34
GreenEnergy	1015	2.64	1.367	38	3.6	2.64	2.64
Age	1053	48.55	14.174	0	.0	48.55	48.55
Gender	1053			0	.0		
MaritalState	1053			0	.0		

Table A.2: Partial correlations and MSA per variable

	D1a	D1b	D1c	D2a	D2b	D2c	D2d	D3a	D3b	D3c	D3d	D3e	D4a	D4b	D4c	D4d	D4e	D5a	D5b	D5c	D6a	D6b	D6c
D1a	.929																						
D1b	.197	.937																					
D1c	.245	.409	.942																				
D2a	.010	-.039	.048	.981																			
D2b	.130	.017	.041	.171	.964																		
D2c	.124	.023	-.035	.122	.256	.964																	
D2d	-.046	-.003	.036	.211	.125	.080	.980																
D3a	.024	-.010	.025	.100	.124	.024	-.018	.974															
D3b	.027	.008	.018	.076	.008	-.071	-.006	.139	.974														
D3c	-.033	.068	-.030	.128	-.031	.020	.041	.111	.242	.974													
D3d	.002	.029	.074	.061	-.022	-.006	-.021	.048	.175	.156	.973												
D3e	-.002	-.023	-.025	.072	-.008	.019	.040	.219	.197	.180	.279	.969											
D4a	.064	.014	-.012	.002	-.012	.013	-.002	.015	.013	-.037	.072	.024	.978										
D4b	.001	.056	-.064	.057	-.038	.062	.022	.004	.031	.003	-.046	-.019	.210	.962									
D4c	-.057	-.011	.151	.025	.137	.101	.033	.081	-.054	.098	-.015	.126	.060	.297	.971								
D4d	.028	.012	.035	-.012	-.049	-.034	.067	.007	-.026	.110	.025	-.029	.197	.220	-.002	.944							
D4e	.005	-.035	.017	-.005	-.020	-.045	.029	.029	.037	-.045	.053	-.070	.138	.068	.040	.439	.940						
D5a	-.003	.043	.038	.031	.006	.109	.176	.048	.019	.052	-.038	.023	.066	.120	.111	-.044	.007	.974					
D5b	.059	.006	-.061	.024	.003	-.062	-.044	.110	.042	.076	.181	-.036	.001	-.038	.025	-.018	.027	.046	.968				
D5c	-.142	.009	.076	.029	-.024	.018	.098	-.107	.211	-.046	.129	-.036	.018	-.026	-.009	.091	-.017	.308	.357	.950			
D6a	.021	-.002	.013	-.051	-.026	.134	.010	.001	.040	-.013	-.031	.049	.042	.129	-.074	.034	.010	.013	.056	-.041	.945		
D6b	.055	.044	-.048	-.030	-.010	.003	.068	.191	.015	-.075	-.047	-.055	.016	-.028	.017	-.039	.051	-.010	.019	.052	.187	.921	
D6c	.005	-.069	.014	.020	.032	-.049	.051	-.118	-.011	.134	-.043	.088	.058	.049	-.046	-.007	-.025	-.102	.057	.172	.307	.294	.929

Table A.3: Total variance explained

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	12.007	52.206	52.206
2	1.490	6.479	58.685
3	1.231	5.353	64.038
4	1.191	5.177	69.216
5	.911	3.959	73.175
6	.606	2.634	75.809
7	.585	2.545	78.354
8	.517	2.250	80.604
9	.468	2.034	82.637
10	.448	1.946	84.584
11	.399	1.735	86.319
12	.383	1.666	87.985
13	.364	1.582	89.567
14	.346	1.506	91.073
15	.293	1.274	92.348
16	.278	1.209	93.557
17	.265	1.152	94.709
18	.259	1.125	95.834
19	.234	1.019	96.853
20	.201	.875	97.728
21	.198	.860	98.588
22	.172	.748	99.336
23	.153	.664	100.000

Table A.4: Factor loadings for varimax rotated 4 factor model

	Factor			
	1	2	3	4
D1a	.065	.111	.789	.140
D1b	.236	.195	.714	.049
D1c	.352	.229	.688	.056
D2a	.755	.209	.322	.150
D2b	.526	.022	.532	.124
D2c	.460	.096	.509	.183
D2d	.635	.275	.290	.230
D3a	.733	.210	.297	.186
D3b	.806	.261	.187	.211
D3c	.803	.269	.211	.183
D3d	.799	.276	.190	.138
D3e	.815	.201	.211	.185
D4a	.395	.654	.229	.224
D4b	.428	.623	.270	.216
D4c	.619	.387	.370	.099
D4d	.336	.795	.165	.151
D4e	.213	.820	.103	.124
D5a	.687	.328	.291	.161
D5b	.723	.234	.095	.272
D5c	.735	.306	.112	.285
D6a	.209	.236	.153	.731
D6b	.154	.079	.141	.788
D6c	.363	.163	.030	.740

Table A.5: Factor loadings for oblimin rotated 4 factor model

	Component			
	1	2	3	4
D1a	-.154	.829	.051	.123
D1b	.079	.698	.117	-.024
D1c	.222	.633	.126	-.043
D2a	.808	.117	-.003	-.015
D2b	.538	.424	-.170	.021
D2c	.415	.409	-.068	.094
D2d	.614	.107	.103	.094
D3a	.775	.093	.001	.032
D3b	.869	-.052	.045	.040
D3c	.865	-.024	.058	.008
D3d	.873	-.043	.074	-.042
D3e	.903	-.023	-.027	.015
D4a	.179	.085	.629	.098
D4b	.228	.123	.580	.084
D4c	.570	.199	.252	-.063
D4d	.084	.027	.826	.013
D4e	-.066	-.006	.898	.007
D5a	.680	.094	.159	.000
D5b	.774	-.130	.035	.130
D5c	.756	-.120	.115	.134
D6a	-.030	.055	.128	.754
D6b	-.058	.064	-.049	.846
D6c	.221	-.119	.004	.741

Table A.6: Factor loadings for varimax rotated 5 factor model

	Component				
	1	2	3	4	5
D1a	.042	.097	.255	.764	.146
D1b	.280	.158	.098	.785	.055
D1c	.369	.201	.180	.721	.060
D2a	.646	.229	.487	.177	.136
D2b	.325	.071	.726	.258	.109
D2c	.234	.156	.773	.197	.168
D2d	.517	.302	.483	.128	.217
D3a	.668	.215	.367	.212	.176
D3b	.805	.247	.198	.189	.203
D3c	.775	.264	.268	.176	.173
D3d	.811	.257	.161	.213	.131
D3e	.775	.199	.301	.160	.175
D4a	.355	.660	.221	.176	.219
D4b	.334	.646	.374	.141	.207
D4c	.499	.412	.494	.212	.087
D4d	.329	.792	.107	.160	.148
D4e	.218	.816	.039	.114	.123
D5a	.609	.340	.388	.188	.150
D5b	.758	.211	.073	.145	.266
D5c	.744	.291	.141	.127	.277
D6a	.176	.243	.183	.092	.728
D6b	.149	.078	.105	.116	.786
D6c	.376	.156	.069	.033	.736

Table A.7: Factor loadings for oblimin rotated 5 factor model

	Component				
	1	2	3	4	5
D1a	-.199	.809	.003	.108	-.135
D1b	.142	.821	.027	-.037	.093
D1c	.232	.723	.059	-.053	.012
D2a	.584	.026	.054	-.007	-.369
D2b	.154	.142	-.069	.027	-.710
D2c	-.008	.067	.061	.098	-.778
D2d	.383	-.023	.176	.097	-.384
D3a	.634	.079	.027	.037	-.220
D3b	.842	.056	.033	.045	.001
D3c	.787	.034	.065	.014	-.088
D3d	.865	.088	.052	-.037	.045
D3e	.805	.018	-.015	.023	-.134
D4a	.098	.050	.673	.087	-.071
D4b	.045	-.006	.661	.077	-.259
D4c	.328	.065	.326	-.062	-.384
D4d	.060	.040	.859	.000	.060
D4e	-.064	.008	.931	-.008	.109
D5a	.515	.045	.203	.003	-.250
D5b	.821	.030	.003	.133	.129
D5c	.758	-.007	.105	.136	.054
D6a	-.057	.008	.151	.741	-.080
D6b	-.028	.063	-.055	.833	-.005
D6c	.272	-.060	-.004	.732	.070

Table A.8: Factor loadings for varimax rotated 6 factor model

	Component					
	1	2	3	4	5	6
D1a	.102	.134	.228	.125	.468	.710
D1b	.222	.141	.157	.079	.833	.114
D1c	.325	.190	.228	.078	.740	.145
D2a	.640	.230	.498	.141	.155	.060
D2b	.337	.076	.722	.105	.146	.231
D2c	.219	.145	.786	.172	.160	.079
D2d	.473	.282	.519	.234	.208	-.135
D3a	.700	.237	.353	.168	.090	.254
D3b	.805	.256	.207	.208	.173	.069
D3c	.773	.271	.279	.179	.163	.056
D3d	.814	.268	.169	.136	.194	.083
D3e	.786	.212	.301	.175	.111	.116
D4a	.343	.660	.239	.224	.168	.038
D4b	.299	.632	.406	.219	.187	-.075
D4c	.475	.404	.519	.097	.229	-.004
D4d	.319	.794	.124	.153	.156	.030
D4e	.230	.829	.039	.120	.055	.117
D5a	.553	.317	.435	.173	.305	-.180
D5b	.741	.211	.094	.277	.189	-.043
D5c	.691	.273	.187	.301	.267	-.224
D6a	.153	.234	.200	.735	.107	-.019
D6b	.167	.090	.092	.779	.016	.196
D6c	.350	.148	.088	.745	.079	-.082

Table A.9: Factor loadings for oblimin rotated 6 factor model

	Component					
	1	2	3	4	5	6
D1a	.027	.435	.084	.090	-.115	.662
D1b	-.033	.931	-.017	-.002	.028	.027
D1c	.100	.790	.029	-.025	-.041	.056
D2a	.562	.023	.055	-.005	-.377	-.021
D2b	.199	.027	-.060	.018	-.723	.151
D2c	-.053	.060	.031	.099	-.818	-.008
D2d	.238	.119	.127	.115	-.426	-.221
D3a	.740	-.072	.076	.025	-.193	.190
D3b	.838	.046	.055	.050	.013	.001
D3c	.777	.032	.081	.018	-.081	-.015
D3d	.869	.073	.078	-.033	.059	.018
D3e	.837	-.036	.015	.021	-.116	.049
D4a	.081	.056	.663	.093	-.087	-.017
D4b	-.047	.084	.620	.089	-.298	-.145
D4c	.256	.120	.300	-.054	-.415	-.088
D4d	.057	.044	.854	.004	.048	-.014
D4e	.011	-.070	.949	-.015	.116	.093
D5a	.331	.236	.147	.027	-.297	-.270
D5b	.755	.100	.005	.148	.131	-.103
D5c	.572	.203	.061	.166	.023	-.300
D6a	-.127	.048	.125	.760	-.104	-.063
D6b	.038	-.062	-.029	.837	.009	.172
D6c	.185	.015	-.024	..755	.058	-.124

A.2 Family entertainment data set

Table A.10: Univariate statistics

	Original data			Missing		Imputed data	
	N	Mean	Std. Dev.	Count	Perc.	Mean	Std. Dev.
ID	2122				.0		
Company	2122				.0		
Wave	2122				.0		
D1a	2114	8.5828	1.22175	8	.4	8.5791	1.22373
D1b	2091	8.2315	1.42600	31	1.5	8.2277	1.42398
D1c	2114	8.1447	1.33533	8	.4	8.1454	1.33422
D1d	2121	8.3810	1.36641	1	.0	8.3808	1.36610
D1e	2096	8.3068	1.46653	26	1.2	8.3057	1.46378
D1f	2098	8.1473	1.37757	24	1.1	8.1446	1.37404
D1g	2099	8.4216	1.25907	23	1.1	8.4178	1.25747
D1h	2038	7.3140	1.97219	84	4.0	7.3018	1.94915
D1i	2040	7.7598	1.64719	82	3.9	7.7512	1.63596
D1	2112	8.3504	1.20529	10	.5	8.3455	1.21304
D2a	2018	8.3642	1.48389	104	4.9	8.3573	1.46068
D2b	2098	8.1249	1.40020	24	1.1	8.1218	1.39507
D2c	1976	8.1771	1.51081	146	6.9	8.1723	1.47718
D2d	2101	8.6516	1.22765	21	1.0	8.6475	1.22477
D2e	2114	7.6921	1.88883	8	.4	7.6902	1.88633
D2f	2088	8.3563	1.32299	34	1.6	8.3574	1.31743
D2g	2114	8.1500	1.32461	8	.4	8.1488	1.32439
D2	2114	8.3434	1.16517	8	.4	8.3429	1.16402
D3a	2066	7.8591	1.52969	56	2.6	7.8529	1.51901
D3b	2041	7.9005	1.53246	81	3.8	7.8834	1.51698

Continued on next page

continued from previous page

	Original data			Missing		Imputed data	
	N	Mean	Std. Dev.	Count	Perc.	Mean	Std. Dev.
D3c	2036	8.1346	1.47729	86	4.1	8.1179	1.46423
D3d	2110	8.0640	1.48461	12	.6	8.0646	1.48191
D3	2046	8.1417	1.37439	76	3.6	8.1209	1.37364
D4a	2119	8.7074	1.18110	3	.1	8.7076	1.18067
D4b	2071	8.5171	1.28144	51	2.4	8.5169	1.27394
D4c	2085	8.3046	1.44270	37	1.7	8.3046	1.43624
D4d	2086	8.0959	1.42314	36	1.7	8.0956	1.41558
D4e	1982	8.5646	1.39286	140	6.6	8.5653	1.36331
D4f	2079	8.1554	1.49989	43	2.0	8.1530	1.49434
D4g	1122	7.7638	1.70019	1000	47.1		
D4h	2103	8.2853	1.41872	19	.9	8.2842	1.41429
D4	2107	8.3484	1.12730	15	.7	8.3480	1.12529
D5a	1978	7.6749	1.70808	144	6.8	7.6854	1.68425
D5b	2044	8.1732	1.44785	78	3.7	8.1753	1.43733
D5c	1957	7.5565	1.72658	165	7.8	7.5546	1.69889
D5d	1797	7.8598	1.58055	325	15.3	7.8499	1.52570
D5e	1795	7.7755	1.71011	327	15.4	7.7838	1.64253
D5f	1829	8.1755	1.50975	293	13.8	8.1753	1.46170
D5	1985	7.8685	1.46647	137	6.5	7.8721	1.45447
D6a	2081	7.8808	1.53116	41	1.9	7.8801	1.52402
D6b	2094	8.0162	1.41884	28	1.3	8.0100	1.41711
D6c	2088	7.9277	1.49076	34	1.6	7.9223	1.48478
D6	2081	7.9044	1.44148	41	1.9	7.8977	1.43733
D7a	2106	8.5760	1.31542	16	.8	8.5734	1.31389
D7b	2088	8.3755	1.34788	34	1.6	8.3661	1.34770
D7c	2052	8.4703	1.24619	70	3.3	8.4505	1.24563
D7d	2081	8.1951	1.38758	41	1.9	8.1887	1.38310
D7e	1856	8.5210	1.42166	266	12.5	8.4946	1.37357
D7	2101	8.4446	1.22178	21	1.0	8.4421	1.21955

Continued on next page

continued from previous page

	Original data			Missing		Imputed data	
	N	Mean	Std. Dev.	Count	Perc.	Mean	Std. Dev.
D8a	2039	6.5238	2.03123	83	3.9	6.5239	2.00703
D8b	1895	5.9288	2.37945	227	10.7	5.9539	2.29375
D8c	1626	6.1507	2.21437	496	23.4		
D8d	1805	6.4681	1.96271	317	14.9	6.4739	1.90214
D8e	1947	6.4751	1.90574	175	8.2	6.4857	1.85644
D8f	2040	6.3613	1.93047	82	3.9	6.3672	1.90781
D8g	1974	6.4772	1.89679	148	7.0	6.4851	1.87418
D8	2081	6.6060	1.78089	41	1.9	6.6029	1.77581
D9a	1372	7.7004	1.80199	750	35.3		
D9b	1952	7.7618	1.62947	170	8.0	7.7531	1.59117
D9c	1943	8.0160	1.50805	179	8.4	8.0022	1.47266
D9d	1835	7.3428	1.77794	287	13.5	7.3295	1.71982
D9e	1876	7.6844	1.69599	246	11.6	7.6732	1.64994
D9	1980	7.7753	1.45455	142	6.7	7.7692	1.44103
Satisfaction	2122	8.2469	1.26473		.0	8.2469	1.26473
ValueForMoney	2084	7.6526	1.57734	38	1.8	7.6472	1.57061
Recommendation	2105	8.3957	1.46722	17	.8	8.3942	1.46682
Revisit	2067	8.2346	2.18319	55	2.6	8.2318	2.16401

Table A.11: Partial correlations and MSA per variable

	D1a	D1b	D1c	D1d	D1e	D1f	D1g	D1h	D1i	D2a	D2b	D2c	D2d	D2e	D2f	D2g	D3a
D1a	.966																
D1b	.245	.974															
D1c	.244	.099	.979														
D1d	.053	.031	.093	.960													
D1e	.029	.026	.037	.363	.965												
D1f	.054	.080	.048	.296	.157	.977											
D1g	.267	.118	.011	.023	.063	.107	.978										
D1h	.020	.059	.042	.038	.035	.039	-.059	.973									
D1i	-.011	.092	.107	.041	.072	.130	.061	.234	.977								
D2a	.040	.111	.010	-.050	-.017	.040	.039	-.013	-.016	.976							
D2b	-.024	.043	.067	.008	-.048	.031	.053	.035	.023	.048	.970						
D2c	.027	-.046	-.013	.011	.011	.017	.024	.040	-.006	.038	.288	.969					
D2d	.077	-.045	-.028	.005	.042	.002	-.014	-.079	.014	.096	.140	.054	.974				
D2e	-.022	.014	-.014	-.071	.053	-.037	-.004	.027	.004	.026	.091	.045	.033	.948			
D2f	.050	.016	.000	.005	-.022	-.004	.029	-.002	-.049	.130	.037	.100	.077	.233	.967		
D2g	.001	.043	.176	.054	-.041	-.003	.066	.022	.000	.176	.128	.102	.159	.022	.128	.975	
D3a	-.037	.036	.049	-.019	-.005	-.001	.054	.026	-.003	-.005	.008	.011	-.026	.010	-.008	-.025	.968
D3b	.013	-.015	.012	-.028	-.035	.027	-.012	.018	.046	-.022	-.009	.043	.039	.017	.087	-.046	.210
D3c	.051	-.004	.007	.027	.039	-.018	-.028	.036	-.006	.083	-.032	.015	.002	-.010	-.089	.104	.207
D3d	.025	-.017	-.017	.016	-.049	-.028	.008	-.026	-.002	-.055	-.005	-.014	.035	-.008	.053	.003	.231
D4a	.072	-.016	-.052	.156	.029	.027	.024	-.020	-.057	.006	.042	-.025	.007	-.040	.038	-.008	.013
D4b	.003	-.012	.065	-.071	.011	.010	-.004	-.008	.016	-.003	.026	.037	.097	.006	.002	-.040	-.018
D4c	-.044	.052	-.011	.043	-.063	.008	-.012	.023	.028	-.022	-.027	-.016	.118	.020	-.043	-.012	.005
D4d	-.036	.002	-.015	.007	.024	.016	.052	.033	.031	-.033	.033	-.011	.011	.037	.031	-.003	.065
D4e	-.043	-.006	.020	-.004	.034	.011	.023	.011	-.080	.010	-.036	.067	.009	.085	.019	-.003	-.029
D4f	-.028	.041	.020	.020	.025	.019	-.022	.020	.040	.071	.004	.043	-.026	.011	-.017	-.006	-.024
D4h	.017	-.048	-.011	-.037	-.003	.003	.030	.031	-.023	-.027	.023	-.049	.047	-.005	.055	.025	.034
D5a	.011	.029	-.006	.060	-.007	-.016	-.044	-.005	.049	.049	-.009	-.046	.027	-.020	.011	-.026	-.005
D5b	.023	.049	.012	-.052	.042	-.034	.054	.005	-.054	.000	.004	.009	.031	-.014	-.005	-.031	-.007
D5c	.060	-.041	.019	.014	-.016	-.023	-.020	.022	.030	-.061	.039	.020	.025	.022	-.036	.016	.023
D5d	-.033	.067	.012	-.019	.059	.000	-.044	-.057	.045	.018	.020	-.028	-.032	-.077	.008	.019	.019
D5e	.001	-.031	.008	.007	-.061	.033	-.017	.048	-.023	.024	.001	.005	-.071	.107	-.008	-.019	-.030
D5f	.034	-.099	-.035	.024	-.052	.013	.082	.019	-.006	-.015	-.020	.001	.055	-.045	.046	-.004	.009
D6a	.000	-.014	.015	.027	.003	.022	-.027	-.024	-.023	.046	.031	.021	.006	.002	-.045	-.015	-.013
D6b	.042	-.068	.024	-.086	.067	-.018	.044	.026	.047	-.001	.014	-.039	-.066	.051	.006	.058	-.046
D6c	-.028	.063	.007	.008	-.042	.028	-.018	.002	-.009	.056	-.032	.010	.068	-.076	.052	-.010	.073
D7a	.049	.004	-.023	.027	.030	-.059	.027	.046	-.035	.027	-.025	-.025	.045	-.022	-.002	-.002	-.041
D7b	-.029	.067	.051	-.037	-.046	.011	.068	.048	.055	.025	-.045	-.061	.017	-.004	.026	.015	-.052
D7c	-.022	-.033	.036	-.008	-.014	.077	-.017	-.012	.035	-.047	.019	.066	.001	.045	.003	-.016	-.029
D7d	-.019	-.012	-.031	.034	.075	-.001	-.030	-.034	.018	-.032	.018	-.009	.000	.096	-.004	.058	.066
D7e	-.011	.016	-.037	.004	.020	.002	.018	-.108	-.004	-.019	.000	.125	-.050	-.009	-.011	.011	.063
D8a	-.023	.024	.005	.020	.041	-.009	.036	-.031	.005	.032	.019	-.022	.013	.074	.007	-.010	.011
D8b	.022	.019	.005	.020	-.017	-.009	.022	.041	.014	.022	.005	.053	-.072	.005	-.023	-.037	.015
D8d	-.071	.007	-.042	-.028	.010	.056	-.027	.020	-.016	.030	-.015	.023	-.012	.069	.016	.007	.041
D8e	-.001	-.003	-.030	-.010	.002	-.019	-.025	-.004	.006	-.026	.005	.013	.012	-.012	-.014	.013	-.022
D8f	.029	.031	.028	.002	-.016	-.009	-.023	-.025	-.016	-.034	.013	.039	-.037	-.008	-.068	.051	.023
D8g	-.012	-.014	.012	.029	-.012	.016	.010	-.009	.035	.017	-.032	-.053	.038	.017	.064	-.021	.002
D9b	-.026	.066	.058	.007	.008	-.022	.060	.022	-.022	.004	-.026	.021	.008	.052	-.037	-.033	.030
D9c	.046	-.060	-.044	.031	.041	-.025	.046	-.014	.086	.024	.041	.008	.046	-.077	.016	-.013	.011
D9d	-.079	.025	.018	.002	-.028	.069	-.018	.108	.037	.036	-.031	.025	-.043	.004	-.014	.041	.004
D9e	.065	-.012	-.008	-.007	.040	-.019	-.038	.048	-.028	-.028	.036	.004	.035	-.034	.031	.015	.022

Table A.11: Partial correlations and MSA per variable [continued]

	D3b	D3c	D3d	D4a	D4b	D4c	D4d	D4e	D4f	D4h	D5a	D5b	D5c	D5d	D5e	D5f	D6a
D1a																	
D1b																	
D1c																	
D1d																	
D1e																	
D1f																	
D1g																	
D1h																	
D1i																	
D2a																	
D2b																	
D2c																	
D2d																	
D2e																	
D2f																	
D2g																	
D3a																	
D3b	.957																
D3c	.367	.955															
D3d	.145	.239	.958														
D4a	-.034	.029	-.023	.962													
D4b	-.037	.035	.073	.108	.974												
D4c	.022	.029	-.039	.319	.151	.959											
D4d	-.025	-.001	-.006	.031	.118	.077	.981										
D4e	.024	-.008	.021	.156	.103	.068	.085	.976									
D4f	-.033	-.055	.224	.071	.107	.007	.139	.090	.971								
D4h	.051	-.029	-.021	.110	.063	.033	.104	.050	.134	.975							
D5a	-.076	.022	.053	.006	-.019	-.016	-.012	-.012	-.031	.010	.967						
D5b	-.024	.068	.053	.058	.005	-.044	.025	.067	-.046	.022	.193	.976					
D5c	.024	-.020	-.047	-.030	.037	.001	.041	-.020	.059	-.033	.325	.198	.964				
D5d	.032	.002	-.033	-.013	-.010	-.040	.020	-.033	.015	.034	.076	.091	.185	.969			
D5e	.061	-.028	-.006	.014	.037	-.045	-.027	-.014	-.004	-.018	.126	.036	.111	.179	.967		
D5f	-.031	.036	.058	.045	-.009	.133	-.040	.025	.008	-.041	.125	.187	.049	.291	.255	.964	
D6a	.041	.011	.021	-.057	-.029	.036	.003	.038	-.018	-.018	-.045	.009	.069	.019	-.012	.035	.950
D6b	.034	.022	.005	.026	.046	-.010	.052	.008	-.042	.018	.031	.001	-.037	.036	.013	-.020	.378
D6c	-.052	-.019	-.008	.023	-.034	-.018	-.006	.062	.046	.009	-.007	.020	-.007	.025	.010	.024	.369
D7a	-.031	-.008	.059	.042	.031	.000	.023	-.010	-.010	.034	-.014	.018	-.066	.000	-.002	.007	.036
D7b	.006	.035	.013	.011	-.034	.041	-.008	-.018	-.010	.032	.007	-.008	.063	.015	-.043	-.018	.013
D7c	.019	.012	-.032	.015	-.003	-.003	.004	-.001	-.003	.039	.022	.023	.001	-.022	-.018	.007	.016
D7d	.033	-.008	-.046	-.009	.045	.064	-.026	.041	.068	.014	.034	.016	-.033	-.046	.161	.014	-.018
D7e	.039	-.001	.023	-.004	.055	-.029	.027	.032	-.041	-.035	-.011	.002	-.029	.101	.094	-.037	-.007
D8a	-.009	-.027	.005	.033	.060	-.041	.007	.030	.025	.016	-.042	-.001	-.023	.007	-.018	.016	-.060
D8b	-.008	-.010	.041	-.041	.002	.005	-.040	.100	-.041	.009	.058	-.045	.046	.058	-.059	-.033	.030
D8d	.018	-.014	-.044	-.017	-.031	.027	.016	.002	.003	.014	.093	.008	.050	-.002	.100	.031	-.010
D8e	.035	.059	.008	.005	.007	.001	.008	-.013	.016	-.033	-.052	.007	-.005	.000	-.008	.024	.052
D8f	.022	-.057	.013	-.015	-.018	.014	.009	.025	.038	-.012	-.031	.029	-.014	-.003	.001	-.012	-.026
D8g	-.006	.017	-.011	-.017	-.008	.021	.023	-.070	-.006	.024	.052	-.005	.014	-.004	-.034	-.035	.012
D9b	-.037	.033	-.018	-.025	.005	-.012	-.021	.009	-.033	.095	.045	-.074	.060	-.040	.018	.013	-.016
D9c	.039	-.068	.116	.029	.003	.023	-.071	.038	.041	-.061	-.017	.038	-.035	-.012	.025	-.012	.049
D9d	.053	-.013	.039	-.018	-.067	.001	.059	-.044	.005	.046	-.007	-.019	.098	.012	.072	-.055	.030
D9e	-.023	.012	-.053	-.028	.053	-.007	.068	.036	.001	-.018	-.009	.011	-.076	.057	-.029	.050	.012

Table A.11: Partial correlations and MSA per variable [continued]

	D6b	D6c	D7a	D7b	D7c	D7d	D7e	D8a	D8b	D8d	D8e	D8f	D8g	D9b	D9c	D9d	D9e
D1a																	
D1b																	
D1c																	
D1d																	
D1e																	
D1f																	
D1g																	
D1h																	
D1i																	
D2a																	
D2b																	
D2c																	
D2d																	
D2e																	
D2f																	
D2g																	
D3a																	
D3b																	
D3c																	
D3d																	
D4a																	
D4b																	
D4c																	
D4d																	
D4e																	
D4f																	
D4h																	
D5a																	
D5b																	
D5c																	
D5d																	
D5e																	
D5f																	
D6a																	
D6b	.949																
D6c	.347	.952															
D7a	.000	-.016	.961														
D7b	-.033	.001	.303	.962													
D7c	.023	.002	.309	.317	.960												
D7d	-.011	-.002	.113	.161	.212	.971											
D7e	.028	.022	.092	.097	.137	.149	.974										
D8a	-.003	.012	-.028	.043	.016	-.057	-.008	.968									
D8b	.007	-.008	-.028	-.022	-.028	.045	-.038	.248	.963								
D8d	-.007	.005	.043	-.005	-.041	-.048	.014	.102	.077	.956							
D8e	.074	.041	-.005	-.032	.015	.009	.023	.155	.129	.172	.974						
D8f	-.003	.046	.047	-.020	.017	-.015	-.008	.048	.024	.332	.115	.946					
D8g	-.010	-.028	-.028	.062	-.036	.052	-.001	.060	.071	.288	.184	.426	.949				
D9b	-.025	.052	-.010	.014	.019	-.028	.007	.028	-.008	.007	.024	-.003	-.001	.963			
D9c	.056	-.018	-.052	.016	.061	-.002	.005	.015	-.032	-.022	-.008	.033	.029	.252	.970		
D9d	-.040	.009	.061	.007	-.030	-.066	.017	.032	.029	-.071	.049	-.001	.011	.187	.135	.955	
D9e	.034	-.047	.000	-.022	.025	.044	-.047	-.027	.069	.057	-.054	-.012	.007	.255	.182	.373	.952

Table A.12: Total variance explained

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	17.782	34.867	34.867
2	3.093	6.064	40.931
3	2.222	4.357	45.288
4	1.931	3.786	49.074
5	1.709	3.351	52.425
6	1.555	3.049	55.474
7	1.457	2.857	58.331
8	1.430	2.805	61.136
9	1.333	2.614	63.749
10	.885	1.735	65.484
11	.833	1.634	67.118
12	.771	1.512	68.630
13	.755	1.481	70.111
14	.706	1.384	71.495
15	.670	1.313	72.808
16	.646	1.268	74.076
17	.617	1.210	75.285
18	.592	1.162	76.447
19	.573	1.123	77.570
20	.558	1.094	78.664
21	.537	1.053	79.717
22	.508	.995	80.712
23	.493	.968	81.680
24	.482	.945	82.624
25	.461	.905	83.529

Continued on next page

continued from previous page

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
26	.456	.893	84.422
27	.447	.876	85.298
28	.435	.852	86.151
29	.427	.836	86.987
30	.407	.798	87.785
31	.398	.781	88.566
32	.388	.761	89.327
33	.382	.750	90.076
34	.373	.731	90.807
35	.366	.717	91.524
36	.352	.691	92.215
37	.344	.675	92.891
38	.333	.654	93.544
39	.316	.619	94.163
40	.290	.568	94.731
41	.285	.560	95.291
42	.283	.555	95.846
43	.272	.534	96.379
44	.266	.521	96.901
45	.256	.502	97.402
46	.247	.485	97.887
47	.240	.470	98.358
48	.235	.461	98.819
49	.227	.446	99.265
50	.195	.382	99.647
51	.180	.353	100.000

Table A.13: Factor loadings for varimax rotated 9 factor model

	Factor								
	1	2	3	4	5	6	7	8	9
D1a	.682	.026	.178	.141	.100	.280	.084	.154	.117
D1b	.673	.155	.110	.095	.122	.237	.113	.109	.070
D1c	.652	.092	.131	.110	.126	.273	.150	.155	.126
D1d	.733	.123	.152	.268	.142	.062	.119	.063	.042
D1e	.680	.116	.091	.243	.155	.049	.129	.033	.082
D1f	.710	.156	.135	.223	.168	.127	.136	.077	.098
D1g	.617	.059	.124	.197	.160	.282	.100	.125	.106
D1h	.455	.131	.118	.106	.089	.081	.360	.131	.020
D1i	.607	.166	.151	.090	.183	.077	.242	.142	.066
D2a	.321	.120	.125	.082	.045	.540	.080	.104	.211
D2b	.288	.074	.128	.175	.068	.623	.142	.072	.085
D2c	.203	.111	.096	.151	.131	.595	.162	.154	.091
D2d	.236	.032	.150	.344	.125	.505	.078	.120	.101
D2e	-.039	.249	.049	.148	.205	.569	.054	.034	-.051
D2f	.145	.103	.110	.178	.128	.659	.065	.100	.067
D2g	.406	.108	.111	.123	.145	.576	.109	.152	.143
D3a	.172	.190	.162	.151	.083	.114	.148	.717	.079
D3b	.137	.161	.139	.109	.172	.173	.129	.742	.111
D3c	.240	.094	.193	.154	.145	.142	.084	.740	.119
D3d	.126	.103	.179	.252	.107	.129	.132	.705	.099
D4a	.311	.024	.172	.683	.161	.100	-.010	.059	.073
D4b	.128	.057	.137	.640	.149	.196	.077	.136	.052
D4c	.189	.108	.146	.655	.174	.063	.032	.088	.060
D4d	.180	.178	.110	.541	.112	.145	.179	.108	.095
D4e	.104	.119	.112	.601	.108	.205	.092	.083	.206

Continued on next page

continued from previous page

	Factor								
	1	2	3	4	5	6	7	8	9
D4f	.213	.166	.109	.544	.086	.146	.138	.177	.049
D4h	.093	.105	.052	.535	.178	.147	.177	.088	.031
D5a	.196	.224	.762	.113	.126	.090	.118	.088	.057
D5b	.179	.140	.704	.204	.145	.135	.057	.165	.133
D5c	.185	.224	.748	.121	.086	.117	.163	.109	.090
D5d	.179	.179	.737	.105	.125	.085	.134	.134	.174
D5e	.081	.174	.723	.107	.224	.141	.151	.117	.088
D5f	.149	.124	.751	.233	.129	.119	.093	.156	.144
D6a	.162	.157	.206	.132	.139	.129	.155	.146	.785
D6b	.165	.176	.175	.162	.135	.145	.131	.129	.777
D6c	.171	.193	.204	.171	.110	.143	.112	.116	.774
D7a	.227	.082	.106	.227	.759	.125	.108	.073	.092
D7b	.277	.115	.145	.181	.752	.139	.126	.097	.067
D7c	.232	.065	.138	.216	.776	.164	.145	.085	.106
D7d	.184	.109	.218	.261	.687	.174	.069	.120	.041
D7e	.105	.091	.211	.122	.638	.155	.047	.191	.137
D8a	.158	.681	.048	.176	.035	.148	.109	.050	.026
D8b	.149	.652	.127	.074	-.033	.074	.162	.087	.092
D8d	.077	.808	.294	.101	.104	.126	.090	.085	.073
D8e	.080	.752	.143	.103	.093	.094	.087	.156	.219
D8f	.119	.817	.169	.098	.123	.100	.086	.090	.088
D8g	.138	.823	.186	.094	.133	.105	.102	.094	.062
D9b	.216	.169	.143	.130	.110	.133	.725	.093	.089
D9c	.246	.127	.135	.185	.136	.138	.623	.166	.184
D9d	.218	.178	.172	.090	.098	.112	.748	.145	.084
D9e	.202	.147	.156	.176	.092	.153	.754	.081	.090

Table A.14: Factor loadings for oblimin rotated 9 factor model

	Factor								
	1	2	3	4	5	6	7	8	9
D1a	.630	-.068	-.103	.004	.100	.018	.002	.063	.188
D1b	.634	.095	-.016	.050	.052	-.020	-.043	.015	.146
D1c	.585	.003	-.030	.041	.097	-.061	-.039	.073	.176
D1d	.704	.063	-.069	.062	-.008	-.026	.170	-.020	-.068
D1e	.652	.065	.005	.093	-.040	-.048	.151	.038	-.075
D1f	.662	.090	-.031	.092	.000	-.040	.101	.043	.000
D1g	.548	-.026	-.030	.079	.063	-.007	.068	.054	.187
D1h	.365	.046	-.031	.019	.078	-.340	-.003	-.048	-.022
D1i	.540	.089	-.047	.123	.082	-.172	-.052	-.001	-.048
D2a	.207	.027	-.047	-.055	.034	-.003	-.051	.186	.513
D2b	.150	-.031	-.070	-.038	-.004	-.089	.055	.037	.610
D2c	.042	.006	-.008	.040	.095	-.109	.018	.038	.570
D2d	.086	-.078	-.083	.016	.052	-.007	.255	.052	.458
D2e	-.183	.206	.013	.164	-.024	-.011	.055	-.105	.580
D2f	-.006	.010	-.048	.039	.037	-.003	.062	.020	.658
D2g	.277	.005	-.012	.052	.086	-.024	-.034	.095	.532
D3a	.010	.077	-.005	-.036	.803	-.036	.010	-.025	-.028
D3b	-.041	.041	.035	.072	.828	-.011	-.057	.012	.034
D3c	.081	-.036	-.039	.029	.825	.049	-.006	.019	-.006
D3d	-.054	-.028	-.031	-.019	.783	-.022	.128	.000	-.015
D4a	.204	-.054	-.106	.053	-.015	.100	.688	.025	-.025
D4b	-.028	-.034	-.057	.037	.079	-.011	.635	-.004	.090
D4c	.064	.040	-.065	.076	.023	.047	.662	.010	-.063
D4d	.029	.098	-.006	.004	.037	-.125	.519	.044	.029
D4e	-.059	.031	-.010	-.006	.002	-.026	.593	.182	.099

Continued on next page

continued from previous page

	Factor								
	1	2	3	4	5	6	7	8	9
D4f	.074	.089	-.009	-.031	.131	-.072	.523	-.013	.031
D4h	-.059	.032	.045	.101	.028	-.145	.524	-.019	.047
D5a	.056	.079	-.837	.004	-.032	-.020	-.020	-.052	-.014
D5b	.022	-.017	-.750	.014	.063	.058	.076	.039	.023
D5c	.030	.069	-.813	-.049	-.011	-.072	-.012	-.015	.013
D5d	.022	.017	-.787	.000	.017	-.033	-.037	.085	-.030
D5e	-.099	.013	-.776	.123	-.002	-.065	-.042	-.015	.039
D5f	-.022	-.047	-.809	-.012	.044	.014	.111	.047	.003
D6a	-.022	.017	-.033	.049	.022	-.056	-.010	.846	-.015
D6b	-.013	.046	.003	.045	.005	-.030	.029	.841	.005
D6c	-.002	.066	-.037	.012	-.012	-.005	.043	.836	.004
D7a	.064	-.001	.045	.822	-.025	-.030	.056	.037	-.019
D7b	.116	.027	.002	.809	.001	-.042	-.009	-.001	-.008
D7c	.051	-.036	.014	.834	-.020	-.070	.029	.046	.018
D7d	.011	.014	-.102	.716	.030	.023	.096	-.036	.039
D7e	-.064	-.007	-.090	.673	.124	.047	-.057	.080	.030
D8a	.074	.708	.075	-.034	-.011	-.028	.112	-.029	.070
D8b	.067	.658	-.021	-.114	.028	-.089	-.002	.042	-.009
D8d	-.045	.808	-.198	.025	-.004	.024	-.010	-.004	.025
D8e	-.041	.755	.013	.021	.090	.027	-.005	.178	-.020
D8f	.014	.839	-.037	.062	.015	.026	-.009	.023	-.005
D8g	.031	.842	-.057	.071	.017	.010	-.020	-.010	-.003
D9b	.005	.021	-.031	.026	-.012	-.785	.008	.018	.021
D9c	.034	-.027	.001	.046	.072	-.648	.058	.125	.009
D9d	.003	.023	-.060	.009	.050	-.806	-.044	.004	-.006
D9e	-.022	-.011	-.049	-.003	-.032	-.822	.063	.017	.042

Table A.15: Factor loadings for varimax rotated 10 factor model

	Factor									
	1	2	3	4	5	6	7	8	9	10
D1a	.672	.046	.176	.130	.118	.297	.162	.117	.097	-.144
D1b	.665	.163	.112	.087	.132	.248	.115	.120	.066	-.002
D1c	.644	.096	.135	.103	.133	.282	.161	.148	.126	.046
D1d	.731	.128	.156	.259	.149	.071	.070	.123	.041	.005
D1e	.681	.114	.097	.237	.157	.054	.039	.117	.090	.083
D1f	.707	.155	.141	.215	.172	.134	.083	.127	.104	.075
D1g	.607	.078	.123	.187	.177	.296	.132	.130	.088	-.121
D1h	.459	.094	.135	.113	.069	.067	.133	.270	.072	.502
D1i	.607	.146	.162	.089	.173	.073	.146	.190	.096	.299
D2a	.305	.142	.122	.076	.064	.551	.108	.113	.188	-.137
D2b	.275	.079	.132	.177	.078	.623	.075	.134	.086	.087
D2c	.189	.114	.100	.153	.139	.591	.156	.151	.093	.103
D2d	.221	.064	.144	.336	.151	.515	.126	.132	.067	-.246
D2e	-.045	.220	.062	.162	.193	.544	.032	-.025	-.010	.440
D2f	.131	.109	.112	.181	.139	.656	.102	.059	.066	.066
D2g	.390	.124	.111	.119	.162	.582	.156	.125	.129	-.045
D3a	.164	.190	.164	.146	.085	.111	.719	.137	.084	.064
D3b	.127	.160	.141	.104	.173	.168	.743	.115	.116	.076
D3c	.230	.100	.194	.146	.152	.144	.743	.089	.114	-.027
D3d	.117	.109	.180	.245	.113	.128	.708	.134	.096	-.012
D4a	.310	.043	.170	.674	.179	.105	.065	.027	.054	-.185
D4b	.127	.062	.140	.637	.157	.190	.141	.079	.052	.003
D4c	.189	.124	.145	.647	.188	.063	.094	.062	.046	-.144
D4d	.184	.159	.121	.545	.104	.130	.110	.128	.125	.284
D4e	.104	.121	.115	.600	.115	.199	.087	.088	.209	.029

Continued on next page

continued from previous page

	Factor									
	1	2	3	4	5	6	7	8	9	10
D4f	.215	.154	.118	.546	.083	.136	.180	.103	.070	.197
D4h	.094	.099	.058	.536	.179	.135	.091	.156	.043	.130
D5a	.190	.225	.763	.108	.129	.088	.091	.113	.058	.024
D5b	.171	.148	.704	.196	.153	.137	.168	.067	.126	-.061
D5c	.180	.219	.752	.118	.086	.113	.111	.144	.099	.098
D5d	.173	.179	.739	.099	.128	.084	.137	.128	.176	.024
D5e	.075	.167	.727	.104	.222	.132	.118	.127	.099	.123
D5f	.142	.134	.751	.225	.138	.121	.160	.106	.135	-.074
D6a	.158	.156	.208	.127	.141	.134	.148	.149	.788	.013
D6b	.162	.171	.178	.159	.135	.148	.131	.117	.785	.059
D6c	.167	.194	.206	.166	.113	.149	.118	.111	.774	-.014
D7a	.219	.090	.107	.218	.766	.116	.076	.116	.087	-.016
D7b	.269	.120	.146	.172	.758	.129	.100	.128	.065	.018
D7c	.225	.068	.140	.209	.781	.153	.088	.142	.107	.048
D7d	.178	.108	.221	.256	.689	.161	.122	.056	.048	.088
D7e	.097	.095	.211	.115	.643	.146	.193	.048	.135	.006
D8a	.152	.687	.049	.172	.041	.141	.052	.112	.022	.011
D8b	.146	.650	.130	.071	-.033	.067	.089	.150	.097	.081
D8d	.070	.809	.295	.097	.107	.114	.086	.083	.075	.051
D8e	.074	.757	.143	.098	.097	.086	.158	.088	.217	.004
D8f	.112	.823	.169	.092	.128	.091	.092	.090	.083	.000
D8g	.131	.829	.186	.087	.138	.095	.096	.104	.059	.012
D9b	.207	.181	.145	.123	.118	.131	.099	.737	.081	.014
D9c	.235	.148	.133	.175	.151	.142	.172	.655	.164	-.102
D9d	.211	.177	.177	.087	.098	.105	.150	.731	.092	.160
D9e	.194	.157	.158	.171	.100	.150	.087	.760	.085	.047

Table A.16: Factor loadings for oblimin rotated 10 factor model

	Factor									
	1	2	3	4	5	6	7	8	9	10
D1a	.620	-.037	-.094	.020	.105	-.031	-.012	.039	.155	-.173
D1b	.631	.103	-.015	.053	.056	-.028	-.050	.011	.131	-.028
D1c	.587	.000	-.034	.040	.101	-.053	-.043	.081	.166	.017
D1d	.708	.064	-.073	.064	-.002	-.025	.155	-.016	-.075	.007
D1e	.662	.050	-.005	.087	-.035	-.021	.141	.057	-.071	.088
D1f	.669	.078	-.039	.087	.004	-.019	.092	.059	-.001	.068
D1g	.539	.005	-.022	.094	.067	-.054	.053	.032	.157	-.154
D1h	.398	-.042	-.067	-.022	.081	-.195	.014	.037	.031	.496
D1i	.558	.036	-.069	.097	.085	-.084	-.045	.050	-.021	.297
D2a	.192	.066	-.033	-.035	.035	-.064	-.059	.158	.474	-.215
D2b	.152	-.027	-.075	-.031	.000	-.090	.058	.046	.598	-.004
D2c	.042	.009	-.012	.045	.097	-.107	.022	.046	.560	.013
D2d	.067	-.012	-.064	.052	.057	-.110	.234	.003	.410	-.323
D2e	-.157	.149	-.018	.145	-.022	.095	.080	-.038	.618	.349
D2f	-.006	.020	-.051	.051	.040	-.012	.066	.024	.645	-.034
D2g	.268	.031	-.006	.068	.089	-.060	-.038	.080	.503	-.127
D3a	.009	.072	-.007	-.041	.806	-.025	.006	-.020	-.027	.043
D3b	-.042	.033	.033	.067	.831	.003	-.058	.019	.035	.046
D3c	.076	-.029	-.036	.032	.830	.036	-.013	.013	-.016	-.047
D3d	-.058	-.019	-.028	-.015	.788	-.035	.118	-.005	-.022	-.035
D4a	.204	-.015	-.102	.080	-.006	.036	.654	.004	-.047	-.196
D4b	-.019	-.023	-.063	.051	.088	-.027	.614	.002	.088	-.028
D4c	.063	.075	-.060	.100	.031	-.009	.630	-.010	-.081	-.156
D4d	.056	.053	-.031	-.011	.044	-.051	.513	.100	.060	.263
D4e	-.047	.031	-.016	.001	.008	-.028	.573	.198	.098	-.002

Continued on next page

continued from previous page

	Factor									
	1	2	3	4	5	6	7	8	9	10
D4f	.096	.062	-.028	-.038	.139	-.024	.513	.026	.052	.176
D4h	-.046	.021	.033	.103	.034	-.125	.511	.003	.059	.104
D5a	.056	.079	-.837	.003	-.031	-.017	-.022	-.050	-.016	.003
D5b	.020	-.005	-.746	.021	.065	.038	.069	.032	.012	-.085
D5c	.035	.054	-.819	-.057	-.010	-.047	-.011	.001	.019	.074
D5d	.023	.010	-.787	-.005	.016	-.025	-.039	.092	-.031	.007
D5e	-.093	-.005	-.783	.115	-.003	-.035	-.037	.004	.049	.093
D5f	-.025	-.033	-.804	-.004	.046	-.011	.101	.038	-.010	-.097
D6a	-.018	-.008	-.030	.032	.015	-.033	-.018	.876	-.022	.004
D6b	-.005	.012	.002	.025	-.002	.008	.023	.880	.004	.048
D6c	.002	.046	-.033	-.002	-.019	.010	.032	.862	-.006	-.026
D7a	.052	.016	.052	.832	-.027	-.052	.045	.023	-.029	-.040
D7b	.105	.039	.008	.816	-.002	-.054	-.017	-.011	-.015	-.008
D7c	.043	-.031	.017	.839	-.023	-.074	.022	.043	.015	.018
D7d	.010	.012	-.105	.719	.030	.034	.092	-.029	.043	.057
D7e	-.072	.001	-.084	.677	.121	.038	-.060	.074	.025	-.023
D8a	.064	.729	.084	-.027	-.013	-.048	.100	-.046	.059	-.028
D8b	.062	.659	-.017	-.119	.024	-.080	-.008	.042	-.010	.055
D8d	-.055	.819	-.190	.026	-.009	.020	-.017	-.014	.020	.010
D8e	-.053	.768	.025	.021	.084	.016	-.015	.166	-.030	-.027
D8f	.000	.861	-.024	.067	.009	.006	-.020	.003	-.017	-.036
D8g	.016	.863	-.044	.075	.012	-.008	-.030	-.030	-.013	-.026
D9b	-.017	.047	-.014	.032	-.017	-.827	-.009	-.015	.000	-.024
D9c	.008	.015	.024	.059	.068	-.718	.035	.081	-.024	-.136
D9d	-.008	.019	-.054	.001	.045	-.800	-.051	-.001	-.008	.128
D9e	-.040	.009	-.035	.001	-.037	-.856	.047	-.009	.024	.007

A.3 Banking product data set

Table A.17: Univariate statistics

	Original data			Missing		Imputed data	
	N	Mean	Std. Dev.	Count	Perc.	Mean	Std. Dev.
Satisfaction	1038	7.27	1.792		.0	7.27	1.792
Disconfirmation	983	2.65	.713	55	5.3	2.65	.697
Recommendation	1038	3.76	1.112		.0	3.76	1.112
Usage	949	1.67	.751	89	8.6	1.67	.721
Value for money	969	2.62	.967	69	6.6	2.64	.943
D1	1005	3.56	.879	33	3.2	3.55	.873
D1a	1026	3.77	1.036	12	1.2	3.77	1.032
D1b	995	3.20	1.195	43	4.1	3.19	1.177
D1c	980	3.80	.913	58	5.6	3.79	.897
D1d	895	3.55	1.030	143	13.8	3.55	.976
D1e	920	3.74	.973	118	11.4	3.71	.945
D1f	949	2.84	1.429	89	8.6	2.83	1.391
D1g	968	3.74	1.034	70	6.7	3.73	1.010
D1h	1011	4.03	.876	27	2.6	4.03	.869
D1i	993	3.16	1.192	45	4.3	3.17	1.176
D1j	962	3.19	1.135	76	7.3	3.19	1.112
D2	976	2.39	1.048	62	6.0	2.43	1.031
D2a	959	2.43	1.222	79	7.6	2.45	1.193
D2b	973	2.36	1.205	65	6.3	2.39	1.187
D2c	896	2.49	1.199	142	13.7	2.51	1.148
D2d	928	2.27	1.172	110	10.6	2.30	1.142
D2e	968	3.11	1.295	70	6.7	3.12	1.263
D2f	944	2.34	1.236	94	9.1	2.37	1.202

Continued on next page

continued from previous page

	Original data			Missing		Imputed data	
	N	Mean	Std. Dev.	Count	Perc.	Mean	Std. Dev.
D3	1012	3.77	.769	26	2.5	3.76	.763
D4	1009	3.89	.642	29	2.8	3.89	.637
D4a	878	3.59	1.001	160	15.4	3.59	.953
D4b	969	4.04	.766	69	6.6	4.05	.746
D4c	1025	4.11	.763	13	1.3	4.11	.759
D4d	1006	3.33	1.146	32	3.1	3.33	1.134
D4e	1021	3.89	.925	17	1.6	3.88	.920
D4f	1031	3.76	1.070	7	.7	3.76	1.067
D4g	1019	3.42	1.157	19	1.8	3.41	1.150
D4h	1021	4.23	.861	17	1.6	4.22	.856
D4i	980	3.80	.921	58	5.6	3.80	.904
D5	665	3.70	.815	373	35.9		
D5a	105	3.55	1.248	933	89.9		
D5b	112	4.06	.852	926	89.2		
D5c	92	3.14	1.339	946	91.1		
D5d	114	3.21	1.379	924	89.0		
D5e	118	3.73	1.114	920	88.6		
D5f	108	3.89	1.026	930	89.6		
D5g	116	3.31	1.435	922	88.8		
D5h	120	4.08	1.124	918	88.4		
D5i	118	3.66	1.249	920	88.6		
D5j	111	3.77	1.061	927	89.3		
D6	657	3.76	.825	381	36.7		
D6a	604	3.81	1.183	434	41.8		
D6b	455	3.83	1.010	583	56.2		
D6c	532	3.57	1.269	506	48.7		
D6d	526	3.80	1.086	512	49.3		
D6e	508	3.57	1.189	530	51.1		
D6f	296	2.91	1.303	742	71.5		

Continued on next page

continued from previous page

	Original data			Missing		Imputed data	
	N	Mean	Std. Dev.	Count	Perc.	Mean	Std. Dev.
D6g	518	4.00	.894	520	50.1		
D6h	510	3.97	.907	528	50.9		
D6i	555	4.19	.903	483	46.5		
D6j	232	3.86	1.275	806	77.6		
D6k	231	4.04	1.042	807	77.7		
D6l	228	3.05	1.306	810	78.0		
D6m	224	4.18	.861	814	78.4		
D6n	229	4.13	.884	809	77.9		
D6o	232	4.28	.885	806	77.6		
D7	875	3.37	.994	163	15.7	3.39	.936
D7a	883	3.57	1.218	155	14.9	3.59	1.147
D7b	899	2.69	1.448	139	13.4	2.68	1.353
D7c	819	3.97	.919	219	21.1		
D7d	799	3.89	1.005	239	23.0		
D7e	847	2.24	1.289	191	18.4	2.24	1.179
D7f	856	3.25	1.186	182	17.5	3.27	1.112
D8	796	3.55	.878	242	23.3		
D8a	231	2.79	1.430	807	77.7		
D8b	91	3.24	1.285	947	91.2		
D8c	87	2.93	1.283	951	91.6		
D8d	129	3.09	1.305	909	87.6		
D9	893	3.27	1.037	145	14.0	3.27	.991
D9a	916	3.50	1.154	122	11.8	3.48	1.110
D9b	730	3.64	1.007	308	29.7		
D9c	789	3.57	.983	249	24.0		
D9d	734	3.50	1.174	304	29.3		
D9e	908	3.49	1.231	130	12.5	3.48	1.173
D9f	863	3.33	1.152	175	16.9	3.30	1.095
D9g	947	2.10	1.148	91	8.8	2.11	1.104

Continued on next page

continued from previous page

	Original data			Missing		Imputed data	
	N	Mean	Std. Dev.	Count	Perc.	Mean	Std. Dev.
D9h	981	4.25	.793	57	5.5	4.24	.775

Table A.18: Partial correlations and MSA per variable

	D1a	D1b	D1c	D1d	D1e	D1f	D1g	D1h	D1i	D1j	D2a	D2b	D2c	D2d	D2e	D2f	D4a
D1a	0.955																
D1b	0.142	0.943															
D1c	-0.001	0.033	0.941														
D1d	0.144	-0.022	0.132	0.945													
D1e	0.199	0.106	0.063	0.106	0.934												
D1f	-0.004	0.139	0.138	0.066	-0.073	0.953											
D1g	0.086	-0.028	0.214	0.022	0.054	0.070	0.947										
D1h	0.196	0.015	0.145	0.024	0.280	-0.006	0.090	0.929									
D1i	0.005	0.227	0.026	-0.072	0.114	0.116	0.047	0.078	0.956								
D1j	0.110	0.206	0.022	0.043	0.024	0.178	0.130	0.050	0.070	0.956							
D2a	0.021	0.112	0.009	0.057	-0.068	-0.001	0.022	-0.018	0.022	0.018	0.927						
D2b	0.043	-0.057	-0.057	0.030	-0.047	0.135	-0.002	-0.002	-0.002	0.051	0.310	0.936					
D2c	-0.013	0.077	0.066	-0.087	0.066	-0.046	-0.014	-0.040	-0.008	0.048	0.196	0.191	0.928				
D2d	0.026	-0.031	-0.049	-0.016	0.031	0.061	-0.008	0.009	0.003	0.017	0.288	0.108	0.302	0.918			
D2e	0.055	0.022	-0.032	-0.013	0.090	0.030	-0.021	-0.011	0.035	-0.013	0.123	0.094	0.068	-0.072	0.942		
D2f	-0.025	0.062	-0.079	-0.004	0.100	0.106	0.077	0.011	-0.006	-0.061	0.123	0.173	-0.011	0.199	0.120	0.932	
D4a	0.049	0.008	0.091	0.043	-0.070	-0.037	0.088	-0.042	0.007	0.102	-0.009	0.064	-0.060	-0.015	0.021	0.044	0.922
D4b	-0.013	-0.062	0.029	0.069	0.006	0.019	-0.050	0.104	-0.034	0.065	-0.013	-0.026	-0.018	0.026	-0.026	-0.019	0.158
D4c	0.034	0.005	0.001	-0.004	0.047	0.062	-0.005	0.042	-0.048	-0.051	0.016	0.002	0.003	-0.019	0.048	-0.061	0.084
D4d	-0.047	0.032	0.066	-0.010	0.081	-0.011	-0.021	0.016	0.014	-0.027	0.050	0.069	-0.025	-0.013	-0.052	-0.039	0.290
D4e	0.020	-0.022	0.063	-0.003	-0.038	-0.039	0.013	-0.070	0.030	0.007	-0.016	0.008	-0.037	0.014	-0.009	0.078	0.121
D4f	-0.015	0.074	0.080	-0.094	-0.036	-0.022	0.003	0.014	-0.025	-0.030	0.020	0.026	0.009	0.003	-0.009	-0.056	0.141
D4g	0.013	0.040	0.001	0.025	-0.013	0.043	0.007	-0.024	0.009	0.021	0.062	-0.012	0.035	0.032	-0.025	0.037	0.075
D4h	0.043	-0.042	0.006	0.074	0.047	0.024	0.023	-0.014	0.037	-0.029	-0.039	0.047	0.028	-0.054	0.070	0.005	-0.044
D4i	0.024	0.046	0.021	0.047	-0.030	0.018	0.027	0.036	-0.030	-0.061	-0.076	0.002	0.080	0.073	0.048	-0.034	0.116
D7a	0.059	0.101	-0.032	-0.029	0.034	-0.040	0.044	0.029	0.016	-0.028	0.024	0.009	-0.064	-0.037	0.258	-0.070	-0.094
D7b	-0.039	-0.005	-0.016	0.021	-0.029	0.015	-0.025	0.059	0.026	0.003	0.040	-0.007	0.032	-0.063	-0.007	0.008	0.014
D7e	0.013	-0.022	-0.048	-0.024	-0.029	0.001	0.098	-0.025	-0.041	0.028	0.000	-0.090	0.016	0.025	-0.018	0.067	-0.001
D7f	-0.061	-0.036	-0.003	0.031	0.043	0.026	-0.100	-0.029	0.054	0.074	0.044	0.001	0.004	0.020	-0.042	0.096	0.076
D9a	-0.023	-0.081	0.051	0.056	-0.025	0.071	-0.033	0.019	0.131	0.068	0.039	-0.077	0.022	0.043	0.117	-0.023	0.151
D9e	-0.051	0.160	-0.031	0.010	0.010	-0.087	0.006	-0.024	0.038	-0.040	-0.086	0.074	-0.069	0.067	0.016	-0.084	-0.066
D9f	0.059	-0.060	0.010	-0.004	-0.010	0.037	0.034	-0.031	0.015	0.033	0.048	-0.037	0.067	-0.115	-0.065	0.156	0.010
D9g	0.016	0.059	0.009	0.020	-0.003	-0.005	-0.050	-0.054	0.093	-0.057	-0.101	0.056	-0.040	0.051	0.037	-0.069	0.038
D9h	0.068	0.032	-0.023	0.019	0.043	-0.037	-0.024	0.017	-0.020	-0.034	-0.055	0.009	0.065	-0.038	-0.063	-0.052	0.054

Table A.18: Partial correlations and MSA per variable [continued]

	D4b	D4c	D4d	D4e	D4f	D4g	D4h	D4i	D7a	D7b	D7e	D7f	D9a	D9e	D9f	D9g	D9h
D1a																	
D1b																	
D1c																	
D1d																	
D1e																	
D1f																	
D1g																	
D1h																	
D1i																	
D1j																	
D2a																	
D2b																	
D2c																	
D2d																	
D2e																	
D2f																	
D4a																	
D4b	0.907																
D4c	0.155	0.920															
D4d	-0.018	0.068	0.904														
D4e	0.055	0.260	0.083	0.924													
D4f	0.031	0.155	0.086	0.105	0.919												
D4g	-0.008	0.016	0.063	0.087	-0.019	0.936											
D4h	0.155	0.000	-0.058	0.097	0.080	0.009	0.903										
D4i	0.081	0.057	-0.027	0.031	-0.019	0.298	0.139	0.922									
D7a	0.107	-0.021	-0.020	0.072	0.054	-0.040	-0.011	0.030	0.820								
D7b	-0.017	0.013	0.018	-0.062	-0.024	0.037	-0.035	-0.031	-0.087	0.678							
D7e	0.084	-0.037	0.090	0.022	-0.056	-0.001	0.056	-0.053	-0.088	0.191	0.511						
D7f	-0.056	-0.002	0.080	-0.074	0.020	-0.035	0.075	0.036	0.559	0.055	0.112	0.833					
D9a	-0.061	-0.039	-0.089	0.092	0.044	0.020	-0.060	-0.017	0.026	-0.077	0.041	0.044	0.912				
D9e	0.117	-0.005	0.023	0.073	-0.010	-0.004	0.011	0.028	-0.059	-0.059	-0.024	0.039	0.183	0.798			
D9f	0.011	0.054	0.110	-0.009	-0.023	-0.013	0.001	-0.010	0.044	0.005	0.024	-0.019	0.249	0.423	0.859		
D9g	-0.044	0.002	-0.013	-0.020	0.058	0.000	-0.053	0.022	-0.021	0.098	0.225	0.026	0.019	0.077	0.078	0.568	
D9h	0.027	0.111	-0.051	-0.015	-0.025	-0.001	0.162	0.039	0.108	0.088	-0.092	-0.014	0.133	0.059	0.040	-0.082	0.867

Table A.19: Total variance explained

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	9.194	27.041	27.041
2	2.666	7.842	34.883
3	1.791	5.267	40.150
4	1.587	4.668	44.818
5	1.561	4.590	49.409
6	1.275	3.750	53.159
7	1.130	3.324	56.483
8	.923	2.714	59.197
9	.907	2.667	61.864
10	.865	2.545	64.409
11	.799	2.351	66.760
12	.761	2.238	68.998
13	.712	2.093	71.091
14	.699	2.055	73.146
15	.674	1.981	75.127
16	.664	1.953	77.080
17	.605	1.780	78.860
18	.588	1.731	80.591
19	.577	1.698	82.288
20	.542	1.593	83.882
21	.538	1.582	85.464
22	.507	1.490	86.955
23	.493	1.451	88.405
24	.471	1.386	89.792
25	.442	1.300	91.092
26	.421	1.238	92.330
27	.405	1.192	93.523
28	.386	1.136	94.659
29	.363	1.068	95.726
30	.334	.981	96.707
31	.321	.945	97.653
32	.299	.880	98.533
33	.270	.794	99.327
34	.229	.673	100.000

Table A.20: Factor loadings for varimax rotated 4 factor model

	Component			
	1	2	3	4
D1a	.275	.679	.213	-.022
D1b	.443	.561	.143	.164
D1c	.128	.490	.383	-.024
D1d	.115	.481	.247	-.037
D1e	.214	.725	.107	-.012
D1f	.524	.392	.162	.080
D1g	.260	.479	.229	-.030
D1h	.109	.701	.151	-.113
D1i	.322	.547	.065	.286
D1j	.425	.532	.162	.103
D2a	.837	.190	.111	.050
D2b	.790	.165	.166	.046
D2c	.765	.141	.099	.004
D2d	.813	.095	.093	.028
D2e	.458	.405	.086	.134
D2f	.728	.184	.057	.100
D4a	.240	.186	.667	.182
D4b	-.049	.250	.572	.031
D4c	.043	.169	.683	.033
D4d	.221	.048	.500	.226
D4e	.113	.092	.690	.149
D4f	.088	.065	.555	.098
D4g	.383	.057	.470	-.020
D4h	.008	.316	.400	-.062
D4i	.218	.208	.539	-.048
D7a	.125	.533	.101	.230
D7b	.037	-.150	-.143	.127
D7e	.078	-.176	-.052	.390
D7f	.251	.381	.066	.353
D9a	.160	.331	.284	.534
D9e	-.064	.198	.270	.626
D9f	.092	.242	.283	.637
D9g	-.018	-.093	-.039	.556
D9h	-.158	.351	.308	.067

Table A.21: Factor loadings for oblimin rotated 4 factor model

	Component			
	1	2	3	4
D1a	.686	-.132	-.088	-.064
D1b	.571	-.314	.102	-.011
D1c	.441	-.006	-.082	-.298
D1d	.466	-.007	-.085	-.151
D1e	.771	-.068	-.072	.065
D1f	.366	-.438	.021	-.076
D1g	.459	-.156	-.085	-.131
D1h	.734	.028	-.169	.008
D1i	.591	-.187	.236	.074
D1j	.531	-.306	.042	-.041
D2a	.128	-.806	-.011	-.069
D2b	.089	-.759	-.014	-.136
D2c	.079	-.748	-.050	-.069
D2d	.027	-.804	-.025	-.073
D2e	.408	-.368	.083	.012
D2f	.146	-.695	.048	-.011
D4a	.028	-.138	.122	-.683
D4b	.137	.148	-.014	-.566
D4c	.010	.046	-.018	-.708
D4d	-.080	-.154	.184	-.532
D4e	-.077	-.029	.099	-.732
D4f	-.072	-.023	.058	-.592
D4g	-.087	-.346	-.071	-.501
D4h	.246	.080	-.104	-.362
D4i	.075	-.145	-.103	-.540
D7a	.576	.009	.188	.028
D7b	-.127	-.066	.143	.122
D7e	-.168	-.079	.402	.022
D7f	.413	-.140	.317	.032
D9a	.317	-.022	.495	-.218
D9e	.192	.190	.607	-.234
D9f	.227	.041	.607	-.237
D9g	-.062	.054	.569	.031
D9h	.329	.262	.038	-.250

Table A.22: Factor loadings for varimax rotated 5 factor model

	Component				
	1	2	3	4	5
D1a	.241	.683	.173	.146	.143
D1b	.421	.542	.124	.275	.026
D1c	.057	.611	.365	.019	-.049
D1d	.072	.532	.221	.055	.051
D1e	.190	.695	.059	.187	.179
D1f	.481	.463	.165	.101	-.083
D1g	.187	.615	.220	-.009	-.099
D1h	.062	.741	.105	.051	.134
D1i	.302	.510	.050	.374	-.049
D1j	.378	.587	.152	.163	-.058
D2a	.838	.192	.124	.082	.025
D2b	.793	.160	.177	.083	.047
D2c	.768	.145	.110	.035	.046
D2d	.810	.122	.113	.024	-.017
D2e	.498	.243	.055	.327	.260
D2f	.723	.194	.073	.111	-.035
D4a	.192	.261	.675	.156	-.103
D4b	-.062	.224	.544	.133	.161
D4c	.029	.157	.663	.110	.147
D4d	.172	.145	.525	.131	-.221
D4e	.102	.077	.684	.183	.063
D4f	.083	.042	.548	.137	.084
D4g	.346	.152	.483	-.053	-.043
D4h	.030	.202	.353	.138	.353
D4i	.213	.189	.517	.055	.199
D7a	.204	.224	.036	.548	.455
D7b	-.022	.025	-.096	-.086	-.434
D7e	.011	-.004	.013	.115	-.601
D7f	.308	.146	.034	.546	.220
D9a	.172	.198	.277	.602	-.053
D9e	-.042	.038	.267	.663	-.096
D9f	.094	.134	.290	.639	-.182
D9g	-.057	-.029	.012	.347	-.524
D9h	-.117	.165	.253	.297	.368

Table A.23: Factor loadings for oblimin rotated 5 factor model

	Component				
	1	2	3	4	5
D1a	.695	-.093	.012	.042	-.096
D1b	.499	-.310	-.031	.192	.006
D1c	.636	.093	.260	-.080	.104
D1d	.557	.053	.114	-.027	-.008
D1e	.731	-.041	-.117	.094	-.136
D1f	.411	-.399	.051	.018	.116
D1g	.643	-.056	.103	-.102	.149
D1h	.815	.101	-.055	-.048	-.081
D1i	.489	-.188	-.111	.318	.074
D1j	.568	-.259	.006	.075	.098
D2a	.039	-.843	.051	.005	-.014
D2b	.001	-.799	.116	.006	-.036
D2c	.003	-.783	.054	-.035	-.037
D2d	-.029	-.832	.064	-.044	.025
D2e	.138	-.460	-.056	.269	-.254
D2f	.068	-.721	-.002	.049	.044
D4a	.151	-.101	.645	.071	.140
D4b	.168	.151	.519	.062	-.130
D4c	.059	.045	.661	.031	-.116
D4d	.051	-.113	.514	.077	.247
D4e	-.053	-.044	.690	.114	-.039
D4f	-.066	-.042	.557	.082	-.066
D4g	.044	-.313	.481	-.133	.071
D4h	.149	.036	.311	.071	-.332
D4i	.088	-.156	.497	-.031	-.171
D7a	.146	-.139	-.093	.512	-.457
D7b	.070	.022	-.088	-.060	.437
D7e	.000	-.001	.012	.151	.601
D7f	.045	-.266	-.078	.525	-.226
D9a	.098	-.090	.174	.579	.059
D9e	-.050	.108	.198	.676	.090
D9f	.038	-.016	.203	.635	.184
D9g	-.041	.080	-.012	.398	.516
D9h	.132	.189	.196	.259	-.357

Table A.24: Factor loadings for varimax rotated 7 factor model

	Component						
	1	2	3	4	5	6	7
D1a	.235	.669	.083	.081	.240	.151	-.053
D1b	.390	.555	.141	.192	.017	.236	-.020
D1c	.054	.628	.363	.073	.074	-.059	-.042
D1d	.104	.506	.048	.077	.340	-.033	.048
D1e	.164	.679	.015	.055	.183	.254	-.056
D1f	.477	.480	.151	.114	.009	.042	.033
D1g	.191	.631	.199	.057	.038	-.083	.012
D1h	.046	.727	.058	-.025	.184	.124	-.050
D1i	.267	.528	.092	.292	-.058	.256	.032
D1j	.361	.604	.157	.144	.004	.103	.019
D2a	.825	.208	.118	.038	.003	.147	-.026
D2b	.789	.171	.143	.053	.060	.124	-.038
D2c	.781	.150	.028	.043	.097	.055	-.032
D2d	.820	.135	.060	.045	.032	.034	-.005
D2e	.453	.229	.061	.094	.135	.461	-.077
D2f	.720	.203	.045	.078	.015	.117	.048
D4a	.197	.275	.665	.167	.180	.015	.080
D4b	-.030	.185	.368	.087	.509	.060	.030
D4c	.042	.147	.586	.089	.355	.056	-.078
D4d	.145	.174	.647	.095	-.060	.062	.154
D4e	.118	.084	.624	.214	.267	.022	-.078
D4f	.039	.062	.679	.030	.023	.204	-.098
D4g	.405	.140	.295	.061	.350	-.186	.072
D4h	.065	.134	.098	.009	.655	.181	-.010
D4i	.266	.149	.269	.063	.566	-.010	-.015
D7a	.108	.189	.125	.094	.184	.810	-.103
D7b	-.005	-.014	-.123	-.184	.065	-.056	.658
D7e	.020	-.024	.039	.023	-.036	-.011	.763
D7f	.216	.115	.152	.094	.101	.764	.133
D9a	.187	.219	.190	.662	.099	.145	-.028
D9e	.004	.054	.108	.806	.159	.038	-.006
D9f	.132	.157	.164	.769	.106	.042	.069
D9g	-.063	-.024	.080	.287	-.143	.075	.568
D9h	-.074	.112	-.019	.253	.582	.153	-.134

Table A.25: Factor loadings for oblimin rotated 7 factor model

	Component						
	1	2	3	4	5	6	7
D1a	.659	-.033	.025	-.098	-.090	-.038	.137
D1b	.487	.037	.135	-.265	-.171	-.019	-.095
D1c	.652	.294	.016	.097	.120	-.042	-.046
D1d	.516	-.059	.048	-.011	.083	.062	.276
D1e	.691	-.089	-.004	-.014	-.204	-.038	.086
D1f	.413	.053	.066	-.396	.022	.030	-.086
D1g	.653	.111	.010	-.066	.144	.011	-.068
D1h	.786	-.033	-.084	.111	-.078	-.032	.088
D1i	.479	-.011	.249	-.130	-.196	.030	-.167
D1j	.567	.053	.090	-.237	-.037	.018	-.108
D2a	.037	.038	-.015	-.831	-.091	-.029	-.061
D2b	-.004	.064	.003	-.801	-.069	-.040	.000
D2c	-.009	-.060	.010	-.812	-.002	-.034	.053
D2d	-.030	-.023	.010	-.855	.020	-.011	-.014
D2e	.096	-.003	.034	-.396	-.428	-.063	.076
D2f	.059	-.040	.039	-.722	-.067	.046	-.038
D4a	.163	.625	.099	-.097	.043	.075	.076
D4b	.120	.318	.045	.098	-.027	.049	.463
D4c	.049	.567	.031	.031	-.016	-.070	.281
D4d	.085	.656	.025	-.056	-.028	.143	-.148
D4e	-.048	.597	.168	-.056	.029	-.081	.184
D4f	-.038	.722	-.046	.042	-.182	-.100	-.058
D4g	.028	.219	.033	-.411	.238	.074	.307
D4h	.060	.029	-.025	-.035	-.161	.023	.645
D4i	.035	.189	.028	-.251	.056	.002	.529
D7a	.076	.106	.013	.007	-.806	-.075	.128
D7b	.027	-.120	-.195	-.013	.019	.670	.128
D7e	-.028	.033	.008	-.008	-.011	.764	-.002
D7f	-.020	.138	.010	-.122	-.766	.156	.060
D9a	.086	.064	.668	-.094	-.067	-.043	-.001
D9e	-.067	-.023	.852	.062	.034	-.025	.086
D9f	.026	.025	.796	-.050	.038	.049	.016
D9g	-.051	.060	.289	.108	-.078	.557	-.144
D9h	.053	-.112	.258	.114	-.121	-.111	.562

Table A.26: Factor loadings for varimax rotated 8 factor model

	Component							
	1	2	3	4	5	6	7	8
D1a	.675	.227	.092	.085	.162	.152	-.040	.159
D1b	.557	.378	.123	.188	.249	.067	-.041	-.073
D1c	.632	.041	.348	.072	-.047	.111	-.052	-.036
D1d	.517	.072	.036	.086	.007	.308	.048	.152
D1e	.682	.172	.040	.057	.247	.038	-.033	.184
D1f	.483	.455	.121	.113	.064	.128	.010	-.121
D1g	.633	.179	.184	.057	-.072	.089	.004	-.050
D1h	.729	.061	.091	-.021	.109	.018	-.015	.210
D1i	.529	.255	.070	.287	.269	.007	.000	-.133
D1j	.605	.355	.146	.143	.108	.045	.008	-.062
D2a	.207	.834	.127	.040	.132	.030	-.011	.002
D2b	.173	.792	.149	.055	.118	.085	-.024	.028
D2c	.151	.786	.043	.048	.046	.090	-.008	.086
D2d	.137	.812	.052	.047	.038	.116	-.002	-.030
D2e	.235	.438	.052	.091	.479	.126	-.089	.040
D2f	.203	.721	.048	.082	.113	.050	.056	-.010
D4a	.281	.178	.653	.175	.033	.222	.068	-.008
D4b	.198	-.048	.387	.104	.086	.352	.055	.328
D4c	.154	.050	.623	.101	.048	.191	-.042	.279
D4d	.171	.160	.660	.102	.040	-.043	.155	-.087
D4e	.094	.096	.614	.220	.046	.275	-.088	.075
D4f	.062	.051	.691	.028	.187	.000	-.096	-.007
D4g	.159	.316	.208	.066	-.084	.609	.013	-.103
D4h	.153	.026	.104	.025	.235	.493	.010	.407
D4i	.174	.168	.189	.069	.106	.728	-.068	.065
D7a	.195	.103	.132	.089	.821	.060	-.114	.122
D7b	-.023	.045	-.038	-.150	-.109	-.175	.737	.234
D7e	-.025	-.001	.023	.046	.014	.059	.741	-.172
D7f	.119	.209	.154	.094	.776	.040	.115	.021
D9a	.223	.183	.190	.663	.152	.067	-.049	.031
D9e	.059	-.001	.113	.810	.049	.090	-.025	.092
D9f	.158	.143	.186	.776	.033	.010	.066	.095
D9g	-.020	-.126	-.006	.292	.148	.161	.472	-.435
D9h	.119	-.032	.084	.273	.117	.111	-.039	.688

Table A.27: Factor loadings for oblimin rotated 8 factor model

	Component							
	1	2	3	4	5	6	7	8
D1a	.654	-.029	.091	.030	-.103	-.020	.091	.094
D1b	.483	.019	.244	.120	-.185	-.060	-.021	-.120
D1c	.649	.276	-.104	.009	.111	-.060	.031	-.080
D1d	.515	-.085	-.030	.049	.043	.054	.284	.066
D1e	.684	-.064	.031	.006	-.203	-.003	-.023	.128
D1f	.411	.028	.364	.052	.004	-.011	.055	-.157
D1g	.651	.102	.059	.005	.138	-.002	.017	-.087
D1h	.779	.004	-.078	-.066	-.067	.026	-.038	.162
D1i	.475	-.033	.111	.231	-.211	-.037	-.087	-.183
D1j	.562	.050	.232	.084	-.039	-.002	-.042	-.100
D2a	.032	.069	.839	-.005	-.070	.009	-.031	.016
D2b	-.009	.087	.800	.013	-.055	-.004	.030	.034
D2c	-.013	-.026	.818	.027	.016	.022	.050	.097
D2d	-.032	-.014	.838	.014	.025	.007	.072	-.022
D2e	.091	-.027	.359	.020	-.454	-.090	.082	-.008
D2f	.055	-.018	.721	.047	-.056	.066	-.001	-.011
D4a	.156	.605	.080	.100	.032	.056	.145	-.063
D4b	.115	.315	-.116	.066	-.054	.087	.343	.241
D4c	.041	.594	-.007	.058	-.004	.000	.148	.241
D4d	.075	.677	.091	.037	.001	.148	-.125	-.097
D4e	-.052	.569	.029	.167	.011	-.093	.218	.028
D4f	-.047	.724	-.022	-.046	-.165	-.091	-.065	-.012
D4g	.034	.100	.272	-.003	.144	-.032	.602	-.196
D4h	.057	-.002	-.026	-.014	-.221	.048	.524	.294
D4i	.039	.056	.094	-.010	-.061	-.099	.740	-.066
D7a	.068	.078	-.036	-.005	-.838	-.113	.021	.044
D7b	.018	.001	.113	-.122	.079	.793	-.137	.219
D7e	-.032	.023	-.010	.019	-.032	.696	.072	-.248
D7f	-.029	.113	.094	-.003	-.795	.101	.001	-.060
D9a	.081	.067	.093	.673	-.068	-.081	-.030	-.013
D9e	-.070	-.015	-.061	.864	.032	-.061	.012	.047
D9f	.020	.063	.081	.821	.059	.043	-.086	.060
D9g	-.049	-.046	-.204	.247	-.152	.355	.142	-.531
D9h	.044	-.005	-.038	.327	-.085	.059	.112	.648

Table A.28: Factor loadings for final varimax rotated 7 factor model

	Component						
	1	2	3	4	5	6	7
D1a	.688	.236	.101	.089	.124	.209	-.048
D1b	.564	.366	.138	.200	.270	-.045	-.040
D1c	.627	.039	.368	.077	-.054	.026	-.042
D1d	.515	.107	.078	.084	-.053	.318	.058
D1e	.699	.172	.025	.059	.219	.173	-.050
D1g	.634	.183	.204	.064	-.095	.002	.019
D1h	.739	.054	.066	-.022	.096	.173	-.039
D1i	.531	.241	.082	.298	.292	-.103	.009
D1j	.596	.349	.156	.148	.125	-.016	.012
D2a	.217	.835	.129	.045	.124	-.002	-.020
D2b	.177	.800	.162	.060	.095	.055	-.030
D2c	.169	.794	.046	.053	.038	.068	-.026
D2d	.149	.830	.072	.054	.035	-.002	-.004
D2f	.207	.728	.057	.086	.090	.020	.056
D4a	.287	.182	.677	.177	.028	.099	.075
D4b	.187	-.006	.409	.089	.015	.520	.051
D4c	.149	.052	.616	.090	.010	.351	-.062
D4d	.175	.140	.637	.098	.086	-.106	.147
D4e	.093	.116	.646	.218	.003	.230	-.076
D4f	.058	.045	.679	.025	.192	.016	-.099
D4h	.146	.088	.151	.014	.137	.666	.008
D7a	.202	.106	.121	.092	.812	.195	-.122
D7b	-.022	.009	-.116	-.175	-.060	.093	.676
D7e	-.032	.029	.044	.035	-.017	-.030	.771
D7f	.117	.219	.140	.092	.812	.106	.107
D9a	.219	.173	.195	.665	.139	.086	-.040
D9e	.058	.002	.116	.807	.033	.144	-.017
D9f	.150	.136	.168	.769	.032	.114	.064
D9g	-.010	-.099	.078	.305	.112	-.241	.536
D9h	.113	-.039	.015	.246	.126	.638	-.116

B Empirical validity of D-LIED implementation: additional results

B.1 Energy data set

Table B.1: Energy data set: mean satisfaction for different expectation and different disconfirmation levels

(a) Disconfirmation of Availability dimension

		Disconfirmation		
		-	0	+
Expectation	Low	3.51	$3.91^{\{a\}}$	3.92
	High	$3.98^{\{a,b\}}$	$4.42^{\{a,b\}}$	$4.49^{\{b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(b) Disconfirmation of Employees dimension

		Disconfirmation		
		-	0	+
Expectation	Low	2.60	3.63	3.90
	High	3.52	$4.35^{\{b\}}$	$4.39^{\{b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(c) Disconfirmation of Services dimension

		Disconfirmation		
		-	0	+
Expectation	Low	2.79	$3.88^{\{a\}}$	$4.44^{\{a\}}$
	High	$3.48^{\{b\}}$	$4.42^{\{a,b\}}$	$4.99^{\{a,b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

Table B.1: Energy data set: mean satisfaction for different expectation and different disconfirmation levels [continued]

(d) Disconfirmation of Information dimension

		Disconfirmation		
		-	0	+
Expectation	Low	3.53	$4.03^{\{a\}}$	$4.51^{\{a\}}$
	High	$4.04^{\{b\}}$	$4.53^{\{a,b\}}$	4.78

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(e) Disconfirmation of Invoices dimension

		Disconfirmation		
		-	0	+
Expectation	Low	2.71	$3.75^{\{a\}}$	$4.09^{\{a\}}$
	High	$3.52^{\{b\}}$	$4.32^{\{a,b\}}$	$4.56^{\{a,b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

B.2 Family entertainment data set

Table B.2: Family entertainment data set: mean satisfaction for different expectation and different disconfirmation levels

(a) Disconfirmation of Overall Experience dimension

		Disconfirmation		
		-	0	+
Expectation	Low	6.69	7.73[a]	8.18[a]
	High	7.80[b]	8.70[a,b]	8.93[a,b]

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(b) Disconfirmation of Product A dimension

		Disconfirmation		
		-	0	+
Expectation	Low	6.69	7.78[a]	8.12[a]
	High	8.24[b]	8.71[a,b]	8.77[b]

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(c) Disconfirmation of Product B dimension

		Disconfirmation		
		-	0	+
Expectation	Low	7.76	7.88	7.91
	High	8.52[b]	8.68[b]	8.64[b]

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

Table B.2: Family entertainment data set: mean satisfaction for different expectation and different disconfirmation levels [continued]

(d) Disconfirmation of Accommodation dimension

		Disconfirmation		
		-	0	+
Expectation	Low	7.59	7.84	7.89
	High	8.17	$8.72^{\{a,b\}}$	$8.67^{\{b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(e) Disconfirmation of Product C dimension

		Disconfirmation		
		-	0	+
Expectation	Low	7.62	7.88	8.06
	High	$8.31^{\{b\}}$	$8.72^{\{a,b\}}$	$8.76^{\{b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(f) Disconfirmation of Product D dimension

		Disconfirmation		
		-	0	+
Expectation	Low	7.75	7.96	7.80
	High	$8.39^{\{b\}}$	$8.76^{\{a,b\}}$	$8.61^{\{b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

Table B.2: Family entertainment data set: mean satisfaction for different expectation and different disconfirmation levels [continued]

(g) Disconfirmation of Personnel dimension

		Disconfirmation		
		-	0	+
Expectation	Low	7.63	7.85	7.90
	High	8.18	8.59[b]	8.76[b]

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(h) Disconfirmation of Prices dimension

		Disconfirmation		
		-	0	+
Expectation	Low	7.80	8.09[a]	7.75
	High	8.55[b]	8.93[a,b]	9.50[b]

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(i) Disconfirmation of Communication dimension

		Disconfirmation		
		-	0	+
Expectation	Low	7.63	7.90[a]	8.12
	High	8.40[b]	8.77[a,b]	8.67[b]

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

B.3 Banking product data set

Table B.3: Banking product data set: mean satisfaction for different expectation and different disconfirmation levels

(a) Disconfirmation of Image dimension

		Disconfirmation		
		-	0	+
Expectation	Low	2.74	3.67$^{\{a\}}$	4.17$^{\{a\}}$
	High	3.22	4.11$^{\{a,b\}}$	4.45$^{\{a,b\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(b) Disconfirmation of Price dimension

		Disconfirmation		
		-	0	+
Expectation	Low	3.72	4.18	4.52
	High	3.93	4.48$^{\{a,b\}}$	4.96$^{\{a\}}$

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(c) Disconfirmation of Product Performance dimension

		Disconfirmation		
		-	0	+
Expectation	Low	2.88	3.78$^{\{a\}}$	3.96
	High	3.67$^{\{b\}}$	4.13$^{\{b\}}$	4.17

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

Table B.3: Banking product data set: mean satisfaction for different expectation and different disconfirmation levels [continued]

(d) Disconfirmation of Product Reliability dimension

		Disconfirmation		
		-	0	+
Expectation	Low	2.86	3.45	3.88
	High	3.49	3.82	4.20 [a,b]

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(e) Disconfirmation of Invoices dimension

		Disconfirmation		
		-	0	+
Expectation	Low	3.57	3.81	4.07
	High	3.61	4.13 [a,b]	4.24

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

(f) Disconfirmation of Communications dimension

		Disconfirmation		
		-	0	+
Expectation	Low	3.62	3.73	4.05 [a]
	High	3.78	4.09 [b]	4.30 [b]

[a] Stat. sign. different at 5% from the cell to the left

[b] Stat. sign. different at 5% from the cell above

C Interaction effects between product dimensions in performance regression analysis: additional results

C.1 Company 2

Table C.1: Regression coefficients: company 2

		Reduced Model	Extended Model
Intercept	α_0	0.17	2.67
Main effects	β_1	0.30	-1.01
	β_2	0.26	-0.33
	β_3	-0.05	0.02
	β_4	0.15	0.07
	β_5	0.12	0.11
	β_6	-0.04	0.88
	β_7	-0.05	1.46
	β_8	0.03	-1.03
	β_9	0.06	0.31
Interaction effects	γ_{72}		-1.58
	γ_{87}		0.35
	γ_{79}		0.73
	γ_{64}		-0.68
	γ_{82}		0.41
	γ_{21}		-0.95
	γ_{49}		-0.56
Adjusted R^2		0.28	0.34

Coefficients significant at 5% are marked in grey

C.2 Company 3

Table C.2: Regression coefficients: company 3

		Reduced Model	Extended Model
Intercept	α_0	0.20	1.15
Main effects	β_1	0.49	-0.23
	β_2	0.10	0.12
	β_3	-0.03	1.01
	β_4	0.03	-0.09
	β_5	0.05	0.29
	β_6	-0.04	0.49
	β_7	0.11	-0.63
	β_8	0.09	-0.58
	β_9	-0.05	0.36
Interaction effects	γ_{84}		0.77
	γ_{29}		0.20
	γ_{54}		-1.05
	γ_{52}		0.77
	γ_{68}		-0.09
	γ_{48}		0.20
	γ_{32}		-0.55
	γ_{38}		-0.16
	γ_{75}		0.16
	γ_{85}		-0.29
	γ_{61}		-0.61
	γ_{47}		-0.61
	γ_{67}		0.33
Adjusted R^2		0.33	0.43

Coefficients significant at 5% are marked in grey

C.3 Company 4

Table C.3: Regression coefficients: company 4

		Reduced Model	Extended Model
Intercept	α_0	0.32	0.66
Main effects	β_1	0.30	0.86
	β_2	0.34	0.31
	β_3	0.01	-0.34
	β_4	-0.22	-0.25
	β_5	-0.01	0.00
	β_6	0.00	-0.18
	β_7	0.05	0.25
	β_8	0.14	-0.16
	β_9	0.01	0.02
Interaction effects	γ_{28}		-0.29
	γ_{38}		0.19
	γ_{68}		0.10
	γ_{12}		-0.42
	γ_{87}		0.15
Adjusted R^2		0.29	0.35

Coefficients significant at 5% are marked in grey

C.4 Company 5

Table C.4: Regression coefficients: company 5

		Reduced Model	Extended Model
Intercept	α_0	0.07	-0.38
Main effects	β_1	0.57	1.02
	β_2	0.32	-0.44
	β_3	-0.17	-0.20
	β_4	-0.03	0.86
	β_5	-0.13	0.84
	β_6	0.09	-0.10
	β_7	0.11	-0.94
	β_8	0.12	-0.37
	β_9	0.05	0.13
Interaction effects	γ_{52}		-0.72
	γ_{74}		1.16
	γ_{14}		-0.66
	γ_{87}		0.45
	γ_{12}		0.32
	γ_{46}		-0.03
	γ_{58}		0.03
Adjusted R^2		0.40	0.51

Coefficients significant at 5% are marked in grey

C.5 Company 6

Table C.5: Regression coefficients: company 6

		Reduced Model	Extended Model
Intercept	α_0	0.27	0.93
Main effects	β_1	0.31	-0.05
	β_2	0.28	-0.10
	β_3	-0.03	0.58
	β_4	0.06	0.18
	β_5	-0.01	-0.37
	β_6	-0.04	0.22
	β_7	0.00	0.10
	β_8	0.11	0.08
	β_9	-0.01	-0.03
Interaction effects	γ_{31}		-0.63
	γ_{41}		0.75
	γ_{63}		-0.16
	γ_{42}		-0.87
	γ_{52}		0.65
	γ_{51}		-0.37
Adjusted R^2		0.32	0.40

Coefficients significant at 5% are marked in grey

C.6 Company 7

Table C.6: Regression coefficients: company 7

		Reduced Model	Extended Model
Intercept	α_0	0.21	0.45
	β_1	0.79	1.07
	β_2	0.18	-0.14
	β_3	0.03	0.40
	β_4	-0.20	-0.12
Main effects	β_5	0.18	-0.26
	β_6	-0.07	-0.33
	β_7	-0.11	-0.39
	β_8	0.05	0.82
	β_9	-0.09	-0.18
	γ_{53}		-0.61
	γ_{84}		-0.67
	γ_{54}		1.50
Interaction effects	γ_{15}		0.74
	γ_{42}		-0.69
	γ_{36}		-0.36
	γ_{46}		1.28
	γ_{14}		-0.79
	γ_{86}		-0.36
	γ_{69}		0.37
	γ_{19}		-0.22
	γ_{25}		-0.54
	γ_{76}		-0.54

Continued on next page

continued from previous page

	Reduced Model	Extended Model
γ_{83}		0.51
γ_{81}		-0.42
γ_{85}		0.38
γ_{71}		0.83
γ_{39}		-0.20
γ_{41}		-0.52
Adjusted R^2	0.39	0.49

Coefficients significant at 5% are marked in grey

D D-LIED IPA for family entertainment data set: additional results

D.1 Company 1

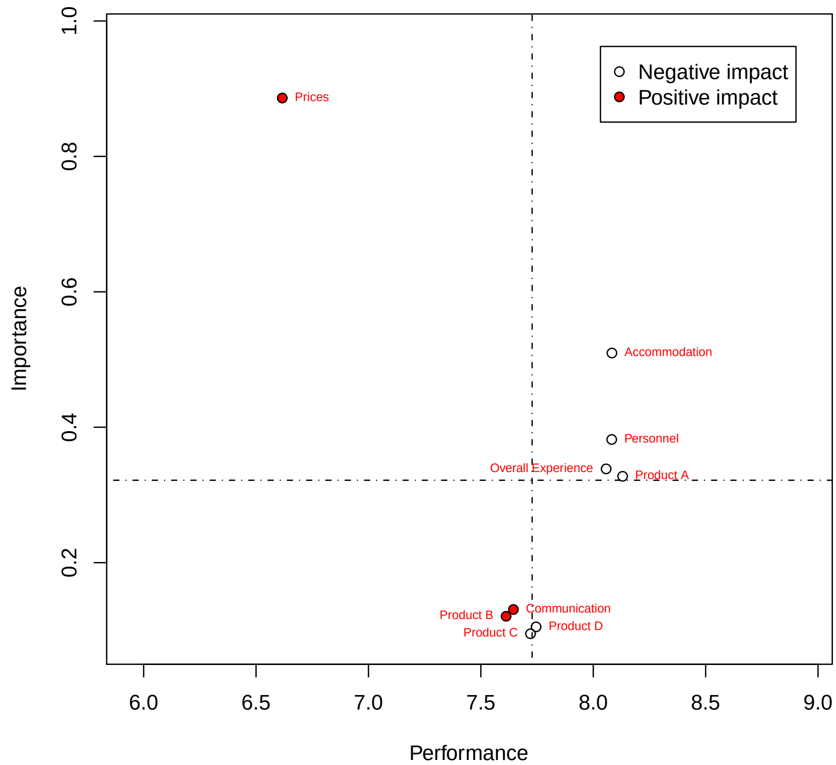


Figure D.1: D-LIED IPA based on disconfirmation impact: company 1.

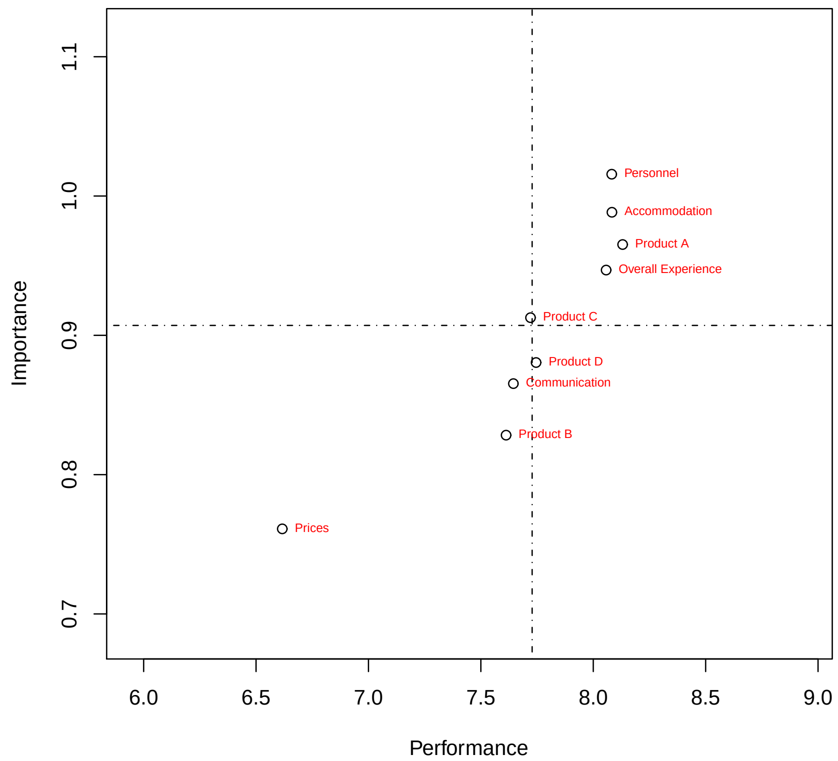


Figure D.2: D-LIED IPA based on marginal impact: company 1.

D.2 Company 2

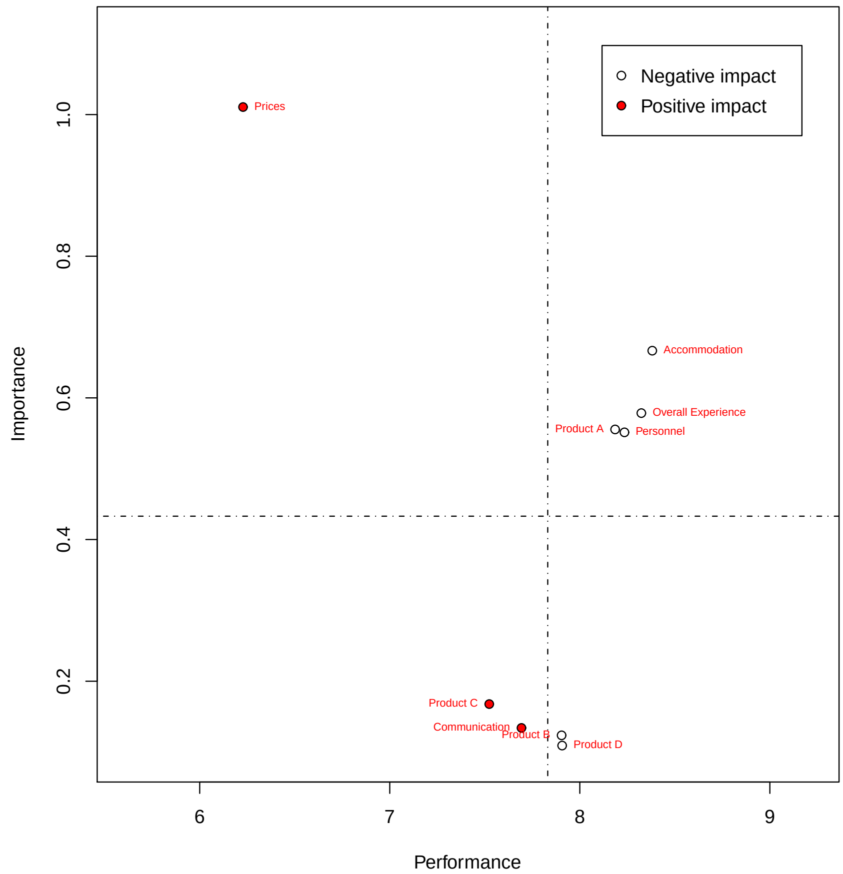


Figure D.3: D-LIED IPA based on disconfirmation impact: company 2.

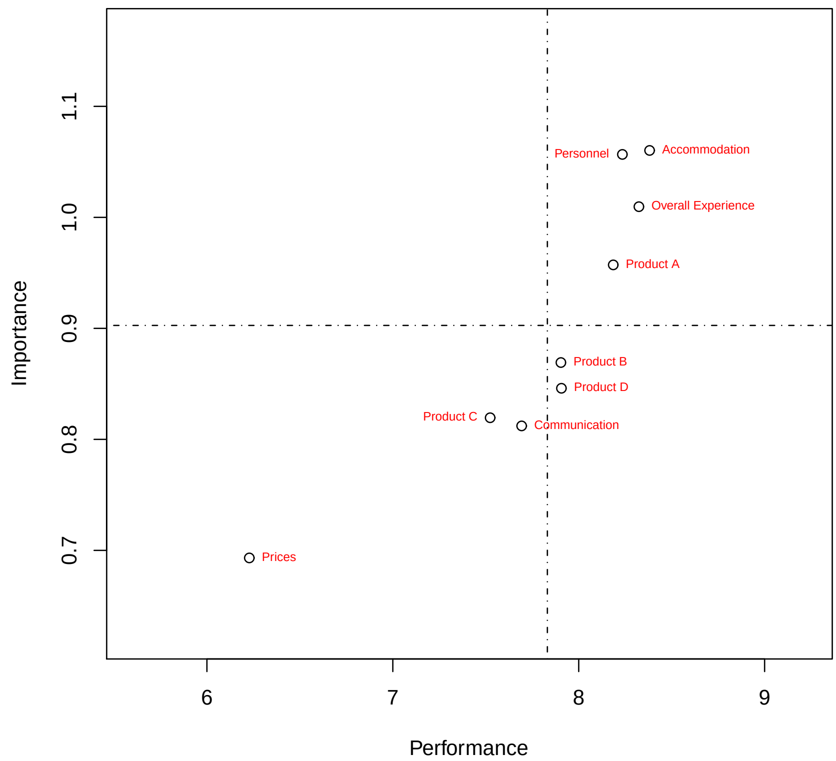


Figure D.4: D-LIED IPA based on marginal impact: company 2.

D.3 Company 3

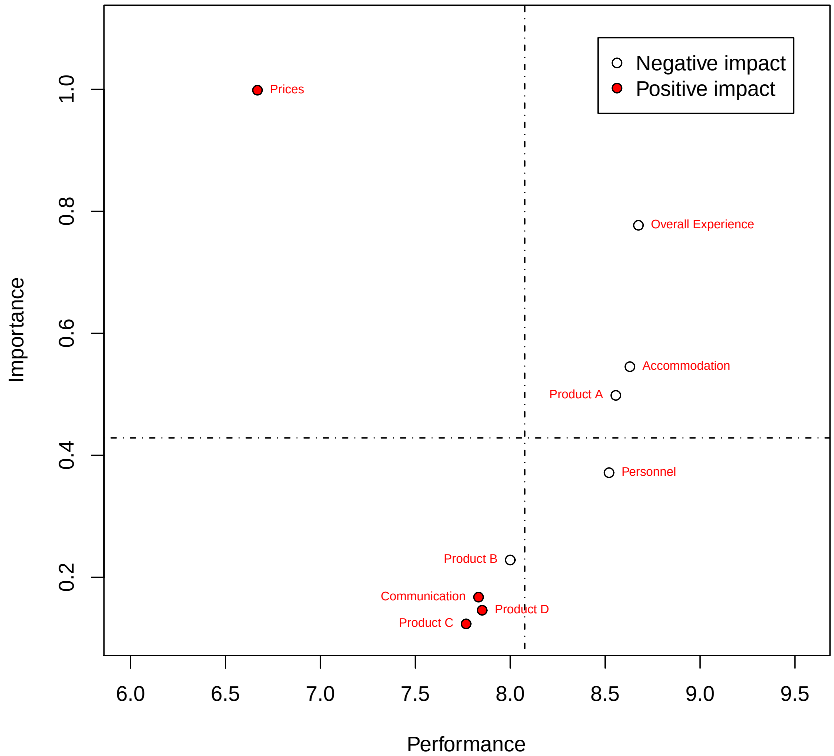


Figure D.5: D-LIED IPA based on disconfirmation impact: company 3.

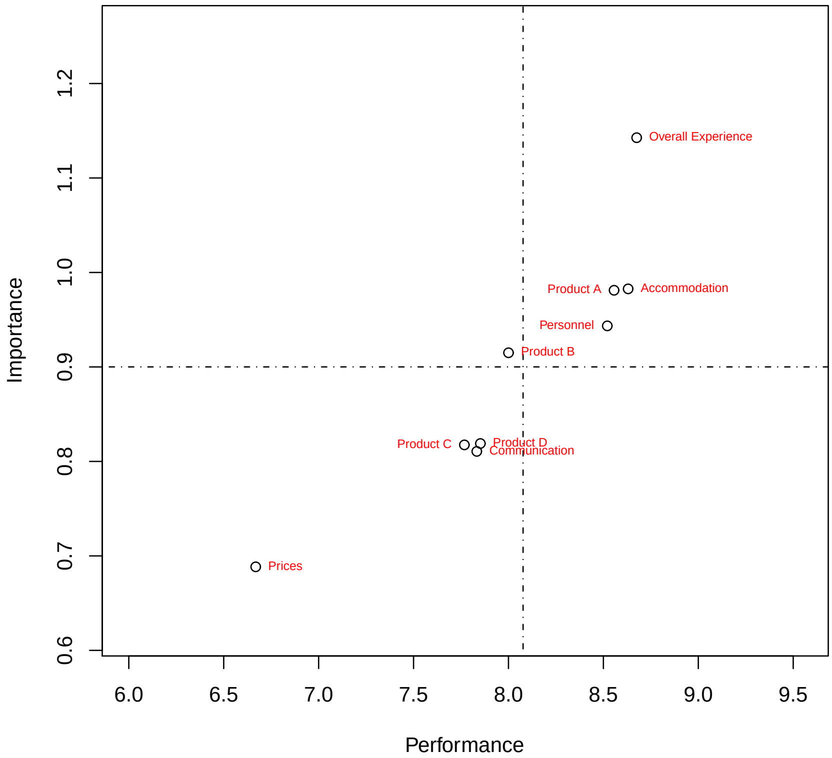


Figure D.6: D-LIED IPA based on marginal impact: company 3.

D.4 Company 5

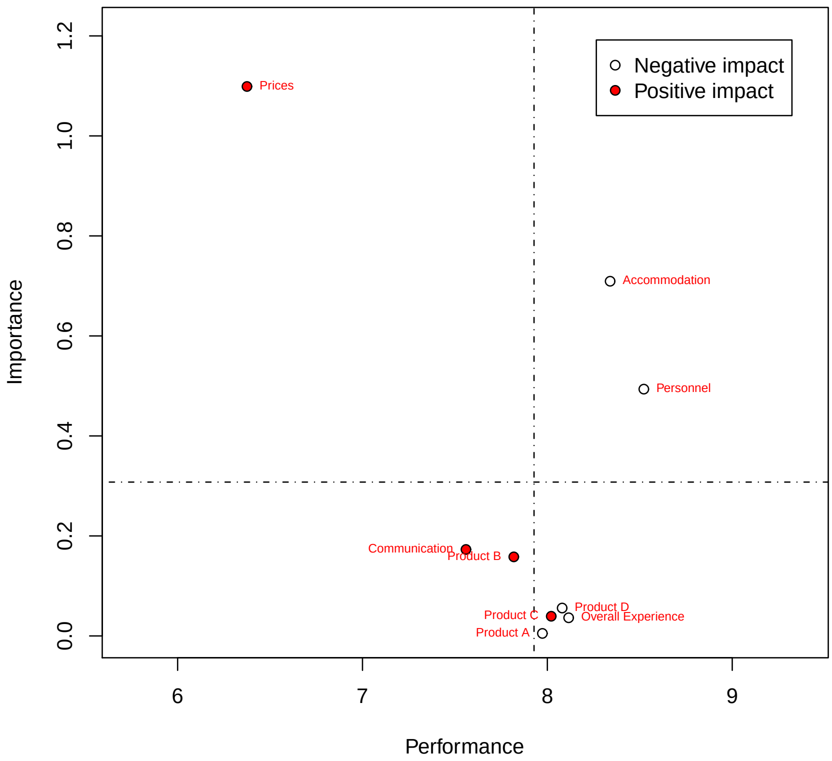


Figure D.7: D-LIED IPA based on disconfirmation impact: company 5.

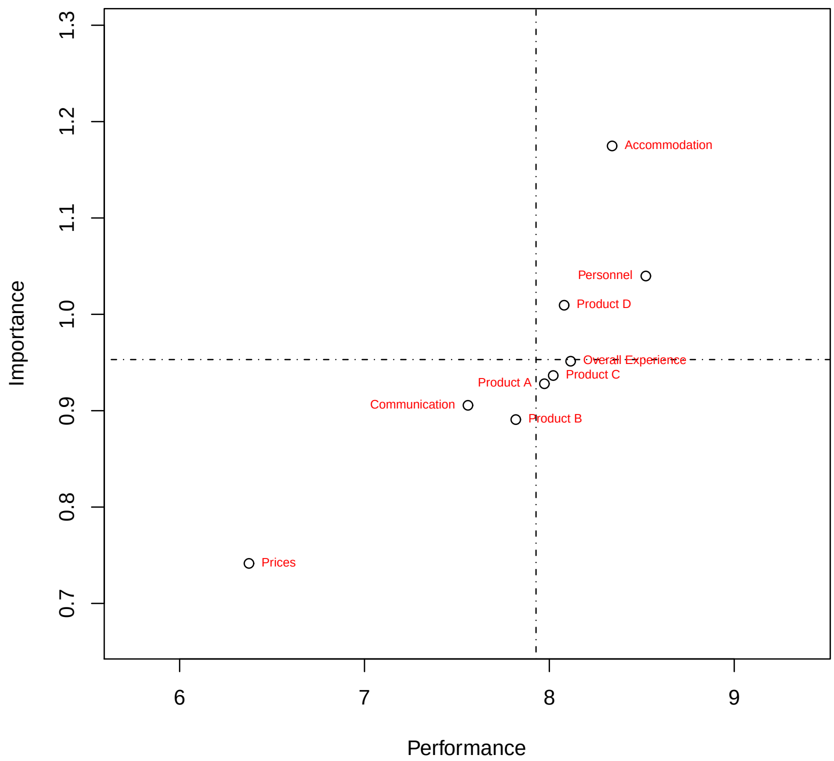


Figure D.8: D-LIED IPA based on marginal impact: company 5.

D.5 Company 6

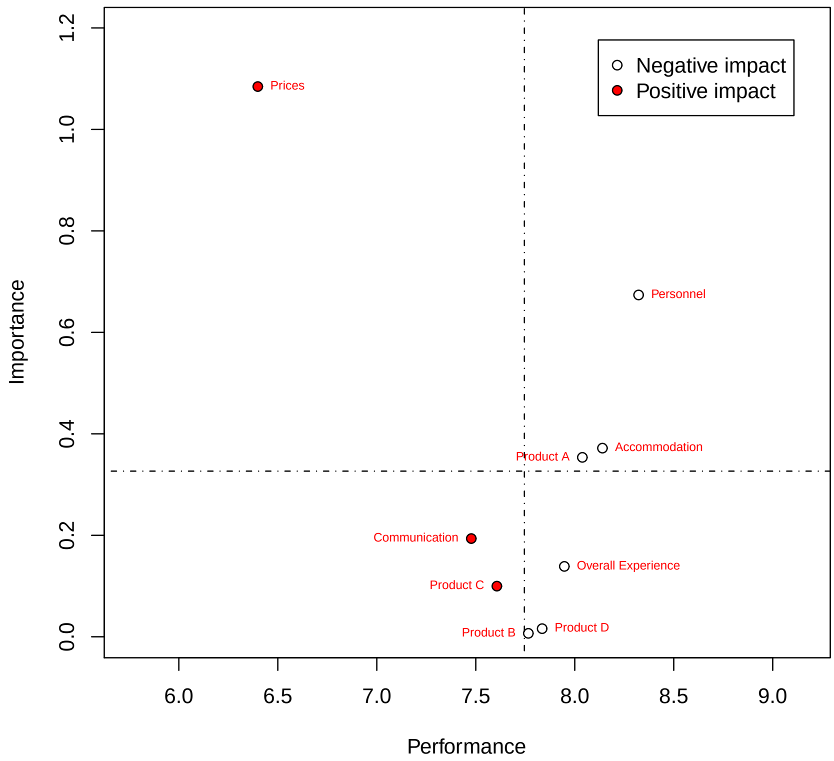


Figure D.9: D-LIED IPA based on disconfirmation impact: company 6.

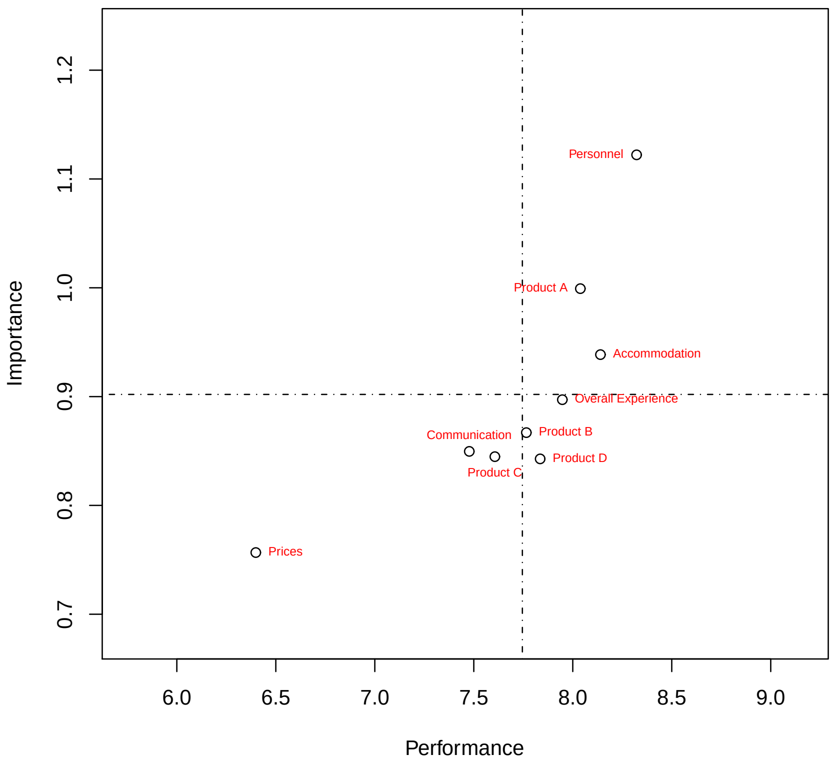


Figure D.10: D-LIED IPA based on marginal impact: company 6.

D.6 Company 7

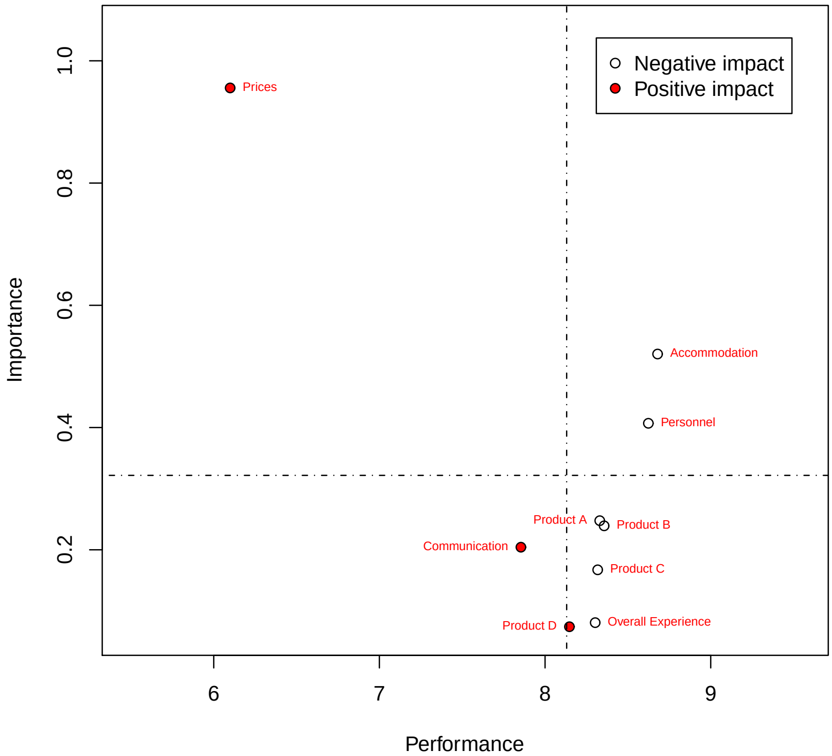


Figure D.11: D-LIED IPA based on disconfirmation impact: company 7.

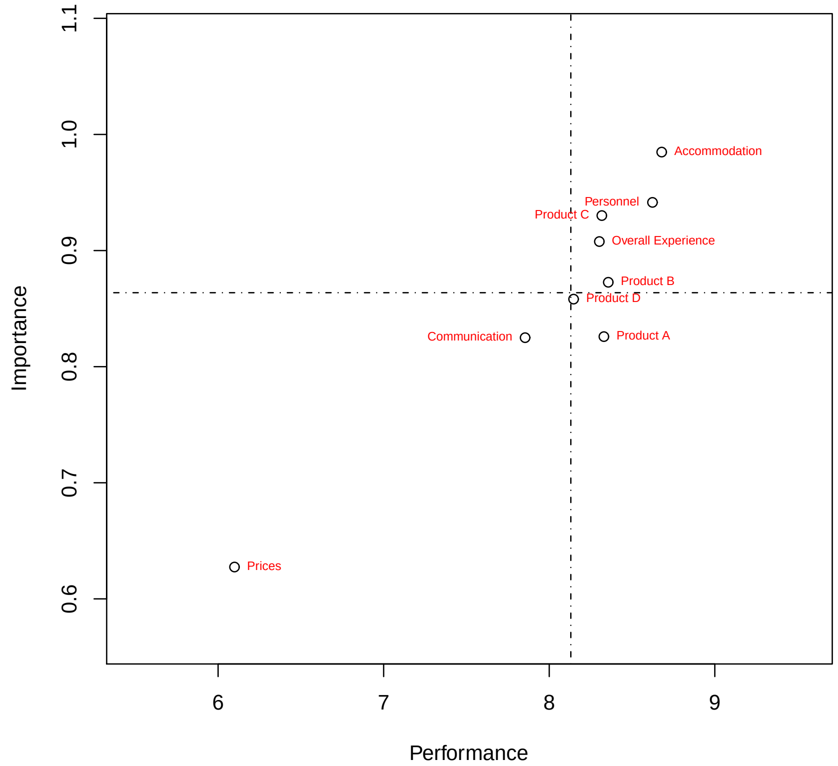


Figure D.12: D-LIED IPA based on marginal impact: company 7.

Bibliography

[1] Coordination action for ubiquitous knowledge discovery.

[2] Akaike, H. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on 19*, 6 (Dec 1974), 716–723.

[3] Anderson, E. W. Cross-category variation in customer satisfaction and retention. *Marketing Letters 5* (Jan 1994), 19–30.

[4] Anderson, E. W., Fornell, C., and Lehman, D. Customer satisfaction, market share and profitability: Findings from sweden. *Journal of Marketing 58* (Jul 1994), 63–66.

[5] Anderson, E. W., Fornell, C., and Mazvancheryl, S. K. Customer satisfaction and shareholder value. *Journal of Marketing 68* (Oct 2004), 172–185.

[6] Anderson, E. W., and Sullivan, M. W. The antecedents and consequences of customer satisfaction for firms. *Marketing Science 12*, 2 (1993), 125–143.

[7] Anderson, R. E. Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance. *Journal of Marketing Research 10* (Feb 1973), 38–44.

[8] Arabie, P., and Hubert, L. Cluster analysis in marketing research. In *Advanced Methods of Marketing Research*. Blackwell, Oxford, 1994, pp. 160–189.

[9] Atkinson, A. *Plots, Transformations, and Regression.* Oxford University Press, 1985.

[10] Ayad, H., and Kamel, M. Finding natural clusters using multi-cluster combiner based on shared nearest neighbors. In *Proceedings 4th International Workshop on Multiple Classifier Systems* (2003), pp. 166–175.

[11] Bearden, W. O., and Teel, J. E. Selected determinants of consumer satisfaction and complaint reports. *Journal of Marketing Research (JMR) 20*, 1 (1983), 21–28.

[12] Beliakov, G. Definition of general aggregation operators through similarity relations. *Fuzzy Sets and Systems 114*, 3 (Sep 2000), 437–453.

[13] Beliakov, G., Calvo, T., and Pradera, A. Handling neutral and absorbent information in aggregation processes. In *Fourth International Summer School on Aggregation Operators* (2007), pp. 31–36.

[14] Beliakov, G., Pradera, A., and Calvo, T. *Aggregation Functions: A Guide for Practitioners*, vol. 221/2007. Springer Berlin / Heidelberg, 2007.

[15] Bezdek, J. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum Press, New York, 1981.

[16] Bratton, D., and Kennedy, J. Defining a standard for particle swarm optimization. In *Proceedings of the IEEE Swarm Intelligence Symposium (SIS 2007)* (2007), pp. 120–127.

[17] Breusch, T., and Pagan, A. A simple test for heteroscedasticity and random coefficient variation. *Econometrica 47* (1979), 135–143.

[18] Brown, T. A. *Confirmatory Factor Analysis for Applied Research.* The Guilford Press, New York, 2006.

[19] Cadotte, E. R., Woodruff, R. B., and Jenkins, R. L. Expectations and norms in models of consumer satisfaction. *Journal of Marketing Research 24*, 3 (Aug 1987), 305–314.

[20] Calvo, T., Mayor, G., and Mesiar, R. *Aggregation Operators: new trends and applications.* Physica-Verlag, 2002.

[21] Cantillo, V., and de Dios Ortzar, J. A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transportation Research Part B: Methodological 39*, 7 (2005), 641–657.

[22] Cardozo, R. N. An experimental study of customer effort, expectation, and satisfaction. *Journal of Marketing Research (JMR) 2*, 3 (1965), p244 – 249.

[23] Carlsmith, J. M., and Aronson, E. Some hedonic consequences of the confirmation and disconfirmation of expectancies. *Journal of Abnormal and Social Psychology 66* (1963), 151–156.

[24] Choquet, G. Theory of capacities. *Annales de l'institut Fourier 5* (1953-54), 131–298.

[25] Churchill, G. A. J. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research 16* (1979), 64–73.

[26] Churchill, G. A. J., and Suprenant, C. An investigation into the determinants of customer satisfaction. *Journal of Marketing Research 19* (Nov 1982), 491–504.

[27] Cohen, J., and Cohen, P. *Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences.* Lawrence Erlbaum Associates, 1983.

[28] Cohen, J. B., and Goldberg, M. E. The dissonance model in post-decision product evaluation. *Journal of Marketing Research (JMR) 7*, 3 (1970), p315 – 321.

[29] Cohen, P. R. *Empirical Methods for Artificial Intelligence.* The MIT Press, 1995.

[30] Conover, W. J. *Practical Nonparametric Statistics*, third ed. John Wiley & Sons, Inc., 1999.

[31] Curtin, F., and Schulz, P. Multiple correlations and bonferroni's correction. *Biological Psychiatry 44* (1998), 775–777.

[32] Cutello, V., and Montero, J. Nondeterministic aggregation operators and systems. *International Journal of Intelligent Systems 13*, 2–3 (1998), 181–192.

[33] Depaire, B., Vanhoof, K., and Wets, G. Managerial opportunities of uninorm based importance–performance analysis. *WSEAS TRANSACTIONS on BUSINESS and ECONOMICS 3*, 3 (Mar 2006), 101–108.

[34] Depaire, B., Wets, G., and Vanhoof, K. Traffic accident segmentation by means of latent class clustering. *Accident Analysis & Prevention 40*, 4 (2008), 1257–1266.

[35] Dhillon, I., and Modha, D. A data-clustering algorithm on distributed memory multiprocessors. In *7th International Conference on Knowledge Discovery and Data Mining (KDD)* (1999), pp. 245–260.

[36] Dick, A. S., and Basu, K. Customer loyalty: Toward an integrated conceptual framework. *Journal of the Academy of Marketing Science 22*, 2 (1994), 99–113.

[37] Dolinsky, A. L., and Caputo, R. K. Adding a competitive dimension to importance-performance analysis: An application to traditional health care systems. *Health Marketing Quarterly 8*, 3/4 (1991), 61–79.

[38] Dombi, J. Basic concepts for the theory of evaluation: The aggregative operator. *European Journal of Operational Research 10*, 3 (Jul 1982), 282–293.

[39] Dubois, D., and Prade, H. On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems 142*, 1 (Feb 2004), 143–161.

[40] Dunn, J. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Cybernatics and Systems 3* (1973), 32–57.

[41] Efron, B. *The Jackknife, the Bootstrap, and Other Resampling Plans.* CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial & Applied Mathematics, 1982.

[42] Efron, B., and Tibshirani, R. *An Introduction to the Bootstrap.* Chapman & Hall/CRC, 1993.

[43] Elrod, T., Johnson, R. D., and White, J. A new integrated model of noncompensatory and compensatory decision strategies. *Organizational Behavior and Human Decision Processes 95*, 1 (2004), 1–19.

[44] Etienne, K. *Introduction to the Basic Principles of Fuzzy Set Theory and Some of Its Applications.* Communication & Cognition, Gent, 1991.

[45] Everitt, B., Landau, S., and Leese, M. *Cluster Analysis.* Arnold, London, 2001.

[46] Falcón, R., Jeon, G., Bello, R., and Jeong, J. Learning collaboration links in a collaborative fuzzy clustering environment. In *MICAI* (2007), pp. 483–495.

[47] Festinger, L. *A Theory of Cognitive Dissonance.* Harper & Row, New York, 1957.

[48] Fishbein, M. Attitude and the prediction of behavior. In *Readings in Attitude Theory and Measurement*, M. Fishbein, Ed. John Wiley & Sons, Inc., New York, USA, 1967.

[49] Fishbein, M., and Ajzen, I. *Belief, Attitude, Intention and Behavior.* Addison-Wesley, Reading, Massachusetts, USA, 1975.

[50] Fodor, J., and De Baets, B. Uninorm basics. In *Fuzzy Logic A Spectrum of Theoretical & Practical Issues*, P. Wang, D. Ruan, and E. E. Kerre, Eds. Springer, Heidelberg, 2007, pp. 51–66.

[51] Fodor, J. C., Yager, R. R., and Rybalov, A. Structure of uninorms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 5*, 4 (Aug 1997), 411–427.

[52] Fox, J. *Applied Regression Analysis, Linear Models, and Related Methods.* Sage Publications, Inc., 1997.

[53] Fox, J. *An R and S-Plus Companion to Applied Regression.* Sage Publications, Inc., 2002.

[54] Gilbride, T. J., and Allenby, G. M. A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science 23* (2004), 391–406.

[55] Grabisch, M., Murofushi, T., and Sugeno, M., Eds. *Fuzzy Measures and Integrals. Theory and Applications.* Physica-Verlag, Heidelberg, 2000.

[56] Green, P. Practice makes perfect. *Marketing Research 16*, 2 (2004), 8–14.

[57] Griffen, A., and Hauser, J. The voice of the customer. *Marketing Science 12*, 1 (1993), 1–27.

[58] Gujarati, D. N. *BASIC ECONOMETRICS.* McGraw-Hill, Inc., 1995.

[59] Hair, J. F. J., Black, W. C., Babin, B. J., Anderson, R. E., and Tatham, R. L. *Multivariate Data Analysis*, 6 ed. Pearson, Prentice Hall, 2006.

[60] Hartigan, J. *Clustering Algorithms.* Wiley, 1975.

[61] Helson, H. Current trends and issues in adaption-level theory. *American Psychologist 19* (1964), 26–38.

[62] Hill, T. *MANUFACTURING STRATEGY. TEXT AND CASES.* MACMILLAN PRESS LTD, 1995.

[63] Homburg, C., Koschate, N., and Hoyer, W. D. Do satisfied customers really pay more? a study of the relationship between customer satisfaction and willingness to pay. *Journal of Marketing 69* (April 2005), 84–96.

[64] Hoppner, F., Klawonn, F., Kruse, R., and Runkler, T. *Fuzzy Cluster Analysis*. John Wiley, Chichester, 1999.

[65] Hosmer, D., and Lemeshow, S. *Applied Logistic Regression*. Wiley, New York, 1989.

[66] Howard, J. A. The structure of buyer behavior. In *Consumer Behavior: Theory and Applications*, J. U. Farley, J. A. Howard, and L. Winston Ring, Eds. Allyn & Bacon, Boston, USA, 1974.

[67] Howard, J. A., and Sheth, J. N. *The Theory of Buyer Behavior*. John Wiley & Sons, Inc., New York, USA, 1969.

[68] Ilgen, D. R. Satisfaction with performance as a function of the initial level of expected performance and the deviation from expectations. *organizzational Behavior and Human Performance 6* (Jan 1971), 345–361.

[69] Inc., S. Pasw missing values 17.0. available at `http://support.spss.com/ProductsExt/Statistics/Documentation/17.0.2/UserManuals/English/PASWMissingValues17.0.pdf`. Accessed April 22, 2009.

[70] Januzaj, E., Kriegel, H.-P., and Pfeifle, M. Dbdc: Density based distributed clustering. In *Proceedings of the 9th International Conference on Extending Database Technology* (Heraklion, Greece, 2004), pp. 88–105.

[71] Jiang, D., Tang, C., and Zhang, A. Cluster analysis for gene expression data: A survey. *IEEE Transactions on knowledge and data engineering 16* (2004), 1370–1386.

[72] Johnson, E., and Kargupta, H. Collective, hierarchical clustering from distributed, heterogeneous data. In *Large-Scale Parallel Data Mining* (2000), vol. 1759/2000 of *Lecture Notes in Computer Science*, pp. 803–803.

[73] Kano, N. Attractive quality and must be quality. *Hinshitsu (Quality) 14*, 2 (1984), 147–156.

[74] Kennedy, J., and Eberhart, R. Particle swarm optimization. In *Proceedings of the 1995 IEEE International Conference on Neural Networks* (Piscataway, NJ, 1995), vol. 4, IEEE Press, pp. 1942–1948.

[75] Kerre, E. E., and Mordeson, J. N. A historical overview of fuzzy mathematics. *New Mathematics and Natural Computing 1*, 1 (Mar 2005), 1–26.

[76] King, G., Honaker, J., Joseph, A., and Scheve, K. Analyzing incomplete policitcal science data: An alternative algorithm for multiple imputation. *American Political Science Review 95*, 1 (2001), 49–69.

[77] Klement, E., Mesiar, R., and Pap, E. *Triangular Norms.* Kluwer Academic Publishers, 2000.

[78] Komorníková, M. Aggregation operators and additive generators. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9* (2001), 205–215.

[79] LaBarbera, P. A., and Mazursky, D. A longitudinal assessment of consumer satisfaction/dissatisfaction: The dynamic aspect of the cognitive process. *Journal of Marketing Research 20* (Nov 1983), 393–404.

[80] LaTour, S. A., and Peat, N. C. Conceptual and methodological issues in consumer satisfaction research. *Advances in Consumer Research 6*, 1 (1979), 431–437.

[81] LaTour, S. A., and Peat, N. C. The role of situationally-produced expectations, others' experiences, and prior experience in determining customer satisfaction. *Advances in Consumer Research 7*, 1 (1980), 588–592.

[82] Martilla, J. A., and James, J. C. Importance-performance analysis. *Journal of Marketing 41*, 1 (Jan 1977), 77–79.

[83] Matzler, K., Bailom, F., Hinterhuber, H. H., Renzl, B., and Pichler, J. The asymmetric relationship between attribute-level performance and overall customer satisfaction: a reconsideration of the importance-performance analysis. *Industrial Marketing Management 33*, 4 (May 2004), 271–277.

[84] Matzler, K., Sauerwein, E., and Heischmidt, K. A. Importance-performance analysis revisited: The role of the factor structure of customer satisfaction. *The Service Industries Journal 23*, 2 (Mar 2003), 112–129.

[85] Menard, S. *Applied Logistic Regression Analysis.* Quantitative Applications in the Social Sciences. Sage Publications, Inc., 2002.

[86] Merugu, S., and Ghosh, J. A privacy-sensitive approach to distributed clustering. *Pattern Recognition Letters 26*, 4 (Mar. 2005), 399–410.

[87] Miller, J. A. Exploring some alternative measures of consumer satisfaction. In *Marketing: 1776-1976 and Beyond*, K. L. Bernhardt, Ed. American Marketing Association, Chicago, 1976, pp. 661–664.

[88] Miller, R. G. *Beyond Anova, Basics of Applied Statistics.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., 1986.

[89] Mittal, V., and Kamakura, W. Satisfaction, repurchase intention, and repurchase behavior: Investigating the moderating effect of customer characteristics. *Journal of Marketing Research 38* (Feb 2001), 131–142.

[90] Mittal, V., T. Ross, Jr., W., and Baldasare, P. M. The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *Journal of Marketing 62*, 1 (Jan 1998), 33–47.

[91] Monti, S., Tamayo, P., Mesirov, J., and Golub, T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning 52* (2003), 91–118.

[92] Moustaki, I., and Papageorgiou, I. Latent class models for mixed variables with appications in archaeometry. *Computational Statistics & Data Analysis 48*, 3 (Mar 2005), 659–675.

[93] Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. *Applied Linear Statistical Models*, fourth ed. Irwin, 1996.

[94] Nunnally, J. C., and Bernstein, I. H. *Psychometric Theory*, 3 ed. Series in Psychology. McGraw-Hill, 1994.

[95] Oliver, R. L. Effect of expectation and discontinuation on postexposure product evaluations: An alternative interpretation. *Journal of Applied Psychology 62*, 4 (1977), p480 – 486.

[96] Oliver, R. L. A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research 17*, 4 (Nov 1980), 460–469.

[97] Oliver, R. L. *Satisfaction: A Behavioral Perspective on the Consumer*. McGraw-Hill, 1996.

[98] Oliver, R. L. Processing the satisfaction response in consumption: A suggested framework and research propositions. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior 2* (1998), 1–16.

[99] Oliver, R. L., and DeSarbo, W. Response determinants in satisfaction judgments. *Journal of Consumer Research 14*, 4 (Mar 1988), 495–507.

[100] Olshavsky, R. W., and Miller, J. A. Consumer expectations, product performance, and perceived product quality. *Journal of Marketing Research (JMR) 9*, 1 (1972), p19 – 21.

[101] Olson, J. C., and Dover, P. Effects of expectation creation and disconfirmation on belief elements of cognitive structure. *Advances in Consumer Research 3*, 1 (1976), p168 – 175.

[102] Parasuraman, A., Berry, L. L., and Zeithaml, V. A. Understanding customer expectations of service. *Sloan Management Review 32* (1991), 39–48.

[103] Parasuraman, A., Zeithaml, V. A., and Berry, L. L. Servqual: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing 64* (1988), 12–40.

[104] Pedrycz, W. Collaborative fuzzy clustering. *Pattern Recogn. Lett. 23*, 14 (2002), 1675–1686.

[105] Pedrycz, W., and Rai, P. Collaborative clustering with the use of fuzzy c-means and its quantification. *Fuzzy Sets Syst. 159*, 18 (2008), 2399–2427.

[106] Peter, J. P. Reliability: A review of psychometric basics and recent marketing practices. *Journal of Marketing Research 16*, 1 (Feb 1979), 6–17.

[107] Rubin, D. B. Inference and missing data. *Biometrika 581–592* (1976), 581–592.

[108] Rudas, I. J., and Fodor, J. Information aggregation in intelligent systems using generalized operators. *International Journal of Computers, Communications and Control 1*, 1 (2006), 47–57.

[109] Sammour, G., Depaire, B., Vanhoof, K., and Wets, G. Identifying homogenous customer segments for low risk email marketing experiments. In *Proceedings 11th International Conference on Entreprise Information Systems: Software Agents and Internet Computing* (Milan, May 2009), INSTICC, pp. 89–94.

[110] Sampson, S. E., and Showalter, M. J. The performance-importance response function: Observations and implications. *The Service Industries Journal 19*, 3 (Jul 1999), 1–25.

[111] Scheffer, J. Dealing with missing data. In *Research Letters in the Information and Mathematical Sciences* (2002), pp. 153–160.

[112] Seaman, J. W. J., Walls, S. C., Wide, S. E., and Jaeger, R. G. Caveat emptor: rank transform methods and interactions. *Trends Ecol. Evol. 9* (1994), 261–263.

[113] Sethi, V., and King, R. Nonlinear and noncompensatory models in user information satisfaction measurement. *Information Systems Research 10*, 1 (1999), 87–96.

[114] Sheskin, D. J. *Handbook of Parametric and Nonparametric Statistical Procedures.* CRC Press, Boca Raton, FL, 1997.

[115] Slack, N. The importance-performance matrix as a determinant of improvement priority. *International Journal of Operations & Production Management 14*, 5 (1994), 59–75.

[116] Smith, P. C., Kendall, L. M., and Hulin, C. L. *The Measurement of Satisfaction in Work and Retirement.* Rand McNally, Chicago, USA, 1969.

[117] Spreng, R. A., MacKenzie, S. B., and Olshavsky, R. W. A reexamination of the determinants of consumer satisfaction. *Journal of Marketing 60*, 3 (1996), 15–32.

[118] Steiger, J. H. Tests for comparing elements of a correlation matrix. *Psychological Bulletin 87*, 2 (1980), 245–251.

[119] Strehl, A., and Ghosh, J. Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res. 3* (2003), 583–617.

[120] Sugeno, M. *Theory of Fuzzy Integrals and Applications.* PhD thesis, Tokyo Institute of Technology, Tokyo, 1974.

[121] Swan, J. E., and Trawick, F. I. Disconfirmation of expectations and satisfaction with a retail service. *Journal of Retailing 57* (1981), 49–67.

[122] Szymanski, D. M., and Henard, D. H. Customer satisfaction: A meta-analysis of the empirical evidence. *Journal of the Academy of Marketing Science 29*, 1 (2001), 16–35.

[123] Tasoulis, D., and Vrahatis, M. Unsupervised distributed clustering. In *Proceedings of the IASTED International Conference on Parallel and Distributed Computing and Networks* (Innsbruck, Austria, Feb 2004), IASTED/ACTA Press 2004, pp. 347–351.

[124] Thibaut, J., and Kelly, H. *The Social Psychology of Groups.* Wiley, New York, 1959.

[125] Toothaker, L. E. *Multiple Comparison Procedures.* Quantative Applications in the Social Sciences. Sage Publications, Inc., 1993.

[126] Topchy, A., Jain, K., and Punch, W. Clustering ensembles: Models of consensus and weak partitions. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27* (2005), 1866–1881.

[127] Tse, D. K., and Wilton, P. C. Models of consumer satisfaction formation: An extensive. *Journal of Marketing Research 25*, 2 (May 1988), 204–212.

[128] Vaidya, J., and Clifton, C. Privacy-preserving k-means clustering over vertically partitioned data. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA, 2003), ACM, pp. 206–215.

[129] Vanhoof, K., Brijs, T., and Wets, G. An indirect measurement for customer expectation. In *Principles of Fuzzy Preference Modelling and Decision Making*, B. D. Baets and J. Fodor, Eds., no. ISBN 90-382-0567-8. s.l. Academia Press 2003, 2003, pp. 109–122.

[130] Vanhoof, K., Pauwels, P., Dombi, J., Brijs, T., and Wets, G. Penalty–reward analysis with uninorms: A study of customer (dis)satisfaction. In *INTELLIGENT DATA MINING. Techniques and Applications*, D. Ruan, G. Chen, E. E. Kerre, and G. Wets, Eds., no. ISBN 3-540-26256-3. Springer Berlin / Heidelberg, 2005, pp. 237–252.

[131] Venables, W. N., and Ripley, B. D. *Modern Applied Statistics with S*, 4 ed. Springer, 2002.

[132] von Neumann, J., and Morgenstern, O. *Theory of games and economic behavior.* Princeton University Press, Princeton, NJ, 1947.

[133] Voss, G. B., Parasuraman, A., and Grewal, D. The roles of price, performance, and expectations in determining satisfaction in service exchanges. *Journal of Marketing 62* (1998), 46–61.

[134] Woodruff, R. B., Cadotte, E. R., and Jenkins, R. L. Modeling consumer satisfaction processes using experience-based norms. *Journal of Marketing Research 20*, 3 (1983), p296–304.

[135] Xu, R., and Wunsch, D. I. Survey of clustering algorithms. *IEEE Transactions on Neural Networks 16*, 3 (May 2005), 645–678.

[136] Xu, X., Jäger, J., and Kriegel, H.-P. A fast parallel clustering algorithm for large spatial databases. *Data Mining and Knowledge Discovery 3* (1999), 293–290.

[137] Yager, R. An ordered weighted averaging operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics 18* (1988), 183–190.

[138] Yager, R., and Kacprzyck, J. *The ordered weighted averaging operator, Theory and Applications.* Kluwer Academic Publisher, Boston Dordecht London, 1997.

[139] Yager, R., Ovchinnikov, S., Tong, R., and Nguyen, H., Eds. *Fuzzy Sets and Applications: Selected Papers by L.A. Zadeh.* John Wiley & Sons, Inc., New York, 1987.

[140] Yager, R. R. A model of participatory learning. *IEEE Transactions on Systems, Man and Cybernetics 20* (1990), 1229–1234.

[141] Yager, R. R. Uninorms in fuzzy systems modeling. *Fuzzy Sets and Systems 122*, 1 (Aug 2001), 167–175.

[142] Yager, R. R. Participatory learning. a paradigm for building better digital and human agents. *Law, Probability and Risk 3* (2004), 133–145.

[143] Yager, R. R., and Kacprzyk, J. *The Ordered Weighted Averaging Operators.* Kluwer Academic Publishers, 1997.

[144] Yager, R. R., and Rybalov, A. Uninorm aggregation operators. *Fuzzy Sets and Systems 80*, 1 (May 1996), 111–120.

[145] Yi, Y. A critical review of consumer satisfaction. In *Review of Marketing 1990*, V. A. Zeithaml, Ed. American Marketing Association, Chicago, 1990, pp. 68–123.

[146] Zadeh, L. A. Fuzzy sets. *Information and Control 8* (1965), 338–353.

[147] Zimmermann, H.-J. *Fuzzy Set Theory and its Applications.* Kluwer Academic Publishers, 1991.