Limburg University Center
Department of Applied Economics

Retail Market Basket Analysis: A Quantitative Modelling Approach

Dissertation submitted to obtain the degree of

Doctor in Applied Economic Sciences

at the Limburg University Center, to be defended by

Tom BRIJS

Promotor:
Prof. dr. G. Swinnen

Co-promotor:
Prof. dr. K. Vanhoof

2002

SUMMARY

Market basket analysis is a generic term for methodologies that study the composition of a basket of products (i.e. a shopping basket) purchased by a household during a single shopping trip. The idea is that market baskets reflect interdependencies between products or purchases made in different product categories, and that these interdependencies can be useful to support retail marketing decisions. Recently, a number of advances in data mining (association rules) and statistics (mixture models) offer new opportunities to analyse such data. In this dissertation, the focus is therefore on the development and application of such techniques for two specific problems where product/category interdependencies play an important role, i.e. in product selection and in behaviour-based customer segmentation.

From a marketing perspective, the research is motivated by the fact that some recent trends in retailing pose important challenges to retailers in order to stay competitive. In fact, on the level of the retailer, a number of trends can be identified, including concentration, internationalization, decreasing profit margins and an increase in discounting. This growing trend of concentration and increase in scale has a significant impact on the relation with the consumer and presents important challenges for today's retailers, including the battle against decreased customer service and loyalty. Indeed, the rise of large retail stores and the fact that customers are getting used to self-service resulted in a loss of personalized customer service and creates new challenges to gain and keep customer loyalty, for instance through personalization. Indeed, in today's large grocery stores, most consumers do not know the manager or checkout clerks, and store personnel hardly know their customers. Additionally, some trends on the level of the consumer can also be identified, such as a decrease in loyalty and a slowdown in consumer spending. Indeed, three out of four customers shop at multiple supermarkets and the average loyalty towards the first store of preference is only about twenty percent. With respect to

consumer spending, statistics indicate that the proportion of the household budget spent on groceries has fallen by thirty percent in the last two decades. As a result of these trends in the retail sector, large retailers are currently faced with the continuous balancing act to further personalize marketing communications with their customer whilst maintaining the variety and efficiency of the supermarket formula. This is not at all an easy task: retailers are fighting over the consumer's "share of wallet" and satisfying the diverse wants and needs of the consumer forces the retailer to offer a wide variety of products in an environment where shelf space is limited and there is a pressure to stock new products every day. In this dissertation, we therefore argue that an increased focus on the customer is necessary to face these challenges and that market basket analysis can provide a useful set of techniques to better understand the customer. Indeed, the analysis of the shopping baskets of customers may reveal interesting knowledge about their purchase behaviour that can be used effectively to set up customized marketing campaigns.

From a methodological perspective, it is the objective of this dissertation to present a set of new techniques to model the important concept of 'interdependency'. Indeed, as a result of the trend for one-stop-shopping, consumers typically make interdependent purchases in multiple product categories and failing to consider those interdependencies may lead to marketing actions with disappointing results. For example, research has shown that promotions may influence sales beyond the promoted product line and that product interdependence effects also play an important role in the context of shelf space allocation, product placement and product mix decisions. For instance, retailers are interested in adding items to the assortment whose sales will not be made at the expense of currently stocked items (cannibalism) but may help increase the sales of other items due to complementarity effects. Product interdependency effects should therefore be taken into account when constructing quantitative marketing models to support such product mix decisions.

A number of recent techniques in data mining (association rules) and statistics (mixture models) provide excellent opportunities to take such interdependencies into account. More specifically, the contributions in this dissertation are concentrated around two topics where interdependency plays an important role, i.e. in product selection and in behaviour-based customer segmentation.

Models for product selection

In the 1980's, researchers already stressed the importance of taking into account product interdependence effects for product assortment decisions, such as product addition and deletion. However, at that time, the models suffered from one important limitation, i.e. the implementation of such models on the level of individual product items. Indeed, as a result of the high number of possible interdependencies, models for product selection were only practically feasible on the product category level, thus only incorporating cross-selling effects between product categories.

In this dissertation, we show that the data mining framework of association rules provides interesting information about cross-selling effects between individual product items that can be used to build a product selection optimization model on the SKU or brand level. More specifically, the building blocks and dimensions of a product assortment are discussed, and an overview is provided of the existing methods for product assortment analysis. From a theoretical point of view, our contribution involves the development of a new model for product selection, named PROFSET that incorporates cross-selling effects between products. Hereto, an integer-programming model for product selection is developed that has the objective to maximize profits from cross-selling between products subject to a number of constraints that reflect retailer specific domain knowledge. First of all, a generic framework will be developed from which different specific models can be built according to the marketing problem at hand. More specifically, two product selection problems in marketing are tackled.

The first model makes an attempt towards solving the following marketing problem: an increasing number of retail distribution chains, such as Carrefour, SPAR and Delhaize, recently provide additional convenience shopping facilities besides their traditional stores to serve time-pressured convenience customers. For instance, the Shop'n Go (Delhaize), GB-Express (Carrefour) and Shop24 (SPAR) are examples of this increasing trend for fast convenience shopping. Typically, these convenience stores are located nearby gas stations, train stations, hospitals, or outside the regular store, although some retailers (e.g. Carrefour and Albert Heijn) also provide specific shop-in-a-shop concepts within the traditional supermarket for time-pressured and convenience shoppers. However, since the typical retail surface is limited (15-150m²), it is of crucial importance to select the right products in order to maximize the profitability of the convenience store. The objective of the first product selection model is therefore to find the optimal set of products to put in such a convenience store. The idea is to maximize the profitability from cross-selling effects between the selected products in the convenience store, based on the discovered cross-selling effects from a traditional store. This way, information about existing cross-selling effects in the regular store can be used to optimize the product composition of the convenience store.

The second model for product selection makes an attempt towards solving another well-known marketing problem: retail stores want to maximize their share of the customer's wallet by stimulating cross-selling of their products inside the store. Typically, there are a limited number of attractive shelf positions available in the store, such as end-of-aisle locations, product positions at the outer walking circuit in the store, shelf positions at eye-level, etc. The optimization problem then arises which products to put at those positions, such that customers will not only buy products at those attractive positions, but that they will also go inside the aisles or inner walking circuits of the store to purchase other items too. The crucial idea is that not only the profit of the selected set of products should now be maximized (like in the first problem), but also the profit resulting from cross-selling with other products located at regular positions in the store.

Both models are implemented on real sales transaction data from a Belgian supermarket store. From a practical point of view, these models must enable retailers to more carefully evaluate the contribution of each product within the total product assortment, taking into account existing purchase complementarity effects between products.

Models for behaviour-based segmentation

As discussed before, today's competition forces consumer goods manufacturers and retailers to differentiate from their competitors by specializing and by offering goods/services that are tailored towards one or more subgroups or segments of the market. The retailer in the FMCG sector is however highly limited in his ability to segment the market and to focus on the most promising segments, since the typical attraction area of the retail store is too small to afford neglecting a subgroup within the store's attraction area. Nevertheless, if different customer segments, in terms of their shopping behaviour, can be identified, these segments could then be treated differently in terms of marketing communication (pricing, promotion, etc.) to achieve greater overall effect.

From a theoretical point of view, a new methodology is introduced for behaviour-based customer segmentation by using the method of model-based clustering to discover hidden customer segments based on their purchase rates in a number of product categories. More specifically, several models for customer segmentation are introduced and developed based on (multivariate) Poisson mixtures. The multivariate nature of the models is imposed to account for the interdependency effects that may exist between the category specific purchase rates. The simplest model assumes no interdependencies between the product category purchase rates, whereas the most advanced model will allow purchase rates to be freely correlated. The main contribution, however, lies in the integration of prior knowledge (via a marginal analysis of the interdependencies) into the multivariate Poisson mixture model in order to limit the variance/covariance structure of the mixture model as much as possible whilst still accounting for most of the existing variance in the data. This will enable the specification of a parsimonious and thus much simpler restricted multivariate Poisson mixture model, compared to the fully-saturated model, yet still theoretically sound since relevant purchase associations are allowed to be freely correlated. Additionally, an expectation-maximization (EM) algorithm is proposed to estimate the parameters of the model. The models are empirically validated on real data and the results of the different models are compared and discussed with respect to their optimal parameter values.

From a practical point of view, segmentation of this kind must enable the retailer to optimise his merchandising strategies by customizing marketing actions in the light of differences in shopping behaviour.

ACKNOWLEDGEMENTS

Thanks to many people who contributed to my work over the many years, working on this dissertation has been an extremely instructive and exciting experience to me.

First of all, I would like to thank my promotor Prof. Dr. Gilbert Swinnen, my co-promotor Prof. Dr. Koen Vanhoof and Prof. Dr. Geert Wets for their encouragement, support and high standards that were vital to the successful completion of this thesis. I was lucky to get all of them.

Furthermore, I am grateful to my external examiners Prof. Dr. Els Gijsbrechts (Professor of quantitative marketing at the university of Tilburg, the Netherlands), Prof. Dr. Dimitris Karlis (Lecturer of statistics at the Athens University of Economics and Business, Greece) and Prof. Dr. Patrick Van Kenhove (Professor of marketing at Ghent University, Belgium) for their useful comments and advice to improve the quality and legibility of this text.

I would also like to thank my colleague Ph.D. students for the fruitful research discussions and the excellent atmosphere in our group. A special word of thanks also goes to Peter van den Troost, Koen Wauman and Prof. Dr. Puneet Manchanda (university of Chicago) for providing me the data to carry out my research.

I also wish to acknowledge the Fund for Scientific Research – Flanders (FWO - Vlaanderen) for providing me the opportunity to carry out this research.

Finally, I like to thank my family for encouraging my doctoral research, and especially my wife Liesbeth. I can't express my gratitude for her endless patience and the great care that she took of our little son 'Mattias' many evenings and nights while I was reading exciting papers, writing on my dissertation or playing with my 'mistress' (my Dell notebook). Without you, it would not have been possible to produce this effort.

Many thanks to you all!

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

This chapter provides the wider retailing context in which the contributions of this research should be placed. First, some recent evolutions and challenges that retailers face today will be highlighted and then it will be argued that the availability and analysis of detailed customer sales transaction data can provide new opportunities to stay competitive in this environment. In this context, we introduce the concept of retail market basket analysis as the collection of methodologies that study the composition of a basket of products (i.e. a shopping basket) purchased by a household during a single shopping trip. The idea is that market baskets reflect correlations between purchases made in different products/categories, and that these correlations can be useful to support retail marketing decisions. One particular methodology for market basket analysis, which has gained increased interest since the mid 1990's, and which has received quite a lot of attention in this dissertation too, is based on the data mining technique of association rules. Therefore, a short introduction to this technique will also be given in this chapter.

Finally, we will introduce the objectives of this dissertation and provide an overview of the chapters to follow.

1.1 The Retailing (R)evolution

Today's retailing sector takes up an important position in the economy. In 1994, the retailing sector represented 30 per cent of the businesses, 14 per cent of the working population and contributed to 13 per cent of the added value in Europe [268]. Furthermore, the sector has undergone, and still undergoes, a number of important changes, such as increasing internationalization, lower profit margins, concentration and diversification. In Belgium for instance, the market share of the type 'F1' distributors (large retail distribution) has increased up to 52.2% in 2000, mainly at the expense of the small (F3) retail stores[1], whose market share has dropped from 9.4% to 8.8% compared to the year 1999 [4]. In fact, 785 retail stores have closed down in 2000, especially in the F3 category of small retailers, whereas the large distributors have continuously opened up new stores during the last few years. Especially in Western European countries, this has resulted in a major shift of market share from small to large retailers. Indeed, over the last twenty years, the market share of hypermarkets has almost tripled from 13.0% to 33.9%, whereas the market share of traditional retail stores has dramatically dropped from 26.6% to 5.5% [4]. In 1998, fusions in the distribution sector accounted for 11.8 billion Euro worldwide against just 2.8 billion Euros in 1994 [4].

This growing trend of concentration and increase in scale, depicted above, has a significant impact on the relation with the consumer and presents important challenges for today's retailers, including the battle against decreased customer service and loyalty. With regard to loyalty for instance, POPAI (Point of Purchase Advertising International) reports that in 2001, 75% of the Belgian customers shop at multiple supermarkets [223].

[1] A complete classification of Belgian retail stores into different store types is provided in appendix 1.

Indeed, according to GFK the average Belgian household is, on an annual basis, customer at 14.8 different outlets/channels, of which 6.8 are food outlets and 8 are non-food outlets or special trade [111]. Moreover, GFK reports that consumer loyalty[2] towards Carrefour is only 20.7%, and this is not much different for other supermarket stores [111].

The reasons are clear, before the 1950's, groceries were sold by the corner grocer. The corner grocer knew his customers' preferences and could offer them a customized service accordingly. The emergence of large retail stores and the fact that customers are getting used to self-service, however, resulted in a loss of personalized customer service and created new challenges to gain and keep customer loyalty, for instance through personalization. For instance, hard and soft discounters, typically stores with limited customer service, have aggressively increased their market share from 26.2% in 1999 up to 28.9% in 2002 [111]. POPAI reports that almost 25% of the customer population also shops at discount stores, such as Aldi and Colruyt [223]. As a result of this trend for discounting and scaling-up, in today's large grocery stores most consumers do not know the manager or checkout clerks, and store personnel hardly know their customers. Therefore, today's large retailers are faced with the continuous balancing act to further personalize marketing communications with their customer whilst maintaining the variety and efficiency of the supermarket formula. This is not at all an easy task: retailers are fighting over the consumer's "share of wallet" and satisfying the diverse wants and needs of the consumer forces the retailer to offer a wide variety of products in an environment where shelf space is limited and there is a pressure to stock new products every day[3].

Moreover, the level of competition has aggravated as a result of the slow-down in consumer spending [178] and decreasing profit margins.

---

[2] Consumer loyalty is measured as the amount of spending on FMCG's in a particular supermarket divided by the total amount spent on FMCG's.

[3] While the grocery stores of the 1950s stocked about 3000 items, supermarkets today easily display ten times that number. The amount of sales space devoted to food products has vastly increased; this applies particularly to super- and hypermarkets (for Belgium: 33% between 1986 and 1990) [268].

In fact, in the year 2000, and after correcting for inflation, ACNielsen [4] reported for the first time since 1994 decreasing sales revenues in the food sector. Simultaneously, as a result of the increasing wealth of the consumer, the proportion of the Belgian household budget spent on food has fallen by 30 per cent in the last two decades: food spending accounted for a 21.1% of the household spending in 1979 versus only 14.9% in 1999. Furthermore, according to KBC [154], profit margins have decreased significantly when comparing the evolution of the consumer price index (+28.6%) in the EU countries from 1990 to 1998 against the evolution of the retail food prices (only +20.3%) during the same period[4]. Table 1.1 below shows the evolution of consumer spending[5] in the FMCG in Belgium over the period 1999 till 2001 [111].

	1999	2000	2001
Number of families (x 1000)	4.209	4.237	4.265
Total spending (x € 1.000.000)	17.060	17.895	18.258
Average expenditure per household (€)	4.053	4.223	4.281
Number of shop visits	218	210	206
Expenditure per shopping visit (€)	18.6	20.1	20.8

Source: GFK (2002)

Table 1.1: The weak evolution of Consumer Expenditure in the FMCG

Table 1.1 shows that the total household expenditure in the FMCG increased only 1.4% from 2000 to 2001, in contrast to 4.1% from 1999 to 2000, which is even less than inflation. Furthermore, the number of shopping visits has slightly decreased while the average expenditure per shopping visit has slightly increased.

[4] Unfortunately, the KBC study does report figures about the cost evolution in the retail sector. Possibly, retailers have succeeded in cutting costs too such that the reported decrease in profit margin could therefore be too pessimistic or even non-existing.
[5] In the case where values are expressed in Belgian francs (BEF), the following exchange rate should be used: 1€ = 40.3399 BEF

In this era of rapid changes and new challenges, retailers are looking for innovative strategies to provide (a distinctive) added value to consumers whilst maintaining a profitable long-term business strategy.

Examples of such recent strategies include the introduction of private labels, image products[6], loyalty card programs, efficient consumer response (ECR), supply chain management through EDI, e-commerce initiatives, category management and the development of purchase associations and self-scanning (e.g. Delhaize). This clearly contradicts with the past where the emphasis in retailing was almost solely on the physical distribution aspects, whereas currently, retailers want to pursue a sustainable competitive marketing strategy. Corstjens & Corstjens [82] refer to this as the evolution in the fast moving consumer goods (FMCG) sector from a *selling* orientation (sell what you can stock) towards a *market* orientation (stock what you can sell).

Setting-up and maintaining such strategies, and running one's business in a more intelligent way, however, crucially depend on accurate information, such as information concerning consumer purchase behaviour, product/category profitability, and others. Researchers [36, 146] acknowledge the analysis of market baskets, i.e. the search for patterns of purchase behaviour by means of the analysis of shopping bags, as highly relevant in this context and therefore, market basket analysis will be the central theme of this dissertation.

1.2 Market Basket Analysis

The research in this dissertation is based on, and has the attempt to contribute to, the growing body of work known as market basket analysis. Market basket analysis is a generic term for methodologies that study the composition of a basket of products (i.e. a shopping basket) purchased by a household during a

[6] For instance, Albert Heijn has recently introduced 'biological' products (from Farm to Folk) into its assortment and Delhaize has introduced premium ready-made meals from Pierre Wynants's first-class restaurant 'Comme-chez-soi'.

single shopping trip. The idea is that market baskets reflect correlations between purchases made in different products/categories, and that these correlations can be useful to support retail marketing decisions. Indeed, as a result of the trend for one-stop-shopping, consumers typically make interdependent purchases in multiple product categories. Failing to consider those interdependencies may lead to marketing actions with disappointing results, as illustrated by the following list of examples.

For example, research has shown that promotions may influence sales beyond the promoted product line. For instance, Walters [286] and Manchanda et al. [183] showed that retail price promotions created significant complementary and substitution effects within the store, i.e. a discount off the price of cakemix has a significant positive impact on the sales of cake frosting, and a discount off the price of a particular brand of cakemix has a significant negative impact on the sales of competing cakemix brands. Obviously, effects of this kind should be taken into account when promoting products. The idea of positive or negative sales effects as a result of promoting a particular brand was used by Mulhern and Leone [209] to develop the concept of *implicit price bundling*. They suggest that the price of a product should be based on the multitude of price effects that are present across products without providing consumers with an explicit joint price for the collection of those products. Well-known, in this context, is the practice of *loss leader* pricing in retailing, where a leader brand is positioned as a 'decoy', i.e. it is heavily promoted in order to draw consumers to the store and simultaneously stimulates the sales of other non-promoted items. Recent work [234], however, weakens the strong impact of cross-price elasticities in the context of consumer choice behaviour.

Furthermore, product interdependence effects also play an important role in the context of shelf space allocation. In fact, most commercial shelf space allocation systems (e.g. PROGALI, OBM, COSMOS, etc.) today still do not take product interdependencies into account. In contrast, Corstjens and Doyle [83] were the first to argue that product cross-space elasticities should be incorporated into the demand side of the profit equation. Indeed, significant cross elasticities among products within the assortment may exist due to complementary and substitution relationships between them. This insight has led to a number of academic shelf space allocation systems [65, 66] that take product interdependence effects into account.

Additionally, product interdependence effects play a crucial role in the context of product mix decisions, e.g. for product addition and deletion decisions and for product selection. For instance, it is a common practice among retailers to distinguish between core products and peripheral products in the assortment [275]. Core products should not be deleted from the assortment because they are the core materialisation of the retailer's store formula. By removing those products, the assortment will not meet the basic expectations of customers who visit the store. In contrast, peripheral products are chosen by the retailer to enhance the store image even more and should be selected to maximise cross-sales potential with basic products. Indeed, retailers are interested in adding items whose sales will not be made at the expense of currently stocked items but may help increase the sales of other items [218]. Consequently, peripheral products should be selected based on their purchase affinity with core products. Swinnen [257] argues that if a product group is highly associated with other (profitable) groups, the addition of new items may increase the group's attractiveness and therefore the customer's willingness to do joint purchases. Similarly, the deletion of an item will be less interesting if it is coming from a group with a high radiation effect on other groups.

Finally, knowledge about product purchase interdependencies can be useful for product placement decisions. Indeed, if it is known that two products/categories are often purchased together, the store layout can be adapted accordingly [146], e.g. to stimulate cross-selling, by placing two products closer together. Alternatively, products with high positive interdependence effects could be placed further apart in order to force the consumer to travel through the store and to stimulate picking other items along the way[7].

Yet, such product placement strategies often contradict with the existing way of putting products together. Indeed, current product placement strategies are mostly product attribute based, i.e. products are put together because they are functionally similar. According to Corstjens and Corstjens [82], 'a retailer may have trouble putting fresh steak, frozen chips, prepacked salad and salad dressing next to each other, but it is worth thinking about. Consumers are used to the product category layout initially set up by product manufacturer, but retailers should be free to experiment with layouts influenced by the patterns of purchases made together, or by certain segments of shoppers. A market-oriented store should ask what layout is ideal for its shoppers, ignoring which supplier it gets its products from'.

1.3 Association Rules

Given the numerous marketing decisions where product interdependencies can play an important role (section 1.2), the question remains how those interdependencies can be obtained from market basket data. Although in chapter 3 we will discuss some other techniques to obtain product interdependencies, most of the contributions in this dissertation will be based, directly or indirectly, on the results obtained from applying a particular data mining technology on market basket data, better known as *association rules* [8]. We will discuss the technology of association rules in detail in chapter 4. However, a short introduction to this technique is already provided here.

In fact, it was only in 1993 that the first article on association rules appeared in the ACM SIGMOD conference on management of data [8].

---

[7] Although such strategy would probably not be very much appreciated by customers who are sensitive to convenience and speed of shopping (run shoppers).

Since then, it has become one of the most popular techniques in data mining with currently over hundred papers being published and it remains to be an attractive research topic in the literature. The principle motivation for the technique can be found in the quest, by computer scientists, for efficient techniques to discover consumer purchase patterns in large retail transactional databases.

Association rules are expressed in an IF-THEN propositional rule-based format. For instance, the association rule IF *diapers* THEN *beer* expresses that consumers who buy diapers also tend to buy beer with it. The objective of association rules is then to discover all such purchase relationships in a given transactional data set that satisfy a minimum (user defined) *support* and *confidence* threshold. The support threshold, on the one end, specifies the minimum proportion of retail baskets in the data in which the items of the association rule must occur together in order for the rule to be frequent. The confidence threshold, on the other hand, is an estimator for the expected conditional probability of the rule, i.e. the probability that the consequent occurs, given that the antecedent has occurred. The confidence threshold therefore specifies a lower bound on the confidence of the rules to be generated by the association rules algorithm.

Apart from retailing in which association rules have been mainly adopted, this technique has been used in other contexts as well, such as cross-selling [21], finding co-occurring medical tests from a health insurance information system [283], reducing fall-out in telecommunications systems [18], and identifying latently dissatisfied customers [37, 38].

Nevertheless, since its introduction by computer scientists, the technique of association rules has never been fully taken-up by the academic marketing research field, in contrast to large distributors (e.g. Delhaize, Walmart, Safeway) that have been experimenting with the technique. In our opinion, this has a lot to do with the fact that most of the attention in the field of association rules has been focussed on increasing the algorithmic efficiency to find such rules [see numerous papers 9, 12, 61, 216, 266], and less on the practical utility of the technique in concrete retail settings. However, in this dissertation, it is our attempt to show that the technique of association rules has a number of interesting retailing applications, either alone or in combination with traditional marketing research techniques, as stipulated in the next section.

1.4 Objectives of this Dissertation

In our opinion, the methodologies for market basket analysis research can be divided into two major areas: on the one hand, the classical (parametric) statistical approach involving econometric techniques, and on the other hand, the relatively recent (non-parametric) data mining approach using association rules. Both approaches (one could even say paradigms) provide interesting and useful, yet slightly different contributions and/or insights (in)to the analysis of scanner data. In contrast to the classical quantitative marketing research approach of scanner data, finding its origin in the mid 70's, and which was primarily developed by statistical and econometric oriented scientists [40, 200], the data mining approach has a much more recent history (early 90's) and was primarily developed by computer scientists [8]. Furthermore, a literature overview demonstrates that both paradigms have indeed developed in parallel without much cross-fertilization between both areas. Although it can be observed that both domains are currently converging, and will probably integrate in the future, both domains still have their own conferences and journals[8]. It is, however, our strong belief that both paradigms provide interesting, yet sometimes different viewpoints on the analysis of scanner data, and therefore it is the broader objective of this dissertation to investigate where and how both research domains can benefit from each other.

---

[8] The Advanced Research Techniques (ART) forum of the American Marketing Association is currently probably the only conference which deals with both statistical and data mining approaches to marketing research.

More specifically, there are two topics that are of particular interest to us and each of them is, directly or indirectly, linked to product interdependence effects: 1) the development of a product selection model that takes into account cross-selling information from market basket analysis, and 2) behaviour-based customer segmentation. Both topics will be treated in the chapters 5, 6 and 7 of this dissertation.

1.4.1 Product Selection Taking into Account Cross-Selling

In the 1980's, Swinnen [257] already stressed the importance of taking into account product interdependence effects for product assortment decisions, such as product addition and deletion. However, he mentioned one important limitation, i.e. the implementation of such models on the level of individual product items. Indeed, he argued that as a result of the high number of possible interdependencies, models for product selection were only practically feasible on the product category level, thus only incorporating cross-selling effects between product categories. However, since by definition a product selection model involves decisions to be taken at the SKU level, such models were not really feasible at that time.

In this dissertation, we show that the data mining framework of association rules provides interesting information about cross-selling effects between individual product items that can be used to build a product selection optimization model on the SKU or brand level. More specifically, we will discuss the building blocks and dimensions of a product assortment, and we will provide an overview of the existing methods for product assortment analysis. From a theoretical point of view, our contribution will come from the development of a new model for product selection, named PROFSET[9] that incorporates cross-selling effects between products.

[9] PROFSET stands for 'PROFitability per SET'

Hereto, we will develop an integer-programming model for product selection that has the objective to maximize profits from cross-selling between products subject to a number of constraints that reflect retailer specific domain knowledge. First of all, a generic framework will be developed from which different specific models can be built according to the marketing problem at hand. This will be illustrated by two particular model specifications. The first model makes an attempt at composing an optimal product assortment for a convenience store.

The second model deals with the problem of which products to put at visually attractive positions in a supermarket store. Both models will be implemented on real sales transaction data from a Belgian supermarket store. From a practical point of view, this model must enable retailers to more carefully evaluate the contribution of each product within the total product assortment, taking into account existing purchase complementarity effects between products.

1.4.2 Behaviour-Based Customer Segmentation

Today's competition forces consumer goods manufacturers and retailers to differentiate from their competitors by specializing and by offering goods/services that are tailored towards one or more subgroups or segments of the market. The retailer in the FMCG sector is however highly limited in his ability to segment the market and to focus on the most promising segments since the typical attraction area of the retail store is too small to afford neglecting a subgroup within the store's attraction area [82]. Nevertheless, if different customer segments, in terms of their shopping behaviour, can be identified, these segments could then be treated differently in terms of marketing communication (pricing, promotion, etc.) to achieve greater overall effect.

From a theoretical point of view, it is our objective to introduce a new methodology for behaviour-based customer segmentation by using the method of model-based clustering to discover hidden customer segments based on their purchase rates in a number of product categories. More specifically, we will develop several models for customer segmentation based on (multivariate) Poisson mixtures and again, the concept of product interdependence will play an important role. The simplest model will assume no interdependencies between the product category purchase rates, whereas the most advanced model will allow purchase rates to be freely correlated. Our contribution, however, will come from integrating data mining results into the multivariate Poisson mixture model in order to limit the variance/covariance structure of the mixture model as much as possible whilst still accounting for most of the existing variance in the data. This will enable the specification of a parsimonious and thus much simpler restricted multivariate Poisson mixture model, compared to the fully-saturated model, yet still theoretically sound since relevant purchase associations are allowed to be freely correlated. From a practical point of view, segmentation of this kind must enable the retailer to optimise his merchandising strategies by customizing marketing actions in the light of differences in shopping behaviour. The model is tested on purchase data from four interdependent product categories, cakemix, frosting, fabric detergent and softener.

1.5 Outline of this Dissertation

Chapter 2 introduces the basic concepts and building blocks of this dissertation. More specifically, we will discuss the history and definition of receipt data as a form of scanner data and present its advantages and disadvantages compared to traditional store level data and panel data. Furthermore, we will discuss the history, the scope and quality of loyalty card data and we will identify different types of loyalty card programs. Subsequently, chapter 3 provides a literature overview of the concept of product interdependence and discusses both the sources of interdependence and methods for measuring interdependencies. Special attention will be given to the technique of association rules in chapter 4. Chapters 5, 6 and 7 will contain the main research contributions of this thesis.

In chapter 5, we will develop a constrained optimization framework (PROFSET) for product selection that has the objective to maximize profits from cross-selling between products subject to a number of constraints that reflect retailer specific domain knowledge. But first the building blocks and dimensions of a product assortment will be discussed and we will discuss the complexity of assortment optimization.

In chapter 6, we introduce the idea of behaviour-based customer segmentation within a supermarket-retailing context and we position it within the larger body of literature available on market segmentation in general. Furthermore, as a matter of illustration, we will provide two concrete applications of 'apriori' behaviour-based customer segmentation on retail market basket data. The first application groups customers based on their frequency of shopping and the average amount that they spend per shopping visit. The second application groups customers according to the size of their shopping baskets. In both applications, we show some differences in the shopping behaviour between the discovered segments in terms of their purchases in a number of product categories.

Chapter 7 introduces an 'a posteriori' behaviour-based segmentation method for clustering supermarket shoppers based on their purchase rates in a number of product categories. We will present several multivariate clustering models based on the statistical method of model-based clustering, also called mixture models or latent class cluster analysis. More specifically, we will propose several multivariate Poisson mixture models where special attention will be paid to the treatment of the variance-covariance matrix based on a marginal analysis of the underlying correlations in the data.

Finally, chapter 8 is reserved for conclusions and an overview of topics for future research.

CHAPTER 2

MARKET BASKET DATA

This chapter deals with the kind of data that is referred to by the name *retail market basket data*, or also called *scanner data*. The term market basket data, however, covers a wide range of different meanings and data sources and therefore we believe that in the context of this research it deserves some clarification. More specifically, we will deal with several definitions of scanner data, such as store data, household panel data and receipt data and we will discuss both the advantages and disadvantages of them. Furthermore, we will discuss the concept of loyalty card data. Finally, an overview is provided of the data set that will be used in this dissertation.

2.1 Scanner Data

This first section deals with the history, definitions and pros and cons of scanner data.

2.1.1 The History of Scanning

During the last 30 years, retailing has undergone a revolution in terms of the adoption of advanced technologies for data collection, storage and analysis. From the early 1970s on, laser technology in combination with small computers enabled American retailers to electronically scan the purchases made in their store [107]. In those days, however, the principle motivation to collect sales transaction data was not to support retail marketing decision making, but to save labour costs by speeding up check outs [176] and to facilitate inventory management and management reporting about store and product sales [87]. In fact, the technology of barcode scanning enabled retailers to cancel the manual pricing of individual articles on the shelves, representing a tremendous saving in labour costs. The coupling of the barcode, containing the product identification, to a table of product prices enabled the retailer to quickly change product prices whenever necessary by just altering a single record in the product database. As soon as the price of the product was changed in the database, this price immediately became available at the checkout. The only manual intervention still needed was to put up the price of the article on the shelves.

In Europe, the origin of scanning devices in retail stores only dates from the early 1980s. However, since then European retailers, and Belgian retailers in particular, have largely made up for lost ground (see table 2.1).

Country	Dec 1987	Jan 1990	Jan 1994
North America			
USA	55	62	71
Canada	38	45	50
Europe			
Sweden	22	44	85
Belgium	15	31	83
Denmark	15	37	83
Finland	15	45	80
Great Britain	17	39	76
France	28	43	74
Norway	15	26	58
Spain	7	14	57
Italy	7	17	56
Netherlands	13	25	56
Austria	5	10	53
Germany	10	29	39
Ireland	4	19	39
Switzerland	1	3	10

Source: NIELSEN (1998)

Table 2.1: Evolution of scanning in food stores in % of all commodity volume turnover

Given the non-uniform use of the concept 'scanner data' in the literature, it is however important to distinguish between different forms of scanner data, i.e. store data and panel data [117]. Store data (also called item sales data or aggregate response data) involves individual (UPC) sales and profit figures by store and by day or week. Thus, in store data, the unit of analysis is the article. In contrast, panel data (sometimes also called household scanner data) represent histories of purchases across different stores of individual products or product categories for a particular sample of households, it is therefore a three-dimensional panel dataset. Usually, members of the same family register their purchases in a diary, which is later collected by the researcher.

In our research, however, we even employ a third category of scanner data defined by Hernant and Persson [136] as receipt data. Here, the unit of analysis is the receipt instead of the article. Therefore, within the scope of this dissertation, we define receipt data as:

> *The purchase related basket data, such as EAN[10] (European Article Number), Customer ID, product price and purchase quantity, time of purchase, etc… that is produced by scanning technology at the check out systems (POS) in a retail store and that is stored in a sales transaction database.*

In contrast to panel data, which expresses customer purchase behaviour across different stores, receipt data thus produces information on the level of the retail store and its customers but can not directly be used to analyse customer behaviour across different stores, unless retail outlet identification within a store chain is also stored onto the receipt.

2.1.2 Advantages of Receipt Data

Given the increasing importance of data driven marketing decision making in retailing, the adoption of electronic point of sales (EpoS) scanning devices has been a major milestone for the creation of high quality data. McKinsey [194] concluded that, in general with regard to the marketing strategy, the following benefits could be obtained from the analysis of EpoS transaction data:

---

[10] UPC (Unified Product Code) in the USA. The UPC code consists of 14 or 16 characters, whereas the EAN code consists of 13 characters. The first two characters of the EAN code determine the country of origin of the product (Belgium=54), the next five characters uniquely identify the manufacturer and the final six characters can be freely determined by the manufacturer.

- Immediate feedback can be obtained after adjustments in pricing, product range, display allocations or advertising;
- Experiments involving the manipulation of marketing variables can be more easily and rapidly analysed;
- Store layouts can be improved through the analysis of product purchase patterns;
- Analysis of transaction numbers and sizes by time of day/day of week can provide guidelines for policies regarding opening hours and customer service levels;
- If some form of customer identification is linked to the transaction record, for example if a store card is used, then many additional opportunities are available. The success of each commodity group in attracting specific customer segments can be analysed. Communications can be sent to certain customers to increase their loyalty to the store and/or to encourage them to use different sections of the store.

Furthermore, when compared with store level data, receipt (and panel) data have some important advantages, both to the retailer and to the manufacturer, as explained in the subsequent sections.

2.1.2.1 Disaggregate information (retail is detail)

Receipt and panel data reflect sales of individual product items (at the UPC level) for the individual customer. In contrast to store level data, receipt data enable to understand brand choice behaviour on the level of the individual customer or household. Furthermore, the time intervals over which receipt data are recorded are much shorter when compared with store level audit data. The former is collected per day, whereas the latter is usually collected over a period of several weeks or even months. Therefore, receipt data enable a more fine-grained analysis of promotional actions when compared with store level audit data where promotional effects can be averaged-out over the longer period, depending on the wear-out effect of the action being undertaken.

2.1.2.2 Low cost of acquisition

The cost of obtaining the data is relatively low since the data are easily obtained from the daily transactional operations of the store. The systematic collection and storage of these data in large database systems indeed enables relatively easy and cost-efficient access. However, this does not imply that the cost of obtaining useful information from the raw data would be low. In fact, our experiments with receipt data indicate that substantial pre-processing must be performed before modelling can be carried out (see section 2.3.1.3). Cases in which pre-processing took more than 70% of the total knowledge discovery process are not exceptional.

2.1.2.3 High speed of delivery

Additionally, because of the electronic storage of the data, the delivery speed of the receipt data to users can by very fast. Indeed, individual customer purchases can be easily extracted from the transactional databases. However, high speed of data delivery does not necessarily imply high-speed delivery of the results. Indeed, the cost of transformation and pre-processing in order to prepare the data for retail market basket analysis may be significant, as indicated in the previous paragraph.

2.1.2.4 High reliability and internal validity of the data

The reliability and internal validity of receipt data is dependent on the quality of the measurement system being used. In general, receipt data are considered to be very accurate since they are part of the store's cash collection and accounting process and much of the human element in recording product movement is eliminated as a result of the use of sophisticated barcode scanning systems. Indeed, if a barcode is scanned after all, then it is also correctly scanned. The inability to scan barcodes is often the result of technical shortcomings because the barcode is 1) damaged, 2) badly printed, 3) covered by the packing, 4) overfrozen, etc.

Furthermore, the collection process of (household) receipt data is relatively unobtrusive, bias-free and complete across products when compared to traditional diary panels [121]. It is unobtrusive since the measurement of purchase behaviour is carried out electronically, which causes the effort made by the customer to be small. This is in strong contrast with panel data where panel members are asked to register their purchases and where refusal and attrition rates have been over 50% [80]. Furthermore, the collection process of receipt data is bias-free in so far that all purchases made by households are effectively scanned at the checkout and thus the nonprobability sampling of households is low. Finally, receipt data is complete across products because all SKU's are scanned at the check out.

2.1.2.5 Reflects product competition

Finally, probably the most important advantage of receipt (and panel) data is that they reflect the decisions of customers being made in a competitive environment, i.e. where multiple products compete for market share [117]. This is especially relevant for retail market basket analysis where the objective is to understand why customers prefer certain products over their substitutes: information both useful for the retailer (e.g., for product positioning) as for the manufacturer (e.g., for market positioning). In this sense, the receipt reflects the shopper's natural way of using the retail store [146]. This product competition information can not be inferred from store data where there is no information available on other product's prices or marketing activities impinging on the customer at the time of purchase.

2.1.2.6 Link with loyalty card data

Finally, the collection of household-level purchase data enables a connection with other household data by means of loyalty cards. Indeed, the customer ID can be used to link purchases (behaviour) to socio-demographic and lifestyle characteristics of the customer and hereby offers excellent opportunities for customer segmentation. Indeed, socio-demo and lifestyle data can be used to profile customers segments, which in turn offers opportunities to target these segments with customized marketing campaigns. Loyalty card data will be discussed in more detail in section 2.2.

2.1.3 Disadvantages of Receipt Data

Besides the appealing advantages of receipt data, some disadvantages can be identified as well.

2.1.3.1 No information on consumption

An important weakness of receipt data is that it tracks only purchases and provides no information on consumption or at-home pantry holdings. Yet, this is important information since it enables to determine how much of the incremental sales from a promotion represent new net sales (due to consumption increases) versus stock-piling (due to pantry loading) [64].

2.1.3.2 No information on out-of-stock situations

Out-of-stock situations can not be tracked with receipt data. Compared with traditional store audit data, this is a major disadvantage of receipt data. If out-of-stock situations occur often, competitive market structure analysis on the basis of scanner data may produce biased results. More information on consumer responses to stock-outs can be found in Campo et al. [71].

2.1.3.3 No information on purchases in other stores

In contrast to household panel data, receipt data only show the purchases made within one particular store, and consequently these data do not reflect their total shopping behaviour if they shop in other stores as well. Indeed, consumers may not only shop at other stores because of the occurrence of out-of-stock situations in their preferred store or because of a promotion on identical products in the other store, but they may also simply shop at different stores for different products. For instance, they may shop at a supermarket for their groceries, except for some specific products that they buy from the heavy discount store. For instance, Walters [286] showed that promotions of products in one store significantly decreased sales of substitutes and complements in a competing store. This lack of information about consumer purchases made at competing stores is also an important disadvantage for customer segmentation strategies based on shopping behaviour, since customer segments will only have local, i.e. store-specific validity.

2.1.3.4 No information on perceptions and attitudes

Receipt data do not contain perceptual and attitudinal information from consumers. In the case of segmentation, this is an important drawback since segmentation is often based on consumer perceptions and attitudes. Furthermore, in chapter 5 we will discuss the consequence of the absence of perceptual information on the development of the PROFSET model for product selection. More specifically, it will be shown that due to the absence of this information, it is not possible to exactly predict the underlying shopping motivations for a customer during a shopping trip (see section 5.4.3.2).

2.1.3.5 Different barcodes for the same product

Some manufacturers tend to label the same products with different barcodes. For instance, the same product sometimes obtains two different barcodes, one for the period of non promotion and one for the period during which the product is on promotion, in order to facilitate reporting and quickly assess the influence of promotions on the sales of the product. Although very useful to track the effectiveness of different promotional campaigns, different barcodes may present data analysis problems when this granularity is not needed and there is the danger of treating the same product as many different products.

2.1.3.6 The same barcode for different products

Some manufacturers tend to use old barcodes of eliminated products to label new products. This can cause problems in the case where receipt data analysis is carried out over a relatively long period of time such that two different products can potentially have the same barcode.

2.1.3.7 Scanner databases are big

Scanner databases are typically very big when compared with traditional store audit data [189]. For instance, Wal-Mart loads 20 million POS updates into its central relational database per day [24]. This is mainly due to the fact that store audit data do not report numbers for all products (due to the lengthy nature of the audit process) whereas scanner data reflect sales of all products in the store. Little [176] estimates the volume of receipt data to be 100 to 1000 times as much as store audit data. To analyse these large volumes of data, powerful computers and efficient data analysis techniques are needed.

2.1.3.8 External validity

The external validity of receipt data refers to the transferability of results to households that are not part of the sample being studied. The problem is in fact threefold.

First of all, scanner data are often only collected and stored for customers who possess a loyalty card since only for these customers a detailed transaction history can be composed. However, the purchase behaviour of customers who possess a loyalty card is likely to be different from those who do not have a loyalty card. Inferences about the purchase behaviour based on the analysis of receipt data may therefore not be valid for the customers who do not possess a loyalty card [64]. This is not so much of a problem for our study, since slightly more than 80% of the customers possess a loyalty card.

Secondly, purchase behaviour analysed by means of scanner data is likely to be store dependent. Indeed, both the in-store as the out-of-store environment is different such that inferences about the purchase behaviour of customers from a particular retail store may not be valid for other stores, even within the same distribution chain.

Finally, the external validity of receipt data also refers to the transferability of results to different members of the same household being studied. Indeed, loyalty cards are often used by different individuals within the same household. Consequently, purchase histories collected by means of a particular loyalty card may potentially reflect the behaviour of different individuals within that household. Therefore, one must be careful in inferring conclusions about purchase behaviour of individuals when different individuals of the same household in fact use the same loyalty card, or vice versa, since members of the same household may also have different cards.

2.1.3.9 Privacy

A recent article in Marketing News [185] reports the founding of an internet-based consumer group (http://www.nocards.org) called Consumers Against Supermarket Privacy Invasion and Numbering (CASPIAN). The article reflects the increased consumer distrust towards supermarkets that use loyalty card data to track the purchase behaviour of individuals or groups of individuals. Indeed, targeting (groups of) individuals with customized promotions may harm their personal sphere of life if they feel manipulated by these activities [31, 39]. As a result, careful attention should be paid to preserving the privacy of the individual in the light of 'fair information practices'[11]. In this context, researchers have made attempts to develop privacy preserving data analysis techniques [10, 81, 237], mostly based on the principle of value distortion, i.e. altering the original data by adding random values to the original values. The fact that privacy issues may indeed become a serious concern is illustrated by Identico systems' (http://www.identicosystems.com) True ID® concept.

---

[11] For a discussion of data mining and privacy, the reader should take a look at a report produced by the Office of the Information and Privacy Commissioner (IPC) by Cavoukian [74].

The True ID concept has already been implemented by the US retailer 'Winn-Dixie stores' (http://biz.yahoo.com/bw/020326/262214_1.html), which is one of US largest supermarket retailers with over 1140 stores in 14 states and the Bahamas. True ID is the first identity verification service designed to improve check management and internal verification systems by 'putting a face' on every transaction. With the customer's permission, the photo and information from a valid photo ID is collected and electronically stored in a database. The next time the consumer initiates a transaction, the customer's image is securely and instantly sent to the point of service, where the employee matches the face in the photo to that of the consumer and decides whether to proceed with the transaction.

2.2 Loyalty Card Data

This section deals with the history, the scope and the quality of loyalty card data, as well as the different types of loyalty schemes.

2.2.1 The History of Loyalty Card Data

In our opinion, the focus on the use of information technology in retailing can be characterized as a move from *efficiency* in the 1970s and 1980s towards *effectiveness* in the 1990s. Indeed, as a result of decreasing retail profit margins, the early implementations of scanner technology in retailing were exclusively focussed towards increasing efficiency by facilitating the checkout of customers and by automating the inventory processes. In fact, at that time there was no need to couple customer information to sales transactions.

The 1990's however, can be characterized by an increasing focus on *effectiveness* of marketing and sales. The recognition that generating more business from existing consumers may be cheaper and more effective than simply trying to acquire new customers or win them from the competition has increased the need to better support the (direct) communication with the customer by understanding the wants and needs of the customer.

For instance, Uncles [272] shows that loyal customers tend to purchase more frequently and spend more. Therefore, many retailers today offer their customers frequent shopper programmes, which provide consumers discounts, coupons and gifts in return for purchasing goods in the store. In this respect, frequent shopper clubs represent a strategy designed to turn discrete purchases over time into a continuous customer relationship, a move that is consistent with the evolution from a product-oriented to a more customer-oriented view in retailing.

The loyalty program should make the consumer feel that the retailer is prepared to listen, and is willing to innovate on behalf of the customers. However, to achieve this goal appropriate actions to corroborate this feeling should be undertaken, such as targeted promotions and customized communication with the consumer. Consumers should feel understood by the retailer. Otherwise the loyalty program will degrade to the classical 'loyalty scheme' offering me-too benefits which may be nice to have (most people like to get something for nothing) but which are no guarantee of continued loyalty. In fact, GFK recently reported that, for the Belgian consumer, a loyalty card program is among the least important criteria to determine his store choice [112].

2.2.2 Scope of Loyalty Card Data

Loyalty card data typically consists of one or more of the three following data types.

- *Identification data:* name and address of the cardholder. In order to build a long lasting relationship with the customer by maintaining a transaction history and profile per cardholder, the identity of the cardholder should be known to the retailer.

- *Socio-demographic data:* such as age, sex, income, family status, religion, education and profession. These data are rather of a static nature. Variables of this type are often used by retailers to segment their customer population and/or to obtain a socio-demographic profile of the store's attraction area. (See section 6.2.1)
- *Lifestyle data:* such as hobbies, cultural involvement, etc. These variables are again often used by retailers to profile customer segments into different customer groups that have the same lifestyle pattern. It is implicitly assumed that consumers belonging to the same lifestyle segment respond similarly to particular marketing-mix changes. (See section 6.2.1)

On top of these three basic types of data, there is currently a trend in retailing to attach more functionality to the loyalty card, such as payment systems, credit facilities and marketing relevant customer information by integrating programmable computer chips into the card, i.e. the 'smart card'. Despite the tremendous possibilities of smart cards, for instance in electronic commerce by storing the user's personal digital signature on the card as a security measure for transactions over the internet, both consumers and retailers tend to be reluctant towards their implementation [208]. Retailers complain that smart cards are up to 20 times more expensive to handle as cash, and consumers are often concerned about security of the information held on the cards.

2.2.3 Types of Loyalty Card Programs

Loyalty programs exist in many forms and their scope is often very different according to the type of customer bond being pursued, i.e. financial bond, social bond or structural bond. However, most differences in loyalty programs can be reduced to the following elements [221].

2.2.3.1 With or without loyalty card

Participation in loyalty programs without the use of loyalty cards is free to everyone, i.e. without the obligation of providing any personal information. For the retailer, this type of loyalty program is easy and inexpensive, but the lack of a database limits its usefulness for customized marketing activities. The customer can obtain gifts, coupons or reductions by handing over tickets that can be collected by purchasing. For loyalty programs with the use of cards, participation in the program is subject to the publication of personal information. The loyalty program feeds and is fed by a marketing database which offers the retailer the opportunity to personalize marketing actions and to carry out data-driven customer analysis.

2.2.3.2 Single program or joint program

There is an increasing trend to organize loyalty programs with the support of different partner organizations. For instance, the Belgian 'Happy Days' card is joint initiative of GB, Fortis bank, Shell, Neckermann, Standaard Boekhandel and others. Customers can collect points by buying from one of the partnering organizations and points can be exchanged for a selected number of gifts. The principle motivations to set up joint loyalty programs are twofold: firstly to share costs between partnering organizations and secondly to offer loyal customers the opportunity to collect points in a much shorter period of time.

2.2.3.3 Participation costs or not

Although most of the loyalty programs are free of charges, sometimes the participation is subject to a onetime or recurring entry fee. These costs are mostly motivated by an increased level of service or exclusivity for the participants in the form of purchase evening invitations, or extended product insurances, etc. Participation in the Belgian 'Happy Days' loyalty program costs a one-time fee of €2.5.

2.2.3.4 Money or product/service rewards

Participants of the loyalty program can by rewarded in a number of different ways. In the past, money rewards have been relatively popular, but today rewards are mostly given in the form of gift products or services. The 'Happy Days' loyalty card program offers a relatively large range of products and services that can be obtained in exchange for collected points.

2.2.3.5 Modest or ambitious savings objectives

Savings objectives that are too ambitious so that the participants of the loyalty program can almost not achieve them are of no use. A survey by MarketResponse [279] has revealed that cardholders prefer smaller gifts (such as small household appliances) instead of expensive vacations. Joint loyalty card programs provide an interesting alternative for the consumer to earn gifts in a much shorter period of time.

2.2.3.6 Catalogue or store

Loyalty programs also differ with respect to how the gifts, earned in the loyalty program, can be obtained by the consumer, i.e. via store(s), catalogues, call centres, etc.

2.2.4 Quality of Loyalty Card Data

Although useful for several marketing purposes, the quality of loyalty card data is questionable for several reasons. First of all, the data may not be representative for the total consumer population doing purchases in the store. Indeed, typically only a subgroup[12] of a retailer's customers step into a loyalty card program and these customers may not be representative of all customers

---

[12] Especially in the past, when loyalty programs were emerging, this presented an important problem. However, today, most customers participate in a loyalty card program in the store where they shop most frequently. For instance, in GB, over 90% of the sales are generated by customers who own a loyalty card.

(see section 2.1.3.8). Secondly, loyalty card data contain numerous missing values because consumers do not (want to) take time to correctly fill-out the questionnaire in order to obtain a loyalty card. Especially privacy sensitive questions related to income, religion, etc, are vulnerable to systematic errors and present important problems for data analysis.

Thirdly, loyalty card data are often not updated. Indeed, once registered the consumer is usually not asked to update his profile after a certain period of time. Yet, it is highly probable that changes in the composition of the household, changes in the needs and wants of the consumer, or simply a change in the home address takes place over time.

2.3 Data Available for this Study

Except from chapter 7, where a different dataset will be used, the research and experiments in this dissertation will be based on real scanner data obtained from a Belgian supermarket store of the F2NI type (average size non-integrated distribution). The data are collected over the period between half of December 1999 and the end of November 2000.

Although it was the objective to collect data for this entire period, due to technical circumstances, the data were collected over three shorter periods. The first period runs from half December 1999 to half January 2000. The second period runs from the beginning of May 2000 to the beginning of June 2000. The final period runs from the end of August 2000 to the end of November 2000. In between these periods, no data is available. This results in approximately 5 months of data and a total number of 88163 receipts available for analysis[13].

[13] However, since the data needs to be cleaned in a number of respects (see section 2.3.3), the final amount of receipts available for analysis will be slightly lower.

2.3.1 Data Content

The dataset contains both the receipts, collected at the check-out, and the loyalty card data for those customers who have purchased at least once during the period of data collection. Each receipt is labelled by a separate ID and contains information about the date of purchase, a customer ID (linked to the loyalty card info), the SKU's purchased, the SKU's price per unit and the amount purchased for each SKU.

Over the entire data collection period, the supermarket store carries (listed) 16430 unique SKU's, but some of them only on a seasonal basis, such as special Christmas items. Although most SKU's are identified by a unique product identification (the barcode), some products are grouped into product categories and appear on the receipt by the product category ID to which they belong. This is the case for individual fruit items (e.g., apple, kiwi, …) which do not receive an individual barcode, but which are grouped into the *fresh fruit* category. The same applies for fresh vegetables, cheese and meat items.

All SKU's are categorized into a product taxonomy. A product taxonomy is a hierarchy of product categories into which individual SKU's are grouped together, according to their functional usage. For instance, the SKU *Coca-Cola 20 cl.* (SKU) belongs to the product category *soft drinks*, which in turn belongs to the product group *beverages*, which finally belongs to the *food* department on the highest level of the taxonomy. Retailers tend to construct such product taxonomies in order to derive aggregate sales and profit figures on different levels of the taxonomy, e.g. for purposes of category management. Furthermore, product taxonomy information is also useful to discover multiple level or generalized association rules [128, 254], i.e. product associations that exist both on the same and on different levels of the taxonomy (see later in section 4.5.1.2).

The loyalty card data consists of a limited set of variables, i.e. socio-demographic and lifestyle data, which are collected from the customer and her household when she subscribes the loyalty card program. More specifically, we have information about the profession and the address of the customer, the number and age of the adults and children in the household, which and how many pets are owned by the customer, whether the customer owns a garden, a microwave and a refrigerator, whether the customer is a club member and whether she owns a car or not. For the period under study, a total of 5108 customers possess a loyalty card and have at least purchased once from this supermarket store.

Finally, the data set contains information about the supermarket's weekly promotion folder, such as which products are on offer at what price and when. Sometimes, the folder also includes reduction coupons.

2.3.2 Data Statistics

This paragraph provides some overview statistics of the data. For instance, figure 2.1 shows, for all households, the distribution of the number of distinct items purchased per visit. It shows that the average number of distinct items purchased per visit equals 13 and that most customers buy between 7 and 11 items per shopping visit. Figure 2.2 shows, for all households, the distribution of the average amount spent (in Belgian francs) per shopping visit. The average amount spent, over all households, equals 1276 BEF.

Figure 2.3 shows, for all households, the distribution of the total number of visits over the period of data collection (24 weeks). Although most customers have visited the store from 4 to 24 times over the entire period, the average number of visits to the store equals 25, which corresponds to about once per week.

Figure 2.4, shows the distribution of the profession of the loyalty card holders. It shows that about 25% of the card holders is employee, followed by housewives, which constitute almost 22% of the shoppers in this store.

Furthermore, figure 2.5 shows that most of the customers who shop in this store either have a partner (household size=2), or a partner and two children (household size=4).

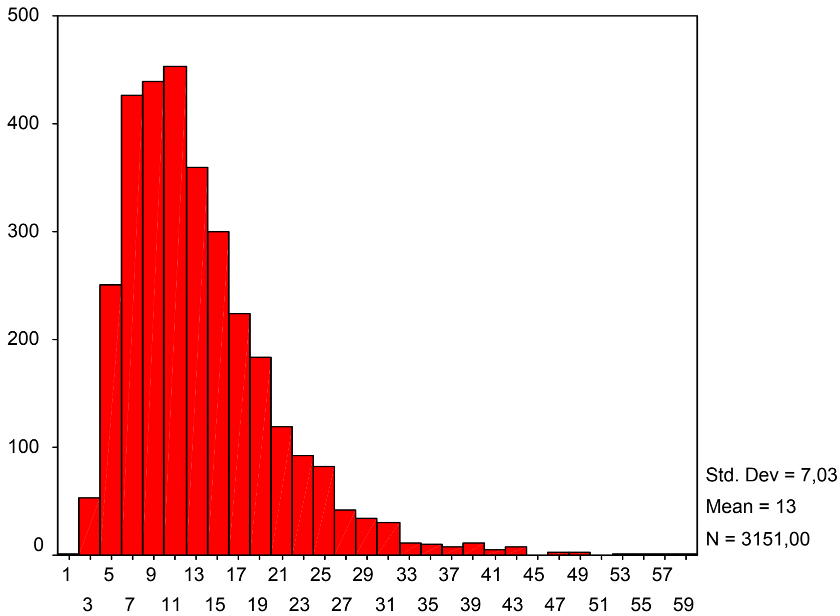


Figure 2.1: Average number of items purchased per visit

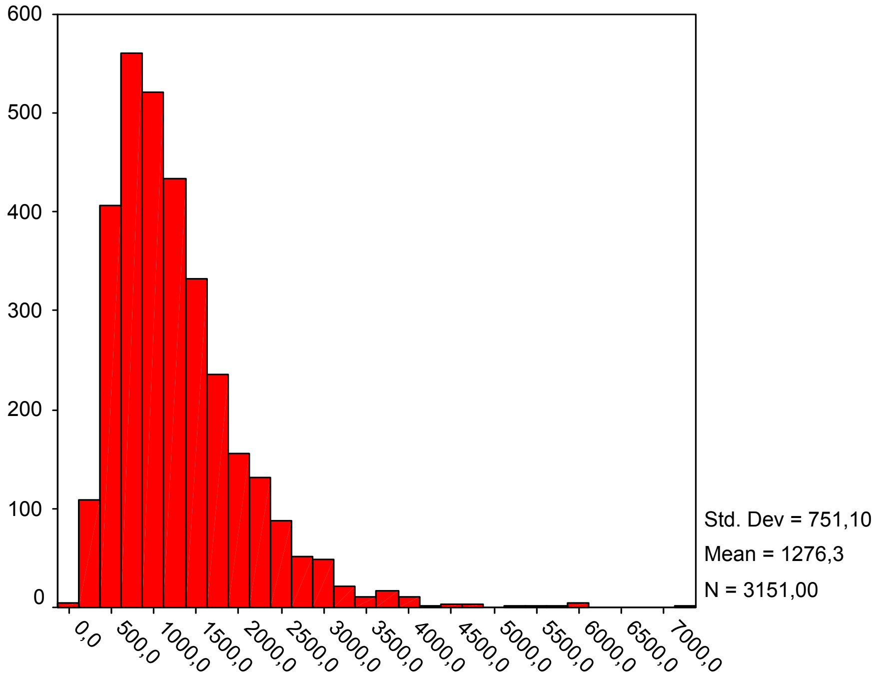


Figure 2.2: Average amount spent (in BEF) per shopping visit

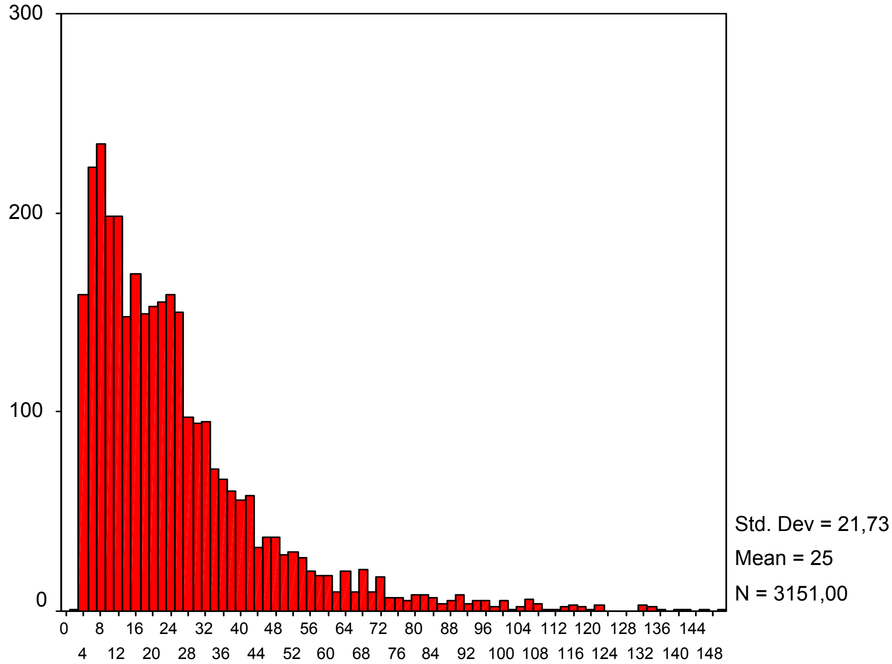


Figure 2.3: Total number of visits over 24 weeks

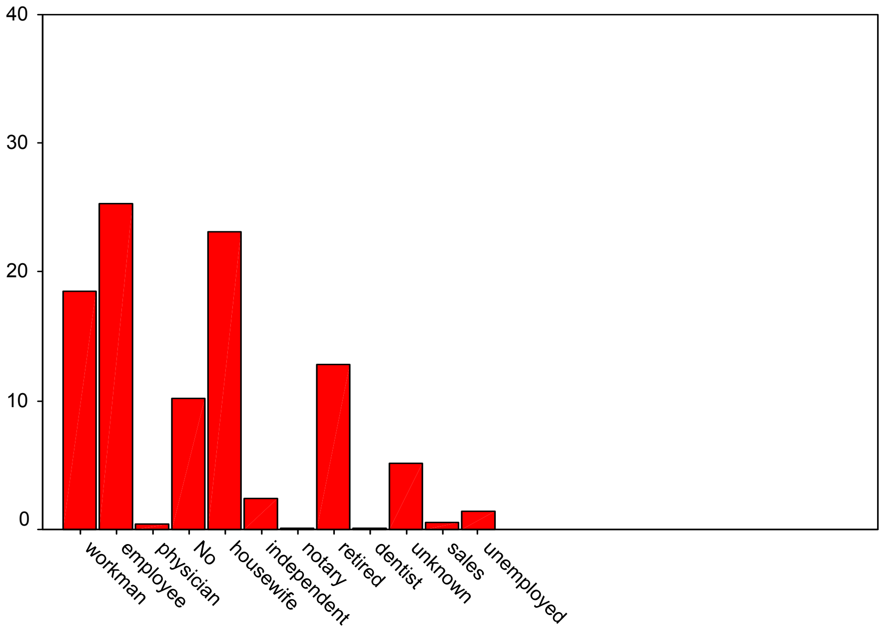


Figure 2.4: Profession of loyalty card holders

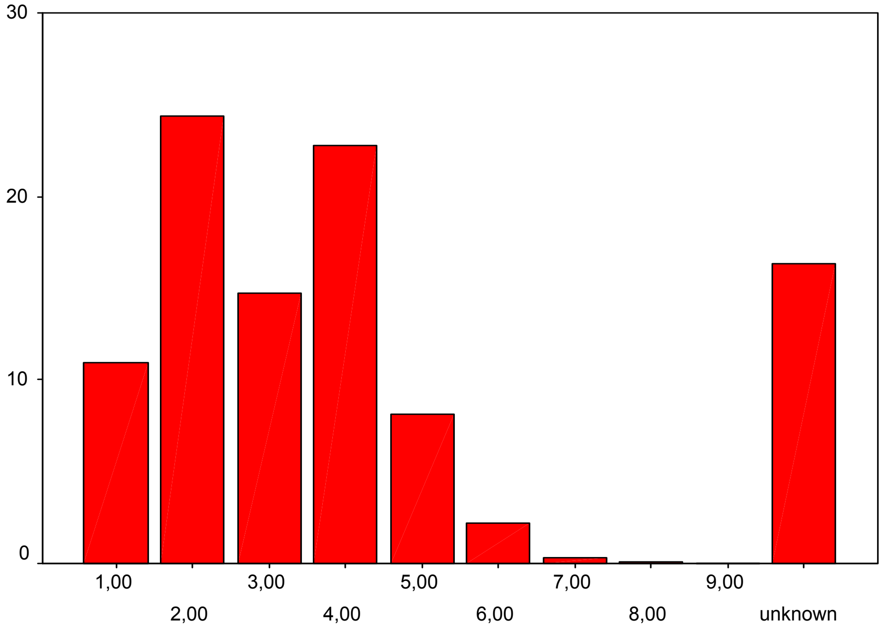


Figure 2.5: Householdsize (including children)

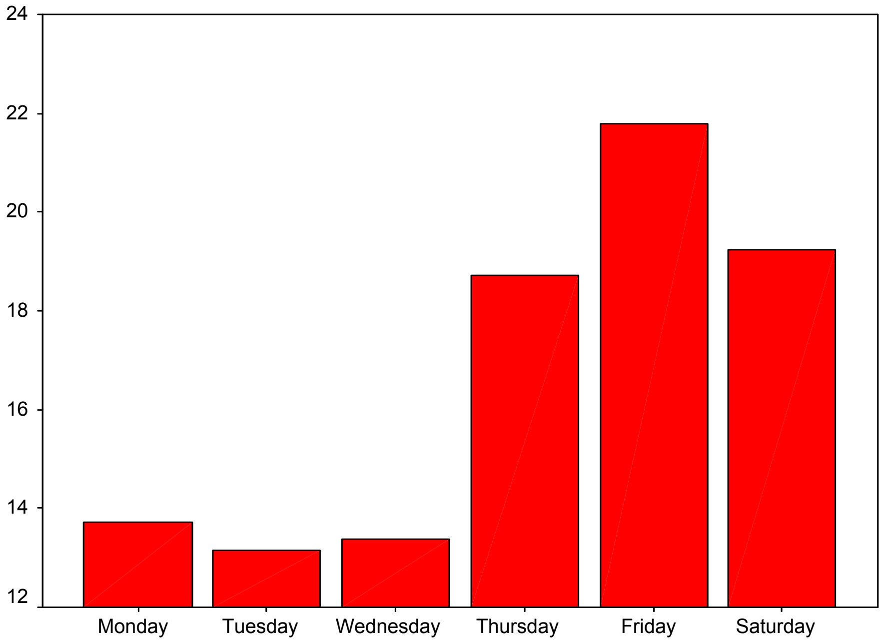


Figure 2.6: Distribution of sales per weekday

Further analysis showed that 89% the items in the assortment are slow moving, i.e. is sold on average less than once per day. Finally, figure 2.6 shows the distribution of the daily visits to the store. From this figure, it is clear that most of the visits to the store take place on Thursday, Friday and Saturday.

2.3.3 Data Preprocessing

Before the data can be used for market basket analysis, a number of data preprocessing (cleaning and transformation) issues need to be addressed. However, the cost of transformation and cleaning in order to prepare the data for analysis is substantial. Although data cleaning can be considered as not being part of the research 'as is', we have decided to report on this topic since our experience with data cleaning in the context of this research was that it turned out to be a very laborious and time intensive, but nevertheless very useful and necessary effort.

Therefore, this paragraph deals with some of the major data cleaning issues that were encountered during the course of this research. It is important to note that the data cleaning issues discussed here are independent of the techniques (see chapter 3 and 4) being used to analyse market baskets. Data preprocessing issues that are dependent on the type of analysis will be discussed separately where appropriate.

2.3.3.1 Typing errors

Since the data were collected over different time periods, a single product (with a unique productID) was sometimes labelled with a different product description, probably as a result of typing errors. In fact, typing errors occurred for 855 (5.2%) out of the 16430 different listed SKU's. Obviously, this poses problems during basket analysis when the product description is used as the unique identifier for a particular product. In that case, an identical product is treated as two different products as a result of a slightly different product description. Therefore, products with the same productID but a different product description were isolated and given an identical description.

2.3.3.2 Granularity

A different problem arises when products carry different productIDs but are labelled with the same description. This occurs for instance with greeting cards or women and men's underwear, where each item possesses its own product identification number, but since the items only differ in colour or size, the retailer assigns them the same product description. According to the marketing research problem, the analyst has to make the choice whether he wants these products being treated as different or not. For our analysis, we decided to group all greeting cards into one generic greeting card product, just as women and men's underwear. The same applies for fresh vegetables and fruit, and for fresh meat and cheese. The supermarket store does not label the different fruit, vegetable, meat and cheese items separately, but assigns one barcode for the entire product category, regardless of which item was purchased within that category.

2.3.3.3 Outliers

In total, 16 receipts with abnormal high volumes (over 80 different products) or prices (over 15000 BEF) were removed from the data. It was our interpretation that these receipts are not representative for the wider customer population and they could badly influence the market basket analysis.

2.3.3.4 Returned products

In 64 baskets, negative volumes and prices were observed on the receipts as a result of returned goods. In the case were the returned product was purchased during a previous shopping visit, and the returned good is exchanged for an identical new one during the current shopping visit, the records with the returned good and the exchanged good were removed from the receipt, just as nothing had happened. In the case where a refund was given and thus the returned good was not exchanged by a new one, the record with the returned good and the original record on the previous receipt were deleted, just as if the purchase of the original product had not taken place.

2.3.3.5 Coupons

In the case where customers returned coupons at the checkout, these coupons are scanned as part of the receipt and they are assigned a unique ID, as if they were products. According to the type of marketing problem being studied, it is up to the analyst to determine whether he wants to include this information into the baskets or not. Since these coupons mostly carry a negative price, they influence the total sales volume of the receipt and as a result they may influence some key statistics about the receipts. Furthermore, our interest does not lie in the analysis of consumer purchase behaviour as a result of coupons. We therefore decided to remove coupons from the baskets.

CHAPTER 3
MEASURING INTERDEPENDENCE

Recent years can be characterized by a rapid increase of powerful techniques to analyse retail market basket data. Both the strong increase in computer power and new developments in the field of data mining and statistics have truly led to a revolution in the possibilities to analyse market basket data. In this chapter, we will provide an overview of the most popular approaches to the analysis of market basket data, and to the measurement of product interdependencies more specifically. The chapter will start with an overview of the causes of product purchase interdependence followed by a number of techniques to measure interdependence based on coefficients of elasticity, the theory of utility, and coefficients of association. To the latter category also belongs the data mining technique of association rules. However, since this dissertation project was conceived as a study of retail market basket analysis with a special focus on data mining techniques, association rules will be treated separately in the next chapter (chapter 4).

3.1 Reasons for Interdependence

The concept of product interdependence has been studied in the literature in a wide range of contexts, e.g. to measure the effect of promotions or displays on the sales of promoted and non-promoted products, to measure the effect of changing shelf space on the sales of the product and/or the sales of other products in the same or other product categories, etc. Furthermore, when reading about product interdependence, several reasons for interdependence can be identified. An excellent reference in this context is the work of Böcker [40] and Merkle [200]. The following sections will therefore mainly draw on their contributions to the field.

Böcker and Merkle argue that although product interdependence effects are usually measured at the consumer side, the origin for interdependence can come both from the producer side, the retailer side and the consumer side.

3.1.1 Producer Side Reasons for Interdependence

From the producer side, there may be technical reasons why products are purchased together, simply because there exists some kind of technical usage affinity between particular products, or products are designed to be usage complements. For instance, many instruments are designed in a modular way such that different parts must be combined in order to obtain a functional end product, such as manual shaving equipment where both a handle and razor blades are needed to shave. Without both components, one can not shave. The same applies for a pocket-torch for which batteries are needed in order to use it, or coffee and coffee filters that go together to make coffee. Some products are thus technically related and therefore implicitly induce purchase complementarity.

3.1.2 Retailer Side Reasons for Interdependence

Many of the purchase interdependencies may also result from choices or decisions made by the retailer, as illustrated in the sections below.

3.1.2.1 Administrative policy

Some products are offered together as a package instead of being sold separately (conditional sale). Although the law actually prohibits conditional sale, it is a common practice in the banking and insurance sector where certain administrative procedures are adopted to force customers to purchase more products at once. However, in grocery retailing, it is probably a less frequent practice. For instance, from personal contacts with the management of the Belgian company Shop24, which sells and composes product assortments for fully-automated convenience stores, we know that they investigate the feasibility of suggesting customers to purchase a food package, including a meal, soft drink and dessert, for a reduced package price instead of purchasing them apart. In fact, the user interface to their automated vending machine enables the shop to communicate with the consumer and make suggestions to him at the moment of purchase.

3.1.2.2 Product and assortment policy

In some cases, the product assortment policy pursued by the retailer may be a reason for observing product purchase interdependencies. In fact, a retailer will typically compose a particular product assortment in line with the strategic positioning of the store and in line with the store formula. This means that the retailer chooses particular products (core assortment) because they reflect the store image. As a result, the retailer implicitly positions these products as interdependent.

3.1.2.3 Pricing policy

Practices such as loss leader pricing and implicit price bundling may also be reasons for observing product purchase interdependencies. Loss-leader pricing means that by reducing the price of a particular product, like spaghetti (even making a loss on it) customers will not only buy more spaghetti, but they will probably also buy more spaghetti sauce. This way, an overall positive profit can be achieved because the high margin spaghetti sauce purchase will compensate for the less profitable spaghetti purchase. In other words, one product will act as a decoy, i.e. spaghetti is positioned in an attractive way in order to pull customers towards the store and to encourage them to purchase spaghetti sauce too.

3.1.2.4 Communication policy

Products may become purchase complements/substitutes as a result of particular promotions adopted by the retailer. In fact, many medium to large size grocery retailers send out weekly promotion folders to advertise a number of products. It is clear that these advertisements have a short-term (and maybe also a longer term) impact on the interdependence relationships between products. For instance, purchase associations between products of the same (national) origin may result from an advertising campaign by the retailer to promote Spanish products or biological products (e.g. Delhaize and Albert Heijn).

3.1.2.5 Display policy

Finally, the location of the products within the store, i.e. shelf space allocation decisions, and more generally the physical environment of the store play an important role in product interdependence effects [258]. Indeed, stronger product purchase interdependence effects can be expected for products that are located close together in the store.

In this context, it should be noted that product interdependencies as a result of retailer-side reasons for interdependence should be handled with care when they are used for retail marketing mix decisions. In fact, the discovered product interdependencies may reflect the result of earlier marketing mix decisions (similar to the identification problem in measuring shelf-space elasticities [83]). In order to avoid this identification problem, it is important for the retailer to carefully store all merchandising parameters (prices, promotions, allocated shelf space, etc) into a database such that market basket analysis results can always be analysed within the correct context.

3.1.3 Consumer Side Reasons for Interdependence

Finally, most of the purchase interdependencies between products probably result from the consumer.

3.1.3.1 Need affinity

Probably the most important reason for observing product interdependencies results from the complex wants and needs of the consumer. Indeed, in general, human wants and needs can not usually be fulfilled by a single product but rather by a bundle of products. These wants and needs, however, are highly dependent on cultural and socio-economical factors resulting in a wide variety of different wants and needs. In other words, consumers are heterogeneous in terms of their response to marketing actions (e.g. as a result of their different socio-demographic/lifestyle background), which determine the utility that they derive from product category purchases.

3.1.3.2 Type of shopping trip

Another important determinant to explain product interdependencies results from the complexity of the decision process to purchase goods (involvement), which in turn depends on the type of good and the subjective risk perception

[258]. This has led to different classification schemes of which the categorization of consumer goods into convenience, shopping and specialty goods is the most well known. Convenience goods are purchased often spontaneously, almost by habit. In this case, the desire by consumers to minimize the costs of their shopping trip causes them to concentrate as much of their purchases as possible during the same shopping trip. In fact, since most grocery shopping trips can be characterized by the purchase of convenience goods, this effect plays very strongly in a supermarket context. In extreme cases, this leads to so-called one-stop-shopping. In the case of shopping goods and specialty goods, the choice evaluation process usually takes more time and products are compared more carefully. Furthermore, since the consumer's financial commitment is higher, this may sometimes cause consumers to purchase from well known, highly marketed brands in order to minimize their perceptual purchase risk.

Another nice example on the Belgian do-it-yourself market is the study by Van Kenhove et al. [278]. They showed that the purpose of the shopping trip (urgent purchase, large quantity purchase, difficult job purchase, regular purchase, and get ideas) influences store attribute saliences and the store choice. Although they have not investigated this in their study, they argue that task definitions may, however, also have an impact on product choice.

Also time pressure [215], habit formation [148], economic reasons (to spread the cost of a trip over many items) and the mood of the customer [97] may explain the existence of joint purchases.

3.1.3.3 Variety-seeking behaviour

Consumers may also purchase products together as a result of variety-seeking behaviour. Indeed, whereas some customers always stick to the same products, other consumers like to try-out new products and experiment with new tastes, colours, packaging, etc. In fact, variety-seeking behaviour has turned out to be one of the most important causes for observing product interdependence in our data. Indeed, although many products occurring in the same basket would traditionally be considered as substitute products (covering roughly the same needs), such as paprika crisps and salty crisps or Cola and Fanta, they tend to be purchased together very often. In fact, our data mining analysis showed that when a customer buys in a particular product category, he often purchases multiple brands and/or varieties within that category.

Despite this variety of reasons for observing product interdependencies, marketing modellers typically only make a difference between purchase complementarity, heterogeneity and co-incidence [183] to explain product or category purchase interdependencies.

Two products/categories are *purchase complements* whenever a particular marketing activity (pricing, promotion or display) influences the consumers' purchase in another category. The reason is that these effects are usually within managers' control and that marketing modellers typically want to separate the effects of controllable versus uncontrollable factors (such as consumer heterogeneity).

Indeed, also *customer heterogeneity* can cause products to occur together in a basket. The idea is that consumers are heterogeneous and that their differences in socio-demographic profile and lifestyle characteristics may influence their intrinsic utilities and responses to marketing actions in each category and that these utilities and responses may be related across categories [183].

Finally, products can end up in the same basket as a result of all other reasons except purchase complementarity and consumer heterogeneity. In that case, the observed co-occurrence is called *co-incidence*. The above reasons for product interdependencies can therefore be summarized as shown in table 3.1.

Purchase complementarity	Co-incidence	Customer heterogeneity
Pricing policy	Usage complements	Different needs/wants
Communication policy	Customer involvement	
Display policy	Time pressure	
Assortment policy	Habit formation	
	Cost of trip minimization	
	Customer mood	
	Store environment	

Table 3.1: Classification of reasons for interdependence

3.2 Measuring Interdependence

The previous section has shown that there are potentially multiple factors that cause product interdependence and that it is difficult to isolate the contribution of each of those factors in the level of observed interdependence (co-occurrence) between a set of products. This section provides a literature overview of the contributions made to measure and/or explain product purchase interdependencies, including micro-economic models of (cross-) elasticity (3.2.1), models of deterministic and stochastic utility theory (3.2.2), and measures to examine co-occurrence such as association coefficients and loglinear analysis (3.2.3). The technique of association rules is treated in a separate chapter (chapter 4).

3.2.1 Cross-Elasticities

A popular way to measure/express product interdependence, both in micro-economic theory as in econometrics and marketing research, is by the use of cross-elasticities. Cross-elasticities measure the effect of the change in the marketing mix action of a particular product/category on the sales of other products/categories.

3.2.1.1 Cross-price elasticity

Most of the work on the measurement of (and causes for) interdependence originates from micro-economic theory. In fact, the theoretical foundations of the problem of product purchase interdependencies were already studied in the beginning of the 20th century in the context of the micro-economic models of price elasticity [214, 238]. Especially, the work of Triffin about cross-price elasticity can be considered as a significant contribution to the field.

Triffin's microeconomic theory of cross-price elasticity [269] dates back to 1940 and defines complementarity (resp. substitution) between two products X and Y, whenever a price decrease (increase) of product X, i.e. (Δp_X), generates higher (resp. lower) sales (ΔS_Y) for Y. Cross-price elasticity in this context is therefore measured by the cross-price elasticity coefficient ε_{XY}:

$$\varepsilon_{XY} = \frac{\Delta S_Y}{S_Y} \bigg/ \frac{\Delta p_X}{p_X} \tag{3.1}$$

A positive value for ε_{XY} indicates a substitution effect between X and Y, whereas a negative value indicates a complementary effect. Although elegant in its simplicity, an important limitation, namely the huge effort to measure all elasticities for a typically wide product assortment, has made implementations within a supermarket environment practically infeasible [200]. More recently, Blattberg and Neslin [35] developed a model for maximizing the profits in a category, taking into account interdependencies between items in the category. The sales of each item are made dependent on the other items' deals. As a result, the category margin and the degree of cannibalization determine the optimal price discount for an item. Also Mulhern and Leone [209] studied the influence of price deals on three groups of products, the promoted item, its purchase complements, and the item's purchase substitutes. They found that a price deal increases the sales volume of the promoted item and the item's purchase complements, but reduces the sales of the item's purchase substitutes.

3.2.1.2 Cross-space elasticity

Later, the concept of cross-elasticities was used by marketeers in the context of shelf-space allocation models to express the impact of shelf-space decisions of one product/category on the sales of other products/categories. For instance, Corstjens and Doyle [83] were the first to include both direct and cross-space elasticities into their shelf-space optimization model. They argue that any shelf-space allocation model to optimize a retailer's profits should take into account both direct and cross-space elasticities. Later, cross-space elasticities were also adopted by Borin, Farris and Freeland [44], Urban [273] and Bultez et al. [65, 66] and Swinnen [257].

Unfortunately, obtaining good estimates of direct and cross-space elasticities for large amounts of products is not straightforward [174]. The literature describes three techniques in this context: experiments, time-series data and cross-sectional data. In-store experiments are probably the most reliable since they experimentally measure the effect of a change in shelf space on the sales. However, since these experiments are very laborious, time consuming and may even be disruptive towards the operation of the store, this method is not used very often. Therefore, Bultez et al. [65, 66] used time-series data to estimate the effect of changing shelf space on sales. Finally, cross-sectional data offers an alternative solution to the measurement problem of elasticities [83, 84]. The idea is that when collecting data from different stores, there is enough variation in the amount of shelf space devoted to products and their resulting sales such that their relation can be estimated by regression techniques. The advantages are speedier results, low cost and no interference with store operations. However, the major drawback is the problem of identification since it is not always clear whether the relation between space and sales is the result of true space elasticity, or whether it merely reflects the retailer's earlier decisions to allocate more space in proportion to past or expected sales.

The above mentioned problems, and the computational difficulties as a result of the non-linear character of the optimization, may explain why many shelf space allocation models do not adopt elasticities [131] or when they do, they usually only consider a very limited number of products/categories.

3.2.1.3 Cross-location elasticity

Also the location of products within the store may have an impact on sales. For instance, Drèze et al. [99] discuss how retailers can boost sales by better managing their available shelf-space through reorganizing the location of the existing products in the assortment. In this context, eye-level is often seen as the best location. It is therefore crucial to carefully think about which products to put at those locations and how the reorganization may affect the sales of other products. By means of experiments, they found out that indeed the location of products has an important effect on the sales of the product and on related products. For instance, they showed that by putting toothbrush at eye-level, instead of toothpaste, the sales of toothbrushes increased by 8% whilst keeping toothpaste sales unaffected.

Chapter 5 in this dissertation is devoted to a discussion of our own optimization model to support such location decisions. The idea is that retailers often put top-selling products at visually attractive locations but that their rule-of-thumb usually does not take into account cross-selling effects with other products. In other words, even though a product is not a top-seller, but belongs to the sub-top selling group, its cross-selling effects with other products may be significant such that overall it can compete with (or exceed) some top-selling products and should therefore deserve an opportunity to be located at an attractive location.

3.2.2 Utility Theory

The theory of utility states that consumers purchase certain (bundles of) products because they derive utility from them. However, a distinction can be made according to how the utility is modelled: deterministic or stochastic.

3.2.2.1 Deterministic utility theory

The classical theory of utility [238] states that a consumer with a limited budget allocates expenditure between different commodities so as to maximise the utility or satisfaction from consumption. Lancaster, however, criticised this theory and developed the microeconomic theory of the household [168, 169], which states that goods are purchased because they represent combinations of certain characteristics that are desired by consumers. Thus, goods themselves are not the immediate objects of preference or utility, such as in the classical theory of utility, but they have associated with them *characteristics* (such as calories, proteins, vitamins, …) that are directly relevant to the consumer. Therefore, the consumer's demand for goods is *derived* from their demand for characteristics.

In this context, the substitutability of one product for another increases as the (perceived) attributes of a set of products become increasingly similar [286]. Product complements, on the other hand, are products that are used in conjunction with one another to satisfy some particular need, i.e. together they provide a set of characteristics that are needed by the consumer to fulfil his utility.

3.2.2.2 Stochastic utility theory

Whereas the classical theory of utility views consumer choice as deterministic, recent developments in marketing research treat consumer choice decisions as stochastic. The latter are therefore referred to as stochastic (or random) utility models instead of deterministic models. Under deterministic choice models, such as those from Lancaster [169] and Luce [179], the consumer is assumed always to assign the same utility to the same choice alternative. The stochastic choice model, however, assumes that the individual draws at random a member of a set of utility functions for each choice occasion. Consequently, utility levels for different alternatives are distributed around mean levels of utility, which depend on the alternatives' attributes (which mostly consist of a constant plus marketing mix effects) [84]. Because of the flexibility and the strong theoretical foundations of the random utility framework, it has been used as the basic underlying framework for the study of brand choice by consumers, better known as brand choice models that have appeared since the early 1980's onwards [e.g. 117, 132, 143, 183]. It would lead us too far to discuss every development in the field of brand choice models. Therefore, in the next paragraphs, only the most recent developments will be highlighted.

The classical brand choice model

This random utility theory assumes that the consumer is a rational decision-maker who aims to maximize the utility from purchasing a (bundle of) product(s). This means that from a set of alternative products, the consumer will pick the product that produces the highest utility for the consumer. Usually, this utility is determined by the sensitivity of the household towards a number of product specific features, such as price, promotion and display. To illustrate this, consider an individual i facing a choice-set of J different (substitute) brands within a certain product category. Then, at shopping occasion t, the utility (U) that he derives from buying brand j can be expressed as:

$$U_{ijt} = V_{ijt} + \varepsilon_{ijt} \tag{3.2}$$

$i = 1, .. , H$ (number of households)
$j = 1, .. , J$ (number of alternative brands)
$t = 1, .. , T$ (shopping visit)

The utility for household i from brand j at shopping visit t is thus composed of a deterministic component $(V_{i,j,t})$ and an error term $\varepsilon_{i,j,t}$, representing for example, the value of a sub utility of unobservable attributes and socioeconomic characteristics. Mostly, $\varepsilon_{i,j,t}$ is assumed to be independently and identically distributed (IID) over alternatives and consumers. The deterministic component, however, consists again of two parts:

$$V_{ijt} = \alpha_{ij} + X_{ijt}\beta_i \tag{3.3}$$

Firstly, there is the intrinsic preference $(\alpha_{i,j})$ of individual i towards brand j. After all, it is believed that the consumer possesses an intrinsic preference towards a brand that can be represented by a constant term (i.e. brand specific intercepts). Secondly, in addition to the intrinsic preference, the deterministic component is influenced by the sensitivity of the consumer with regard to the different marketing-mix variables, such as the price, promotion and display of the brand. This sensitivity is reflected by the c x 1 vector β_i which differs across consumers, but which is mostly indifferent with respect to time and the brand alternative. $X_{i,j,t}$ is a 1 x c vector of explanatory variables that includes the price, promotion and display of brand j at purchase occasion t. In a hierarchical framework, these sensitivities will in turn depend on the socio-demographic and/or lifestyle characteristics of the household/consumer (e.g. see [183]).

It is important to note that the utilities (U_{ijt}) can not be observed directly. Therefore, they are also referred to as *latent* utilities, but they can be mathematically inferred from the choices made by the panellist. The link between the observed behaviour (I_{it}) and the latent utility for any product k can be represented as follows:

$$I_{it} = j \text{ where } U_{ijt} = max_k\,(U_{ikt}) \tag{3.4}$$

In other words, the brand j chosen by panellist i on choice occasion t is the one that represents the highest utility among all J brands in the category being studied. The purchase process is therefore characterized by a discrete choice (i.e. the consumer purchases a product or not) and the alternatives are indivisible. This is different from the classical economical view on utility maximization [168, 169, 284] where the consumer buys proportions of different products, i.e. alternatives are divisible. The latter are called continuous choice models but since they are of minor importance in the literature, they will not be treated further in this overview.

Variants of the classical brand choice model

The classical brand choice model has been extended in a number of ways, as discussed in subsequent paragraphs.

Single versus multiple category choice context

The objective of single category choice models is to study how consumers choose between different competing brands within a single product category [117, 261], whereas the objective of the multi category choice model is to study the purchase behaviour of households in several product categories simultaneously [14, 183, 232, 242]. In the latter framework, it is assumed that a consumer chooses a product from a particular category in the context of a larger choice task [63]. Popular models to study purchases of one product within a single category include the multinomial logit and probit models. The simultaneous purchase of multiple products is often studied by multivariate logit and probit choice models. In this context, the contribution by Manchanda et al. [183] is worth to note since they developed a model for multicategory purchase incidence decisions within a random utility framework that allows for simultaneous, interdependent choice of items.

Dealing with unobserved heterogeneity

A second difference between brand choice models relates to how the model deals with unobserved heterogeneity. Unobserved heterogeneity across households has been widely recognized as a critical research issue in choice modelling [19, 91, 291]. It refers to the fact that the households being studied are assumed to be heterogeneous in nature, which implies that households react differently to the same marketing-mix variables and often have different base preferences for products. However, it is not apriori known how many or how big these different customer segments are, i.e. they are hidden (unobserved) in the data. Consequently, unobserved heterogeneity can be dealt with in basically three ways that relate to the level of aggregation of the choice model.

In *aggregate* level choice models [117], only one response function is estimated for the entire sample of households. In order to allow for unobserved heterogeneity, the parameters of the aggregate response function are defined as stochastic variables following some distribution across the population of study. Subsequently, choice predictions of individuals outside the sample are made by using the aggregate level parameters. Therefore, this approach does not fully capture the individual customer differences in the sample.

In *group* level choice models [291], one response function is estimated per customer segment. These customer segments may be defined apriori (i.e. user defined) or post hoc (either by a traditional clustering approach or a simultaneous approach, i.e. latent class cluster models, where the segments and the response functions within each segment are calculated simultaneously). Subsequently, choice predictions for individuals outside the sample are made conditional on their membership probabilities to one or more of the identified customer segments.

Finally, in *individual* choice models, one response function is estimated for each individual household in the sample and thus a set of response parameters is estimated for each individual household. As a result of the large number of parameters that have to be estimated, this approach requires a large number of observations per individual in order not to jeopardize degrees of freedom and parameter stability. Therefore, in practice, individual choice models are not used very frequently. Yet, from the theoretical point of view, individual response models allow for maximal flexibility in modelling individual consumer choice behaviour. Furthermore, they enable predictions of choice behaviour on the level of the individual.

Besides the level of aggregation of the response function, a further distinction can be made with regard to the type of heterogeneity (see [91]) and its impact on the definition and estimation of the utility equation.

Response heterogeneity: means that individuals have different intrinsic preferences for some products or categories and therefore this type of heterogeneity is typically reflected in the intercept term $(\alpha_{i,j})$ of the utility function.

Structural heterogeneity: means that individuals may respond in a different way to the same attribute values because they assign a different (utility) value to different product attributes according to their individual needs. This type of heterogeneity is typically reflected in the β_i parameters of the utility function. The specification of both response heterogeneity and structural heterogeneity gives rise to different estimation procedures for the preference and response coefficients. One approach is called the *fixed effects* approach, wherein a set of parameters is estimated for each household separately [77, 230] and no particular probability distribution of heterogeneity must be specified. However, the fixed effects model involves estimating a large number of parameters and requires long purchase histories for each household. Therefore, in the case of insufficient observations per panellist[14], the fixed effects model has proven to produce biased and inconsistent estimates not only for the fixed term, but also of the effects of marketing mix variables [141].

[14] Allenby and Rossi [19] report that the average number of household purchases in most product categories is often less than 12 per year.

A more tractable approach to estimating the model parameters is therefore to assume that these parameters vary across households according to some probability distribution. This is referred to in the literature as the *random effects* model [134, 141]. However, again different implementations of the random effects model have been proposed in the literature according to whether the parameters are assumed to follow a predefined probability distribution [78, 114] across the households or no specific parametric distribution is imposed but the distribution is estimated empirically using the underlying data [50], or the coefficients for each household are made dependent on a further set of variables, such as socio-demographic variables [14]. Research has shown that the non-parametric (distribution free) approach produces a better fit of the brand choice model. However, some people argue that the computational difficulties in estimating the model are high [50], although different papers lead to different outcomes [292].

Perceptual heterogeneity: means that individuals may differ in their perceptions, familiarity and/or recall of the underlying attributes utilized in their decision processes. This may be reflected via different values of the $X_{i,j,t}$ observations in the utility function.

Form heterogeneity: means that households may differ according to how their utility is constituted. For instance, the utility function was previously conceived as a linear function. However, it is possible that some customers use another form of utility function, e.g., non-linear. Moreover, some customers may value the product attributes in a compensatory way whereas others do not. This type of heterogeneity is typically reflected in the functional form of the utility function or of the choice model as a whole.

Distributional heterogeneity: individuals may possess higher or lower variance or shape parameters in the utility equation. This implies that the parameters of the error distribution of $\varepsilon_{i,j,t}$ may be different for different households. Furthermore, distributional heterogeneity may be reflected in the type of the error distribution being used (logistic, normal, …).

Time heterogeneity: consumers may differ in their reaction to their past purchase experiences and behaviors. For instance, some customers tend to be very loyal towards a product whereas others may possess more volatile utility functions whose structure changes rapidly over time. This type of heterogeneity may reflect almost any aspect of the utility function. For instance, it may affect the constant $(\alpha_{i,j})$ of the utility function (see formula 3.3) due to habit formation or variety seeking behaviour (see next paragraph), but it may also change the importance that they attach to the different elements of the marketing mix (i.e. the beta-coefficients in formula 3.3).

Zero order versus higher order effects

Finally, most discrete choice models assume that the consumer wants to maximize the utility on each purchase occasion and therefore ignore that the utility of a brand on a particular purchase occasion may be affected by the choice(s) made by the consumer on previous shopping occasions, i.e. state dependence effects or purchase event feedback. State dependence refers to the idea that for some customers the probability to purchase a particular brand increases (decreases) when the same brand has been chosen on previous purchase occasions, i.e. positive (negative) state dependence. Positive state dependence (or habit formation) may result from a routine behaviour [140] of the customer to buy the same brand repeatedly over time. In contrast, negative state dependence may result from a variety seeking behaviour [187] of the customer to try and purchase different brands over time.

State dependence effects can be dealt with in a number of ways. Mostly, a measure of brand loyalty is introduced, such as the most recent purchase [145] or an exponentially weighted sum of all past purchases [117]. Dynamic discrete choice models follow another approach, which in general requires the consumer to solve a dynamic optimisation problem [79]. Furthermore, researchers have investigated whether state dependence effects differ across product categories and if there exists a relationship between marketing mix sensitivities and state dependence and if state dependence can be distinguished from unobserved heterogeneity [1]. Since this is not the focus of this dissertation, we will not elaborate on this, but the interested reader is referred to an excellent discussion of state dependence by Seetharaman, Ainslie and Chintagunta [242].

3.2.3 Measuring Co-occurrence

The approaches to measure product interdependencies discussed so far have one element in common: they *quantify* and *explain* complementarity and substitution effects as the customer's purchase reaction in one product/category as a result of particular marketing actions (price, promotion, space, location) for another product/category, either within a cross-elasticity framework or within a utility-maximizing framework. The approaches that will be discussed in subsequent sections will differ from the earlier approaches in so far that they *measure*, but do *not explain*, interdependency effects between products/categories. We will therefore talk about 'co-occurrence' instead of 'complementarity'. In other words, the measures discussed hereafter will enable to quantify the amount of co-occurrence between products but not the reason for its existence.

3.2.3.1 Association coefficients

One of the earliest attempts to express product purchase relationships was developed by Böcker and Merkle [40, 200] during the late 1970s and the early 1980s. Both authors noticed that there was a lack of theoretical and statistical background with regard to the analysis of product interdependencies. Indeed, at that moment, the only existing theory of product interdependence was the microeconomic theory of cross-price elasticity [269] and Lancaster's household theory [168]. Although very useful from a theoretical point of view, these theories lacked practical relevance since they require the measurement of price elasticities for a large range of products, which has shown to be practically infeasible. The demand for practically feasible measurement systems and the rise of large amounts of consumer purchase data by means of electronic barcode scanning systems have therefore motivated Böcker and Merkle to study product interdependencies and to propose a number of empirically tractable measures of product interdependence, called association coefficients.

Association coefficient	**Calculation**	**Value range**
Tetrachoric	$\cos\left(\frac{\Pi}{1+\sqrt{\frac{ad}{bc}}}\right)$	[-1,+1]
Yule's Q	$\frac{ad-bc}{ad+bc}$	[-1,+1]
Phi	$\frac{ad-bc}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	[-1,+1]
Hamann	$\frac{(a+d)-(b-c)}{a+b+c+d}$	[-1,+1]
Simple matching	$\frac{a+d}{a+b+c+d}$	[0,+1]
Russel-Rao	$\frac{a}{a+b+c+d}$	[0,+1]
Jaccard	$\frac{a}{a+b+c}$	[0,+1]

Table 3.2: Overview of association coefficients

In their overview of the literature, they highlight several association coefficients for nominal data[15] (see table 3.2) based on pairwise contingency

[15] Their discussion also includes interdependence analysis for ordinal and interval scaled data by means of (rank-order) correlation coefficients where the data represent sales volumes or dollar sales per product. However, since such correlation coefficients can potentially be influenced by fluctuations in the unit price of products, Böcker and Merkle favour interdependence analysis on nominal data (purchase or not).

tables (see table 3.3) where *'a'* equals the joint frequency of occurrence of both product(category) i_1 and i_2 in the data and *'b'* equals the frequency of occurrence of i_2 but not i_1 in the data, etc.

Product i_1 / i_2	1	0	Σ
1	a	b	$a+b$
0	c	d	$c+d$
Σ	$a+c$	$b+d$	$n=a+b+c+d$

Table 3.3: Example of a contingency table

As a matter of illustration, the Yule's Q coefficient [155] will be calculated for an example set of observations including 7 market baskets and 6 products (i_1 to i_6) in table 3.4.

TID	i_1	i_2	i_3	i_4	i_5	i_6	items purchased
1	1	0	1	1	0	0	3
2	0	0	0	0	1	1	2
3	0	0	0	0	1	1	2
4	0	1	1	1	1	0	4
5	1	0	1	1	1	0	4
6	1	1	1	1	1	0	5
7	1	0	1	1	0	0	3
Total item sales	4	2	5	5	5	2	23

Table 3.4: Illustration for 7 multiple purchases

Yule's coefficient is symmetric and takes values from –1 to +1, where –1 represents a strong negative interdependency, +1 a strong positive interdependency, and 0 no interdependency between both examined product categories. For instance, according to the Yule's Q coefficient (table 3.2), the association between products i_1 and i_2 in table 3.4 equals

$$Q_{i_1 i_2} = \frac{ad - bc}{ad + bc} = \frac{(1\times2)-(1\times3)}{(1\times2)+(1\times3)} = -\frac{1}{5}$$

Note that Yule's Q coefficient tends to go to +1 or –1 when one of the cells in the contingency table equals 0. Take for instance the association between products i_1 and i_3:

$$Q_{i_1 i_3} = \frac{ad - bc}{ad + bc} = \frac{(4\times2)-(1\times0)}{(4\times2)+(1\times0)} = 1$$

To solve this shortcoming, Agresti [13] suggests adding a small constant value to that cell of the contingency table such that the (false) perfect association is avoided. However, this remains an important problem for the calculation of Yule's coefficient. Indeed, calculating Yule's coefficient for all products (i_1 to i_6) given in table 3.4 and after adjustment for zero values provides the following matrix of association coefficients, as illustrated in table 3.5.

Item	i_1	i_2	i_3	i_4	i_5	i_6
i_1	-					
i_2	-0.2	-				
i_3	0.999	0.999	-			
i_4	0.999	0.999	0.999	-		
i_5	-0.999	0.999	-0.999	-0.999	-	
i_6	-0.999	-0.999	-0.999	-0.999	0.999	-

Table 3.5: Yule's Q coefficients of association (after adjustment)

The example given above is a rather extreme case since the number of instances compared to the number of products is small, which easily leads to a zero frequency in the contingency table with an extreme coefficient as a result. However, even on real datasets, where the amount of SKU's is much larger, this remains an important weakness of the Yule's Q coefficient. To illustrate this, consider dataset 1 (section 2.3) of 88163 baskets. The expected joint frequency of occurrence of two items in a contingency table will be below 1 if the multiplication of the observed row and column count is below 88163. Now, suppose that the row and column counts in the contingency table are equal, then this already happens if both items occur in less than 0.336% of the baskets:

$$\frac{x^2}{88163} < 1 \Leftrightarrow x < \sqrt{88163} \Leftrightarrow x < 296 \Leftrightarrow s(x) < 0.336\%$$

Given that over 89% of the items in our supermarket are slow moving (purchased on average less than once per day), this remains an important problem for the calculation of the Yule's coefficient. In general, Böcker and Merkle illustrate the lack of stability with respect to the cell frequencies of the contingency table for many of such coefficients, which may hinder their practical implementation on real data.

This motivated them to design a new and more robust association coefficient, explained below. In fact, a matrix is built containing the frequencies of simultaneous purchases for all product pairs (see table 3.6) based on the market baskets in table 3.4. However, the matrix rests on the assumption that symmetric and transitive relations exist between product sales. Symmetry implies that purchase relations from product i_1 to product i_2 equal those from i_2 to i_1. The assumption of transitivity was introduced to process the data coming from more than two concurrent purchases, i.e. when a relation exists between $i_1 \Rightarrow i_2$ and between $i_2 \Rightarrow i_3$, then it is assumed that there also exists a relation between $i_1 \Rightarrow i_3$.

Item	i_1	i_2	i_3	i_4	i_5	i_6	Total
i_1	0	1	4	4	2	0	11
i_2	1	0	2	2	2	0	7
i_3	4	2	0	5	3	0	14
i_4	4	2	5	0	3	0	14
i_5	2	2	3	3	0	2	12
i_6	0	0	0	0	2	0	2
Total	11	7	14	14	12	2	60

Table 3.6: Matrix of association frequencies for product pairs

However, practical observations show that these assumptions are highly questionable. Now, since multiple purchases of products (for instance i_1, i_2, i_3 and i_4 are purchased together) are divided into two-way relations (i_1i_2, i_1i_3, i_1i_4, i_2i_3, i_2i_4 and i_3i_4), it can be shown that the number of two-way relations will increase in proportion to the number of products purchased together (m) with a factor $m*(m-1)/2$. Consequently, products with an equal purchasing frequency will be treated unequally if they arise from baskets that differ with respect to the number of products purchased. This can be seen from the tables 3.4 and 3.6, which show that products i_2 and i_6 are included in two purchases that differ in volume. The number of two-way relations adds to 7 for i_2 and 2 for i_6 (last row in table 3.6).

To correct for such unequal treatment, Böcker and Merkle suggest weighting all two-way relations with a factor $1/(m-1)$. The resulting matrix of association frequencies is depicted in table 3.7. In other words, frequency data are normalized in order to take into consideration the unequal total amount purchased for each product.

Item	I_1	i_2	I_3	i_4	i_5	i_6	Total
i_1	0	1/4	1+7/12	1+7/12	7/12	0	4
i_2	¼	0	7/12	7/12	7/12	0	2
i_3	1+7/12	7/12	0	1+11/12	1+11/12	0	5
i_4	1+7/12	7/12	1+11/12	0	11/12	0	5
i_5	7/12	7/12	11/12	11/12	0	2	5
i_6	0	0	0	0	2	0	2
Total	4	2	5	5	5	2	23

Table 3.7: Matrix of association frequencies using weighting factor $1/(m-1)$

Finally, the new association coefficient ($A_{i1,i2}$) for two items, say i_1 and i_2, is calculated from table 3.7 as follows, with the respective results shown in table 3.8:

$$A_{i1i2} = \frac{a}{\min\{b,c\}} \tag{3.5}$$

where a = the frequency of joint purchases of i_1 and i_2

b = the frequency of purchases of i_1

c = the frequency of purchases of i_2

Item	i_1	i_2	i_3	i_4	i_5	i_6
i_1	-					
i_2	0.125	-				
i_3	0.396	0.292	-			
i_4	0.396	0.292	0.383	-		
i_5	0.146	0.292	0.383	0.183	-	
i_6	0	0	0	0	1	-

Table 3.8: Association coefficients calculated from table 3.6

Unfortunately, data storage problems are enormous since calculating all association coefficients for some 15000 items in a supermarket requires the construction of a (15000 x 15000)-matrix! However, when calculated on the category level, association coefficients can provide useful input to construct strategic retail business units (RBU's) or merchandise lines based on in-store shopping patterns [67]. The idea, represented in figure 3.1, consists of grouping together into merchandise lines categories that are highly interdependent. Categories that are contained within the same merchandise lines can then be treated more closely together in promotional campaigns, pricing, display, and others.

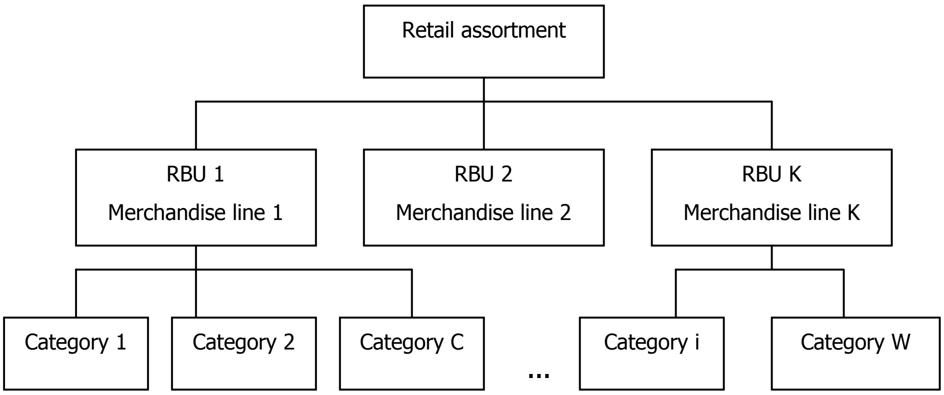


Figure 3.1: Structuring retail assortments into merchandise lines

Association coefficients are useful in this context since they can be considered as measures of similarity between product categories and when translating them into distances, a number of multivariate statistical techniques can be applied, including cluster analysis for finding groups of strongly associated product categories and multi-dimensional scaling for visualization of the association results into a two-dimensional map [125]. Both methods were applied on the 100 best selling product categories in our supermarket data. Hereto, the similarities were first transformed into distances by using the Minkowski distance metric since both methods require distances as input. Once the distances are computed, Ward's hierarchical clustering was applied on the transformed data, as illustrated by the dendogram below (figure 3.2). The dendogram shows which observations and/or clusters are merged during each interation of the clustering algorithm until all observations belong to the same cluster at the top of the graph.

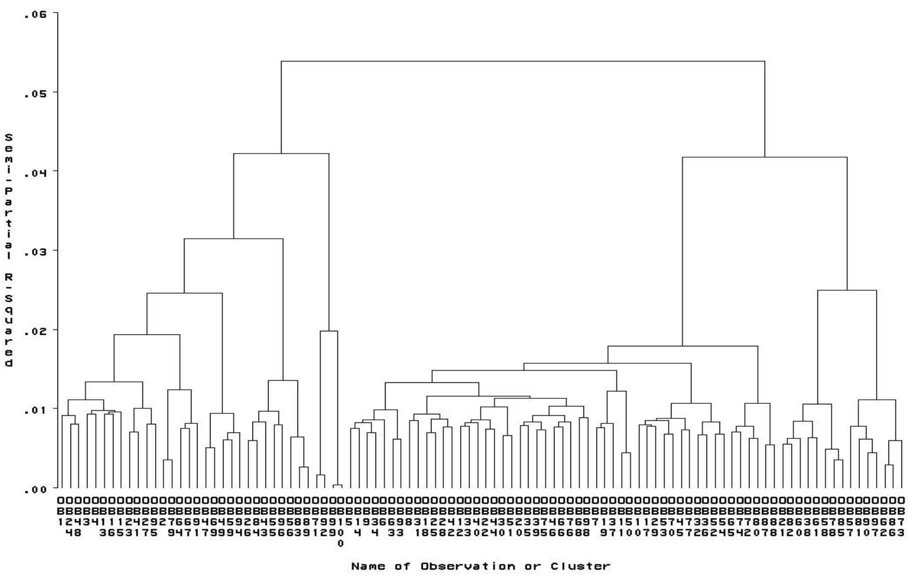


Figure 3.2: Cluster dendogram

From this dendogram, and dependent on where the dendogram is cut, the cluster membership for each product category can be generated. Table 3.9 shows the 10 cluster solution. It can be seen that indeed interesting clusters can be found, which contain product categories that tend to be logically related. For example, cluster 5 contains mainly maintenance products, whereas cluster 10 contains mainly personal healthcare product. Cluster 7, on the other hand, contains mainly frozen food products, whereas cluster 1 and 2 contain fresh food products. A similar pattern can be found in the results from multi-dimensional scaling. Note that figure 3.4 is an enlargement of figure 3.3 to obtain a better view of the cluster of cluttered points in the upper left corner of figure 3.3.

RBU	Product categories	Size
1	Fresh fruit & vegetables, fresh meat	2
2	Bread, fresh cheese, buns, bakeoff products, fresh sandwiches, pastry, pie & biscuit & cake	7
3	Nuts & appetizer biscuits, red wine, white wine	3
4	Soft drinks, waters, heavy beers, light beers	4
5	Maintenance products, toilet paper & kitchen roll, washing powder, dishwashing, maintenance tools, softener, abrasives, liquid detergent	8
6	Dry cookies, confectionery, chocolate, fresh cookies, candy bars, speculaas, gingerbread	7
7	Fresh filling, vegetables, frozen vegetables, frozen soups, sliced vegetables and fruit, ice-cream, frozen potato products, frozen meat products, frozen fish products, frozen ready made meals, frozen pizza	11
8	Margarine spread, baking margarine, yoghurt, whipped cheese, sauces, pasta, milk, canned vegetables, canned fish, flour products, crisps, sandwich filling, coffee, soups, eggs, desserts, grain products, sugar, fruit juices, canned fruit, cheese spread, filter & waste bags, baby food, mayonnaise, biscuit, hard cheese, broth, canned meat, spices, soft cheese, tea, stockings, butter, prepacked bread, rice, dry products, oils, cream, salads, prepacked meat, low calorie products	41
9	Paper ware, stationery, cigarettes, tobacco, newspaper & magazines, catfood, dogfood, electricity, candles, cutlery	10
10	Shampoo, shower gel, sanitary towels, brosserie, toothpaste, makeup, beauty products	7

Table 3.9: Merchandise lines generated by hierarchical clustering

The results of the multi-dimensional scaling analysis were obtained by treating the distances as interval variables. Typically, the *stress-value* is calculated as well to measure the badness-of-fit of the multi-dimensional scaling solution. The stress-value ranges from zero to one and basically computes the difference between the fitted distances on the graph with the real distances in the data. In the case where there is a high correlation between the real and the fitted distances, the stress value is low (close to zero) and the fit is good, otherwise the fit is bad. The stress value for the presented solution equals 0.0143 (after 3 iterations). This is significantly better than the stress value obtained for random numbers. Consequently, the 2-D multi-dimensional scaling graph can be considered as a good representation of the real distances in the data.

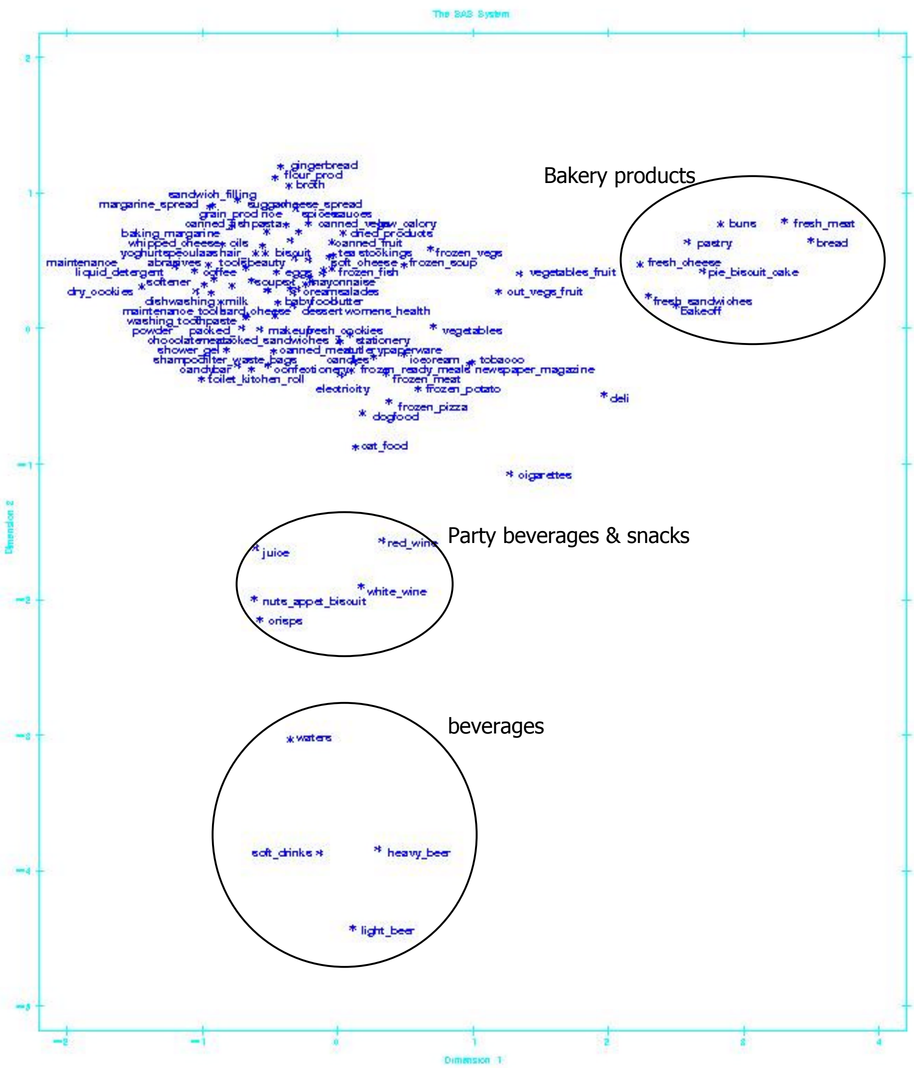


Figure 3.3: Multi-dimensional scaling results

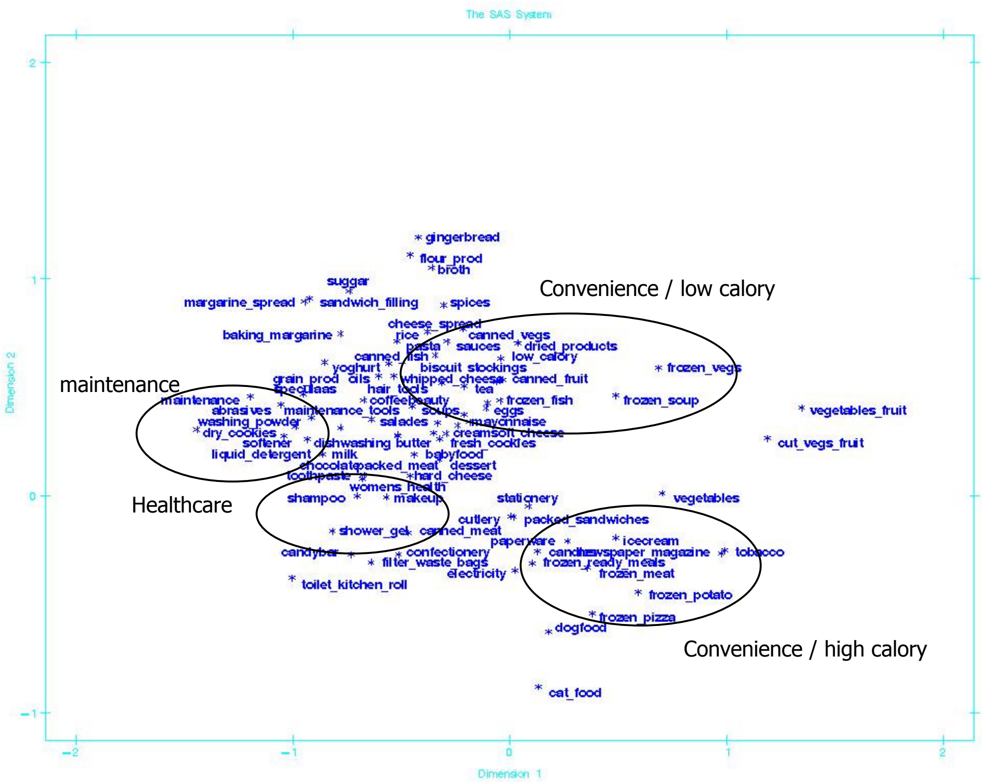


Figure 3.4: Enlargement of cluttered points in figure 3.3

At this point, it is worth mentioning that similar results were found when using the 'interest' measure from association rule (formula 4.1 in section 4.4.4) mining instead of using association coefficients. See appendix 2 for more details.

3.2.3.2 Loglinear analysis

Whereas association coefficients divide higher-order associations into two-way associations in order to calculate the level of interdependence between two products, loglinear analysis provides a sound statistical framework to directly examine higher-order associations.

The loglinear model is one of the specialized cases of generalized linear models for Poisson-distributed data [13, 160]. Intuitively, loglinear analysis can be considered as an extension of the two-way contingency table where the conditional relationship between two or more discrete categorical variables is analysed by taking the natural logarithm of the cell frequencies within the contingency table. Loglinear models are commonly used to evaluate multi-way contingency tables that involve three or more variables. As a result, loglinear models are very well suited to demonstrate association between variables.

The basic strategy in loglinear modelling involves fitting models to the observed frequencies in the cross-tabulation of categorical variables. The models can then be represented by a set of expected frequencies that may or may not resemble the observed frequencies. Models will vary in terms of the marginals they fit, and can be described in terms of the constraints they impose on the associations or interactions that are present in the data. Once expected frequencies are obtained, different models can be compared that are hierarchical to one another. The purpose is then to choose a preferred model, which is the most parsimonious model that fits the data. The choice of a preferred model is typically based on a formal comparison of goodness-of-fit statistics (likelihood ratio test) associated with models that are related hierarchically (i.e. models containing higher order terms also implicitly include all lower order terms).

For instance, for two categorical variables, each with two levels (2x2 table), the following model can be used to evaluate if an association exists between the variables:

$$ln(F_{ij}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \tag{3.6}$$

$ln(F_{ij})$ = is the log of the expected cell frequency of the cases for cell ij in the contingency table

μ = is the overall mean of the natural log of the expected frequencies

λ = represent 'effects' which the variables have on the cell frequencies

A and B = the variables
i and j = refer to the categories within the variables

Therefore:
λ_i^A = the main effect for variable A
λ_j^B = the main effect for variable B
λ_{ij}^{AB} = the interaction effect for variables A and B

The above model is called the *saturated* model because it includes all possible one-way and two-way effects. Given that the saturated model has the same amount of cells in the contingency table as it does have effects, the expected cell frequencies will always exactly match the observed frequencies, with no degrees of freedom remaining [160]. In order to find a more parsimonious model that will isolate the effects best demonstrating the data patterns, a *non-saturated* model must be discovered. This can be achieved by setting some of the effect parameters to zero. For instance, if the effects parameter λ_{ij}^{AB} is set to zero (i.e. we assume that variable A has no effect on variable B, or vice versa), the unsaturated model[16] is obtained:

$$ln(F_{ij}) = \mu + \lambda_i^A + \lambda_j^B \tag{3.7}$$

Moreover, it can be said that the models presented above are hierarchically related to each other, i.e. they are nested. In other words, the unsaturated model is nested within the saturated model.

Once the models have been fitted, it is necessary to decide which of the unsaturated models provides the best fit. As far as the models are nested within each other, this can be carried out using the likelihood ratio test. If F_{ij} represents the observed frequency and f_{ij} the fitted frequency, then the likelihood ratio test [13] is defined as:

[16] Note that for this model, the unsaturated model is analogous to the chi-square analysis, testing the hypothesis of independence between variables *A* and *B*.

$$LR = 2\sum_i \sum_j F_{ij} \log\left(\frac{F_{ij}}{f_{ij}}\right) \tag{3.8}$$

The *LR* test is distributed chi-square with degrees of freedom (df) equal to the number of cells minus the number of non-redundant parameters in the model. In other words, the df equals the number of λ parameters set equal to zero. The df value decreases as the model gets more complex, with the $df = 0$ for the saturated model. As a result, the LR tests the residual frequency not accounted for by the effects in the model (i.e. the λ parameters set equal to zero). Therefore, larger LR values indicate that the model does not fit the data well, and thus the model should be rejected. At this point, the LR test can be used to compare the saturated model with a (smaller) nested model:

$$LR_{difference} = LR_{nested} - LR_{overall} \tag{3.9}$$

The degrees of freedom (df) equal the df of the nested model minus the df of the saturated model. If the $LR_{difference}$ is not significant, it means that the more parsimonious nested model is not significantly worse than the saturated model. Then, one should choose the nested model since it is simpler.

For an illustration of loglinear analysis on a real data set, the reader is referred to section 7.5.2. Briefly, the example illustrates a split-half loglinear analysis on the four-way contingency table 7.1. The likelihood ratio test is used to show that the potentially complex four-way interaction between cakemix, frosting, fabric detergent and softener can be explained by two two-way interactions between cakemix and frosting, and between fabric detergent and softener. In fact, all other two-way, three-way and four-way interactions do not prove to be statistically significant to explain the interaction between the four given variables.

CHAPTER 4

ASSOCIATION RULES

Although the technique of association rules belongs to the category of measures of co-occurrence (see section 3.2.3), it will be treated here in a separate chapter. The reason is that this dissertation project was conceived as a study of retail market basket analysis with a special focus on data mining techniques. Therefore, the subsequent sections will provide a concise overview of the field of data mining, the technique of association rules and its applications in the context of retail market basket analysis. Association rules were first introduced in the context of knowledge discovery in databases and therefore we will first discuss its historical relevance within the data mining community (section 4.1). It is followed by an introduction into the basic concepts and definitions of association rules discovery (section 4.2). Furthermore, an efficient algorithmic implementation to discover all association rules will be provided (section 4.3). Subsequently, we discuss the problem of selecting the most interesting association rules. To this end, an overview of the literature will be given of different paradigms and post-processing methods in order to find the most interesting rules (section 4.4). Moreover, we contribute to the solution of this problem by presenting a novel optimization framework, named SetCover, to reduce the level of redundancy in a set of association rules (section 4.4.2.2). Finally, attention will be paid to some recent syntactic (section 4.5.1) and semantic (section 4.5.2) generalizations of association rules.

Parts of this chapter are based on work reported in [58, 59].

4.1 A Short History

Retail databases (and business databases in general) have grown tremendously in recent years, but the capabilities for analysing such large amounts of data have not developed at the same rate as the capabilities of collecting and storing these data. In 1996, Fayyad et al. [104] claimed that 'our capabilities of collecting and storing data of all kinds have far outpaced our abilities to analyse, summarize, and extract *knowledge* from this data'. As a result, organizations are becoming increasingly concerned with the study of knowledge discovery methods. In this context, the strong increase in computer power has led to new techniques (such as decision trees, artificial neural networks and association rules) for the analysis of large volumes of data. The area of research that is involved in the development of efficient algorithms to search for hidden, interesting and useful information in large amounts of data is known as Data Mining [104], or also Knowledge Discovery in Data (KDD).

More specifically in 1993, and particularly relevant in the context of this dissertation, Agrawal et al. [8] recognized a lack of functionality in the current database systems for users to take advantage of the huge amounts of retail transactional data. Therefore, they were the first to introduce the technique of association rules to mine a large collection of transactions for hidden patterns of consumer purchase behaviour. Their work was however quickly taken up by other researchers in the machine learning/data mining field who understood the applicability and importance of database mining in the retail industry. According to Park et al. [216] 'the analysis of past transaction data can provide very valuable information on customer buying behaviour, and thus improve the quality of business decisions (such as what to put up on sale, which merchandises to place on shelves together, how to customize marketing programs'. Houtsma and Swami [139] corroborated these statements by observing that 'the competitiveness of companies is becoming increasingly dependent on the quality of their decision-making. Hence, it is no wonder that companies often try to learn from past transactions and decisions in order to improve the quality of decisions taken in the present or future. In order to support this process, large amounts of data are collected and stored during business operations. Later these data are analysed for relevant information'.

Data mining researchers soon identified a number of important applications of association rule analysis in retailing. Agrawal and Srikant [12] claim that 'finding all such (association) rules is valuable for cross-marketing and attached mailing applications. Other applications include catalogue design, add-on sales, store layout, and customer segmentation based on buying patterns'.[17] In fact, since the pioneering work of Agrawal et al. [8], more than 100 papers have been published on the topic of association rules (especially on improving the efficiency of finding such rules) and it remains to be amongst the most popular topics at current data mining conferences and journals.

4.2 Definitions

In this paragraph, a number of definitions will be introduced that are needed to formally define the technique of association rules. Most definitions are drawn from the computer science literature [8], however, for many of them, a straightforward retail interpretation is provided.

Definition 4.1: A set of items ($\mathcal{I}$)

Let $\mathcal{I} = \{i_1, i_2, \ldots, i_m\}$ be a set of literals, called items. ■

Example: the product assortment in the retail store. For instance, the first data set in this study contains 16430 unique SKU's.

[17] See chapter 6 and 7 for our contribution to the field of segmentation based on buying patterns.

Definition 4.2: A set of transactions ($\mathcal{D}$)

Let $\mathcal{D}$ be a set of transactions, where each transaction T is a set of items such that $T \subseteq \mathcal{I}$. ■

Example: The data set ($\mathcal{D}$) used in this study contains 88163 receipts where each receipt contains a set of items that are purchased by a particular consumer during a particular shopping visit.

Definition 4.3: An itemset (X)

We say that a transaction T *contains* X, a set of some items in $\mathcal{I}$, if $X \subseteq T$. ■

Example: an itemset is a set of items, such as {diapers, crisps, beer}.

Definition 4.4: An association rule

An association rule is an implication of the form $X \Rightarrow Y$, where X and Y are itemsets (i.e., $X \subset \mathcal{I}$, $Y \subset \mathcal{I}$), and $X \cap Y = \varnothing$. ■

Example: an association rule expresses which products tend to be purchased together during a single shopping trip. For instance, the rule *diapers* $\Rightarrow$ *beer* indicates that people who buy diapers also tend to buy beer. For reasons of interpretation, the consequent of the rule usually contains just a single item, although more items are technically possible.

Definition 4.5: Support of an association rule

The rule $X \Rightarrow Y$ has support *s* in $\mathcal{D}$ if $s\%$ of the transactions in $\mathcal{D}$ contains $X \cup Y$. ■

Example: the support expresses the fraction of transactions in $\mathcal{D}$ that contain both the items in the antecedent and the consequent of the rule. Thus, a support of 20% means that the items in the antecedent and the consequent of the rule occur together in 20% of the total volume of transactions.

Definition 4.6: Confidence of an association rule
The rule $X \Rightarrow Y$ holds in the transaction set $\mathcal{D}$ with confidence c if $c\%$ of transactions in $\mathcal{D}$ that contain X also contains Y. ■

Example: in statistical terms, the confidence is an estimator for the conditional probability of Y given X, i.e. $\mathrm{P}(Y \mid X)$ and it can be calculated as $s(X \cup Y) / s(X)$.

Objective: given a set of transactions, the standard problem of mining association rules is therefore to generate all association rules that have support and confidence greater than the user-specified minimum support (*minsup*) and minimum confidence (*minconf*) thresholds. Section 4.3 will deal with the computational details of the algorithm to discover such associations.

4.3 Algorithms for Association Rule Discovery

A number of algorithms have been developed to discover association rules in receipt data and it still remains to be a highly researched topic in the computer science literature. Briefly, the algorithms differ in how they search the space of possible associations. We will illustrate why this is important by discussing two approaches to the association rules discovery problem: the naive and the intelligent approach. However, before going in to that, we will first introduce the general two-phase methodology that is used by almost all the algorithms that discover association rules.

4.3.1 Two-Phase Methodology

Discovering association rules typically involves the execution of two (sequential) phases.

4.3.1.1 Finding frequent itemsets

The first phase involves looking for so-called *frequent itemsets*, i.e. itemsets for which the support in the database equals or exceeds the minimum support threshold set by the user (definition 4.5). This is computationally the most complex phase because of the number of possible combinations of items that need to be tested for their support.

4.3.1.2 Generating association rules

As soon as all frequent itemsets are known, the discovery of association rules is relatively straightforward. The general idea is that if, say, $ABCD$ and AB are frequent itemsets, then it can be calculated whether the rule $AB \Rightarrow CD$ holds with sufficient confidence by computing the ratio $confidence = s(ABCD) / s(\mathrm{AB})$. If the confidence of the rule equals or exceeds the $minconf$ threshold set by the user, then it is a valid rule. Most research [8, 12, 61, 184, 216, 298] has focussed on the first phase, as this is computationally the most complex phase. The rule generation phase is less complex. For an itemset of size k, there are potentially 2^k-2 confident rules. To illustrate this, two approaches to the discovery of frequent itemsets will be presented: the naive and the intelligent approach.

4.3.2 The Naive Approach

The naive approach to finding all frequent itemsets is exponentially complex. More specifically, for n items, testing all possible combinations for their support involves calculating $2^n - 1$ frequencies. To illustrate this, consider the following small example. Suppose we have a database D containing 4 items $\{i_1, i_2, i_3, i_4\}$, then finding all frequent itemsets involves checking all subsequent combinations:

<u>1-itemsets</u>	<u>2-itemsets</u>	<u>3-itemsets</u>	<u>4-itemsets</u>
$\{i_1\}$	$\{i_1, i_2\}$	$\{i_1, i_2, i_3\}$	$\{i_1, i_2, i_3, i_4\}$
$\{i_2\}$	$\{i_1, i_3\}$	$\{i_1, i_2, i_4\}$	
$\{i_3\}$	$\{i_1, i_4\}$	$\{i_1, i_3, i_4\}$	
$\{i_4\}$	$\{i_2, i_3\}$	$\{i_2, i_3, i_4\}$	
	$\{i_2, i_4\}$		
	$\{i_3, i_4\}$		

It is clear that the naive approach is computationally infeasible already for relatively small numbers of items. Especially in super- or hypermarkets, where product assortments can easily contain (tens of) thousands of items, this approach is clearly not feasible. Therefore, more clever techniques have been developed that allow for the computation of support within a reasonable amount of time. One of them is the *Apriori* algorithm presented in the next section.

4.3.3 The Apriori Algorithm

A more intelligent approach for finding all frequent itemsets is the *Apriori* algorithm [12]. It is a breadth-first search algorithm based on the downward closure principle that states that 'all subsets of a frequent itemset must also be frequent'. In other words, if at least one subset of an itemset is not frequent, the itemset can never be frequent anymore. This principle simplifies the discovery of frequent itemsets considerably because for some itemsets, it can be *apriori* determined that they can never be frequent such that their support does not have to be checked against the data anymore. Figure 4.1 illustrates the *Apriori* algorithm in detail (drawn from [12]).

L_1 := {frequent 1-itemsets};

k := 2; // represents the pass number

while ($L_{k-1} \neq \emptyset$) **do begin**

C_k := New candidates of size k generated from L_{k-1};

for all transactions $T \in \mathcal{D}$ **do begin**

Increment the count of all candidates in C_k that are contained in T.

end

L_k := All candidates in C_k with minimum support.

$k := k + 1$;

end

Answer := $\bigcup_k L_k$;

Figure 4.1: Apriori algorithm

The first pass of the algorithm simply counts item occurrence to determine the frequent 1-itemsets, i.e. itemsets containing just one item. A subsequent pass, say pass k, consists of two phases. First, the frequent itemsets L_{k-1} found in the $(k\text{-}1)$th pass are used to generate the candidate itemsets C_k. To generate these candidate itemsets, the Apriori candidate generation function, described in figure 4.2 below, is adopted. Next, the database $\mathcal{D}$ is scanned and the support of candidates in C_k is verified against the data. These two operations (candidate generation and support counting) continue until, according to the downward closure principle, no candidate itemsets can be generated anymore. The outcome of the algorithm is guaranteed to include *all* frequent itemsets.

We will now focus on the Apriori candidate generation function. In fact, this function consists of two steps: a *join* step and a *prune* step. Given L_{k-1}, the set of all frequent $(k - 1)$-itemsets, the algorithm returns a superset of the set of all frequent k-itemsets.

The *join* step goes as follows:

insert into C_k
select $p.item_1$, $p.item_2$, ... , $p.item_{k-1}$, $q.item_{k-1}$
from L_{k-1} p , L_{k-1} q
where $p.item_1 = q.item_1$, ... , $p.item_{k-2} = q.item_{k-2}$, $p.item_{k-1} < q.item_{k-1}$;

Next, in the *prune* step, all itemsets $c \in C_k$ are deleted for which some $(k\text{-}1)$-subset of c is not in L_{k-1} :

for all itemsets $c \in C_k$ **do**
  for all $(k$ -1)-subsets s of c **do**
    if $(s \notin L_{k-1})$ **then**
      delete c from C_k ;

Figure 4.2: Apriori candidate generation function

Example
Let L_3 be $\{\{1\ 2\ 3\}, \{1\ 2\ 4\}, \{1\ 3\ 4\}, \{1\ 3\ 5\}, \{2\ 3\ 4\}\}$. After the join step, C_4 will contain $\{\{1\ 2\ 3\ 4\}, \{1\ 3\ 4\ 5\}\}$. However, in the prune step, the itemset $\{1\ 3\ 4\ 5\}$ will be deleted because the itemset $\{1\ 4\ 5\}$ is not in L_3. Consequently, C_4 will only contain the candidate itemset $\{1\ 2\ 3\ 4\}$.

4.3.4 Other Algorithms for Discovering Association Rules

Most of the research in the field of association rules has focussed on the discovery of more efficient algorithms to discover association rules. In fact, almost all algorithms that have been developed after Apriori adopt the levelwise candidate generation and pruning principle to generate frequent itemsets. However, they mostly differ with respect to how they generate and count candidate itemsets. For instance, DHP [216] uses a hashing scheme to collect upper bounds on the frequencies of the candidate itemsets for the following pass. Other approaches aim at reducing the number of passes over the database [61, 235, 266]. More recent approaches adopt a depth-first search, which no longer uses the downward closure principle [5, 129]. Finally, there is increased interest in adopting constraints to further reduce the number of candidate patterns as much as possible [69, 109].

It must be said, however, that most commercial software packages still rely on the original Apriori algorithm to discover association rules. Therefore, although a lot of progress has been achieved in computational efficiency in the academic community, the business user will probably not benefit from it.

4.4 Post-Processing of Association Rules

Probably the most important problem with association rules, which so far remains largely unsolved is the 'interestingness' of association rules. Indeed, the main strength of association rule mining is that, since it discovers all association rules that exist in a database, it can reveal valuable and unexpected information. These strengths, however, are also its weakness, i.e. the number of discovered rules can be huge, hundreds or even thousands of rules, which makes manual inspection of those rules practically infeasible. In other words, association rule results sometimes create a new data mining problem of the second order. This makes post-processing of these rules very important, i.e. we need good methods to reduce the number of association rules to the most interesting ones. The reasons for this problem of interestingness can be found in the limitations of the support-confidence framework, adopted by almost all association rule algorithms (see section 4.4.1). Solutions to the problem of interestingness have been approached from different angles, i.e. by reducing the number of rules through redundancy reduction (section 4.4.2), rule clustering (section 4.4.3), or by assessing the quality of individual rules through objective (section 4.4.4) and subjective (section 4.4.5) measures of interestingness. In section 4.4.2.2, we also make a contribution to this discussion by presenting a novel post-processing optimization model, named 'SetCover', to select the least redundant set of association rules.

4.4.1 *Limitations of the Support-Confidence Framework*

Soon after the first implementations of association rules came the insight that the traditional *support – confidence* framework for association rules lacks some theoretical and practical foundations.

First of all, setting good values for the support and confidence parameters in association rule mining is not trivial. For instance, setting the threshold value of the support parameter too low causes the association rules algorithm to produce enormous amounts of rules of which many will be overfitting the data. On the other hand, setting the support threshold too high increases the probability of finding trivial relations and of missing some important associations between products. For example, there is a big chance of missing some interesting associations between products with different purchase cycles, e.g., Camembert cheese with a high support, since it is purchased on a weekly basis, and Bordeaux red wine with a much lower support, since it is purchased less frequently. Yet, the combination of red wine and Camembert cheese may present an important association. Essentially, this observation advocates the use of different support thresholds for different product classes.

With regard to the confidence threshold, researchers have argued that confidence is not a good parameter to discover the most valuable or interesting rules. Indeed, confidence is heavily affected by the apriori frequency of the consequent of the association rule and does not correct for this bias. In other words, association rules that have a frequently purchased product in the consequent of the rule (e.g. Coca Cola) will, in general, have much higher confidence values than rules with infrequently purchased products (like cheese spread or pepper sauce) in the consequent since the prior probability of observing soft drink in a basket will be much higher than observing cheese spread or pepper sauce. As a result, high confidence rules will often be trivial rules with frequently purchased products in the consequent of the rule.

Furthermore, not all rules with high support and confidence are interesting. Some of the discovered rules correspond to prior knowledge or expectations, refer to uninteresting attributes or attribute combinations, or simply represent redundant information. Many researchers have therefore suggested alternative approaches to the interestingness problem, which can be classified into association rule reduction techniques, rule clustering and techniques that measure individual rule quality in terms of objective and subjective measures of interestingness.

4.4.2 Association Rule Reduction

One approach to solving the interestingness problem of association rules is to reduce the overload of information by filtering out redundant rules.

4.4.2.1 Removing redundancy: The RuleCover heuristic

One way to deal with the rule quantity problem is to remove the redundancy in association rules. Indeed, since association rules are a form of unsupervised learning, the discovered rules are not mutually exclusive and collectively exhaustive, unlike for instance decision rules in supervised learning. This means that a particular market basket can be covered[18] by multiple rules and that some baskets may not be covered by any rule. This is illustrated by table 4.1. The matrix shows whether a particular basket (TID_i) is covered by a particular association rule (R_j) or not. Formally, we define:

$$s_{i,j} = \begin{cases} 1 & \text{if basket } i \text{ is covered by rule } j \\ 0 & \text{if basket } i \text{ is not covered by rule } j \end{cases}$$

[18] By 'covered' we mean that the items that appear in the rule also appear in the basket.

TID	R_1	R_2	R_j	R_K
1	1	0		1
2	1	1		1
3	0	0		0
4	0	0		1
5	1	0		0
…				
i	…	…		…
…				
N	0	1		0

Table 4.1: Market baskets covered by association rules

From table 4.1, it is clear that the discovered set of association rules will contain some level of redundancy. This can be handled by pruning or summarizing the discovered rules. In this context, Toivonen et al. [267] proposed the concept of 'rule covers' to prune away redundant association rules. The Rulecover heuristic is shown in pseudocode in figure 4.3.

Input: Set of rules $\Gamma = \{X_i \Rightarrow Y \mid i = 1, ..., n\}$.
Sets of matched rows $m(X_iY)$ for all $i \in \{1, ..., n\}$.
Output: Rule cover Δ.
Method:

```
Δ := ∅;                      // rule cover
s' := ⋃_{i=1}^{n} m(X_i Y_i);  // rows unmatched by cover
for all i ∈ {1, ..., n} do
    s_i = m(X_iY);           // rows of s′ matched by rule i
end;
while s′ ≠ ∅ do
    choose i ∈ {1, ..., n} so that (X_i ⇒ Y) ∈ Γ and | s_i | is largest;
    Δ := Δ ∪ {X_i ⇒ Y};      // add the rule to the cover
    Γ := Γ \ (X_i ⇒ Y);      // remove the rule from the original set
    for all (X_j ⇒ Y) ∈ Γ do
        s_j = s_j \ m(X_iY);  // remove matched rows
        end;
        s′ := s′ \ m(X_iY);   // remove matched rows
end;
```

Figure 4.3: Rulecover heuristic for association rule selection

The Rulecover heuristic is basically a greedy mechanism that uses an original set Γ (containing the entire set of association rules) and then iteratively selects a rule $X_i \Rightarrow Y$ to move it into Δ. In each pass, the rule is selected which covers the maximum number of instances that are left over after having deleted the instances that were covered by the rule that was selected during the previous pass. This process continues until no instances or rules are left over. At the end, Δ is then said to contain the minimum rule cover of Γ.

4.4.2.2 Removing redundancy: The SetCover optimization model

The use of heuristic procedures to solve the redundancy problem inspired us to propose an optimal solution to this problem by using an integer-programming model based on the set covering principle [59]. An important argument for the optimal approach is that the selection of rules for the final ruleset is independent of any ordering of the rules, in contrast to the RuleCover heuristic, where the stepwise selection of a subsequent rule is dependent on which rules have been chosen during the previous steps. Consequently, because of the adoption of heuristic selection criteria, there is a reasonable chance that some of the previously selected rules are not optimal from an overall perspective. An integer programming approach, however, always results in the most optimal selection irrespective of the ordering of the rules, because it adopts a simultaneous selection of rules instead of a stepwise selection. To illustrate this, let i (i=$1..n$) be the index for the baskets and let j (j=$1..K$) be the index for the association rules, and x_j is a Boolean decision variable indicating whether rule j is selected or not, i.e. $x_j \in \{0,1\}$. Then, the Integer Programming (IP) model depicted in figure 4.1 will select the minimum set of association rules covering all initially covered instances.

The target function specifies that the model must select as few association rules as possible. This means that the model searches for association rules that are as far apart as possible in the space of market baskets. The first constraint in the model ensures that the originally covered basket space is not reduced such that the selected ruleset will still cover the same set of baskets that were covered by the original ruleset.

$$Min\left(\sum_{j=1}^{K} x_j\right)$$

s.t.

$$\forall i(i = 1 \rightarrow n) : \textstyle\sum_{j=1}^{K} x_j * s_{i,j} \geq 1$$

$$\forall x_j \in \{0,1\}$$

Figure 4.4: Setcover optimization model for selecting association rules

If this constraint would be neglected, the model would select no rules at all since the objective function forces the model to select as few rules as possible.

An empirical comparison of Rulecover and Setcover on two real datasets turned out that both the methods are able to significantly reduce the total number of rules (from 50-63%). However, when comparing the heuristic Rulecover approach with the optimal Setcover approach, it turned out that for one dataset Setcover produced significantly better results than Rulecover (10% less rules) whereas for the other dataset, the difference was only minor (1%). Consequently, Rulecover does quite a good job, although Setcover sometimes produces even better results depending on the kind of dataset.

Another way of dealing with redundancy was proposed by Liu et al. [177] as rule summarization. The idea is to represent the essential underlying relationships in the data by means of direction setting (DS) rules. More specifically, DS rules provide summarizations of other more specific rules where the direction of correlation of those specific rules can be derived from the direct setting rules. In other words, given the DS rules, all other rules (i.e. the non-DS rules) are not surprising anymore. Limiting the discovery process of association rules to just the DS rules therefore significantly reduces the number of rules.

Finally, Zaki [297] proposes the concept of 'closed' frequent itemsets to solve the redundancy problem in association rules. The idea is that the set of closed frequent itemsets is much smaller but uniquely determines the set of all frequent itemsets.

4.4.3 Association Rule Clustering

Another way of tackling the interestingness problem in association rules is to cluster the discovered rules. Basically, two approaches can be found in the literature.

The first approach [171] aims at finding clustered association rules by combining similar, adjacent association rules. Instead of using attribute equalities to form rules such as $(\text{age}=40)\Rightarrow(\text{own\_home}=\text{yes})$, clustered association rules have value ranges using inequalities $(40\leq\text{age}<42)\Rightarrow(\text{own\_home}=\text{yes})$. The idea is to form the latter rule from the association rules $(\text{age}=40)\Rightarrow(\text{own\_home}=\text{yes})$ and $(\text{age}=41)\Rightarrow(\text{own\_home}=\text{yes})$. The problem, however, is how to efficiently find such clustered association rules. In their paper, Lent et al. [171] provide a geometric-based algorithm for two-dimensional association rules. The granularity of the grid determines the amount of observations in each cell of the grid. Subsequently, clusters of observations are generated in the grid that, in turn, will determine the interval partitions of the attribute values. Once these interval partitions are found, the corresponding clustered association rules can be generated.

The second approach to clustering association rules is based on hypergraph partitioning [126, 127]. Hypergraph partitioning methods were originally used in the semi-conductor industry for VLSI (very large scale integration) design, i.e. to organize a set of components onto integrated electronic circuits. The idea is to group items by means of a hypergraph partitioning algorithm, such that items in the same cluster have many connections with other items inside the cluster and few connections with items outside the cluster. Therefore, hyperedges are constructed from the frequent itemsets since they reflect the multi-way relationship between items. A hypergraph $H=(V,E)$ therefore consists of a set of vertices (V) and a set of hyperedges (E). In the case of association rules, the vertex set corresponds to the distinct items in the database and the hyperedges correspond to the relationship between the items as determined by the frequent itemset in which they appear. For instance, the itemsets $\{A, B, C\}$, $\{B, D, E\}$ and $\{D, E, F\}$ can be represented by hyperedges as shown in figure 4.5.

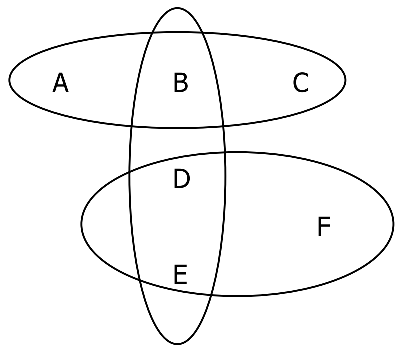


Figure 4.5: Association rule hypergraph

Furthermore, hyperedges usually contain weights representing the strength of the relationship between the vertices in the respective hyperedge. Han et al. [126] suggest using the average of the confidence of all the rules that can be constructed by the hyperedge as the weight for the hyperedge.

Next, a hypergraph-partitioning algorithm is used to partition the hypergraph such that the weight of the hyperedges that are cut by the partitioning is minimized. In other words, the number of relations that are violated by partitioning the items into different clusters is minimized. The weights are used to influence the clustering process to achieve this goal. The clustering itself is carried out by a clever graph partitioning heuristic. The criteria used to evaluate the quality of the clustering solution is based on the within and between cluster connections. Although the authors report several successful applications outside the traditional retail market basket analysis field, experiments with the hypergraph partitioning method were, however, not very successful on our retail data. More specifically, we witnessed the following problems.

Firstly, retail basket data are often characterized by a small number of frequently purchased products (such as soft drinks) that occur in many of the frequent itemsets. Consequently, these items appear in many of the hyperedges, which makes clustering difficult since any partitioning of the data will result in a lot of hyperedges that cross multiple partitions. Obviously, removing such frequent items from the analysis could solve this problem. However, still then the clustering produced bad results.

Secondly, when calculating the connectivity for each cluster solution (i.e. in how many of the hyperedges, inside the cluster, the item appears relative to the total number of hyperedges, inside + outside, in which the item appears), the results showed that most of the vertices/items are not significant within the cluster (with connectivity>0.1). Therefore, the clustering solution eventually boils down to clusters with only very few (one or two) significant vertices.

Thirdly, we experimented with different definitions for the weight of a hyperedge. The idea of the weight is to reflect the strength of the mutual connection. However, it is our opinion that the average of the confidences of the rules that can be generated from the hyperedge, as suggested by the authors of the method, does not reflect the strength of the connection very well. This is easy to illustrate and it is again related to our first remark, namely the effect of some very frequent items on the confidence of the resulting rules. For instance, suppose we have $\{A, B, C\}$ and let C be much more frequent than A or B. Then, the rules $\{A,B\} \Rightarrow \{C\}$ and $\{A,C\} \Rightarrow \{B\}$ will have very different confidences. The confidence of the first rule will be very big, whereas the confidence of the second rule will be very small. Taking the average of both to calculate the weight for $\{A, B, C\}$ will obviously lead to a moderate confidence level. Moreover, this will be the case for most frequent itemsets involving items with strongly different supports. Consequently, experiments on market basket data showed very similar weights for most of the hyperedges such that the weights were not able to influence the cluster solution. Therefore, alternative weight definitions were adopted to obtain better results. More specifically, we used 'interest' (section 4.6.4) since it better reflects the dependency between the items in the hyperedge and since it is a symmetric measure, i.e., it does not depend on the association rules that can be constructed from the hyperedge. Unfortunately, the results were not significantly improved by adopting 'interest' instead of average confidence.

Finally, the authors of the hypergraph partitioning method do not provide good indications of the multitude of parameter settings that can be set to run the clustering method. Therefore, different clustering solutions with different parameter settings were implemented in batch and compared to discover the best parameter settings for our data. Unfortunately, the results did not provide good indications for optimal values of the parameter settings. Furthermore, the user needs to specify the number of clusters and there is no guidance on how to determine whether a *k*-cluster solution is better or worse than the *k+1*-cluster solution.

4.4.4 Objective Rule Interestingness Measures

Objective measures of interestingness are based on the statistical properties of association rules. Amongst others [137], the most well known are interest [7, 60], correlation [60, 206] and intensity of implication [118]. We will illustrate them by using the following real example shown in table 4.2. The association rules algorithm produces the following results on data set 1, containing 88163 receipts:

Itemset	**Support count**	**Support (%)**
{orange juice}	579	0.6573 %
{semi-skimmed milk}	2885	3.273 %
{orange juice, semi-skimmed milk}	140	0.1584 %

Table 4.2: Association rule results

The *interest*, or also called the lift measure, measures the statistical dependence of a product association by relating the observed frequency of co-occurrence $s(A \cup C)$ of the antecedent (A) and the consequent (C) of the rule against the expected frequency of co-occurrence under the assumption of conditional independence of A and C. Interest is therefore defined as:

$$I(A \Rightarrow C) = \frac{s(A \cup C)}{s(A) * s(C)} \tag{4.1}$$

Note that interest is a symmetric measure. An interest value equal to 1 indicates that the observed frequency of the rule in the data (nominator) equals the expected frequency (denominator), given the assumption of conditional independence between the antecedent (A) and the consequent (C) of the rule. An interest value larger than 1 indicates that the combination of A and C occurs more frequently in the data (i.e. positive interdependence) than we would expect. An interest value smaller than 1 indicates less than expected co-occurrence, or thus a negative interdependence. Applied on the example given in table 4.2, the interest measure of the rule orange juice⇒semi-skimmed milk equals

$$I(\text{orange juice} \Rightarrow \text{semi-skimmed milk}) = \frac{0.001584}{0.006573 * 0.03273} = 7.363$$

This is a fairly high value and thus demonstrates highly positive interdependence between orange juice and semi-skimmed milk. The interest measure has therefore been used as guidance for retailers to identify complementarity and substitution effects between products [25]. From a marketing modeller's perspective, this may however not be entirely accurate. They usually define complementarity/substitution in terms of the effect on the sales of a particular product as a result of a marketing action on another product (see figure 3.1). The interest therefore really only measures higher or lower than expected co-occurrence instead of complementarity or substitution. Our own analysis on real data showed that when looking at the association rules with high interest values, the involved products are typically usage complements or result from variety-seeking behaviour by the customer.

Another objective measure of interestingness for association rules is based on the statistical notion of *correlation* between the items in the antecedent and the consequent of the rule [60, 177, 206]. The idea is to construct a contingency table from the association rule results and test the interdependence between the antecedent and the consequent of the rule by means of chi-squared analysis. This is demonstrated below for the example in table 4.3.

	semi-skimmed milk	¬ semi-skimmed milk	Totals
orange juice	140	439	579
¬ orange juice	2745	84839	87584
Totals	2885	85278	88163

Table 4.3: contingency table for orange juice and semi-skimmed milk

A chi-squared analysis on this contingency table produces the following result, with i representing the row index and j the column index:

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{4.2}$$

thus,

$$\chi^2 = \frac{(140 - 579 \times 2885 / 88163)^2}{579 \times 2885 / 88163} + \frac{(439 - 579 \times 85278 / 88163)^2}{579 \times 85278 / 88163} +$$

$$\frac{(2745 - 87584 \times 2885 / 88163)^2}{87584 \times 2885 / 88163} + \frac{(84839 - 87584 \times 85278 / 88163)^2}{87584 \times 85278 / 88163}$$

$$= 804.87 >> 3.84$$

For a p-value of 0.05 with one degree of freedom, the cut-off value is 3.84. Consequently, semi-skimmed milk and orange juice are significantly interdependent at the (1-0.05) 95% confidence interval. However, some important comments can be made on these results.

First of all, it can be noticed that there exists a relation between the chi-squared test for statistical interdependence and the interest value (see formula 4.1). Indeed, the chi-squared distance for the (1,1) cell in the contingency table corresponds closely to the interest measure. In fact, the larger the interest value deviates from 1, the bigger its contribution to the chi-squared statistic. This is fairly easy to proof, changing notation from interest to chi-squared:

$$I(i \Rightarrow j) = \frac{s(i,j)}{s(i) \times s(j)} = \frac{O_{ij}/n}{\frac{O_i}{n} \times \frac{O_j}{n}} = \frac{O_{ij}}{\frac{O_i \times O_j}{n}} = \frac{O_{ij}}{E_{ij}} \qquad (4.3)$$

Now, the rule that maximizes the deviation of the interest from 1 $\left|\frac{O_{ij}}{E_{ij}} - 1\right|$ also maximizes $\left|O_{ij} - E_{ij}\right| / E_{ij}$ and thus also $\left(O_{ij} - E_{ij}\right)^2 / E_{ij}$ which is the chi-squared distance of that cell in the contingency table. In other words, rules with strong negative or positive interdependence as measured by the interest value contribute strongly to the chi-squared statistic. The only difference is that the chi-squared statistic measures the overall interdependence for a set of variables (thus over the entire contingency table), whereas the interest measures the interdependence between a set of events of those variables.

Secondly, the chi-squared test rests on the normal approximation to the binomial distribution. This approximation breaks down when the expected values are small. Moore [205] therefore suggests to use the chi-squared test only when all cells in the contingency table have expected values greater than 1, and at least 80% of the cells have expected values greater than 5. In a real case scenario, however, these requirements will be easily broken. One way to avoid this problem is to set the minimum support threshold high enough, or to use an exact calculation of the probability instead of the chi-squared approximation. The latter, however, turns out to be prohibitively expensive [60].

Finally, it is tempting to use the value of the chi-squared statistic as an indication of the degree of dependence. However, an important limitation of the chi-squared statistic is that it tends to produce larger values when the data set size tends to grow to infinity. While comparison of chi-squared values within the same data set may be meaningful, it is therefore not advisable to compare chi-squared values across different data sets.

Last but not least, *intensity of implication* [118, 256] is also worth mentioning within the context of statistical measures of interestingness of association rules. The idea is to measure the statistical surprise of having so few negative examples on a rule as compared with a random draw. Consider a database $\mathcal{D}$, where $|\mathcal{D}|$ is the number of transactions in the database, and an association rule $X \Rightarrow Y$. Now, let U and V be two sets randomly chosen from $\mathcal{D}$ with the same cardinality, i.e., $s(X)=s(U)$ and $s(Y)=s(V)$, and let $\neg Y$ mean 'not Y' as shown in figure 4.6.

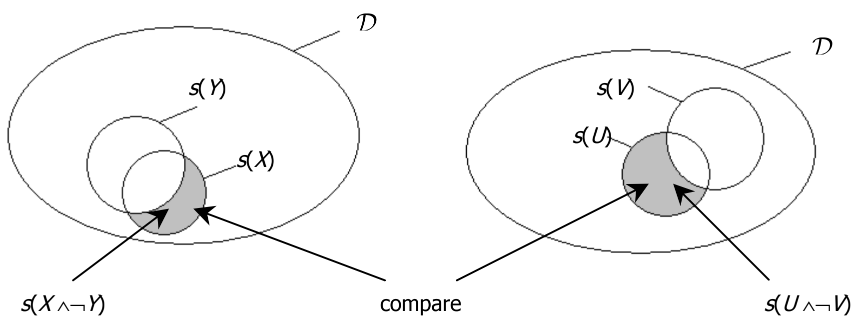


Figure 4.6: Intensity of implication

Let $s(U \wedge \neg V)$ be the random variable that measures the expected number of random negative examples under the assumption that U and V are independent, and $s(X \wedge \neg Y)$ the number of negative examples observed on the rule. Now, if $s(X \wedge\neg Y)$ is unusually small compared with $s(U \wedge\neg V)$, the one we would expect at random, then we say that the rule $X \Rightarrow Y$ has a strong statistical implication. In other words, the intensity of implication for a rule $X \Rightarrow Y$ is stronger, if the quantity $\mathrm{P}[s(U \wedge\neg V) \leq s(X \wedge\neg Y)]$ is smaller. Intensity of implication is then defined as $1 - \mathrm{P}[s(U \wedge\neg V) \leq s(X \wedge\neg Y)]$. The random variable $s(U \wedge\neg V)$ follows the hypergeometric law and therefore the intensity of implication can be written as:

$$1-\sum_{k=\max(0,s(X)-s(Y))}^{s(X\wedge\neg Y)} \frac{C_{s(\neg X)}^{s(\neg Y)-k} \times C_{s(X)}^{k}}{C_{|\mathcal{D}|}^{s(\neg Y)}} \tag{4.4}$$

This formula for intensity of implication is suitable as long as the number of cases in the database, i.e. $|\mathcal{D}|$, is reasonably small. Otherwise, the combination numbers in the above formula explode very quickly. Therefore, Suzuki et al. [256] came up with an approximation of this formula for big datasets. They argue that if $s(U \wedge\neg V)$ is small, which is often the case in rule discovery, then Poisson approximations can be applied. In that case, the above formula for intensity of implication reduces to a much simpler version that is easier to compute:

$$\begin{aligned} &1-\sum_{k=\max(0,s(X)-s(Y))}^{s(X\wedge\neg Y)} \frac{C_{s(\neg X)}^{s(\neg Y)-k} \times C_{s(X)}^{k}}{C_{|\mathcal{D}|}^{s(\neg Y)}} \\ &\approx 1-\sum_{k=0}^{s(X\wedge\neg Y)} \frac{\lambda^{k}}{k!} e^{-\lambda} \\ &\text{with } \lambda \equiv \frac{s(X)\times(|\mathcal{D}|-s(Y))}{|\mathcal{D}|} \end{aligned} \tag{4.5}$$

Nevertheless, the computational burden is still quite high since for every rule, the calculation involves the summation over a relatively large number of calculations.

4.4.5 *Subjective Rule Interestingness Measures*

It has been noted [220] that objective measures of interestingness, although useful in many respects, usually do not capture all the complexities of the pattern discovery process, and that subjective measures are sometimes needed to define the interestingness of a pattern.

One approach to tackle the subjective interestingness problem is based on the specification of constraints that allow the user to put conditions on the associations to be generated. These constraints mostly specify if particular attributes, or attribute classes, should appear in the antecedent or consequent of the rule. The rule is then interesting if it matches the user-provided constraint(s). A further distinction can, however, be made according to whether these constraints are used during the rule discovery process, or as a post-processing step after rule induction. The interested reader is referred to Klemettinen's rule templates [159] for the post-processing approach. Examples where the constraints are immediately taken into account in the rule discovery process itself include 'constrained association queries' by Ng et al. [210] and 'Dense-miner' by Bayardo et al. [27], 'Direct' by Srikant et al. [255] and 'integrated and online filtering' [113].

Finally, *actionability* and *unexpectedness* [248] were also proposed as subjective measure of interestingness for association rules. Actionability refers to the extent to which a particular rule can be acted on to the user's advantage. Actionability in that context constitutes an important measure of interestingness since it permits the user to perform his job better by taking the right decisions in the light of the newly discovered knowledge. In a retailing context, we might think of a pattern being actionable if it provides useful input to the retailer to optimize his current merchandising strategies. For instance, if an association rule shows that red wine and cheese sell well together, the retailer could place the high-margin red wines next to the cheese items, leaving the rest of the lower margin red wine assortment where it is. By making it more convenient for the customer to buy high-profit red wine along with his cheese purchase, this may be an easy way to increase store profits. Unexpectedness, on the other hand, refers to the extent to which a rule is surprising to the user, which means that it contradicts with what the user reasonably expects under his existing system of beliefs. For instance, in a retailing context, we could argue that we are not interested in association rules that are relatively obvious to anyone with some familiarity with the industry. For instance, the discovery that toothpaste sells well with toothbrushes would not surprise the owner of a grocery store and would therefore probably not be useful for promotion purposes. People will by toothpaste with toothbrushes regardless of any promotional campaign encouraging them to do so.

This finally brings us to the view on interestingness of association rules that we believe may be of interest in the particular context of retailing, i.e. the *micro-economic view* on interestingness.

4.4.6 The Micro-Economic View on Interestingness

When reviewing the existing approaches that define interestingness of associations, we realized that there was one important element missing in the entire treatment of interestingness, i.e. the business value of associations. In fact, the existing approaches to the interestingness problem do not explicitly view the value of associations within the micro-economic environment of the retailer. Indeed, they treat the interestingness of associations within a statistical or subjective domain-independent[19] context.

Yet, when confronting the retailer with a list of frequent itemsets and/or association rules, it was often not clear what benefits could be gained from using these rules. In fact, discussions with retailers further increased our belief that the business value of associations was crucial in the treatment of the interestingness problem within a retailing context.

---

[19] Obviously, the unexpectedness framework relies on prior beliefs and those are clearly domain dependent. However, by domain-independent we mean that all of the discussed approaches so far are general enough so as to be applied in any domain of interest.

Quickly grew the idea that the discovery of frequent itemsets and association rules itself are not the final stage in the knowledge discovery process. Indeed, it is our opinion that they may provide useful input for specific marketing optimization problems that use cross-selling information for better decision-making, such as product selection and shelf location decisions. This will be the main topic in chapter 5.

4.5 Association Rule Generalizations

Instead of post-processing association rules to identify the most interesting ones, another way of increasing the interestingness of association rules is to extend the classical association rules framework by enabling richer rule expressions or by introducing additional variables beyond classical retail products. In general, a distinction can be made between *syntactic* and *semantic* generalizations. Syntactic generalizations refer to extensions in terms of the form/construction of the association rule, whereas semantic generalizations refer to the incorporation of other information into the rules, such as loyalty card information, to enrich their practical relevance.

4.5.1 Syntactic Generalizations

Syntactic generalizations can be divided into quantitative association rules, generalized association rules, sequential association rules, constrained association rules and negative association rules, as discussed in subsequent sections.

4.5.1.1 Quantitative association rules

The classical association rules problem discovers relationships between Boolean (0-1) attributes in a relational table and is therefore sometimes also referred to as the *Boolean association rules* problem. However, in many circumstances, real-world databases contain richer attribute types. For instance, in the case where the number of items purchased is of importance, the problem of mining *quantitative association rules* was proposed in [253]. In this case, association rules are discovered that include quantitative (such as age, income) and categorical attributes (such as zip code, car type). This enables the discovery of rules that indicate how a given range of quantitative and categorical attributes may affect the values of other attributes in the data.

Example: Salary [40K - 50K] $\cap$ age [25-30] $\Rightarrow$ Buying Ford car

Essentially, the algorithm works as follows. Each of the quantitative attributes is discretized into a number of disjoint intervals and each of these intervals is consequently mapped to an item which represents that range. Once these pseudo-items are constructed, the classical association rules algorithm [12] can be used to find the association rules. However, partitioning quantitative attributes into intervals must be carried out very carefully because:

- If the number of intervals for a quantitative attribute is large (i.e. small-width intervals), the support for any individual interval may be too low such that some rules involving this attribute may not have sufficient support. As a result, either very few rules will be discovered, or the rules will be nearly as specific as the data itself.
- If the number of intervals for a quantitative attribute is low (i.e. large-width intervals), the support for any individual interval may be high such that some rules involving this attribute in the antecedent of the rule may not have sufficient confidence. As a result, many rules with little information will be generated.

To summarize, if the intervals of the quantitative attribute are too large, a rule may not have sufficient confidence; if the intervals are too small, a rule may not have sufficient support. To solve this catch-22 situation, Srikant and Agrawal [253] make use of a *partial completeness measure* that gives a handle on the information lost by partitioning and as a result produces an answer on how many partitions there should be and where to place the cuts. An alternative approach to the interval-partitioning problem was given by Wang et al. [287] who proposed two measures (i.e. *J*-measure and Surplus measure) to calculate the information loss caused by merging intervals.

4.5.1.2 Generalized (or multi-level) association rules

The concept of generalized association rules was first introduced by Srikant and Agrawal [254] and was inspired by the fact that, in many cases and especially relevant within the area of retailing, taxonomies (*is-a* hierarchies) over items are available (see figure 4.7).

Retailers indeed typically categorize the items in their assortment into a hierarchy of product categories and it might be interesting to discover associations that span different levels of the taxonomy.

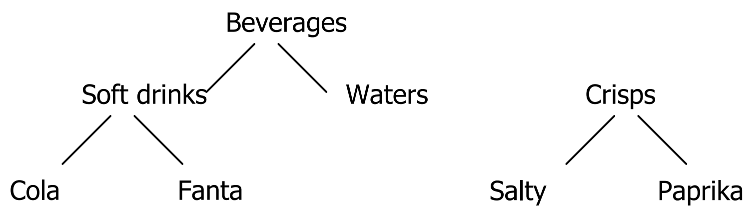


Figure 4.7: Product taxonomy

However, the introduction of a product taxonomy poses some new problems to the discovery of product associations. For instance, we may infer a rule that people who buy Soft drink tend to buy Crisps from the fact that people bought Cola with Crisps and Fanta with Crisps. However, the support for the rule Soft drink $\Rightarrow$ Crisps may not be the sum of the supports for the rules Cola $\Rightarrow$ Crisps and Fanta $\Rightarrow$ Crisps since some people may have purchased Cola, Fanta and Crisps in the same transaction. Also, Soft drink $\Rightarrow$ Crisps may be a valid rule, while Cola $\Rightarrow$ Crisps and Beverages $\Rightarrow$ Crisps may not. The former may not have minimum support, and the latter may not have minimum confidence.

Thus, finding such rules across different levels of the taxonomy is valuable since rules at lower levels (rules expressing associations among items at the SKU level) may not have minimum support. And since a typical supermarket carries thousands of items, the support for rules involving only low-level items in the taxonomy tends to be very small. Therefore, if one wants to find strong associations involving only low level items of the taxonomy, the support threshold must be set very low, which may lead to statistically insignificant results. On the other hand, finding rules at higher levels of the taxonomy may produce rules corresponding to intuitive or expected knowledge [159].

A naive way of discovering generalized association rules is to create an 'extended' transaction T', including all the items in T as well as all the ancestors in the taxonomy of all items in T. This way, the standard Apriori algorithm can be adopted. However, running Apriori on the extended transactions tends to slow down the performance of Apriori considerably, which is obvious given the increased size of the baskets. Therefore, Srikant and Agrawal [254] developed three efficient algorithms to find generalized association rules, i.e. *Stratify*, *Estimate* and *EstMerge* and Han and Fu [128] developed another algorithm in order to find *multiple level association rules*, which is the same as *generalized association rules*.

4.5.1.3 Sequential association rules

The association rule methods discussed so far have not dealt with the notion of 'time', i.e. they express relationships that exist concurrently. However, in some cases there is timing information available (e.g. web clickstream data) or there exists a particular interest in finding time-dependent relationships of behaviour. For example, customers typically apply for a current account and a savings account, and then they apply for a mortgage loan. Note that the purchase of these banking products needs not to be consecutive, other items may be purchased in between. In contrast to the standard association rules problem, the sequential patterns problem, however, brings about some additional difficulties.

Firstly, users often want to specify maximum and/or minimum time gaps between adjacent elements of the sequential pattern, i.e. users are only interested in sequential patterns for where adjacent elements occur within a specified time interval. For instance, we are not interested in knowing that the application for a mortgage loan takes place 30 years after the opening of a banking account, but only for instance within a 10-year period.

Secondly, for many applications it does not matter if items in an element of a sequential pattern were present in two different transactions, as long as the time stamps of those transactions are within some small time window. For instance, if the savings account was opened just a few days after the current account, we treat them as being opened together.

Thirdly, the support of a sequence is defined in a different way as the support of an association rule. The support of a sequence is defined as the fraction of total customers who support the sequence, and not the fraction of all transactions, like in association rules.

Agrawal and Srikant [11, 252] discuss the discovery of *sequential patterns* and propose three algorithms, *AprioriSome*, *DynamicSome* and *AprioriAll* to find such rules.

4.5.1.4 Constrained association rules

The association rules problem finds all rules that satisfy user-specified minimum support and confidence thresholds. However, users are often only interested in a subset of associations, for instance, containing at least one item from a user-defined subset of items. Although this problem can be solved again in a naive way by post-processing the association rules, Srikant et al. [255] developed an adaptation of the Apriori algorithm to directly incorporate these constraints into the associations discovery algorithm. More concretely, they consider constraints that are Boolean expressions over the presence or absence of items in the rules, possibly combined with taxonomy information. For example, (Soft drink $\wedge$ Candy) $\vee$ (descendants(Beverages) $\wedge \neg$ ancestors(Mars)) constrains the rule discover process towards rules that either contain both soft drinks and candy, or contain beverages or any descendants of beverages and do not contain Mars or any ancestors of Mars. Conjunctions of conditions were also considered by Lakshmanan et al. [167] and were extended to contain constraints on the antecedent and consequent and arbitrary Boolean constraints by Goethals and Van den Bussche [113].

4.5.1.5 Negative association rules

Savasere et al. [236] discussed the problem of mining for strong negative associations. Negative associations are defined as rules that express what items a customer is not likely to purchase given that he buys a certain set of items. For example, 60% of the customers who buy potato chips do not buy spaghetti. Although the problem of negative associations, at first, looks like a simple extension of the association rules problem, namely also counting the frequency of candidate itemsets containing non-presence of items, it will become evident that this is computationally intractable for large product assortments. This can be easily understood, given the probability that an item is not present in a particular basket is very high. For instance, it is easy to find a large number of frequent itemsets of the form (Coca-cola, $\neg$ any other item). Moreover, most of the rules resulting from these frequent itemsets will be extremely uninteresting. The problem therefore reduces to finding *strong* negative associations. Therefore, Savasere et al. [236] use taxonomy information to calculate the deviation of the expected support of a candidate negative itemset and compare it with their actual support to decide whether a particular negative large itemset is interesting or not.

4.5.2 Semantic Generalizations

The association rules problem as discussed in paragraph 4.3 aims at discovering purchase associations between items in retail market basket data. In other words, association rules only contain products. Several researchers have been involved in incorporating other information, such as loyalty card information, into association rules in order to enrich their practical relevance. These extensions therefore refer to semantic generalizations of the rules.

4.5.2.1 Profile association rules

The profile association rules problem [6] was introduced to examine associations between customer profile information and behavioural information and it is closely related to the quantitative association rules problem. In the quantitative association rules problem, however, the form of the rules is such that quantitative attributes appear in the antecedent of the rule and a single categorical attribute appears in the consequent. In the profile association rules problem, the antecedent consists of customer profile information, such as age, salary, years of education, marital status, etc, whereas the consequent consists of product purchase information.

Example: Age [30-40] $\cap$ Salary [80K-125K] $\Rightarrow$ (Buying Ford) | (Buying Nissan)

The above rule shows that the common profile or characteristics shared by a group of customers buying *either* a Ford-car *or* a Nissan-car and satisfying the support and confidence requirements, are aged between 30 - 40 and have a salary between 80K - 125K. Note, however, that the above rule does not show that among the customers exhibiting the given profile, most of them buy *both* a Ford-car *and* a Nissan-car. The latter can be discovered by normal association rules. The idea is that some customer segment, i.e. those customers exhibiting the same characteristics, produces a set of behaviours, rather than that each one of that customer segment produces every single behaviour within that set.

4.5.2.2 Virtual items in association rules

Closely related to the problem of discovering profile association rules is the concept of virtual items in association rules. Berry and Linoff [32] introduced the concept of virtual items in order to include other information than products into the rules. This information can relate to aggregate transactional information, such as day, time and location of purchase, method of payment, customer characteristics obtained from loyalty cards, etc.

Example: August $\wedge$ hamburger $\Rightarrow$ barbecue sauce

The example association rule expresses that if people buy hamburgers in August, they also tend to buy barbecue sauce. In this example, August is a virtual item since it is not a product item but relates to the circumstances in which the hamburger and the barbecue sauce were purchased together. When the non-product items are transformed into dummy variables, the Apriori algorithm can again be used to discover association rules containing virtual items. For instance, experiments on the datawarehouse of a clothing chain with the sales location as virtual item revealed that there existed significant differences in the purchases of children's polo's and T-shirts together with children's bermuda shorts for different sales outlets [58].

4.5.2.3 Cyclic and calendric association rules

Typically in association rules, the data is treated as one large segment, with no attention being paid to segmenting the data over different time intervals. In fact, analysis of the data may reveal that certain combinations of items may only be frequent within specific time segments/intervals and do not occur within other time segments [213]. For instance, the selling of potatoes and mayonnaise may primarily occur between 4PM and 6PM. Therefore, if the data are segmented over different time intervals, the support of this itemset will be significantly higher in this time interval compared with other time intervals during the day. Therefore, Özden et al. [213] proposed the notion of cyclic association rules, i.e. association rules with minimum support and confidence at regular time intervals. Unfortunately, their algorithm required the specification by the user of the time granularity (e.g., hour, day, week, ...) and the minimum and maximum cycle lengths l_{min} and l_{max} of interest (e.g. 24 hours). Furthermore, the algorithm fails to capture fuzzyness: a rule that would exhibit a pattern 'most of the time' but not all of the time. Finally, the algorithm fails to handle multiple time units (like hours, weeks and days) simultaneously and to consider numeric attributes instead of Boolean attributes.

Therefore, in [226], the notion of calendric association rules was introduced which included several generalizations to the work of Özden et al. [213]:

- A notion called *calendar algebra* that is used to define and manipulate groups of time intervals;
- The notion of finding *fuzzy* patterns in association rules which allows to find patterns in the data that *approximately* match the user-defined patterns;
- The handling of multiple units of time.

An association rule is called *calendric* if the rule has the minimum support and confidence during every time unit contained in a calendar (modulo a mismatch threshold, which allows for a certain amount of error in the matching).

CHAPTER 5
PROFSET: A Framework for Product Selection

In this chapter, the problem of product selection in retailing is being studied. The contribution of this chapter lies within the formulation of a generic constrained optimization framework for product selection, called PROFSET. The framework is generic because it provides a general container model in which different specific models can be built, according to the particular product selection problem at hand. This is illustrated by two concrete model specifications that will be tested on retail sales transaction data.

The first model makes an attempt towards solving the following marketing problem: an increasing number of retail distribution chains, such as Carrefour, SPAR and Delhaize, recently provide additional convenience shopping facilities besides their traditional stores to serve time-pressured convenience customers [198, 204]. For instance, the Shop'n Go (Delhaize), GB-Express (Carrefour) and Shop24 (SPAR) are examples of this increasing trend for fast convenience shopping. Typically, these convenience stores are located nearby gas stations, train stations, hospitals, or outside the regular store, although some retailers (e.g. Carrefour and Albert Heijn) also provide specific shop-in-a-shop concepts within the traditional supermarket for time-pressured and convenience shoppers. However, since the typical retail surface is limited (15-150m²), it is of crucial importance to select the right products in order to maximize the profitability of the convenience store.

Consequently, the objective of the first product selection model is to find the optimal set of products to put in such a convenience store, based on sales transaction data from the *regular* store. The idea is to maximize the profitability from cross-selling effects between the selected products in the convenience store, based on the discovered cross-selling effects inside the *regular* store. This way, information about existing cross-selling effects in the regular store can be used to optimize the product composition of the convenience store.

The second model for product selection makes an attempt towards solving another well-known marketing problem: retail stores want to maximize their share of the customer's wallet by stimulating cross-selling of their products inside the store. Typically, there are a limited number of attractive shelf positions available in the store, such as end-of-aisle locations, product positions at the outer walking circuit in the store, shelf positions at eye-level, etc. The optimization problem then arises which products to put at those positions, such that customers will not only buy products at those attractive positions, but that they will also go inside the aisles or inner walking circuits of the store to purchase other items too. The crucial idea is that not only the profit of the selected set of products should now be maximized (like in the first application), but also the profit resulting from cross-selling with other products located at regular positions in the store. The model, thus rests on the important assumption that customers will do the effort to search for related products inside the gondolas. This might, however, not be realistic for those shoppers for whom convenience and speed of shopping are of high importance (i.e. the typical shoppers of the first model). In fact, for them, the assignment of products to attractive and regular positions in the store may even disturb their shopping trip since they will have to go inside the gondolas to find their preferred products. For these customers, the convenience store or shop-in-a-shop concept is probably much more appealing.

A general assumption underlying both models is that the use of frequent itemsets to express the level of interdependence between items is only permitted if the assocations are not disturbed or created as a result of environmental circumstances like store layout, promotions, pricing and others (referred to as the *identification problem* in the literature [83]). For instance, if an association between a collection of products exist solely due to location decisions taken by the management, then it is possible that the association disappears when the products are separated into attractive and regular positions. Consequently, for those products or categories where such environmental factors play an important role, the models presented in this chapter should not be used.

From a technical point of view, the contribution of this chapter is the use of data mining results, i.e. frequent itemsets, in combination with linear programming techniques to optimize the above mentioned product selection problems. From a practical point of view, the proposed models aim at taking a step into the right direction by illustrating that product interdependence effects play an important role in evaluating the performance of individual product items for product selection. Indeed, it is argued that the contribution of a product goes beyond its direct impact on the profitability, but that it may also have an indirect impact through purchase interactions with other products and that frequent itemsets can be used as a measure to quantify this effect.

The rest of this chapter is organized as follows. First of all, an overview of the existing marketing literature on product selection is provided. This includes a discussion of the basic building blocks of a product assortment, the dimensions of the product assortment and existing methods for product selection and shelf space management. Secondly, and most important in this chapter, the constrained optimization framework for product selection (PROFSET) will be introduced and two particular model specifications will be discussed and implemented on real data. Finally, this chapter concludes with a sensitivity analysis and discusses the contributions and limitations of the PROFSET model.

This chapter is based on work reported in [52, 53, 54, 55, 56] and has been referred to in [22, 68, 102, 130, 157, 163, 222, 259, 260].

5.1 Introduction

In general terms, the product assortment of a retail store can be considered as a multitude of articles that are provided for sale to the consumer at a particular point in time. However, the product assortment is more than just a 'bunch of products'. In fact, for any type of retailer, the assortment has major consequences towards several aspects of the retail organization and to the consumer, in so far that marketeers often mention the idea of assortment 'policy'. The marketing concept of assortment policy has, throughout the years, however evolved from a rather static to a highly dynamic exercise. Indeed, in the past, retailers saw their job as one of buying products and putting them out for sale to the public. If the products were sold, more of them were ordered. If they did not sell, they were disposed of. Blischok [36] describes retailing in this model as a *product-oriented* business, where talented merchants could tell by the look and feel of an item whether or not it was a winner. In order to be successful, retailing today can no longer be just a product-oriented business. One could say that the composition of a product assortment has become a complex exercise.

Indeed, as a result of larger product assortments, the daily pressure on the retailer to stock new items and the compulsion to satisfy the diverse, complex and changing wants and needs of the consumer, it is of crucial importance to keep an eye on the success of individual product items in order to maintain long term profitability of the retail store. In this context, the ideal product assortment is currently subject to a number of constraints: it should reflect the store's image, generate sufficient profitability for the retailer and at the same time offer a wide variety of choice to the consumer.

Indeed, the product assortment being carried by the retailer has a major impact on how the consumer evaluates the retailer (store image), and therefore ultimately determines the success and the profitability of the store[20].

To this end, the product mix can be considered as a *strategic* and a *dynamic* instrument for the retailer. It is a strategic instrument because it should make clear to the consumer which wants and needs will be satisfied by the retailer and to what respect the retailer is different from the competition [275]. Furthermore, it is a dynamic instrument since the product assortment is subject to continuous changes. Take for instance the food industry which introduces new products on the market at an ever increasing speed and which forces the retailer to stay on the trends in order to satisfy the rapidly changing wants and needs of the consumer. Especially in the sector of the large distributors, this has resulted in a rapid expansion of the product assortment (both in width and depth) and presents a major challenge to the retailer who is confronted with the bottleneck of limited shelfspace. As a result, the management of the product assortment is probably one of the most crucial and difficult tasks for the retailer.

The availability of detailed sales transaction data by means of scanning devices however provides a rich source of information and therefore offers new opportunities to monitor and fine-tune the product mix. In this chapter, a constrained optimization framework will be developed that uses the information in these scanner data to improve product mix decisions, more specifically, for the case of the two particular marketing scenarios depicted above.

[20] In this context, there is recently increased interest in assortment rationalization [64], both in the industry as in academics, as a result of increasing product proliferation which not only adds significant cost to the system but also creates confusion among customers. In fact, a recent study by Broniarczyk [62] showed that eliminating as many as half of the items had no significant impact on consumer perceptions or purchase behaviour as long as shelf space was held constant and most consumers could find their favourate brands.

However, before introducing this framework, a concise overview will be given of the existing marketing literature on the building blocks and dimensions of the product assortment, together with existing methods for product selection and the closely related problem of shelf-space allocation.

Finally, it is important to note that choice decisions (which products to stock) are usually followed by allocation decisions (how much shelf space to allocate to each selected product). However, the PROFSET framework only models choice decisions and not allocation decisions.

5.2 Product Assortment Characteristics

A product assortment can be characterized in terms of its building blocks and its dimensions [29].

5.2.1 Building Blocks of the Assortment

The product assortment is the most general container concept for the collection of products being carried by the retailer. However, since the product assortment should ultimately translate the retailer's store image into a concrete physical product offering, retailers mostly break it down into a number of hierarchical levels, including the assortment group, the department group, the product category, the article and the article variety. Building and keeping such hierarchy (also called product taxonomy) up-to-date is an ample job for the retailer. Furthermore, the number of levels, their definition, and the level of detail in which the product assortment is broken down into smaller segments, will be different for each retailer.

5.2.1.1 The assortment group

The product assortment is first of all divided into assortment groups. Assortment groups are the most general division of products into groups, mostly food and non-food items. Depending on the type of retail store, the relative composition of the product assortment into food and non-food items will be different. For instance, large F1-type distributors, like Carrefour and Delhaize 'Le Lion', typically carry both assortment groups, i.e. a wide range of food product categories and a smaller set of quickly moving non-food product categories. In contrast, the smaller to average size non-integrated retailer, like GB Contact and GB Express, will mainly carry food categories and a very limited set of non-food categories, such as cleaning products and general household goods.

5.2.1.2 The department group

The assortment group is subsequently divided into different department groups, like healthcare, beverages, bakery products, frozen food, fresh food, cereals, etc. A department group combines products that have a physical similarity and which satisfy a 'rather general' consumer want or need. For instance, the need for beverages, or the need for healthcare, etc. Typically, the number of department groups can serve as an indication of the width of the product assortment.

5.2.1.3 The product category

The department group (like healthcare) is subsequently divided into product categories, like shampoo, dental care, body soap, toilet paper, feminine care, etc.

The product category represents a collection of products that satisfy a more 'specific' consumer want or need, compared to the department group[21]. As a result, the combination of the department groups and the product categories determines the total width of the product assortment.

5.2.1.4 The article

Each product category (like shampoo) is composed of articles, also called brands, like Head & Shoulders Wash and Go, Pantene, Palmolive, L'Oreal, etc. Articles satisfy a still more specific consumer want or need. Furthermore, the number of different products in a category ultimately determines the depth of the product category. According to the retailer formula, product categories can be very deep, like in supermarkets, or can be very shallow, like in discount stores that typically carry a wide range of product categories with just a few fast-moving products per category[22].

5.2.1.5 The article variety

Finally, the article variety is the lowest level building block of the product assortment and satisfies the most narrow want or need of the consumer. The article variety can be very different according to the product. It relates to the package size, the volume, the weight, the colour, the taste, the quality level, etc. of the product. For instance, for L'Oreal, different shampoos exist for regular hair, dandruff treatment, sensitive skin, greasy hair, damaged hair, etc. The number of product categories (width of the assortment) combined with the variety of products available within each category (the depth of the assortment)

---

[21] Recently, the 'product category' has received a slightly different definition within the theory of category management. Indeed, in the context of category management, the product category is defined as a collection of products that can be considered as substitutes because they satisfy similar wants or needs of the consumer and which should be treated as a separate strategic business unit (SBU) [3]. This means that the product category must be customized on a store-by-store basis through a separate product mix that is consistent with the local customer needs.
[22] A further difference should be made, however, between hard discounters (such as Aldi and Lidl) and soft discounters (such as Colruyt).

finally determines the size the total product assortment and the product choice available to the consumer.

5.2.2 Dimensions of the Assortment

The dimensions of the product assortment relate to the width and depth. The variety being offered to the consumer is therefore determined by the width and depth of the product assortment. Furthermore, these dimensions are related to the concepts of complementarity and substitution.

5.2.2.1 Width of the product assortment

The width of the product assortment relates to the number of department groups and product categories being carried by the store. A wide product assortment offers many different product categories, whereas a narrow assortment carries a limited number of product categories. The former can, for instance, be found in supermarkets and hypermarkets, whereas the latter can be encountered in specialty stores (e.g. a golf shop, or a baby store). Furthermore, the width of the product assortment directly relates to the level of complementarity between products. A wide product assortment offers opportunities for one-stop-shopping and cross-selling in contrast to a narrow assortment where there is only little coverage of the multitude of wants and needs of the consumer.

5.2.2.2 Depth of the product assortment

The depth of the product assortment relates to the number of different articles and article varieties being offered in each product category. A deep assortment offers many different articles/brands per category, whereas a shallow assortment offers only a limited choice of products in each product category. The former can, for instance, be encountered in supermarkets and specialty stores, whereas the latter is typical for convenience and hard discount stores.

Furthermore, the depth of the assortment is directly related to the concept of substitution effects between products. A deep assortment offers many different products that satisfy similar consumer wants and needs. As a result, they often occur together in the consideration set of the consumer (they compete for similar mindspace) and thus the consumer will make a choice between them based on his preferences, i.e. they are substitutes. A shallow assortment offers lesser choice within a product category, and thus the level of competition between products in the same category will be lower.

5.2.2.3 Core assortment

The core or basic assortment contains those products that should at least be contained in the product assortment since they are most closely connected with the store image [275]. It is like the refrain of a song, it should be recognized by the consumer as unique to the store. Removing the core assortment from the store would confuse the consumer since his idea about the offering of the store would no longer comply with the communicated store image. For instance, the credo of Cash-Fresh is 'fresh, friendly and profitable'. The core product assortment should therefore be consistent with that image in such a way that it fulfils the customer's expectations about the store, for instance by offering a deep choice of fresh fruit, vegetables and meat products. The core assortment should therefore create a *pull* effect that generates traffic towards the store.

In the last few years, the size of the core assortment has increased significantly as a result of the increased expectations of the customer towards the satisfaction of wants and needs by the store. In the past, supermarkets mainly carried grocery items. Today, however, the core assortment of the larger supermarket store contains much more product categories, including fresh bread, meat products, vegetables, etc. Yet, the core assortment represents the smaller part of the product assortment, mostly containing low-margin fast moving products.

5.2.2.4 Peripheral assortment

On top of the core assortment, there is the peripheral assortment containing products that are chosen by the retailer to confirm the store image even more and which should be selected to maximise cross-sales potential with products from the core assortment. Indeed, retailers are interested in adding items whose sales will not be made at the expense of core assortment items but which have a positive radiation effect on those items, i.e. help increase the sales of core items [218]. In fact, Van der Ster and van Wissen [275] distinguish between three types of peripheral assortment products: image increasing products, changing products and profit increasing products.

Image increasing products are added to further increase the store image. Furthermore, they should have a positive radiation effect on other products or on the entire product assortment. Recent examples include bio products from Albert Heijn and Delhaize[23].

Changing products are those products within the peripheral assortment that must prevent the consumer from getting bored with the product assortment. Indeed, life is subject to quickly changing trends and fashions and consumers expect their retail store to innovate for them. Therefore, retailers reserve a small part of their assortment to experiment with new products. Special locations in the store can be reserved for trial products to estimate their market performance, often also sponsored by the manufacturer. Examples include the green (instead of red) ketchup from Heinz and the introduction of 'diet' crisps and 'extra crispy' crisps by Smiths.

---

[23] Wine was clearly an image product for Delhaize in the past: it contributed in a positive way to the image of Delhaize as being a rather exclusive product and was supplementary to other items in the assortment. Today, wine has however lost a great deal of its exclusive character and customers probably perceive it as a crucial product in the core assortment. One could therefore argue whether it still represents an image product for Delhaize. Furthermore, since the entry of Carrefour on the Belgian market, Carrefour also offers a wide variety of wines such that the comparative advantage of wine for Delhaize over other retailers has diluted. This is maybe the reason why Delhaize has recently launched some new initiatives on this domain, including for instance high quality ready-made meals, oriental products, and bio products.

Finally, profit-increasing products are those products in the peripheral assortment with relatively high profit margins. They usually represent the biggest part of the product assortment, but are rather slow-moving, in contrast to the core assortment articles, which usually have low profit margins but are fast-moving. Examples include cutlery, maintenance tools, etc.

5.2.2.5 Supplementary assortment

In some Belgian supermarkets, like Aldi, Lidl and Carrefour, there seems to be a recent trend to have weekly (nation wide) promotions (the so-called special buys) on products (mostly non-food). For instance, hard discounters like Aldi and Lidl have previously offered laptop computers, bicycles, garden equipment, etc. at razor-sharp prices. Usually, one would not expect such products to be sold in a discount store, but rather in a speciality store. These products are allocated to the supplementary assortment since they have nothing to do anymore with the actual store formula and are mostly only available during the promotional period.

5.3 The Problem of Assortment Optimization

The problem of assortment optimization is an important issue in retail marketing (section 5.3.1) and it is closely related to that of shelf-space management. Moreover, it is a complex exercise since both the demand and the cost side should be incorporated into the profit equation of any optimization model (section 5.3.2). However, since all those parameters are not always easy to obtain, most commercial systems do not incorporate all of these effects and often rely on simple heuristics. Academic models, however, have attempted to incorporate both demand and cost sides into their optimization models (section 5.3.3).

5.3.1 The Need for Assortment Optimization

The continuous pressure on the retailer to innovate, to satisfy the diverse wants and needs of the consumer and to differentiate from the competition forces him to keep stocking more and more items. From this perspective, this would lead to an uncontrolled expansion of the product assortment and the search for more retail space to sell those products. However, to remain competitive, retailers must rationalize their assortment. Indeed, retailers know that by adding more and more products, at some moment, a point will be reached where the marginal cost (stocking, handling, etc.) of adding more items to the assortment will exceed the marginal return from selling those products to the consumer. In other words, at some point, stocking more products will lead to lower total profits, or even a loss. Still, retailers are often resistant to eliminate products from the assortment because they fear that this would lower consumer assortment perceptions and decrease the likelihood of choosing their store [62]. This fear is probably increased by the lack of information about the demand and cost implications of product assortment decisions, as discussed in the next section.

5.3.2 The Complexity of Assortment Optimization

The complexity of the assortment optimization problem in retailing stems from the different implications that product decisions have on the sales and costs associated with that decision. In general, one can distinguish between demand and cost side effects.

5.3.2.1 Demand-side effects

Products have different profit margins. As a result, the assortment choice has an important impact on the profitability of the store. Indeed, adding or deleting products from the assortment has a direct and profound impact on the profitability of the product assortment.

Furthermore, it is generally agreed upon in the literature [44, 66, 83] that assortment optimization should take into account product purchase interdependencies, such as cross-selling and substitution effects between products. Indeed, purchase relationships may exist between different products and the addition, deletion or substitution of a product from the assortment, or changing the amount of shelf-space allocated to the product, can increase or decrease the total assortment profitability more than can be explained by the effect of the particular product decision alone, i.e. more than indicated by the *direct space elasticity* of the product. Indeed, an increase in sales may be the result of additional sales due to cross-selling, whereas a zero-effect or decrease in sales can be the result of substitution effects between products, i.e. product cannibalism. The latter effects are usually denoted as *cross-space elasticities*. A decrease in sales, for instance, can occur if a newly added product, or a product of which the shelf space is increased, is considered as a close or perfect substitute of the existing products in the category. Unfortunately, this kind of information is not always readily available and as a result, most commercial assortment planning systems do not take such interdependence effects into account.

Another issue is that of stockout effects on the profitability of a product category. In fact, previous research has shown that out-of-stock situations may prevent customers from buying another alternative from the same product category and that this effect can by quite large (for an overview article see [71]).

Finally, apart from direct or indirect profit implications resulting from product decisions with different gross margins, the retailer should also look at other product related revenues, such as special deals, allowances, etc., in order to get a complete picture of the revenues associated with each product.

5.3.2.2 Cost-side effects

Apart from the purchase cost of the product (already accounted for in the gross margin of the product), there are also costs associated with the flow of the product throughout the distribution channel. These costs relate to distribution centre costs, transportation costs and store costs and should therefore also be allocated to each product, as far as direct allocation is possible. The reason for including these costs is that two products may have a similar/identical gross margin, although their net profitability for the assortment may be significantly different as a result of different inventory or handling costs. For instance, one product may be easier to stock and to replenish than the other.

5.3.2.3 Direct product profitability

Integrating demand- and cost-side effects in order to determine the final profitability per product is better known as the concept of Direct Product Profitability (DPP), as illustrated in figure 5.1.

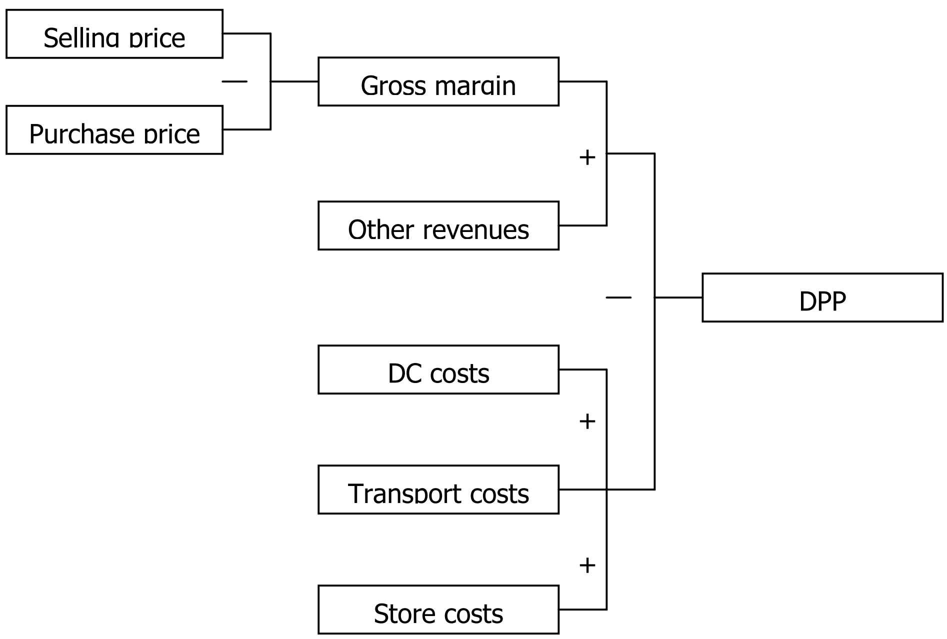


Figure 5.1: Direct Product Profitability (DPP) illustrated

Although the concept of DPP is not new, it already appeared in the 1960's in the context of academic shelf space allocation systems, it has not been used in practice for a long time. According to van der Ster and van Wissen [275], this has to do with the difficulty of obtaining correct product related cost and revenue data. However, since the introduction of barcode scanners and database systems in retailing, parts of these data have become more easily available for decision making, which has given the DPP concept a new impetus in practice. Direct product profitability is usually expressed in terms of the net profitability per unit, or the DPP per unit per week, or the DPP per m² shelfspace, etc.

5.3.3 Methods for Product-Mix Decisions

Based on the demand- and cost-side product information presented above, more accurate product selection and shelf space allocation decisions can be made. Even though the PROFSET model introduced in this dissertation is a product selection model and not a shelf space allocation model, a concise literature overview of both types of models will be provided in this section. The reason is that product selection and shelf space allocation are closely related to each other. One can generally distinguish between three types of models.

The first type of models only treats product selection without dealing with shelf space. A second type of models only treats optimal shelf space allocation (e.g. [65, 66]), provided that the product selection problem is already carried out. Finally, still other models treat product selection and shelf space allocation simultaneously (e.g. [20, 44, 131]). In the first and second case, distinct models are used, first to select an optimal set of products, and second to allocate shelf space to the selected products, whereas in the latter case the problem of product selection and shelf space allocation are treated within one overall model.

Product-mix and shelf space allocation models in general can, however, again be divided into two classes: heuristic models and optimization models.

5.3.3.1 Heuristics for selection/shelf space allocation

As a result of the complexity and the cost to collect all relevant product related demand and cost information, many retailers rely on rather simple heuristics (rules of thumb) to select products or allocate shelf space, based on a number of quantitative criteria, also called indices of SKU productivity [43], such as unit sales, dollar sales, rotation speed, gross margin, contribution per m², DPP and many others. Furthermore, these measures typically do not account for product interdependence effects and therefore they do not reflect the dynamics in the store. Nevertheless, these indices serve as the input for product selection and shelf allocation methods where the focus is on simplicity and ease of use.

Some popular commercial systems include the PROGALI model [182], which allocates shelf space in proportion to total dollar sales. In the OBM model [203], shelf space is allocated proportional to the product's gross profit. Other systems for shelf space allocation have concentrated on minimizing costs of inventory and handling, such as COSMOS [85], SLIM [76] HOPE [100] and ACCUSPACE [172]. CIFRINO [75] and McKinsey [193] combine both product revenues and costs to allocate shelf space in relation to DPP. However, none of them incorporates demand elasticities. Finally, SPACEMAN developed by ACNielsen, is also worth mentioning. On top of shelf space allocation, SPACEMAN visualizes shelf space allocation into store planograms.

With regard to product selection, the method of 'product portfolio analysis' is also worth mentioning. Although originally the idea of portfolio analysis was proposed in the context of multi-product firms, it seems like an interesting approach to evaluate the retailers existing portfolio of products along a number of important performance dimensions like sales, market share, profitability, growth potential, etc. A well-known product portfolio instrument is the Boston Consultancy growth-share matrix. This matrix is built along two dimensions, namely the product's market share and the stage of the product in the product life cycle. The product's market share reflects the cost advantage that a manufacturer has versus his competitors such that products with a high market share generate more cash. The stage of the product's life cycle is usually measured by its sales growth. The idea is that for high growth products, more cash will be needed to consolidate the growth, in contrast to low growth products. Swinnen [257] offers a critical review of the BCGS matrix within the context of a supermarket chain. He argues that it is very uncertain whether the BCGS approach is of practical usefulness. First of all, he argues, the cost advantage for products with relatively high market share is only guaranteed under the assumption that the experience curve holds and this has not yet been studied for distribution firms. Secondly, the objective to generate cash may not be the primary objective of supermarket chains, whose main resource appears to be the availability of shelf space. Thirdly, the product-portfolio approach is a rather strategic approach and it is therefore best suited for decisions on the level of product categories and thus not on the SKU level. To conclude, although the existing product portfolio approach seems less suited for product selection decisions in supermarkets, the BCG matrix may present an interesting approach to evaluate different product groups according to particular dimensions that are relevant to the supermarket retailer, such as image and contribution per m^2.

Finally, ABC analysis rank-orders manufacturer brands according to the perceived price and quality levels, advertising efforts, brand reputation and distribution coverage [275]. 'A'-brands can be described as having a high perceived price and quality, a wide distribution coverage and a premium brand reputation. They primarily serve to support the retailer's store image by taking advantage of the excellent image of the brand, often as the result of large investments in advertising by the manufacturer. 'B'-brands have less reputation, less geographical spread and a lower perceived price and quality reputation than 'A'-brands. They primarily serve to cover the lower-end of the product assortment. Finally, and mainly as the result of increased price competition during the last few years, 'C'-brands are perceived as having a very low price and quality. They are usually distributed by a single retail chain and in most cases the manufacturer does no longer invest into advertising such products. The categorization of manufacturer brands into ABC groups may therefore serve as a guidance to select particular products into each category.

5.3.3.2 Optimization models for selection/shelf space allocation

The objective of constrained optimization models for product selection is to use mathematical or operations research procedures to find a product mix that maximizes a particular objective (like profit) subject to a number of constraints (like limited shelf space) using quantitative data. However, the term 'optimization model' may be somewhat misleading in so far that, as a result of the complexity of the product selection problem, the overall best model probably does not exist. In fact, each model aims at finding the best solution relative to the objectives, constraints and product data available.

Probably one of the first models for product selection was the model by Anderson and Amato [20]. They introduced a mathematical optimization model for simultaneously determining the optimal set of products to choose from a large set of available products, together with the amount of display space to allocate to each selected product. Their model is similar to the model presented in this dissertation in so far that our model also aims at selecting the most profitable subset of products out of a larger set of available products. However, in contrast the PROFSET model, their model does not take into account cross-selling effects between products. On the other hand, they include brand-switching behaviour into their model, which we do not. Unfortunately, they did not test their model on real data.

In 1979, Hansen and Heinsbroek [131] developed a product selection and shelf space allocation model taking into account the space elasticity of sales and a number of constraints related to the minimum amount of shelf space allocated to selected products. Their model, however, does not include substitution and complementarity effects between products. They argue that the information needed to estimate the demand interdependencies for supermarkets was not available at that time and that the large number of such interdependencies would render any systematic treatment of them all impossible in practice. The objective of their model is to maximize profit, taking into account the unit margin of products, product demand in function of allocated shelf space, the unit cost of space and the cost for replenishment of the shelf stock.

The absence of cross space elasticities in most product selection models has motivated Corstjens and Doyle [83] to present a model for optimizing retail space allocations where both main and cross-space elasticities were considered. They used cross-sectional data to estimate the elasticities for a small set of 5 product categories, including chocolate confectionary, toffee, hard-boiled candy, greeting cards and ice cream. The results of their model show that ignoring cross-elasticities may lead to a major suboptimalization in the allocation procedure. Later, the Corstjens and Doyle model was also implemented by Swinnen [257] on cross-sectional data from twenty-seven stores belonging to one supermarket chain in Belgium. The study implemented the model on ten product groups and found significant positive cross-space elasticities between canned tomatoes and spaghetti, and between Knorr Royco dry soup and tomato soups.

Bultez et al. proposed the S.H.A.R.P [65] and S.H.A.R.P II [66] model. They tried to maintain the parsimony of the Corstjens and Doyle model and also incorporated both direct and cross-space effects and modelled costs as a function of sales per unit space.

The issue of product selection was also studied by Borin et al. [44]. They also included main and cross-space elasticity effects into their model. Furthermore, they argued that the effect of stockouts and assortment decisions should also be considered in a profit optimization problem. Indeed, a stockout of a particular item may influence the sales of substitute items in the same category (known as stockout demand), and an item's sales may also increase due to the non-selection of a particular product as a result of switching behaviour by consumers (known as acquired demand). Given the huge number of considered products in our model, stockout demand and acquired demand are not included in our optimization framework.

5.4 PROFSET: An Optimization Framework for Product Selection

This section introduces the PROFSET constrained optimization framework for product selection, based on the use of frequent itemsets (to express product purchase interdependencies) from association rule mining. PROFSET is considered as an optimization framework since it has become a general model that enables the implementation of different specific optimization models for concrete retail assortment optimization problems (see section 5.4.3). However, before introducing the PROFSET framework, we believe that it may clarify things if we compare our contribution against the existing approaches on product selection and/or shelf space allocation presented in the previous section.

5.4.1 What PROFSET Includes

Firstly, the model for product selection presented in this chapter is a product selection model, and not a shelf space allocation model.

Secondly, on the demand side of the profit equation, it includes both main and cross-selling effects between individual product items. We argue that, although DPP provides a very accurate answer to the question 'what is the direct profit of the product to the retailer?', it does not answer the most relevant question for the retailer, i.e, 'what is the total profit generated by the product?' Indeed, the TPP is an estimation of the Total Product Profitability of a product for a retailer and consists of both the DPP and IPP (Indirect Product Profitability). The IPP measures the indirect effect of the product on the sales of other products. The IPP can therefore include both positive and negative interdependence effects between products. However, in this dissertation, we only use positive interdependence (cross-selling) effects.

Thirdly, our model includes a store limitation parameter, similar to the models of Corstjens and Doyle [83] and Borin et al. [44]. This means that the product selection model includes a constraint on how many items can be selected from the total assortment.

Fourthly, our product selection model includes upper and lower bounds on the number of items selected from each product category. This is again similar to the models of Corstjens and Doyle [83] and Borin et al. [44] where the amount of space in a category is expressed in terms of the amount of standard facings available.

Finally, the theoretical development of our model includes cost-side information on product inventory and product handling costs per product. Given the difficulties to obtain such information, especially handling information, cost side information is, however, not included in the empirical development of our model.

5.4.2 What PROFSET Does Not Include

Not the least important concerns the discussion of topics that are not covered by the suggested product selection model in this dissertation.

Firstly, as already mentioned before, the model is not a shelf space allocation model. This is in contrast with other optimization approaches that often combine the product selection and shelf space allocation decision problem [e.g. 83, 131] into the same model. Secondly, and a direct result of the choice not to model shelf space allocation decisions, our model does not include main and cross-space elasticity effects into the demand side of the profit equation.

Finally, the model does not include stockout effects. This means that brand-switching effects as a result of temporary stock-out or exclusion of a particular product from the assortment are not taken into account. This is clearly a limitation on the theoretical completeness of our model. However, given the huge amounts of products considered in the model, parameter estimation for such effects would be practically infeasible.

5.4.3 Overview of Model Specifications

As indicated before, we will introduce several different versions of the PROFSET model according to how the following three criteria are combined (table 5.1).

Criteria		
Optimization Criterion (A)	Inside hitlist (A.1)	Inside + outside hitlist (A.2)
Allocation Rule (B)	Support-based (B.1)	Loglinear based (B.2)
Category Constraints (C)	Yes (C.1)	No (C.2)

Table 5.1: Overview of PROFSET model specifications

First of all, PROFSET enables the specification of different objectives for optimization, which are directly linked to the type of marketing problem that is tackled (section 5.4.3.1). In this dissertation, we will discuss two specific model implementations. A first implementation aims to solve the problem of composing a product assortment for a small convenience store, such as the Shop24 (A.1). A second implementation aims at finding the best products to be positioned at visually attractive positions in a regular supermarket store (A.2).

Secondly, PROFSET models can differ according to the profit allocation heuristic being used to allocate the gross margin from transactions to frequent itemsets (section 5.4.3.2). A first profit allocation heuristic is based on the support criterion (B.1), whereas a second implementation is based on the statistical significance of itemsets obtained by means of loglinear analysis (B.2).

Finally, PROFSET models can differ according to whether product category constraints are included in the model (C.1) or not (C.2), see section 5.4.3.3.

These ideas are developed in detail in the next sections, after which an empirical setup is discussed to test and compare the different models (section 5.5), followed by a discussion of the results of this comparison (section 5.6). Finally, in section 5.7, we will evaluate the sensitivity of the PROFSET model with respect to a different selection of baskets.

5.4.3.1 The optimization criterion (A)

A first decision needs to be taken with respect to the choice of the optimization criterion. In fact, the optimization criterion determines the objective of the model and is dependent on the particular marketing problem that needs to be solved. In both models, the objective is to find an optimal selection of products (a hitlist), yet for a different marketing application.

Optimization inside the hitlist (A.1)

Consider the following assortment optimization problem. Besides the traditional supermarket stores, an increasing number of retail distribution chains like Carrefour, SPAR and Delhaize recently provide additional convenience shopping facilities to serve time-pressured convenience customers [198, 204]. For instance, the Shop'n Go (Delhaize), GB-Express (Carrefour) and Shop24 (e.g. Delhaize and SPAR) are examples of this increasing trend for fast convenience shopping. These convenience stores are typically located nearby high traffic locations such as gas stations, train stations, hospitals, or outside the regular store, although some retailers (e.g. Carrefour and Albert Heijn) also provide specific shop-in-a-shop concepts within the traditional supermarket for time-pressured and convenience shoppers. The idea is that in modern times, with dual earners and flexible working hours, a number of shoppers has only limited time for shopping and they expect and insist on the freedom to choose where to shop, and more importantly when to shop at the time that is most convenient for them. The convenience store aims to fulfil this need for nearness and more flexible shopping hours. In fact, in the case of the automated convenience store Shop24, shopping can be done 24 hours round the clock.

However, since the surface of the typical convenience store is rather small (from 15 to 150 m² depending on the type of convenience store), composing the right product mix is of crucial importance to maximize the store's profitability. The idea developed in this dissertation is to maximize the profitability from cross-selling effects between the selected products in a convenience store, based on the discovered cross-selling effects inside the *regular* store. This way, information about existing cross-selling effects in the regular store can be used to optimize the product composition of the convenience store.

.

Model Formulation

This marketing optimization problem leads to the following specification of the PROFSET model:

$$\max\left(\sum_{X\in L} M(X)P_X - \sum_{i\in \mathcal{I}} Cost_i Q_i\right) \tag{1}$$

s.t.

$$\sum_{i\in \mathcal{I}} Q_i = ItemMax \tag{2}$$

$$\forall X \in L, \forall i \in X : Q_i \geq P_X \tag{3}$$

$$\forall P_X, \forall Q_i \in \{0,1\} \tag{4}$$

PROFSET A.1 Model

Hereafter, the model will be clarified and an optimization routine to calculate its optimum value will be explained.

The Objective Function

The objective function (1) contains two sets of Boolean decision variables, P_X and $Q_i \in \{0,1\}$, whose values determine the total profit to be maximized. Furthermore, the objective function consists of two parts: a profit part that contributes to the profitability and a cost part that decreases the value of the objective function. If L denotes the set of frequent itemsets, and if $P_X = 1$ when itemset X is selected and $P_X = 0$ when itemset X is not selected, then the profit part contains the sum over all frequent itemsets of the profit margin $M(X)$

associated with each itemset $X \in L$. In contrast, if $\mathcal{I}$ denotes the list of all product items, and if $Q_i = 1$ when product i is selected and $Q_i = 0$ when product i is not selected, then the cost part of the objective function contains the sum of all product related costs $Cost_i$ associated with each product $i \in \mathcal{I}$. These costs include all costs that can directly be related to the product, like inventory and handling costs.

Thus, the objective function contains both frequent itemsets X and product items i. For now, it is sufficient to know that margins $(M(X))$ contribute positively to the profit and are associated with frequent itemsets (not individual items). How those margins per frequent itemset are calculated will be discussed later. Furthermore, the costs $(Cost_i)$ decrease the profit of the selected set of products and are expressed as a total cost per product (and not per frequent itemset).

The Constraints

The PROFSET framework is defined as a constrained optimization problem. Besides the Boolean nature of the decision variables P_X and Q_i, the optimization model therefore contains two more constraints. Firstly, there is a constraint on the total number of products allowed to be selected by the model (2) since there is usually limited shelf space available. Secondly, there is a set of constraints (3) that links product items $\mathcal{I}$ to itemsets X. The reason is that the objective function is expressed in terms of two different Boolean decision variables P_X and Q_i between which a relation exists. Indeed, product items i are contained in frequent itemsets X. Consequently, if the model tries to maximize the objective function by selecting frequent itemsets with high profit margins (i.e. the X for which $M(X)$ is large), it implicitly also selects the products contained in those selected itemsets. However, if the selection of frequent itemsets would not have been constrained, all frequent itemsets would be selected since this would produce the maximum achievable value for the positive part of the objective function. Consequently, the maximum number of products allowed under constraint (2) would be exceeded. Therefore, these constraints (3) implicitly link the frequent itemsets in the objective function to their corresponding items on which constraint (2) places a maximum value of items that can be selected ($ItemMax$).

Altogether, the PROFSET constrained optimization model will thus select those product items i that maximize the value of the objective function as a result of direct and cross-selling effects between the selected products, minus their corresponding costs.

Profit Parameters

The profit parameters $M(X)$ in the PROFSET model are associated with frequent itemsets X and express the contribution generated by a particular frequent itemset. However, the calculation of these parameters is not straightforward (see section 5.4.3.2). The reason is that if an itemset $\{i_1, i_2, i_4\}$ is frequent and thus contained in the list of frequent itemsets L, all its subsets $\{i_1\}$, $\{i_2\}$, $\{i_4\}$, $\{i_1, i_2\}$, $\{i_1, i_4\}$ and $\{i_2, i_4\}$ are also members of L. Consequently, when the gross margin generated by a particular sales transaction $M(T)$ must be distributed over the list of frequent itemsets $X \in L$, two particular problems arise[24]. Firstly, the problem of double counting must be avoided. Indeed, $M(T)$ should be distributed over mutually exclusive itemsets ($X, Y \in L$ and $X \cap Y = \emptyset$), i.e. itemsets who do not have any items in common. Secondly, it is not directly clear over which frequent itemsets the profit margin should be distributed. In fact, if the purchase intentions of the customers would have been known, the profit margin could be distributed over the frequent itemsets representing those purchase intentions. Take for instance the above itemset, if a consumer wanted to purchase i_1 and i_4 together, then the margin could be distributed over $\{i_1, i_4\}$ and $\{i_2\}$. However, these purchase intentions are unknown and can

[24] Thus, for each transaction (basket) in the database, the margin generated by the individual items in that transaction is distributed over one or more frequent itemsets. This means that the margins are allocated to the frequent itemset by running through the database record per record. Furthermore, since the margins are not directly available from the receipts, a pre-processing step is carried out to calculate the margin per item for each transaction.

not be inferred from the data with certainty. In fact, these purchase intentions can only be reliably obtained by interviewing the consumer when leaving the supermarket. Therefore, later in section 5.4.3.2, different strategies will be proposed to distribute the profit margin of a sales transaction $M(T)$ over the frequent itemsets contained in that transaction.

Cost Parameters

Finally, the PROFSET objective function contains an aggregate cost parameter $(Cost_i)$ per product calculated over the entire period of data collection. This cost parameter reflects the costs that can be directly attributed to the product, such as product handling and inventory costs. Product handling costs relate to the physical handling of the goods. Inventory costs can include several costs related to stocking physical goods, such as the cost for heating (or cooling) and renting of inventory space. Both handling and inventory costs per product are a function of the rotation speed and the shelf space allocated per product. The higher the sales rotation of a product, and the lower the shelf space allocated per product, the higher the product handling cost. The inventory cost depends on the rotation speed and the shelf space in so far that the product consumes energy (e.g. in freezers) and consumes inventory space that needs to be rented. In practice, these costs are not always easy to obtain, especially product handling costs. In fact, for our experiments, the retailer could not provide these data (unfortunately) and therefore we will not include them into our empirical results[25]. However, if cost data would have been available, they could be included either in the calculation of the net contribution margin per product as a 'unit cost per product', or as an aggregate cost over all sold products of a particular SKU for the entire period of data collection. For reasons of presentation, we decided to include them as an aggregated cost separately into the model specification under the 'cost part' of the objective function.

[25] In our experiments, we will thus use the sales value of items instead of the margins per item. Therefore, the optimization model will maximise the sales value of the selected set of products instead of the profit margin.

An Example

Suppose $\mathcal{I}=\{cola,\ peanuts,\ cheese,\ beer,\ crisps\}$ and association rule analysis identifies the following list L of frequent itemsets[26] $X \in L$ and their associated profit margins $M(X)$:

$X_1 = \{cola,\ peanuts\}$	$M(X_1) = 10$
$X_2 = \{peanuts,\ cheese\}$	$M(X_2) = 20$
$X_3 = \{cola,\ beer\}$	$M(X_3) = 30$
$X_4 = \{beer,\ crisps\}$	$M(X_4) = 25$

Furthermore, in order to enable a clear comparison with the example results of model A.2 (see later in the text), we assume that the cost for each product is equal. As a result, they do not need to be included in the objective function because they do not influence its optimal value. Finally, suppose that $MaxItems = 2$, i.e. only 2 out of 5 items can be selected by the model. Then, the PROFSET framework can be written as:

$$\text{Max } Z = 10P_1 + 20P_2 + 30P_3 + 25P_4$$
$$\text{s.t.}$$
$$Q_1 + Q_2 + Q_3 + Q_4 + Q_5 = 2$$
$$Q_1 \geq P_1 \quad Q_2 \geq P_1$$
$$Q_2 \geq P_2 \quad Q_3 \geq P_2$$
$$Q_1 \geq P_3 \quad Q_4 \geq P_3$$
$$Q_4 \geq P_3 \quad Q_5 \geq P_4$$
$$Q_1, Q_2, Q_3, Q_4, Q_5, P_1, P_2, P_3, P_4 \in \{0,1\}$$

It is not so difficult to see that $Cola = 1$, $Peanuts = 0$, $\text{Cheese} = 0$, $\text{Beer} = 1$, and $Crisps = 0$ maximizes the value of the objective function (Z=30) and satisfies all constraints. In other words, any other combination of products that satisfies the

[26] Note that, for reasons of simplicity, we have not included single itemsets in this example although in practice they would be included.

constraints would yield a lower total profitability from cross-selling (minus the product costs) and is thus not optimal.

The calculation of the PROFSET model is carried out by an extremely efficient Mixed Integer Programming (MIP) solver, called CPLEX 6.5 [34]. CPLEX 6.5 is a commercial operation research software (www.cplex.com) and uses a branch-and-bound (with cuts) algorithm, which solves a series of Linear Programming (LP) subproblems to solve large MIPs. But, since a single MIP generates many LP subproblems, MIPs can be very computer intensive and require significant amounts of physical memory. The reason is that the branch-and-bound tree may be as large as 2^n nodes, where n equals the number of binary variables, such that a problem containing only 30 binary variables (i.e. the number of frequent itemsets in this case) could produce a tree having over one billion nodes. Therefore, typically a stopping criterion is being set, e.g. a time limit or a relative optimality criterion, the latter specifying that the search for the optimal solution is aborted if the current best solution is $x\%$ below the best possible integer solution. In fact, our model contains as many binary variables as there are frequent itemsets discovered during association rule mining. Furthermore, the number of constraints of type (3) equals the number of frequent itemsets multiplied by the number of items contained in each frequent itemset since for each frequent set, there are as many constraints as there are items contained in that set. Finally, there is only one type (2) constraint. Concrete details about the number of variables, constraints and the execution time of the PROFSET model on the data in this dissertation are given in the empirical results section (section 5.6).

Optimization inside + outside the hitlist (A.2)

The PROFSET model specified before maximizes the profitability from main effects + cross-selling effects *between products in the optimal set* (see A.1 in table 5.1). In other words, cross-selling effects that exist from products inside the optimal set (hereafter referred to as the hitlist) to products outside the optimal set do not increase the value of the objective function. This can be illustrated by the example in the previous section. For instance, 'peanuts' is not a member of the hitlist and thus $P_1=0$ such that the margin of the frequent itemset $M(X_1) = 10$ does not increase the value of the objective function. This formulation of the PROFSET model is suitable when the hitlist is isolated from the rest of the assortment. However, we realized that there are circumstances in which not only the profitability from main effects + cross-selling effects between items in the hitlist should be maximized, but that also the cross-selling effects from items in the hitlist towards items outside the hitlist should be taken into account during optimization (see A.2 in table 5.1). To illustrate this, consider the second optimization problem.

The second model deals with the problem of allocating attractive but scarce shelf space to products, e.g. at end-of-aisle locations, at eye-level locations, at locations close to the outer walking paths in the store, etc. Especially because of their visual attractiveness, these locations are often sold to manufacturers who pay high prices to have their products at those locations. However, besides the direct profit from selling those locations to manufacturers, there is an indirect profit or loss attached to those locations as well. Indeed, if consumers would only visit these attractive locations for 'cherry-picking', and would not go inside the gondolas, then there is an alternative loss associated with these locations as a result of decreased cross-selling (i.e. lost sales) with other items inside the gondolas. Consequently, those attractive locations should contain a carefully tuned mix of sponsored products and of products that encourage customers to go inside the gondola and purchase other products too (i.e. products that have a positive radiation effect on other products inside the gondolas). Therefore, in this dissertation, the idea is developed that not only the profit of the hitlist should now be maximized (like in the first model), but also the profit resulting from cross-selling with other products at regular positions in the store. We therefore distinguish between two different products, display products (Q') at visually attractive positions and regular products (Q) at other positions in the store. This specification of the optimization criterion leads to a slightly different specification of the PROFSET model, as illustrated below.

Model Formulation

The PROFSET model for this second application is different from the one specified before in so far that the objective function and the constraints that link the items to the frequent itemsets are slightly different. Note that |X| denotes the number of items in the itemset.

$$\max\left(\sum_{X\in L} M(X)P_X\right) \tag{1}$$

s.t.

$$\sum_{i\in I} Q'_i = ItemMax \tag{2}$$

$$\forall X \in L: \left\lceil \frac{1}{|X|}\sum_{i\in X} Q'_i \right\rceil = P_X \tag{3}$$

$$\forall P_X, \forall Q'_i \in \{0,1\} \tag{4}$$

PROFSET A.2 Model

The Objective Function

The objective function (1) is similar to the one discussed in the PROFSET A.1 model, except from the absence of product cost information. The reason is that with regard to product handling or inventory costs, we assume that it does not matter whether the product is selected as a display product or a regular product. In the end, they all end-up in the store and we assume that the cost for handling the product is identical, regardless of whether it is a display or a regular product. If the costs for handling the product would be different according to the position in the store, then it would be wise to account for product handling costs in the objective function.

The Constraints

Constraint (2) specifies that the number of products to be selected for attractive locations in the store is limited. Furthermore, constraint (3) requires some additional explanation. It specifies that whenever at least one product in the itemset is a display product ($Q'_i = 1$), the itemset is selected for the objective function ($P_X = 1$). This is enforced by the ceiling function '$\lceil \text{arg} \rceil$'. This function returns the closest integer value that is bigger than its argument (e.g. $\lceil 2.5 \rceil = 3$). Therefore, if at least one of the products in a particular itemset is a display product, then the ceiling function forces that itemset to be chosen by the objective function. On the other hand, if none of the products is a display product, then the ceiling function returns a 'zero' value and thus $P_X = 0$. In other words, if an itemset does not contain any display products, then there is no cross-selling gross margin contributing to the value of the objective function. This is illustrated for a two-itemset in the truth table below.

Q'_1	Q'_2	P_X
0	0	0
1	0	1
0	1	1
1	1	1

Table 5.2: Truth table for a two-itemset

The margin of a frequent itemset is therefore only added to the value of the objective function if at least one of the items in the itemset is a display product. Furthermore, products that are not selected by the model are not display products and thus they are regular products.

An Example

Suppose we have the same example again like for the A.1 model, i.e., $\mathcal{I}$ = {*cola, peanuts, cheese, beer, crisps*} and association rule analysis identifies the following list L of frequent itemsets $X \in L$ and their associated profit margins $M(X)$:

X_1 = {cola, peanuts} $M(X_1) = 10$

X_2 = {peanuts, cheese} $M(X_2) = 20$

X_3 = {cola, beer} $M(X_3) = 30$

X_4 = {beer, crisps} $M(X_3) = 25$

Furthermore, suppose that $MaxItems = 2$, i.e. only 2 out of 5 items can be selected by the model as a display product, the other three must be regular products. Then, the model can be written as:

$$\text{Max } Z = 10P_1 + 20P_2 + 30P_3 + 25P_4$$

$$\text{s.t.}$$

$$Q'_1 + Q'_2 + Q'_3 + Q'_4 + Q'_5 = 2$$

$$\lceil \tfrac{1}{2}(Q'_1 + Q'_2) \rceil = P_1$$

$$\lceil \tfrac{1}{2}(Q'_2 + Q'_3) \rceil = P_2$$

$$\lceil \tfrac{1}{2}(Q'_1 + Q'_4) \rceil = P_3$$

$$\lceil \tfrac{1}{2}(Q'_4 + Q'_5) \rceil = P_4$$

$$Q'_1, Q'_2, Q'_3, Q'_4, Q'_5, P_1, P_2, P_3, P_4 \in \{0,1\}$$

It is not so difficult to see that (note the difference with the optimal solution for the A.1 model) $Cola = 0$, Peanuts = 1, $Cheese = 0$, $Beer = 1$ and $Crisps = 0$ maximizes the value of the objective function (Z=85) and satisfies all constraints. Consequently, *peanuts* and *beer* will be display products and cola, cheese and crisps will be regular products. Any other combination of display products that satisfies the constraints would yield a lower total profitability from cross-selling since it leaves at least one itemset unselected and thus does not maximize the value of the objective function. In other words, by putting any other combination of products in the attractive display area, there will be less cross-selling to the non-display area and thus the solution would be sub-optimal. In this model specification, there are as many type (3) constraints and decision variables in the objective function (P_X) as there are frequent itemsets in the model.

5.4.3.2 The gross margin allocation rule (B)

A second decision criterion leading to alternative PROFSET model specifications was already shortly introduced before. It concerns the problem of distributing the margin of sales transactions to frequent itemsets. We will distinguish between two allocation rules, i.e. support-based allocation and dependency-based allocation.

Support-based allocation (B.1)

The idea is that some purchase combinations occur more frequently than others because consumers consider them as purchase complements. In that case, frequent itemsets can be interpreted as frequent purchase intentions of consumers. This raises the question whether it is possible to identify which and how many purchase intentions are embedded in each sales transaction. In fact, a single transaction T can contain multiple frequent itemsets $X \subseteq T$ such that it is not straightforward to determine which purchase intentions have played at the time of purchase. We therefore define the concept of a *maximal frequent subset* of a transaction.

Definition 5.1: Maximal frequent subset
Let F be the collection of all frequent subsets of a transaction T. Then $X \in F$ is called maximal, denoted as X_{max} , if and only if $\forall Y \in F : |Y| \leq |X|$. ■

Using this definition, the following rationale will be adopted to allocate the margin of a sales transaction $M(T)$. If there exist only one maximal frequent subset $X_{max} \subseteq T$, then we allocate the proportion of $M(T)$ that is attributable to X_{max} (i.e. $m(X_{max})$) to $M(X)$. However, if multiple maximal frequent subsets exist, then one maximal frequent subset X_{max} (see definition) will be drawn from all maximal frequent subsets according to the probability distribution Θ_T with

$$\Theta_T(X_{\max}) = \frac{Support(X_{\max})}{\sum_{Y_{\max} \in T} Support(Y_{\max})} \qquad (5.1)$$

The margin of the selected maximal frequent subset $m(X_{max})$ of T is then assigned to $M(X)$ and the process is repeated for $T \setminus X_{max}$. Or, in pseudocode:

for every transaction $T \in \mathcal{D}$ **do** {
  while ($T \neq \varnothing$) or ($\exists X \subseteq T$) **do** {
    if $\exists!\, X_{max} \subseteq T$
      then $M(X) := M(X) + m(X_{max})$;
        else draw X_{max} from all maximal frequent subsets
        using probability distribution Θ_T ;
        $M(X) := M(X) + m(X_{max})$
        with $m(X_{max})$ the profit margin of X_{max} in T ;
    $T := T \setminus X_{max}$;
  }
}
return all $M(X)$;

Figure 5.2: Pseudocode for support-based margin allocation

This support-based allocation is illustrated by the following example. Say during gross margin allocation, we are given a transaction T = {cola, peanuts, cheese} and assume that table 5.3 contains all frequent subsets of T.

Frequent sets	Support	Maximal	Unique
{cola}	10%	No	No
{peanuts}	5%	No	No
{cheese}	8%	No	No
{cola, peanuts}	2%	Yes	No
{peanuts, cheese}	1%	Yes	No

Table 5.3: Frequent subsets of transaction T

In this example, there is no unique maximal frequent subset of T. In fact, there are two maximal frequent subsets, namely {cola, peanuts} and {peanuts, cheese}. Consequently, it is not clear to which maximal frequent subset the gross margin $M(T)$ should be allocated. Moreover, we would not allocate the entire gross margin to the selected itemset, but rather the proportion $m(X)$ that corresponds to the items contained in the selected maximal frequent subset.

Now, how can one determine to which of the two maximal frequent subsets the profit margin should be allocated? The crucial idea here is that it really depends on what has been the purchase intention of the consumer at the time of purchase. However, since we do not possess this important piece of information, we use the support of the maximal frequent subsets of T as a probabilistic estimation. Indeed, if the support of a frequent subset is an indicator for the probability of occurrence of this purchase combination, then according to the data, customers (on average) purchase the maximal subset {cola, peanuts} two times more frequently than the maximal subset {peanuts, cheese}. Consequently, it could be argued that it is more likely that the consumer's purchase intention has been {cola, peanuts} instead of {peanuts, cheese}. This information is used to construct the probability distribution Θ_T, reflecting the relative frequencies of the maximal frequent subsets of T. Now, each time that a sales transaction T = {cola, peanuts, cheese} is encountered in the data ($\mathcal{D}$), a random draw from this probability distribution Θ_T will provide the 'most probable' purchase intention (i.e. frequent subset) for that transaction. Consequently, on average, in two of the three times that this transaction occurs maximal frequent subset {cola, peanuts} will be selected and $m(\{\text{cola, peanuts}\})$ will be allocated to $M(\{\text{cola, peanuts}\})$. After this, T is split up as follows: $T := T \setminus \{\text{cola, peanuts}\}$ and the process of assigning the remaining margin is repeated as if the new T were a separate transaction, until T is empty or does not contain a frequent set anymore.

Dependency-based allocation (B.2)

The idea of dependency-based allocation is similar to that of support-based allocation except for the use of a statistical methodology to allocate the gross margin $M(T)$ to a number of frequent subsets $X \subseteq T$. More specifically, the idea is that given a list of frequent itemsets L, we want to know which of those frequent itemsets are statistically significant in order to use only statistically significant itemsets for gross margin allocation. In order to illustrate the idea, consider the following example. Say, we have a database $|\mathcal{D}|$=1000 and a particular transaction T = {cola, peanuts, cheese, crisps} and we are given a list of frequent subsets of T in table 5.4.

Frequent sets	Support	Count
{cola}	10%	100
{peanuts}	5%	50
{cheese}	8%	80
{crisps}	3%	30
{cola, crisps}	1.5%	15
{cola, peanuts}	2%	20
{peanuts, cheese}	2.5%	25
{cola, cheese}	3%	30
{peanuts, crisps}	2%	20
{cola, peanuts, cheese}	1%	10
{cola, peanuts, crisps}	1%	10

Table 5.4: Frequent subsets of transaction T

The question is now how to distribute $M(T)$ over the different frequent subsets of T. For instance, should the gross margin be allocated to the single itemsets, the 2-itemsets or even the 3-itemset? We will use loglinear analysis to discover whether indeed the existing dependency in the 3-itemset can be declared from the k-itemsets with k<3. For instance, suppose a particular product combination {cola, peanuts, cheese} is (very) frequent. Is that just because {cola, peanuts} and/or {peanuts, cheese} and/or {cola, cheese} are (very) frequent, or is there something special about the triple that all three occur frequently in transactions? In the former case, the mining of the triple hasn't really found anything that couldn't have been deduced from the results of examining the pairs. In the latter case, the triple adds some new insight into the problem of identifying frequent co-occurring product combinations. It is the idea to use loglinear analysis (section 3.2.3.2) to examine this. This idea of using loglinear analysis to test the statistical significance of frequent itemsets was also adopted by DuMouchel and Pregibon [101].

Loglinear analysis, however, uses contingency tables to calculate the multi-way dependencies, which are clearly not given by the association rule analysis. Yet, any multi-way contingency table for a k-itemset can be derived from the support values of the k-itemset and its subsets. For the itemset {cola, peanuts, cheese}, this is illustrated in table 5.5.

Cola	**Peanuts**	**Cheese**	**Calculation**
0	0	0	#Transactions – sup{cola} – sup{peanuts} – sup{cheese} + sup{cola, peanuts} + sup{cola, cheese} + sup{peanuts, cheese} – sup{cola, peanuts, cheese}
1	0	0	sup{cola} – sup{cola, cheese} – sup{cola, peanuts} + sup{cola, peanuts, cheese}
0	1	0	sup{peanuts} – sup{cola, peanuts} – sup{peanuts, cheese} + sup{cola, peanuts, cheese}
0	0	1	sup{cheese} – sup{cola, cheese} – sup{peanuts, cheese} + sup{cola, peanuts, cheese}
1	1	0	sup{cola, peanuts} – sup{cola, peanuts, cheese}
1	0	1	sup{cola, cheese} – sup{cola, peanuts, cheese}
0	1	1	sup{peanuts, cheese} – sup{cola, peanuts, cheese}
1	1	1	support{cola, peanuts, cheese}

Table 5.5: Building a contingency table for {cola, peanuts, cheese} from support frequencies

A similar table, which is not shown here, can be constructed for the 3-itemset {cola, peanuts, crisps} and all the 2-item subsets in table 5.4. The significant interactions, as a result of loglinear analysis on those tables, are given in table 5.6. This table shows that both 3-fold interactions are not significant, but that all the 2-itemsets are significant at the $p<0.001$ level.

Interaction	**Chi-squared value**	**p-value**
{cola, peanuts}	40.46	<0.001
{cola, cheese}	57.33	<0.001
{cola, crisps}	37.53	<0.001
{peanuts, cheese}	78.36	<0.001
{peanuts, crisps}	92.87	<0.001
{cola, peanuts, cheese}	10.0	0.0016
{cola, peanuts, crisps}	4.02	0.0451

Table 5.6: Significant interactions identified by loglinear analysis

Given the significant itemsets and their respective chi-squared values, the distribution of the gross margin $M(T)$ of T = {cola, peanuts, cheese, crisps} is now carried out as follows.

Definition 5.2: Maximal significant subset
Let F be the collection of all significant subsets of a transaction T. Then $X \in F$ is called maximal, denoted as X_{max} , if and only if $\forall Y \in F : |Y| \leq |X|$. ■

If there exist only one maximal significant subset $X_{max} \subseteq T$, then we allocate the proportion of $M(T)$ that is attributable to X_{max} (i.e. $m(X_{max})$) to $M(X)$. However, if multiple maximal significant subsets exist, then one maximal significant subset X_{max} (see definition) will be drawn from all maximal significant subsets according to the probability distribution Θ_T with

$$\Theta_T(X_{\max}) = \frac{Chisq(X_{\max})}{\sum_{Y_{\max} \in T} Chisq(Y_{\max})} \tag{5.2}$$

The margin of the selected maximal significant subset $m(X_{max})$ of T is then assigned to $M(X)$ and the process is repeated for $T \backslash X_{max}$. Or, in pseudocode:

For every transaction $T \in \mathcal{D}$ **do** {
  while $(T \neq \varnothing)$ or $(\exists X \subseteq T)$ **do** {
    if $\exists! \, X_{max} \subseteq T$
      then $M(X) := M(X) + m(X_{max})$;
        else draw X_{max} from all significant subsets using probability distribution Θ_T;
    $M(X) := M(X) + m(X_{max})$;
    with $m(X_{max})$ the profit margin of X_{max} in T ;
    $T := T \setminus X_{max}$;
  }
}
return all $M(X)$;

Figure 5.3: Pseudocode for dependency-based margin allocation

Thus, if there is a unique significant maximal subset $X_{max} \subseteq T$, then the gross margin attributable to the items in X_{max} (i.e. $m(X_{max})$) will be allocated to the margin of the frequent itemset X. If, however, multiple significant maximal subsets of T exist, then all maximal significant subsets of T are included in the probability distribution Θ_T in proportion to the value of their chi-squared statistic, and one of them is drawn from this distribution as the winner[27].

[27] Since chi-squared values of different significant itemsets can be very close, the model might be over-sensitive to small differences in the chi-squared values. However, by using the probability distribution Θ_T we minimize the maximum error of being wrong, at least if we draw multiple times from the distribution. In other words, the error of doing a false allocation is averaged out over all draws from the distribution. Indeed, in the limit, the number of times that each significant itemset is selected will be proportional to the strength of its dependence in relation to the other significant itemsets. For the experiments on our data, we may expect this effect of proportionality to happen since the absolute support of each itemset is higher than or equal to 42 (see section 5.6.2.1).

In this example, there is no unique significant maximal subset of T, and thus the choice is between the 2-itemsets contained in table 5.6. Suppose that the subset {peanuts, crisps} is drawn from Θ_T. In that case, the gross margin m({peanuts, crisps}) within transaction T is added to the gross margin M({peanuts, crisps}) of the itemset {peanuts, crisps}. The analysis is then repeated for $T \setminus$ {peanuts, crisps} until T is empty or there are no statistically significant subsets of T left.

5.4.3.3 The Category Constraints (C)

Finally, the PROFSET model can be specified with or without product taxonomy information.

Without category constraints (C.1)

The most general PROFSET model, as specified in section 5.4.3.1, does not take into account category constraints on the items in the model. In other words, the decisions about which products to include in the hitlist are taken on the lowest level of detail (SKU) without any product taxonomy information taken into account.

Although, from a practical point of view, this may not be very wise, it enables the PROFSET model to fully exploit the existing main and cross-selling effects between products without any constraints. The model will therefore always yield the highest possible value of the objective function and will show the full potential of the model in identifying significant cross-selling effects between products.

However, the solution may not be very realistic from a retailer point of view since we expect some product categories to be over-represented in the optimal solution (hitlist). This is straightforward since some product categories are visited very frequently and thus appear in most of the baskets, such as waters and soft drinks. The conclusion from the model might therefore be to sacrifice some product categories (reduce the width of the assortment) in exchange for a more product alternatives in popular product categories (deeper product categories). For instance, the model might sacrifice the 'men's shaving' category in exchange for another brand of water, although the hitlist already contains a few water brands, because the water brand is overall more profitable than any product in the men's shaving category.

With category constraints (C.2)

None of the models presented in the preceding sections take into account category constraints on the items in the model. However, in practice, retailers often use product taxonomies to categorize items into higher-level product categories (see section 4.5.1.2). Furthermore, assortment decisions are often taken first on a more strategic level, i.e. higher in the product taxonomy, by specifying the depth and width of the product assortment in relation to the store formula and the strategic positioning of the retail store, before translating the choices on the product category level down to the individual products. For instance, the retailer may want to balance the assortment of the convenience store or the shelf space available for display products such that some product categories are included and others are not, or some product categories obtain more shelf space than others. For example, the retailer may want to compose the assortment such that it contains a well-balanced choice of core convenience products (beverages, ready-made meals, toilet paper, napkins) together with a well-tuned set of impulse products (candy bars, biscuits, tea-stoves) that must increase the consumer's total expenditure when buying from the automated shop. Indeed, retailers are interested in adding items whose sales will not be made at the expense of the core items but may help increase the sales of those core items (sales complements) [218]. In the case of allocating products to attractive display locations in the store, the retailer may want to balance the assortment such that it is as much in line with the store image as possible. This enables the specification of constraints both on the category level and on the individual product level. For instance, a minimum and/or maximum number of products per category are allowed to enter the hitlist, hereby putting more weight on product categories that are crucial to the store image. Additionally, if the retailer evaluates that particular products should be included anyway (e.g. core assortment products), even though they are not very profitable from a cross-selling point of view, such constraints on the individual product level can be easily included by forcing the model to select them anyway.

5.5 Empirical Setup

In order to test the effect of the different alternative options (optimization criterion, margin allocation rule and category constraints) on the results of the PROFSET model, four different model implementations were created, as illustrated in table 5.7 below.

5.5.1 Model Specifications

The empirical setup is conducted such that each option alternative is tested exactly once against the other by constructing 4 different models.

For instance, the results of model 1 and model 2 will be compared on their difference with respect to the inclusion or exclusion of category constraints, ceteris paribus the other options (optimization criterion and allocation rule).

Model Specification	**Optimization Criterion**	**Allocation Rule**	**Category constraints**
Model 1	A.1	B.2	C.2
Model 2	A.1	B.2	C.1
Model 3	A.2	B.2	C.1
Model 4	A.2	B.1	C.1

Table 5.7: Overview of experimental setup

The results of model 2 and 3 will be compared on their difference with respect to the optimization criterion being used, ceteris paribus the type of allocation rule and the exclusion of model constraints. Finally, the results of model 3 and model 4 will be compared on their difference with respect to the type of allocation rule being used, ceteris paribus the optimization criterion and the exclusion of category constraints. This way, all alternative options are compared exactly once against each other whilst holding all other options constant. Furthermore, the number of brands selected by each model will be constrained to 150. This again makes comparison across models straightforward[28]. However, before discussing the empirical results, we will present each of the 4 models in detail.

5.5.1.1 Model 1

The first model describes the optimization problem where the objective is to compose a product assortment for the automated convenience store and where the retailer specifies a number of category constraints that must be satisfied by the model. Furthermore, the allocation rule to distribute the gross margin from sales transactions to frequent itemsets is based on the statistical analysis of dependencies between the items in the transactions by means of loglinear analysis. The model therefore combines the A.1, B.2 and C.2 option alternatives (see figure 5.4).

The variables and parameters of the model are identical to those introduced in the general formulation of the PROFSET model (section 5.4.3.1), except from a new set of parameters $C_1, \ldots, C_n$ that represent the product categories. Each product category contains a set of items and it is assumed that the allocation of items to those product categories is determined in advance by the

[28] The choice of selecting only 150 brands could be debated, especially for the A.2 application where there are probably much more attractive positions in the store. However, in order to make a fair comparison between the suggested models, this number should be the same for all models. Indeed, from a technical point of view, it is not possible to make a fair comparison between the A.1 and A.2 model if the number of products selected is different. On the other hand, from the marketing point of view, both applications are different and thus in practice one would select a different number of products for each application. Yet, we have opted for a technical comparison.

retailer. As a result, an extra set of constraints (4) appears specifying the minimum and/or maximum number (or an exact number) of items to be selected for each product category (see appendix 3). This way, categories can be included and given an importance weight in the final product assortment, or can be excluded from the automated shop assortment overall. Additionally, these constraints enable the flexibility to introduce qualitative marketing domain knowledge about the weight that should be given to different product category types, including convenience products and impulse products.

$$\max\left(\sum_{X \in L} M(X) P_X - \sum_{c=1}^{n} \sum_{i \in C_c} Cost_i Q_i\right) \tag{1}$$

s.t.

$$\sum_{c=1}^{n} \sum_{i \in C_c} Q_i = ItemMax \tag{2}$$

$$\forall X \in L, \forall i \in X : Q_i \geq P_X \tag{3}$$

$$\forall C_c : ItemMin_{C_c} \leq \sum_{i \in C_c} Q_i \leq ItemMax_{C_c} \tag{4}$$

$$\forall P_X, \forall Q_i \in \{0,1\} \tag{5}$$

Figure 5.4: Specification of Model 1

5.5.1.2 Model 2

Model 2 is very similar to model 1, except from the fact that it does not contain product category constraints that limit the optimization model to freely select as much products form each category as needed to maximize the value of the objective function. The other decision criteria (A and B) are identical, i.e. the optimization criterion and the gross margin allocation rule are identical to model 1. This specification therefore leads to the most general definition of the PROFSET model as introduced before, and repeated here for the sake of clarity.

$$\max\left(\sum_{X\in L} M(X)P_X - \sum_{i\in I} Cost_i Q_i\right) \tag{1}$$

s.t.

$$\sum_{i\in I} Q_i = ItemMax \tag{2}$$

$$\forall X \in L, \forall i \in X : Q_i \geq P_X \tag{3}$$

$$\forall P_X, \forall Q_i \in \{0,1\} \tag{4}$$

Figure 5.5: Specification of Model 2

It can be seen that the model does not contain any product category constraints in contrast to the specification of model 1 in the previous section.

5.5.1.3 Model 3 and model 4

Model 3 and model 4 can be treated together since they only differ with respect to the gross margin allocation rule, i.e. the calculation of the $M(X)$ parameters is different, which is not shown in the model specification. In other words, the values of the $M(X)$ parameters will be different in both model specifications. In both models, the optimization criterion is A.2 and there are no product category constraints considered in the model (C.1). Both models can therefore be presented as illustrated in figure 5.6.

$$\max\left(\sum_{X\in L} M(X)P_X\right) \quad (1)$$

s.t.

$$\sum_{i\in I} Q'_i = ItemMax \quad (2)$$

$$\forall X \in L: \left\lceil \frac{1}{|X|}\sum_{i\in X} Q'_i \right\rceil = P_X \quad (3)$$

$$\forall P_X, \forall Q'_i \in \{0,1\} \quad (4)$$

Figure 5.6: Specification of Model 3 and 4

5.6 Empirical Results

This section discusses the empirical results of the different models introduced in section 5.5.1.1 to 5.5.1.3. We have chosen, however, not to report all the details for each individual model since many of the work carried out to estimate each model is very similar across all models. Furthermore, each model outputs a selection of 150 products such that a detailed comparison would quickly lead to confused results. We therefore think that it will contribute to the legibility of the text by discussing the results of model 1 in detail, and subsequently highlighting some interesting differences between the different models as suggested before in table 5.7 (section 5.6.3 and further). The interested reader, however, can find the detailed results for each model in the appendices 4 to 7 at the end of this text.

Before going into the detailed results for model 1, we will first discuss some important issues related to the data that have been used for this study.

5.6.1 Data Preparation

The empirical analysis carried out in this chapter relies on the supermarket data that were already introduced and used previously in this text. The raw data (see dataset in section 2.3) consists of 82497 retail sales transactions, collected over a period of 21 weeks[29], from a Belgian supermarket store. On the lowest level of detail, the data contains 16404 different SKU's. However, in order to make a fair comparison between the different models, some choices need to be made with respect to the data such that the results are comparable across the different experiments. More specifically, choices need to be made with regard to 1) the depth of analysis, 2) the width of the analysis, and 3) the scope of the analysis.

5.6.1.1 Depth of the analysis

A first choice with regard to the data needs to be made with respect to the depth of the analysis, i.e. whether the data should be analysed on the level of the SKU, the brand, or even the category.

The choice is not really clear. Probably a multi-phased approach is most preferred where an analysis is first carried out on the category level to determine the most important categories, followed by an analysis on the brand level and finally on the SKU level. This multi-phased approach also follows the logic of building retail assortments where choices are made in a hierarchical top-down way. For practical reasons, however, we have chosen to carry out the analysis on the brand level and we assume that the retailer has already made choices with regard to which product categories that will be contained in the convenience store. Hereto, all SKU's were regrouped into different brands irrespective of their package size. This regrouping of SKU's into brands reduced the number of products from 16404 SKU's to 6866 different brands.

---

[29] The weeks surrounding the Christmas period were not included in the analysis because the purchase behaviour in these weeks is rather different from the rest of the data collection periods.

5.6.1.2 Width of the analysis

A second choice regarding the data should be made with respect to the width of the analysis, i.e. whether all products should be included in the analysis or not. We decided to exclude brands from product categories that are usually not contained in an automated convenience store, like the Shop24. It concerns products like fuels, clothing, bedding, round games, garden equipment, frozen food, and a few others. Frozen food products are excluded for the obvious reason that the automated shop is refrigerated but not cold enough to contain frozen products. Furthermore, tobacco products and spirits or distilled based beverages (such as Bacardi Breezer) were removed from the analysis because the Belgian legislation does not permit tobacco products and spirits or spirit-based beverages to be sold in vending machines that are accessible to under-aged people. In contrast, and not at all straightforward in terms of legislation, fermented beverages below 21% alcohol are allowed, including for instance light beers, heavy beers, wines, ports, champagne, sherry, etc. For the A.2 application, where the objective is to reorganize the location of products in a traditional store environment into attractive and regular shelf positions, the exclusion of the product categories may not be needed. However, in order to make a fair comparison between models 2 and 3, it is better to exclude them from the analysis since otherwise the effect of including cross-selling effects outside the hitlist can not be evaluated on the same data. Furthermore, four product categories were additionally removed from the analysis, i.e., fresh fruit/vegetables, fresh meat and fresh cheese. The motivations are twofold: firstly, for these product categories, no detailed SKU data was available (see section 2.3.3.2) and secondly, they appear in almost any basket (they have extremely high support) such that they would bias the analysis.

5.6.1.3 Scope of the analysis

A final choice regarding the data should be made with respect to whether all retail baskets will be included in the analysis or whether a more targeted selection should be made.

Firstly, in our experiments, we decided to include all retail baskets. Although this is probably a reasonable choice for the selection of products for attractive locations in the store, it is probably a false assumption within the context of selecting products for a convenience store. Indeed, the entire collection of retail baskets will reflect the purchase behaviour of different types of customers, such as weekend stock-up shoppers who buy large amounts of products, week fill-in shoppers (buying only a few items), emergency shoppers, etc. Clearly, the behaviour of time-pressured convenience store shoppers, such as those who shop in a convenience store, will be better reflected by fill-in baskets than by stock-up baskets. Therefore, in section 5.7, a sensitivity analysis will be carried out where only fill-in baskets are selected for the analysis instead of all the baskets that are used in the subsequent experiments.

Secondly, outlier baskets were removed. Outlier baskets were identified as those that contain over 80 different products.

In the subsequent sections, we will discuss the results of the 4 models based on these data.

5.6.2 Model 1 in Detail

This section presents the detailed results and steps taken to estimate model 1. More specifically, we will discuss the extraction of frequent itemsets from the prepared data, significance analysis of the frequent itemsets using loglinear analysis, the construction of the model including category constraints, and finally the discussion of the empirical results.

5.6.2.1 Association rule analysis

Frequent itemsets were generated with the Apriori algorithm (section 4.3.3) on the prepared data with an absolute support count=42, which equals a support percentage of 0.05%. Therefore, a brand or a set of brands is considered frequent if it occurs at least in 42 baskets, which corresponds to a brand or set of brands to be purchased at least twice per week. This resulted in 5875 frequent sets of size 1 to 4, as illustrated by table 5.8.

Itemset size	1	2	3	4
Count	2894	2775	171	35

Table 5.8: The number of frequent sets for different sizes

The table shows that the majority of the frequent sets are of size 1 and 2 and that bigger sets are rather exceptional. The amount of time needed to generate these frequent sets of brands equals 54 seconds on a standard Pentium III 450 Mhz machine. At this point, the choice for the (low) support threshold could be disputed. This is partially justified. However, an even lower support (0.009%) was used by De Schamphelaere, Van den Poel and Van Kenhove [92] in the context of association rule mining for do-it-yourself stores. Furthermore, it turns out that the model selection results are quite robust with regard to relatively small changes in the support parameter. This is not so surprising. On the one hand, setting the support threshold too low will lead to the generation of more frequent itemsets. But even if they would be statistically significant, their support is usually too low to have a significant influence on the value of the objective function of the product selection model. For instance, the lowest-support item (Tomato soup with meat balls private label) that is contained in the optimal set of products (see appendix 4) has an absolute support count of 102, which is still far above the minimum support threshold of 42. Itemsets with very low support therefore do not usually influence the product selection. On the other hand, the higher the support threshold, the fewer items will appear in the itemsets such that the product assortment from which the optimization model must choose its optimal items will be too small and will include only category winners. We therefore decided to allow a product (or product combination) to participate in the competition if it appears at least twice a week.

To conclude, the absolute support threshold was selected by means of trial-and-error exercises (absolute support ranging from 20 to 50) and according to common sense logic (a minimum of 2 purchases per week was interpreted as acceptable).

5.6.2.2 Loglinear analysis and gross margin allocation

The loglinear analysis on the 2981 multi-item frequent sets (k=2 to 4) was carried out by a batch program in SAS. For each frequent itemset, the program finds the most unsaturated model that fits the data well enough and it returns the significance values of the interactions for this most unsaturated model. Overall, from the 2981 frequent itemsets only 1758 (60%) of the multi-item associations (k=2 to 4) show statistical significance (table 5.9).

Itemset size	1	2	3	4
Count	2894	1688	35	35

Table 5.9: Statistically Significant Associations

The chi-square and p-values for the associations are used to distribute the gross margin of all the transactions $T \in \mathcal{D}$ over the significant itemsets according to the dependency-based allocation heuristic. The result is a list of significant frequent itemsets with their respective sales revenues. This information serves as the input for the PROFSET model.

5.6.2.3 PROFSET model implementation

Based on the list of significant frequent itemsets the objective function of the PROFSET model contains 1758 multi-item and 2894 single itemsets and thus it contains 4652 Boolean decision variables (P_X) altogether. Furthermore, the model contains 6515 (i.e., 1x2894 + 2x1688 + 3x35 + 4x35) constraints of type 3. The number of items to enter the automated convenience store (constraint 2) was limited to 150, as was discussed before.

Finally, for the 126 selected product categories, model 1 contains 2 additional constraints per category[30] (i.e. 2 x 126 = 252) to specify the minimum and maximum number of products allowed to be selected by PROFSET (see appendix 3). In practice, these product category constraints will typically strongly depend on the retailer's preferences and/or store image considerations. However, we have made a 'reasonable' selection.

5.6.2.4 Empirical results

Figure 5.7 presents the iteration log of the calculation of model 1 and shows some interesting results.

Firstly, it shows that the optimal value of the objective function for this model equals 13640052 BEF. The value of the objective function should however be interpreted with great care. First of all, the objective value is not a valid measure to estimate the expected sales revenue of the selected set of products. The reason is that the model does not take into account brand switching effects in the case of stockouts. Furthermore, the objective value is based on the traffic intensity in the traditional store, which will be higher than for the automated shop.

---

[30] Note, however, that the model is not forced to select at least one product for each category (see appendix 3). In fact, for 44 out of the 126 product categories, the model is free to select null or more products. Only for the other 82 categories, the model should select at least one or more products. This is important since it determines the model's remaining 'degrees of freedom' to choose the most profitable products.

Iter:1 Dual objective = 44910609.00
Iter:149 Dual objective = 44509255.00
Iter:574 Dual objective = 43516381.00
Iter:815 Dual objective = 42220230.00
Iter:1160 Dual objective = 40233673.00
Iter:1537 Dual objective = 37335204.00
Iter:1982 Dual objective = 34306357.00
Iter:2372 Dual objective = 31742338.00
Iter:2737 Dual objective = 28503703.00
Iter:3185 Dual objective = 24795936.00
Iter:3621 Dual objective = 20887650.00
Iter:3995 Dual objective = 18830347.00
Iter:4401 Dual objective = 16587898.00
Iter:4725 Dual objective = 14907491.00
Iter:4951 Dual objective = 14011061.50
Iter:5121 Dual objective = 13654554.00
Root relaxation solution time = 2.41 sec.
Fixing integer variables, solving final LP.
Tried aggregator 1 time.
LP Pres. Elim. 6415 rows and 11338 cols.
All rows and columns eliminated.
Presolve time = 0.17 sec.
Proven optimal solution.
MIP Solution: 13640052 (5130 iterations)
Final LP: 13640052 (0 iterations)
Best integer solution possible 13540052
Absolute gap: 0.0000
Relative gap: 0.0000

Figure 5.7: PROFSET calculation iteration log

The expected sales revenue of the selected set of items is therefore probably significantly lower and can not be inferred from the objective value of the optimization model.

A better way to evaluate the value of the objective function is to think of the optimal value as the sales revenue that would be achieved by reducing the assortment of the traditional store down to the smaller selected assortment, given that none of the existing customers would switch to another store. Obviously, this is very unlikely since they will have to shop in another store to complement their purchase with items that are no longer stocked in the given store, which will probably cause most of the customers to switch entirely from one store to another where they can still do one-stop-shopping.

Another useful evaluation is the comparison of the objective value of the optimization model against the value obtained by selecting products according to a rule of thumb. Indeed, a good rule of thumb would be to select the highest revenue brands from each category (subject to the category constraints discussed before) and to calculate the value of the objective function for this set of brands. In fact, by selecting brands according to the rule of thumb, the objective value equals 13433793 BEF, which is 206259 BEF (or 1.5%) lower than the optimal value from the PROFSET model. At first sight, this increase in profitability due to the optimization model may not be very impressive. However, in a sector where profit margins are extremely low, and where these results can be multiplied over a large number of stores, this increase in profitability due to the optimization model is probably worth the effort.

Secondly, the iteration log shows that the obtained solution for the model is proven to be the overall optimal solution, since the MIP solution and the LP solution equal the best integer solution possible.

More important, however, are the values of the decision variables in the model, which determine the 150 brands to be selected for the automated shop assortment. In fact, we expect the set of selected items by the model to be somewhat different compared to the solution that would be obtained by using simple business heuristics, like selection based on the product's total generated sales revenue. Indeed, the key idea of the model is to exploit cross-selling effects to improve assortment selection compared with the selection based on ranking individual brands according to their individual total sales revenue within the product category. More specifically, we expect two phenomena in the model results.

Firstly, we expect some products with relatively low DPP (in our results total sales revenue) within the product category to be contained in the hitlist because of important cross-selling effects (IPP) with other products in the automated shop assortment. In other words, we expect some brands that would not be profitable enough according to their DPP yet to be selected by the model because they produce high cross-selling effects (IPP) with other products contained in the optimal set.

Secondly, we expect some products with relatively high DPP within the product category not to be contained in the optimal selection because other products in the category, with lower DPP but with higher cross-selling effects (IPP), will be more profitable overall and therefore drive away these products from the optimal set. Indeed, product i will be selected over j, if

$$\begin{gathered} TPP_i > TPP_j \\ DPP_i + IPP_i > DPP_j + IPP_j \end{gathered} \tag{5.3}$$

This can happen either when the difference in DPP is greater than the difference in IPP,

$$DPP_i - DPP_j > IPP_j - IPP_i \tag{5.4}$$

or when the difference in IPP is greater than the difference in DPP

$$IPP_i - IPP_j > DPP_j - DPP_i \tag{5.5}$$

The former happens very often when a product category is dominated by a single or small number of high market share products. In that case, the difference in DPP between the first and the second product in the category is often very big such that the IPP of the secondly ranked product is usually not big enough to compensate for the difference in DPP between both products. In that case, the model will obviously choose the first product.

The latter will happen particularly when two products are comparable with regard to their DPP, but the IPP makes the difference between both products. In other words, this will often happen in product categories where the market share is not too much skewed towards a single or few products, but market shares are distributed rather evenly between the brands.

Appendix 4 presents the selected products for model 1, together with the category that the products belong to, their sales revenue position within the category and their sales revenue position overall (i.e. within the total assortment). The following conclusions can be drawn from these results.

It can be observed from the results that model 1 mostly chooses the top products with respect to total sales revenue in each product category. At first sight, this looks a little disappointing since the PROFSET model was designed to incorporate cross-selling effects between products and they do not seem to be very relevant when looking at the results. However, it turned out that when ranking the products according to their total sales revenue in each product category, the 80-20 rules seems to apply for two thirds (2/3) of the product categories. In other words, 80% of the sales in a particular product category are generated by 20% (or even less) of the products in that category which fits in with other reported studies [86]. In other words, in many product categories, a small set of brands tends to dominate the sales of the product category. To illustrate this, take for instance the product category 'cacao products'.

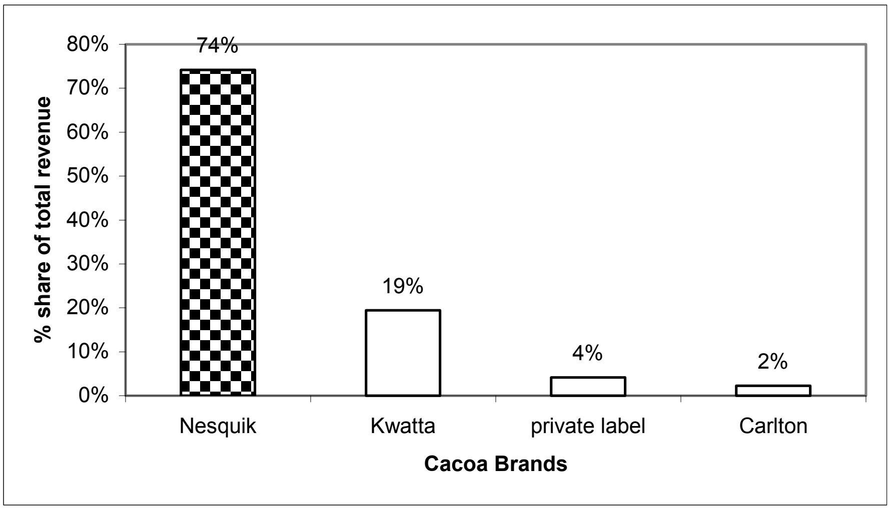


Figure 5.8: Share of category per cacao brand

Figure 5.8 illustrates the dominance of the Nesquik brand, which accounts for almost 80% of the sales revenue in that product category. As a consequence, for the other brands to be selected by the PROFSET model, they should have significantly higher cross-selling profits with other products in the hitlist than the Nesquik brand in order to get selected, which in this case is very unlikely, given the dominance of the Nesquik brand.

The PROFSET model will therefore only make a difference in product categories that are not dominated by a small number of brands, or where the sales revenue of the top brands is comparable such that the difference in cross-selling effects for those brands can play a differentiating role. For model 1, this is the case for 13 (8,7%) out of the 150 products, indicated in bold in appendix 4. However, instead of discussing all 13 brands separately, we will illustrate the idea for the 'sandwich filling' product category where 'chocolate confetti Meurisse' is selected instead of 'choco Boerinneke' and 'chocolate confetti Kwatta', as illustrated in figure 5.9.

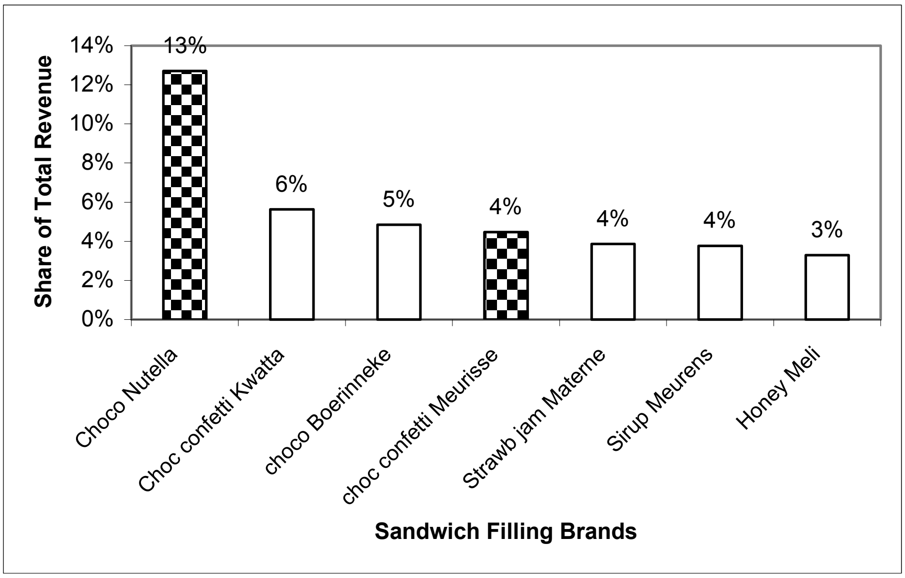


Figure 5.9: Share of category per sandwich filling brand

Figure 5.9 shows that the chocolate spread and chocolate confetti brands account for the highest total sales revenue within the category. However, except for the 'chocolate spread Nutella' brand, the difference between the total sales revenue for the different chocolate spread/confetti brands is within an acceptable range (4-6%), i.e., they are all close competitors when a second product must be chosen from this category. In fact, the results in appendix 4 show that model 1 selects two products from the sandwich filling category: 'choco Nutella' and 'chocolate confetti Meurisse'. Thus, the PROFSET model selects the 'chocolate confetti Meurisse' brand instead of 'chocolate confetti Kwatta' or 'choco Boerinneke', although the latter have a higher total sales revenue within the product category. The reason must be that 'chocolate confetti Meurisse' exhibits higher cross-selling effects with other products in the selected itemlist than the other two brands such that the overall sales revenue from main effects (DPP) + cross-selling (IPP) is higher for the 'chocolate confetti Meurisse' than for the others. This is shown in the tables of association rules below (table 5.10-5.12).

Sup	Conf	Dep.	I(R)	Antecedent	Consequent
0.1176	15.57	+	2.04	Kwatta	Baking margarine Solo
0.0776	10.27	+	1.79	Kwatta	Fresh eggs
0.1006	13.32	+	1.69	Kwatta	Coca cola

Table 5.10: Rules involving Chocolate confetti Kwatta

Sup	Conf	Dep.	I(R)	Antecedent	Consequent
0.0921	13.33	+	7.14	Meurisse	Choco Nutella
0.1539	22.28	+	2.92	Meurisse	Baking margarine Solo
0.0727	10.53	+	2.51	Meurisse	Still Water Spa
0.0691	10	+	2.38	Meurisse	Semi-skimmed milk Inza
0.0994	14.39	+	1.82	Meurisse	Coco cola

Table 5.11: Rules involving Choco confetti Meurisse

Sup	Conf	Dep.	I(R)	Antecedent	Consequent
0.1079	13.88	+	1.82	Boerinneke	Baking margarine Solo

Table 5.12: Rules involving Choco Boerinneke

From the association rules, it can be concluded that 'chocolate confetti Meurisse' appears in more rules than 'choco Boerinneke' or 'chocolate confetti Kwatta'. Furthermore, 'chocolate confetti Meurisse' has a highly significant cross-selling effect with 'choco Nutella', which is the top selling brand in the sandwich filling category. Moreover, the associations between 'chocolate confetti Meurisse' and the other products that also appear in the 'choco Boerinneke' and 'chocolate confetti Kwatta' rules have generally higher interest values I(R)). Indeed, the association with 'baking margarine Solo' has an interest value equal to 2.92 compared to 1.82 and 2.04 for 'choco Boerinneke' and 'chocolate confetti Kwatta' respectively. Furthermore, even though 'chocolate confetti Meurisse' has a lower support (count=570) than 'chocolate confetti Kwatta' (count=623) and 'choco Boerinneke' (count=641), the support with ‘baking margarine Solo’ is higher for ‘chocolate confetti Meurisse’ than for ‘choco Boerinneke’ and ‘chocolate confetti Kwatta’, as illustrated below.

Count(Meurisse, Solo) = 127
Count(Kwatta, Solo) = 97
Count(Boerinneke, Solo) = 89

Moreover, when looking at the sales revenue of the itemsets, similar conclusions can be drawn. For instance, the sales revenue allocated to the combination ‘baking margarine Solo’ and ‘chocolate confetti Meurisse’ is significantly higher than for ‘chocolate confetti Kwatta’ or ‘choco Boerinneke’.

M(Meurisse, Solo) = 7128 BEF
M(Kwatta, Solo) = 2322 BEF
M(Boerinneke, Solo) = 2942 BEF

These results therefore illustrate why PROFSET has chosen ‘chocolate confetti Meurisse’ instead of ‘choco Boerinneke’ or ‘chocolate confetti Kwatta’ for inclusion in the convenience store.

Similar results can be found for the other 11 products. For instance, model 1 selects three products from the category ‘dry biscuits’. However, it selects the product ‘Center wafers LU’ although it is ranked only at the fifth place in that product category. In fact, when analysing the market shares within the ‘dry biscuits’ category (see figure 5.10), it can be observed that ‘Center wafers LU’ is chosen instead of ‘Pokemon energy wafel’ and ‘Pick up Bahlsen’ which are ranked respectively third and fourth in the category according to the total sales revenue.

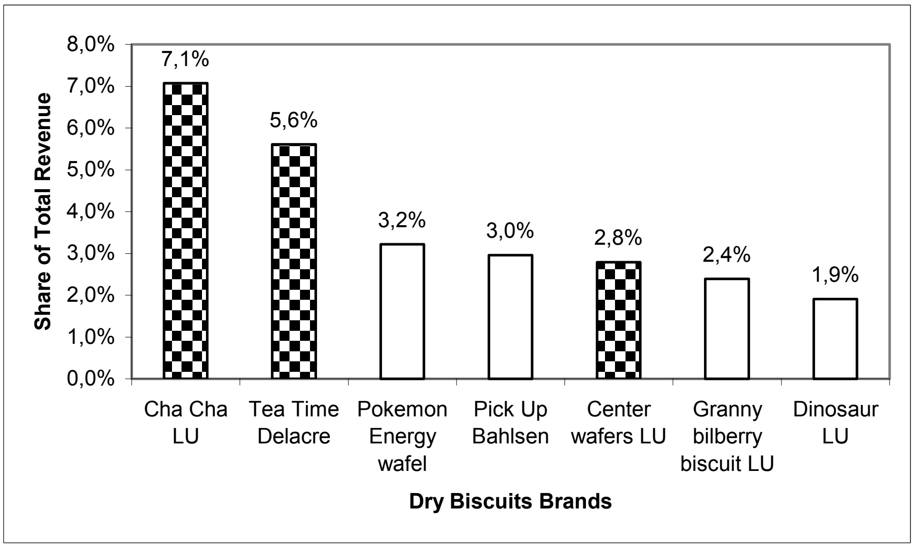


Figure 5.10: Share of category per dry biscuit brand

The fact that model 1 selects 'Center wafers LU' is therefore again due to its stronger cross-selling effects with other products in the hitlist such that overall, it is more profitable for the retailer than 'Pokemon Energy wafel' or 'Pick Up Bahlsen'.

Finally, a few words about the product selection in the 'champagne and sparkling wines' category. Model 1 selects the 'Samson Bubbles' sparkling wine for children, although it is only ranked at the sixth place in that category. In this case, this is not due to higher cross-selling effects than for the other products that are ranked higher in the category. However, the preceding products in that category, although having a higher total sales revenue, have infrequent support and therefore they can not be selected since they are not included in the model. This is an illustration where PROFSET may fail to select the correct products, i.e. when the support is too low to be included in the model, but when the unit sales is very high such that overall the product may still be interesting to select. To put it a little different, if you were supermarket retailer and certain to sell just a few Rolex watches per year, would you include them into the assortment? Probably yes!

5.6.3 Comparison of Model 1 Against Model 2

The objective of the comparison between model 1 and model 2 is to analyse how important product category constraints are in the product selection. In fact, from the 150 products selected, model 1 and 2 only have 97 brands in common (see appendix 4 and 5). More important, however, model 2 (without category constraints) selects only 50 product categories from the 126 available, whereas model 1 (with category constraints) selects 86 from the 126 available product categories. This demonstrates that retail domain knowledge plays an important role to obtain a reasonable selection of products. On the other hand, the inclusion of category constraints into the model also involves a 'cost', i.e. the value of the objective function decreases from 16196024 for model 2, to 13640052 for model 1; a decrease in sales revenue of almost 16%. The explanation for this decrease is straightforward.

By adding category constraints into model 1, the model is forced to select a minimum number of products from certain product categories that do not generate much sales. Excluding these constraints, like in model 2, in some sense offers the model more degrees of freedom to choose the most highest sales revenue products and this clearly contributes positively to the value of the objective function. Again, however, the value of the objective function should rather serve as an indicative value and not as an expected value.

Additionally, the increase in sales revenue of the PROFSET model over the selection of brands based on the rule of thumb is less impressive than in model 1. In fact, the total sales revenue of the set of brands based on the rule of thumb, i.e. in the case where products are selected with the highest product specific sales revenue, equals 16071524, which is 124500 BEF (or 0.77%) lower than the objective value of the PROFSET model.

Furthermore, model 2 selects 14 products that are not top-sellers, i.e. when ranked according to their total sales revenue, they are not contained in the 150 best set of products. Especially worth mentioning here is the selection of 3 candy bar brands that are, according to total sales revenue, not positioned within the top 150 brands, and are yet selected by the model. It concerns the following brands, with their respective total sales revenue position between brackets, 'Bounty pack' (215), 'Milky Way pack' (194) and 'Snickers pack' (179). However, when examining the data mining results, it becomes clear why model 2 has selected them for the hitlist. It turns out that candy bar brands are often purchased together. Not only do the lower sales revenue candy bar brands sell well together, they also sell well together with candy bar brands that are positioned much higher according to the total sales revenue, such as 'Leo pack' (32), 'Twix pack' (82), 'Mars pack' (83) and 'M&M's pack' (101). Together, they create a significant cross-selling effect, which makes PROFSET select these lower ranked brands too. This is illustrated by a selection of association rules in table 5.13 below.

Sup	**Conf**	**Dep.**	**I(R)**	**Antecedent**	**Consequent**
0.0655	36.24	+	50.6	Bounty, Mars	Snickers
0.0558	32.86	+	45.9	Bounty, Twix	Snickers
0.0594	31.82	+	44.4	Milky way, Twix	Snickers

Table 5.13: Rules involving Snickers, Bounty and Milky Way

Note the high values of confidence and interest for the rules. Consequently, the strong interdependencies between candy bar brands contribute strongly to the total sales revenue of the candy bar brands such that particular brands that have a rather weak ranking according to DPP may still be interesting products to select for the hitlist.

5.6.4 Comparison of Model 2 Against Model 3

Model 2 and 3 differ quite fundamentally with regard to the objective of their optimization. Model 2 tackles the problem of selecting a limited number of products for inclusion in an automated shop assortment (A.2). Model 3 on the other hand, tackles the problem of assigning products to a limited number

(150) of attractive positions in a traditional store environment (A.1). As theoretically argued in section 5.4.3.1, the objective of optimization in model 2 therefore involves the maximization of profitability within the selected set of products, whereas in model 3, the objective of optimization involves the maximization of profitability both within and outside the selected set of products. If both optimization problems are different with respect to their optimization criterion, then this should also be observed in the selection of products by both models.

In fact, the optimal selected set of products of model 2 and model 3 have 118 (78.6%) out of 150 items in common (see appendix 5 and 6). Furthermore, the objective value of model 3 equals 18498040 whereas for model 2 it is only 16196024. Model 3 thus seems to capture more cross-selling effects (+14,2%) than model 2, which is in the line of expectations. Moreover, the objective value obtained by selecting products for attractive positions based on the rule of thumb equals 18346387, which is 151653 BEF (or 0.82%) lower than the value obtained by the PROFSET model.

The fact that both models have quite a large proportion of brands in common may look somewhat disappointing in the sense that we would expect both models to select a more distinct set of products. Yet, the effect of high concentration of market share in 2/3 of the product categories again hampers the model to make a significantly different selection. Since all the models use the same data, this is a problem that is common to all models and which makes it more difficult to spell out the differences between the formulated models.

Our focus is on those products that are different to both models, such as 'Mascarpone cheese'. This brand is selected by model 3 to be located at an attractive position in the shop, although it is not selected by model 2 to be part of the assortment of the automated shop. When examining the data mining results, it turns out that there is a strong interdependence effect between 'Mascarpone cheese' and 'Boudoirs LU', as illustrated by the association rule in table 5.14 below.

Sup	Conf	Dep.	I(R)	Antecedent	Consequent
0.0958	28.83	+	55.97	Mascarpone	Boudoirs LU

Table 5.14: Rule involving Mascarpone cheese

Clearly, 'Mascarpone cheese' and 'Boudoirs LU' are both needed to prepare the famous dessert 'Tiramissou'. Furthermore, this is the only association in which 'Mascarpone cheese' appears. This explains why model 2 does not take Mascarpone cheese into the hitlist. Indeed, if model 2 would select 'Mascarpone cheese' for inclusion in the hitlist, then Boudoirs LU must be chosen too, otherwise 'Mascarpone cheese' does not add any sales to the objective function. However, since the amount of space in the automated shop is limited, the sales revenue resulting from this combination of brands is not high enough to sacrifice two entries for it. On the other hand, in model 3, the sales revenue from the combination of 'Mascarpone cheese' and 'Boudoirs LU' will add to the value of the objective function if at least one of both products is selected for the hitlist. Model 3 therefore has to sacrifice only one entry in order to capture the sales revenue from this product combination. This is the reason why model 3 selects 'Mascarpone cheese' whereas model 2 does not. Another interesting example, which nicely illustrates the difference in product selection between model 2 and model 3, is again the candy bars example. In the previous section, it was explained that, although 'Bounty pack', 'Milky Way pack' and 'Snickers pack' were not among the 150 best products according to their individual sales revenue, they were yet selected by model 2 since they exhibit high cross-selling effects between them and with higher positioned candy bar brands, such as 'M&M's pack', 'Twix pack' and 'Mars pack'. The logical consequence must then be that model 3 will not select 'Bounty pack', 'Milky Way pack' and 'Snickers pack' since the higher positioned candy bar brands are already selected in the hitlist. Indeed, consider the association rules in table 5.13 again.

Clearly, if 'Mars pack' and 'Twix pack' are already selected, then the sales revenue of the above cross-selling combinations already adds-up to the objective value of model 3. Therefore, model 3 will not select 'Bounty pack', 'Milky Way pack' and 'Snickers pack' anymore since the model would then allocate three entries without gaining any additional revenue.

The idea of model 3 is therefore that if 'Mars pack' and 'Twix pack' are positioned at attractive locations in the store, the given cross-selling effect with 'Bounty pack', 'Milky Way pack' and 'Snickers pack' will make customers go inside the gondolas to purchase them too.

5.6.5 Comparison of Model 3 Against Model 4

Model 3 and model 4 are different in so far that they adopt a different heuristic to allocate the sales revenue of transactions to frequent itemsets (see section 5.4.3.2), ceteris paribus the absence of category constraints (C.1) and the selection of products for attractive locations in the store (A.2). Model 3 allocates the sales revenue of transactions to frequent itemsets based on the notion of support (B.2), whereas model 4 allocates the sales revenue to frequent itemsets based on the notion of statistical interdependency (B.1).

The empirical results show that model 3 and 4 have 136 out of the 150 brands in common (see appendix 6 and 7). Furthermore, the objective value of both models differs only slightly. The objective value of model 3 equals 18498040, whereas the objective value of model 4 equals 18524100, corresponding to a difference of only 0.14%. This means that the gross-margin allocation rule (or in this case: the sales revenue allocation rule) does not have a very significant influence on the revenue of the product selection. Thus, although 14 (9.3%) brands are different between both models, this does not translate into a significant impact on the sales revenue of the selected hitlist. When comparing the optimal value of the objective function with the objective value obtain from assigning brands to attractive positions based on the rule of thumb, it turns out that the latter equals 18365642. Consequently, the PROFSET model beats the rule of thumb by 158458 BEF, or 0.86%

However, in this case it is more difficult to find examples that illustrate why model 3 and 4 have made a different decision with respect to a particular product. It has to do with the different allocation of the sales revenue of a transaction to the collection of given itemsets based on the support or interdependency between the items in the transaction. The different allocation rule may in fact lead to a higher or lower allocation of sales revenue to a particular itemset such that the itemset obtains a different total revenue in the objective function of the model. As a result of this different total revenue of an itemset in the objective function, the model may or may not select the itemset.

5.7 Sensitivity Analysis

This sensitivity analysis has the objective of evaluating the impact of a different basket selection on the results of the presented PROFSET model. Indeed, especially for application 1, there are several reasons to believe that the entire collection of market baskets of the retail store under study are not necessarily a good representation of the behaviour of time-pressured convenience shoppers. Indeed, their behaviour is probably better reflected by fill-in shopping trips instead of major (stocking-up) shopping trips. Therefore, it is the objective of this sensitivity analysis to select a more representative sample of market baskets, which better reflects so-called fill-in shopping trips.

The sensitivity analysis is split into two stages. The first stage consists of a literature overview and an exploratory data analysis to evaluate whether some differences can be found in the behaviour or characteristics of regular shoppers versus time-pressured convenience shoppers. In the second stage, the output of this preliminary analysis will be used to make a selection of baskets from the regular store which we think better reflects the behaviour of convenience customers. However, at this point, it is important to emphasize that this selection will merely serve as a proxy and has not the intention to present an exact solution to the problem of selecting the right baskets for this exercise.

5.7.1 Towards Better Basket Selection

In the marketing literature one typically distinguishes between two types of shopping trips: *major* shopping trips and *fill-in* shopping trips. Kahn and Schmittlein [148] specify that fill-in trips satisfy more urgent needs and generally involve smaller effort and time commitments when compared to major shopping trips. Based on market basket scanner data, they propose two criteria to discriminate between such shopping trips, i.e. the amount spent per shopping trip and the inter shopping trip time. They argue that fill-in trips can be characterized by a lower expenditure per shopping trip and shorter inter shopping trip times than major shopping trips. The decision where to cutoff low versus high expenditure and low versus high inter-purchase times is typically based on the examination of histograms where, in the case of unimodal histograms, the mode of the histogram is used as the cutoff value and, in the case of bimodal histograms, the midpoint between the two modes is taken as the cutoff point. This means that the cutoff value may be different from dataset to dataset such that a 'one-size fits all' cutoff value for any application does not exist. For example, in Popkowski and Timmermans [224], fill-in shoppers are identified as those with an expenditure of less than $7.5 and less than 4 days since their previous shopping trip. However, in Kim and Park [158] and Bell and Latin [28], the expenditure cutoff value equals $20. In the 'Marsh Super Study' [225], stock-up shoppers are those customers who purchase more than 35 items (they account for only 16% of the customer population), routine shoppers buy 11 to 34 items and account for 41%, and fill-in shoppers buy 10 or fewer items and account for 43% of the customer population. Moreover, in terms of coupon usage, Kahn and Schmittlein [148] show that the likelihood of using coupons is lower during fill-in shopping trips than during major shopping trips.

In terms of socio-demographic profiles of major shopping trip customers versus fill-in customers, the 'Marsh Super Study' [225] found that the larger the shopping trip, the more likely it is that a woman is pushing the cart. Women account for 69% of fill-in shoppers and nearly 90% of stock-up shoppers. Men, in contrast, are just the opposite. The larger the shopping trip, the less likely it is that a man is the primary shopper. Furthermore, Kim and Park [158] found that major shopping trips can be characterized by customers who are relatively more educated, working fulltime, more likely to have young children and have a higher income than fill-in shoppers. According to Kim and Park, the reason is that these households are typically more time-pressured and tend to concentrate their shopping activities during a single shopping trip per week. Some of these findings are corroborated by Popkowski and Timmermans [224] who found that large households and highly educated shoppers are less likely to be fill-in shoppers. Finally, in her study of Shop24 customers, Haeck [124] found that most of the Shop24 shoppers are aged between 18 and 35. In fact, 70% of the customers are aged below 45 and over half of the customers do not have children. Moreover, 80% of the customers have less than 2 children and housewives also less use the shop24.

To summarize, for the application presented in this dissertation, we are primarily interested in market baskets containing a small number of items, with relatively short inter shopping trip times from households with 2 children or less from relatively young families where the oldest member of the household is below the age of 45.

In order to determine the cutoff values for the number of items and the inter shopping trip times, we will look back at two histograms (figure 2.1 and 2.3) that were provided in section 2.3.2 of chapter 2. The first histogram (figure 2.1) shows the distribution of the average number of items per visit for each household in the sample. The graph shows a peak at 7 to 11 items per basket. More specifically, the mode=11 and the average number of items per basket across all consumers equals 13. It is worth mentioning that the mode of the distribution closely resembles the typical cutoff value for fill-in baskets as reported in the literature [225].

The second histogram (figure 2.3) showed, for each household, the distribution of the total number of visits over the period of observations (i.e. 24 weeks). It turns out that there is a significant proportion of customers who visit the store just once in 2 to 4 weeks, and that 60% of the customers visit the store about once or less per week. The average number of visits over the entire period of 24 weeks equals 25, or slightly more than once per week, with some exceptional customers who visit the store even more than once per day. A reasonable cutoff value for the total number of visits would therefore be 30. In that case, a customer visits the store on average 1.25 times per week. Furthermore, since he will probably shop at other stores too, the inter shopping trip time will probably be even smaller than reported by our data. However, since we do not have data of purchases made at other stores, we think that this cutoff value is severe enough.

Based on the above histograms for our data and based on the typical socio-demographic profile of fill-in shoppers found in the literature, fill-in baskets could then be defined as baskets for which :

1) the number of distinct items<10,
2) and the average weekly shopping frequency>1.25,
3) and the shopper's age<45,
4) and the shopper has less than 3 children.

However, when consecutively applying the above criteria on the total data set, it turns out that the number of fill-in baskets that are left over (see table 5.15) becomes rather small (15383) for further analysis.

Total dataset	#items<=10	#visits>=30	Age<45	Children<=2
88163	43841	26726	23187	15383

Table 5.15: Evolution fill-in baskets when consecutively applying the given criteria

Therefore, we decided to apply only the first criterion (maximum number of items per basket=10) in order to retain a sufficient number of baskets for this sensitivity analysis. After all, it is not the intention to present the most precise solution to approximate fill-in baskets, but to select baskets that reflect more or less the idea of fill-in baskets. The dataset finally retained thus contains 43841 retail baskets of a maximum size=10. However, the average number of items per basket will be much lower, as illustrated by the histogram in figure 5.11. Indeed, the modal basket contains 2 distinct items with only 25% of the baskets containing 5 distinct items or more.

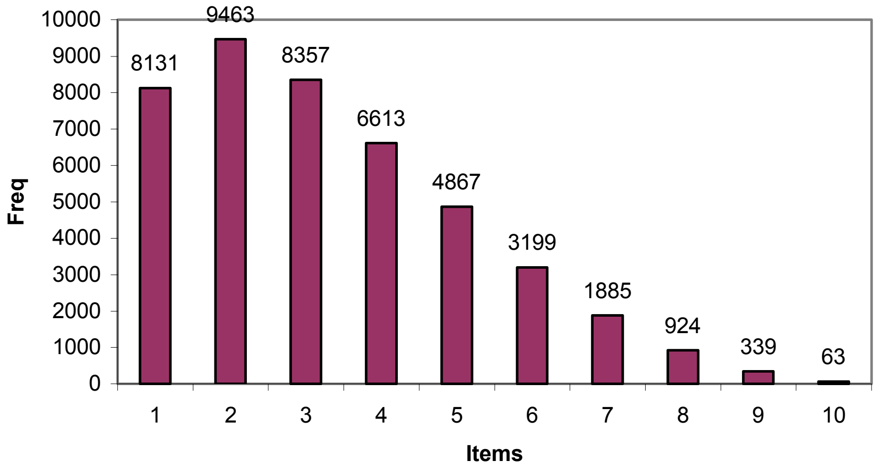


Figure 5.11: Histogram of items per basket

5.7.2 Sensitivity Analysis Results

The objective of this sensitivity analysis is *not* to recalculate each of the models presented before, but to compare the results for one of the previously discussed models (using all baskets) with those obtained for the same model, but instead using fill-in baskets. We decided to select model 2 (see table 5.7 and the associated results in appendix 5) as a benchmark since 1) from a practical perspective, model 2 describes the selection of products for a small convenience store and the use of fill-in baskets is particularly relevant in this context and since 2) from a technical perspective, model 2 is not subject to any category constraints such that the optimal selection of products will not be biased by category size limitations. Indeed, this way a change in the optimal set of products will be entirely due to a different selection of retail baskets with all other possible influences (like category constraints) held constant.

To summarize, it is our goal to evaluate to what extent model 2 leads to a different brand selection based on fill-in baskets compared with the 150 brands selected based on all available baskets. Comparing the new results (based on fill-in baskets: appendix 8) with those previously obtained (based on all baskets: appendix 5), some interesting conclusions can be drawn.

The 'old' and the 'new' model results have 119 (80%) of the 150 selected brands in common. However, despite this large overlap, some noticeable changes in sales rankings can be observed for several categories of brands, both for brands that are common to the 'old' model as for brands that are uniquely selected by the 'new' model. In fact, the *position* of a brand (see last column in appendix 5 and 8) refers to the sales ranking of that particular brand within the total assortment. Indeed, it can be observed that, in general, beers, spirits, wines and bakery products have considerably improved their sales position in the new results based on fill-in baskets compared to the results based on all the baskets, as illustrated by table 5.16.

The reverse, however, is true for washing powders, dishwashing and washing-up liquid, candy bars, dry biscuits, crisps and waters (except for private label waters which have improved their sales position and Spa waters that have remained almost the same) as shown by table 5.17 below.

Therefore, even though the new model (based on fill-in baskets) selects many of the same products as those selected by the old model (based on all the baskets), the relative sales position of most brands changes significantly.

Category	Brand	Position all	Position fill-in	Selected by old model 2	Selected by new model 2
Regular and heavy Beers	Cristal Alken	27	13	Yes	Yes
	Duvel	48	22	Yes	Yes
	Grimbergen dubbel	279	152	No	Yes
	Kasteelbier	478	122	No	Yes
	Leffe blond	99	72	Yes	Yes
	Leffe brown	81	47	Yes	Yes
	Palm	33	18	Yes	Yes
	Stella Artois	234	88	No	Yes
	Westmalle dubbel	189	91	No	Yes
	Westmalle tripel	166	43	No	Yes
Spirits	Jonge Bols	144	76	No	Yes
	Sherry dry	111	94	Yes	Yes
Wines	Soave classico	326	138	No	Yes
Bakery products	Apple bun	287	68	No	Yes
	Cocolate bun	175	98	No	Yes
	Glacé	110	56	Yes	Yes
	Sausage roll	451	96	No	Yes
	Double sausage roll	167	36	No	Yes
	Biscuit 3 fruits	392	113	No	Yes
	Bo bread	92	64	Yes	Yes
	Bo French bread	44	23	Yes	Yes
	Bo Kaiser bread	78	54	Yes	Yes
	Grey bread	9	5	Yes	Yes
	Multi-grain bread	28	11	Yes	Yes
	Rye bread	164	74	Yes	Yes
	Sandwiches	12	9	Yes	Yes
	Whole-meal bread	117	51	Yes	Yes
	White bread	15	8	Yes	Yes
	Bran bread	91	49	Yes	Yes

Table 5.16: Brands that have gained sales position

Category	Brand	Position All	Position fill-in	Selected by old model 2	Selected by new model 2
Washing powder, dishwashing liquid, washing up liquid	Biotex blue	118	175	Yes	No
	Coral intens	84	129	Yes	Yes
	Dash futur	74	109	Yes	Yes
	Dash scoops	6	12	Yes	Yes
	Dixan doses	41	60	Yes	Yes
	Dreft compact	17	26	Yes	Yes
	Dreft household	49	58	Yes	Yes
	Dreft dishwashing	132	117	Yes	Yes
Candy bar packs, dry biscuits	Bounty pack	215	446	Yes	No
	Cha cha LU	35	65	Yes	Yes
	Center wafers LU	134	290	Yes	No
	Leo pack	32	52	Yes	Yes
	M&M's pack	101	146	Yes	Yes
	Mars pack	83	133	Yes	No
	Milky Way pack	194	395	Yes	No
	Pick Up Bahlsen	125	189	Yes	No
	Pokemon Energy	109	131	Yes	Yes
	Snickers pack	179	245	Yes	No
	Tea time Delacre	53	29	Yes	Yes
	Twix pack	82	136	Yes	No
Waters	Water still Contrex	114	142	Yes	Yes
	Water still Evian	67	95	Yes	Yes
	Water spark. Priv.l.	113	107	Yes	Yes
	Water still priv. l.	66	41	Yes	Yes
	Water spark. Spa	13	14	Yes	Yes
	Water still Spa	4	4	Yes	Yes
	Water still Vittel	73	85	Yes	Yes
Crisps	Paprika Croky	168	367	Yes	No
	Paprika Smiths	88	182	Yes	Yes
	Salty Croky	124	257	Yes	No
	Salty Smiths	86	158	Yes	Yes

Table 5.17: Brands that have lost sales position

In some cases this results in a different product selection. For instance, for some bakery products (like apple bun, chocolate bun, sausage roll, double sausage roll and biscuit 3 fruits) the increase in their sales position within the fill-in baskets makes them attractive for the model to select, whereas in the total collection of baskets their position is not profitable enough to get selected. For other brands, the reverse is true, as indicated before. Examples include candy bar packs, such as 'Center wafers LU', 'Mars pack', 'Twix pack' and others. The reason here is that they have a considerably lower sales position within fill-in baskets such that they are not attractive enough to be selected.

5.8 Conclusions

Based on the empirical comparison between the presented models, and the subsequent sensitivity analysis, we can now discuss the contributions and limitations of the model and the data that we used to estimate the models.

5.8.1 Model Contributions

Our first contribution lies in accounting for interdependence within the product selection problem. Indeed, previous research has made clear that interdependence plays an important role and that failing to consider those interdependencies may lead to marketing actions with disappointing results [183, 209]. The model presented in this dissertation makes an attempt at taking into account such interdependency effects for the problem of product selection in retail marketing. The empirical results on real scanner data have shown that the model is indeed capable of using cross-selling information during the product selection process. However, to be honest, we expected the impact of cross-selling effects to be more pronounced on the product selection than it eventually turned out. We believe that the reasons for this are twofold and both are related to the data that we used to estimate the model. First of all, as already discussed in the empirical section, the purchases in each product category are highly biased towards a very limited number of (highly advertised) brands. This phenomenon is therefore also reflected in the product selection results. Although we have not tested this, we believe that the model will probably be better able to exploit cross-selling effects in retail assortments where there is less market concentration, such as for do-it-yourself (DIY) stores[31]. Secondly, our data did not contain any cost information, unfortunately. For proprietary reasons, the retailer could not share gross margin information with us and information related to inventory and product-handling costs were simply not available. This obviously limits the potential of the model to evaluate the difference in DPP and IPP of the given products. The empirical results were therefore expressed in terms of sales revenues, although the theoretical development of the PROFSET model discussed the case where cost information is indeed available.

Our second contribution lies in the presentation of an optimization framework for product selection instead of a rule of thumb based on a simple heuristic. Indeed, the approaches presented in this dissertation are based on an optimization framework, instead of on heuristics. As a result, the impact of product replacement decisions on the rest of the assortment and on the objective value of the model can be quickly evaluated. As Doyle and Gidengil [98] put it: 'stores are not interested in calculations of isolated demand elasticities but are rather interested in the effect of product mix changes on overall store profitability'. In fact, since the model is an integer-programming model, product replacement decisions will induce a dynamic reselection problem where the model will re-calculate the optimal solution for the model as a result of a deletion or replacement of a particular product. In other words, by deleting a particular product from the hitlist, the model will automatically re-calculate the entire solution and determine the next-best brand to enter the hitlist. In this context, we have argued that the value of the objective function

[31] De Schamphelaere et al. [92] note however that highly significant cross-selling effects are less apparent in DIY stores and that the support of itemsets is usually very low.

serves only as a means to evaluate the impact of alternative product selection decisions but that the value itself should not be used as an estimate of the potential profitability of the selected assortment. In other words, the value of the objective function itself has no meaning, it only provides a means to evaluate alternative product selection decisions.

Our third contribution lies in accounting for the micro-economic reality of the retailer. The model is capable of dealing with retail domain knowledge, such as product category information. Furthermore, the model is flexible such that additional constraints that might be relevant to the retailer can be easily included into the model. For instance, restrictions towards the composition of the assortment in terms of convenience, impulse and emergency products, or in terms of core assortment items and peripheral assortment items can easily be integrated into the model by means of additional constraints.

Fourthly, the contributions of the model should be evaluated within the context of model completeness and complexity versus simplicity of use. As Simkin [249] states: 'there has to be a compromise between completeness and simplicity: validity versus usability'. Obviously, the product selection problem in marketing is a problem with many faces. Researchers have therefore made an attempt to include as many of these dimensions of product selection into their models as possible. This has, however, increased the computational and operational complexity of such models and therefore also limited their practical usability in real situations. We believe that our model, although less complete from a theoretical point of view, is probably simpler to use in real situations since it directly uses information from receipt data to estimate the cross-selling effects.

Fifthly, we believe that the PROFSET framework is easily adaptable to the problem of product assortment rationalization, i.e. to reduce the number of SKU's in order to limit or prevent product proliferation within certain product categories. However, as Bucklin and Gupta put it in a recent paper [64] 'Given the agreement on the need to rationalize product assortment and reduce the number of SKU's, the next question is how to decide which items to eliminate'.

Furthermore, they state that 'discussion with practitioners suggested that they follow a simple, and somewhat naive, procedure of deleting, say, the bottom third (in terms of sales or profits) of the items in a category'. We believe that the PROFSET model enables to approach this problem in a more scientific way, i.e. it enables to optimally select products from product categories taking into account the cross-selling effects with items in other product categories.

Finally, the sensitivity analysis (section 5.7) has shown that basket selection should be carried out with great care. Indeed, it was shown that fill-in baskets probably best reflect the purchase behaviour of time-pressured convenience shoppers and that this selection of baskets has an influence on the optimal selection of products by the PROFSET model. However, more research would be needed to select those baskets that are most suited for the product selection problem at hand.

5.8.2 Model Limitations

Besides the contributions discussed before, the PROFSET model is based on a set of assumptions, which should be taken into account when applying the model.

A first important limitation of the current implementation of the PROFSET model is the absence of dealing with product substitution effects. Indeed, the frequent itemset framework on which the PROFSET implementation is based, is one of counting events, and not of non-events. This means that currently PROFSET only increases the profitability of items if they have significant cross-selling effects. However, the model does not decrease a product's profitability if there are strong negative interdependencies with other items. The calculation of non-events by means of negative association rules (see section 4.5.1.5) could potentially provide a solution to this problem. Yet, the efficient calculation of such associations still remains one of the most difficult challenges in frequent pattern mining today.

Secondly, the choice of the support threshold in combination with the dependency based profit allocation heuristic could be subject of discussion. In fact, the dependency based profit allocation heuristic is not able to identify all significantly dependent product associations since it only uses frequent itemsets as input. Indeed, the input for dependency calculation is dependent on the choice of the support threshold. As a result, if this threshold is too high, some significant product interdependencies with relatively low support may not be discovered and thus can not play a role in the objective function of the PROFSET model. This is, however, not really a problem since the support of these missing interdependencies is too small to influence the value of the objective function, unless there would exist very infrequent but significantly interdependent associations with extremely high profitability (see for instance the 'Samson Bubbles' example in the last paragraph of section 5.6.2.4). However, in grocery retailing, where gross margins are mostly within reasonable ranges, this is often not the case. In fact, sensitivity analysis experiments showed that indeed the product selection is very robust (insensitive) with regard to the support threshold parameter.

Thirdly, the treatment of product deletion decisions by means of the PROFSET model currently does not take into account the effects of brand switching by consumers. In other words, the effect of a deletion of a product from the optimal set currently produces an upper bound on the maximum loss of sales due to this deletion and its resulting loss of cross-selling effects with other products. However, previous research [71] has shown that at least some portion of consumers is willing to substitute their favourite brand for an alternative brand when it is no longer available. The sales loss due to product deletion may therefore be less profound than indicated by the percentage change in the value of the objective function. Indeed, although some customers will switch stores and will do their purchase in another store, other customers will buy the other things in the current store and purchase the missing product in another store. Additionally, some customers will not purchase the missing product at all, and finally still other customers will purchase the other things in the current store and substitute some other product for the missing one.

Fourthly, as indicated in the prologue of this chapter, it is not clear to what extent the 'identification problem' is present in the data. Indeed, the use of frequent itemsets to express the level of product interdependency is only permitted in so far that the product associations do not exist due to environmental factors, such as product placement, pricing, promotion, etc. However, as far as we know, there is currently no research available on how strong this identification problem affects the discovered product associations and whether such environmental effects are category specific or not. Suitable strategies to find out whether the identification problem is indeed present could be to use cross-sectional data from different stores or to set up experiments where, for instance, the location of two products is changed in order to investigate the effect on the strength of the product association. For those categories where research would indicate that the identification problem has a strong impact, the PROFSET model should not be used.

Fifthly, the retail store may have other goals than profit maximization (e.g. maximizing traffic, maximizing awareness). Such goals are not included in the current PROFSET model.

CHAPTER 6
BEHAVIOUR-BASED CUSTOMER SEGMENTATION

Today's competition forces consumer goods manufacturers and retailers to differentiate from their competitors by specializing and by offering goods/services that are tailored towards one or more subgroups or segments of the market. The retailer in the FMCG sector is, however, highly limited in his ability to segment the market and to focus on the most promising segments since the typical attraction area of the retail store is too small to afford neglecting a subgroup within the store's attraction area [82]. This does not mean, however, that the retailer has no options available to maximize his share-of-wallet from the customer. Indeed, retailers are seeking ways to customize the communication with the customer in order to offer the customer tailored product offers or promotions based on his/her past purchase behaviour. More specifically, the availability of huge amounts of transactional and loyalty card data provides excellent opportunities to segment shoppers based on past purchase behaviour.

This chapter provides a literature overview of the topic of customer segmentation in general, with an emphasis on behaviour-based segmentation more specifically. In fact, the definition of segmentation and the different bases (of which behaviour is one) and methods for segmentation will be discussed together with criteria to evaluate the quality of cluster solutions. Furthermore, we will provide two illustrations of behaviour-based segmentation on our data.

6.1 The Concept of Market Segmentation

Segmentation of the consumer market (or market segmentation) is of great importance to retail marketing. This is not at all surprising. Consumer goods manufacturers have realized that the concept of 'one size fits all' is very difficult or even impossible to maintain in a highly competitive market. In fact, for many consumer products a manufacturing-oriented strategy, illustrated by Henry Ford's famous statement 'People can have the Model T in any colour, so long as it's black' is no longer successful in the current market situation. Today, competition forces consumer goods manufacturers to differentiate from their competitors, i.e. to specialize and to offer goods/services that are tailored towards one or more subgroups or segments of the market. Manufacturers realize that shoppers are heterogeneous in nature and that they possess different wants and needs. Lunn [180] attributes this increased heterogeneity in consumer behaviour to fundamental changes in society. He claims that purchasing has become much more discretionary and less concerned with the necessities of daily life. This is the result of an increased wealth and level of education of the consumer, accompanied by increasing social mobility and by erosion of traditional class-determined patterns of behaviour: the typist, like the movie star, can afford to express her personality through the latest fashions, and the factory worker may take his holiday in Greece [166]. This is where *market segmentation* comes into play.

Since Smith's pioneering article [250] in 1956, many definitions of market segmentation have been proposed, but essentially they can be summarized as the partitioning of the market into homogeneous sub-markets in terms of customer demand, resulting in the identification of groups of customers that respond differently to the marketing mix. The benefits to be gained from adopting a segmentation strategy are appealing: marketing resources can be allocated more effectively and efficiently and consumers are presented with customized product or service offers.

However, despite the importance of market segmentation and its associated benefits, Corstjens [82] argues that retailers in the FMCG sector (in contrast to manufacturers) are highly limited in their ability to segment the market and to focus on the most promising segments. Indeed, the typical attraction area of the retail store is too small to afford neglecting a subgroup within the store's attraction area. In fact, the supermarket should appeal to as much of the heterogeneous public in its attraction area as possible. However, the heterogeneity that is present in the consumer behaviour of customers within the store's attraction area provides opportunities for *customer* segmentation [135] which in turn can be used to feed customized merchandising strategies towards those segments. For instance, differences in shopping behaviour between customers may be used to target customers with different promotions, e.g., by printing customized promotional offers on their receipts, or by distributing segment-specific promotional leaflets. Therefore, if carefully carried out, both retailers and consumers can benefit from *customer* segmentation: specialisation by the retailer to offer added-value in particular segments will act as a barrier to entry for other retailers if the competences of the retailer are unique and difficult to copy (i.e. sustainable advantages), and consumers are offered customized product or service offers.

6.2 Segmentation Research Overview

Despite the extensive body of literature about market segmentation (both theoretical and practical), a general consensus about an optimal segmentation methodology is not available in the literature. The reason maybe is that, instead of a single concept, over the years segmentation has evolved into an umbrella topic covering a diversity of issues [180]. Indeed, segmentation can be viewed from different perspectives, i.e., segmentation as a strategy, and segmentation as a methodology.

Segmentation as a strategy is concerned with the targeting of products to a selection of customer groups, whereas segmentation as a methodology is more concerned with the techniques and the methods.

In this chapter, we are primarily interested in techniques and methodologies for segmentation. In this context, the work by Frank, Massy and Wind [106] can be considered as the leading reference. They claim that differences in segmentation can be distinguished along two major dimensions, i.e. the basis for segmentation and the (analytical) segmentation technique being adopted. However, since their synthesis, dating back 29 years now, the field has witnessed a number of new developments in information and data analysis technology. This has been the motivation for a new book on market segmentation by Wedel and Kamakura [291] and since this book provides the latest up-to-date overview of the domain, our synthesis in the following sections will mainly draw on their work.

6.2.1 Segmentation Bases

The measures most frequently used for segmentation are typically drawn from either one or a combination of the following: demographics, behaviour, benefits, and psychographics. However, the concrete choice for one or a combination of these segmentation bases largely depends on the business question under study [294]. The idea is that segmentation places customers in groups on the basis of their similarity on a chosen set of variables. Afterwards, members of different segments will be treated differently in marketing communications to achieve different marketing objectives with greater overall effect. Segmentation bases can be classified according to two criteria, 1) general or product-specific, and 2) observable or non-observable. *General* bases for segmentation are independent of products, services or circumstances, whereas *product-specific* bases for segmentation are related to the product, the customer or the circumstances. *Observable* segmentation bases can be measured directly, whereas *non-observable* bases must be inferred. The combination of both results in the classification of segmentation bases is shown in table 6.1.

	General	Product-specific
Observable	Cultural, demographic, geographic and socio-economic variables	Usage frequency, brand loyalty, store loyalty and patronage, usage situations, purchase moment
Non-observable	Psychographics, values, personality and life-style	Psychographics, benefits, perceptions, elasticities, attributes, preferences, intentions

Source: Wedel and Kamakura [291]

Table 6.1: classification of segmentation bases

In the next four sections, we will discuss each of the given bases for segmentation separately and, where available, examples within the supermarket retailing literature will be used to illustrate the concepts. Furthermore, since there is an increased trend to combine several of the given bases for segmentation, a separate section will be devoted to the means-end chain approach, which aims at integrating several bases for segmentation and which has recently gained increased interest in the marketing literature.

6.2.1.1 General observable bases

General observable bases for segmentation are probably the oldest and most frequently used in segmentation research [110]. Popular examples of this kind are demographic variables (age, gender, ethnicity, marital status, household composition, religion, …), geographic variables (area of residence, country, …), socio-economic variables (level of income, education, profession, …) and cultural variables (interest in books, vacation, concerts, …). Their popularity has even led to some well-known (commercially available) systems, such as the geo-demographic ACORN™ (A Classification Of Residential Neighbourhoods) segmentation system [119] and the Belgian socio-demographic MOSAIC typology [251] which groups 160.000 street segments in Belgium into one of thirty MOSAIC types, such as:

- Elite: highest social class and high income
- High welfare: owners of relatively recent houses with modern comfort
- Semi-rurals: mixture of white and blue colour jobs
- Metropolitans: concentrated in larger cities and agglomerations
- Middle class: especially white colour jobs, medium to lower comfort houses
- Labourer class: better labourer class, working in modern industries
- Industrial area: lower education, working in basic industries
- Rural area: living in rural areas, large families with many children
- Foreigner majority area: high concentration of foreigners living in big agglomerations and basic industry areas, low education and low socio-economic status.

General observable bases for segmentation are mostly available from public sources and thus relatively easy to collect, stable and reliable.

Some recent examples of the use of general observable bases for segmentation include the study by Segal and Giacobbe [243], which carry out a Ward's hierarchical cluster analysis on 10.000 customers in a large metropolitan area in the USA on the basis of a set of demographic variables, such as occupation, household composition, income, housing, etc. Their analysis revealed 4 demographic segments for which they subsequently analysed and found profound differences in the market share of 4 major supermarket chains. Gensch [110] clustered 700 individuals on 19 demographic variables, such as income, age, marital status, number of cars, etc. in order to test the advantages of disaggregate choice models. Gensch shows that meaningful segmentation can increase the predictive fit of choice models and can lead to different managerial actions and strategies when the assumption of homogeneity is too restrictive. Hortman et al. [191] carry out a demographic segmentation based on the age of the head of household, number of working adults in the household, and years in residence at the current address. Their cluster analysis produced three distinct segments, i.e. a baby boomers segment, a middle-aged family group and an elderly segment. These segments were subsequently used to study the differences in the respondent's emphasis on price, selection, and convenience when selecting supermarkets.

Despite the above-mentioned studies, researchers often criticize that segmentation on the basis of these measures lacks a direct link with purchase behaviour [88, 108]. In other words, it is not theoretically clear whether differences in socio-demographic background produce significant differences in purchase behaviour. Best [33] calls this the *demographic trap* and refers to Greenberg and Schwartz [116] who state that indeed 'demographic segmentation seldom provides much guidance for product development or message strategies'. While some studies found some small significant differences in responsiveness to marketing variables [288], other studies conclude that these differences are in fact too small to be relevant for practical purposes [106, 188].

6.2.1.2 Product-specific observable bases

Product-specific observable bases include (purchase) behaviour-based variables such as user status [106], usage frequency [271], brand loyalty [117], store loyalty [117], store patronage [106] and usage situations [94].

Segmentation based on user status [106] divides customers into groups according to their status as a user (user, non-user, past user, potential user, …), but does not take into account the frequency of usage. Segmentation based on usage frequency divides customers into groups based on their intensity of buying a product(s), such as light – medium – and heavy buyers

[271]. Recently, the existence of huge amounts of scanner data provides a new impetus for segmentation on the basis of *purchase* frequencies (not necessarily *usage* frequencies). Indeed, because scanner data provide information on the purchase history of consumers on a very detailed level, i.e. mostly on the SKU level, these data provide interesting segmentation opportunities.

For instance, Dillon and Kumar [96] introduced a mixture-clustering model based on the purchase frequency of candy. They estimate a Poisson mixture model on the data and identify six underlying segments of light (from 0 to 1.08 packs per week), heavy users (from 3.21 to 7.53 packs per week), and extremely heavy users (from 11.1 to 13.6 packs per week) of candy. The segmentation model developed in the next chapter of the dissertation is built on the principles of the Dillon and Kumar model, but extends it in terms of the number of product categories being taken into account and how it deals with correlated purchases. In fact, whilst the Dillon and Kumar model clusters consumers based on the purchase rate in just one product category, the models developed in the next chapter will take into account multiple product categories (see section 7.7).

Cadez et al. [68] also used a mixture framework based on fitting a multinomial mixture distribution to model sales transaction data. For a clothing retailer, they found two types of transactions, the first involving mostly men's clothing, and a second mostly involving women's clothing items, but not both.

Another approach was taken by Ordonez et al. [212]. They used a mixture of Normal distributions to fit a sparse data set of binary vectors corresponding to the raw market baskets in a sales transaction database. Ordonez et al., do not take correlations between product purchases into account by assuming diagonal covariance matrices.

Reutterer [227] uses Kohonen Self-Organizing Maps (SOM) to cluster supermarket shoppers into 9 segments based on their preference for different brands within the margarine product category. Although preference is in fact an unobservable basis for segmentation (see section 6.2.1.4), preference in this context was measured by the relative purchase frequency of a brand within its category. He found distinct segment differences in the preference for private label and national brands.

Other product specific observable bases for segmentation include for instance brand loyalty. In fact, Guadagni and Little [117] were among the first researchers to include brand loyalty into brand choice models. They modelled purchase probability by introducing past purchase behaviour as an explanatory variable through loyalty attributes. More specifically, brand loyalty was operationalized as the exponentially weighted average of past purchases of the brand.

Store loyalty has also been suggested as a useful basis for segmentation since store-loyal customers tend to exhibit different sensitivities to changes in prices, promotion, and other marketing variables than do non-loyal customers since they devote less effort to shopping and are less concerned with finding the best bargains [106].

With regard to store patronage, consumer response to price for frequently purchased food products was found to be different between shoppers of two major chains and the shoppers of independent stores [105]. In another study, Kopp et al. [164] applied a k-means clustering on a group of 1.650 supermarket shoppers based on a vector of share-of-shopping-visits for eight competitive retailer groups. They identified six distinct segments and subsequently tested their difference in profile on the basis of a number of descriptor variables, such as fashion lifestyles, attribute importance, demographics, etc. Moreover, McCurley Hortman et al. [191] carried out an a-priori segmentation of shoppers based on whether the majority of their shopping trips were made to a discount-image store or to a non-discount image store.

Also usage situations have been identified as important discriminators for consumer behaviour. The idea is that the utility for products may be different for the same person depending on the usage situation. For instance, suntan lotion may be used in different situational contexts, such as for boat sunbathing, home sunbathing, sunlamp bathing, snow skiing, etc. which may influence the utility for a particular product in the context of the situation in which the product will be used. Dickson [94] even claims that benefit segmentation (see section 6.2.1.4) can be viewed within the person-situation segmentation framework since the benefits that consumers seek in products are basically person- and situation-based. De Pelsmacker and Van Kenhove [90] studied the influence of the usage situation on the purchase of cookies. They discovered that more than 50% of the variance in the data could be attributed to the usage situation dimension (i.e. are the cookies purchased for self-consumption or not).

Finally, the moment of purchase maybe a useful basis for segmentation to enable the distinction, for instance, between regular stock-up shoppers versus emergency top-up shoppers, or lunch-time shoppers versus evening shoppers, or home and daytime shoppers versus work and weekend shoppers, or fast-checkout customers versus regular checkout customers, etc [82]. Van Kenhove et al. [277, 278] analysed the influence of purchase situations on the choice of DIY (Do-It-Yourself) stores. They found that customers have different needs and use different criteria for evaluating store choice according to the usage situation.

6.2.1.3 General non-observable bases

Lifestyle segmentation variables refer to the consumer's (AIO) *activities* (work, hobbies, social events, vacation, leisure time, clubs, shopping, sport), the consumer's *interests* (family, home, job, community, leisure, fashion, food, media, culture), and *opinions* (about him/herself, social matters, politics, economy, business, education, upbringing, products, future) [23]. Some popular lifestyle segmentation studies include VALS (value and lifestyle segmentation) [202] and LOV (list of values) [147]. For instance, Gutman and Mills [123] cluster fashion merchandise shoppers based on a number of fashion lifestyle related variables, such as fashion leadership, fashion interest, etc. Their study revealed 7 fashion segments, such as leaders, followers, independents, uninvolveds, etc. which were subsequently profiled on a set of demographic variables. An important motivation to use these specific fashion related lifestyle variables is that other studies have argued that 'general' lifestyle variables provide a poor discrimination towards purchase behaviour.

6.2.1.4 Product-specific non-observable bases

Segmentation variables of this kind include product-specific psychographics, product-benefit perceptions and importances, brand attitudes, preferences and behavioural intentions [291]. Product-specific psychographics are defined as value orientations (e.g. attaches high value to healthy products, or environmental friendly products, …), role perceptions (e.g. perceives himself as social entertainer, or creative person) and buying style (e.g. repeat buyer, impulsive, variety seeker, …) towards a (set of) products [93].

Benefit segmentation aims at discovering clusters of consumers that seek similar benefits when evaluating and choosing or purchasing (in) products or retail stores. Benefits may be measurable, such as economical or durable, but may also be rather abstract concepts, such as contemporary or old-fashioned. The difference in importance that consumers attribute to these benefits offers an interesting basis for segmentation since they reflect the needs that consumers have. For example, Miller and Granzin [201] discover segments of consumers on the basis of a number of benefits from fast-food retail chains, such as speed of delivery, price, friendliness of employees, taste of the food, etc. Rudolph [231] uses k-means clustering to discover segments based on the benefits that German, British and Swiss customers seek to determine their store choice. He identified 5 clusters: 1) the dissatisfied service(rendering) customer, 2) the environmentally interested grumbler, 3) the happy senior customer, 4) the communication critical convenience customer, and 5) the price sensitive regular customer. The importance weights that customers attach to the different product/store attributes can be discovered directly or indirectly. Since customers tend to find all attributes important when scoring a rating scale, the preferred method of measuring importance weights is by means of indirect measurement, e.g. through the derivation of part-worths in conjoint analysis.

The attribute importance values can then be used for segmenting customers into groups who seek similar benefits in a product/store.

6.2.1.5 Means-end chain theory: an integrated framework

Although the different bases for segmentation discussed in the previous four sections provide valuable alternatives for market segmentation, marketing researchers have realized that each of them separately provides a rather myopic view on the complexity of consumer behaviour. They have therefore looked for ways to integrate different bases for segmentation. The means-end chain theory provides such an integrated framework to explain (differences in) consumer behaviour and therefore also provides excellent opportunities for segmentation.

The idea behind means-end chain theory [122, 211, 299] is based on the relationship between *product attributes*, *benefits* and *consumer values*, as illustrated in figure 6.1.

Product attributes $\rightarrow$ Product benefits $\rightarrow$ Consumer values

Figure 6.1: The means-end chain

In fact, it is often assumed that products and services can be characterized in terms of their concrete (tangible) attributes and that consumer behaviour can be explained based on the consumer's perception of the level of performance of each product on these attributes and how important each attribute is to the consumer. For instance, discrete choice models (section 3.2.2.2) are based on this assumption. In fact, in many product categories, this paradigm seems to work well. However, in some circumstances, for instance in markets where the performance of products has become very similar (e.g. detergents, toothpaste), or in markets with complex products where there is too much information for the consumer to handle (e.g. cars, computers), consumers will rely on simplified constructs, such as reliability, performance and excitement, to evaluate and distinguish between products [285]. These higher-level constructs are also referred to as *benefits*, which are frequently used within the context of benefit segmentation (section 6.2.1.4). In other words, concrete product attributes gain their relevance since they allow the consumer to achieve certain benefits. Benefits can therefore be considered as higher level (intangible) constructs made of combinations of concrete (tangible) product attributes. In turn, these benefits become important since they are linked to higher-order personal values, such as happiness, security, sense-of-belonging, and achievement [239, 240]. Thus, in order to achieve those higher-order personal values, consumers have a higher need for some benefits than others which may explain their behaviour to choose one product over another. To summarize, in the means-end theory, product attributes are viewed by the consumer as a 'means' to achieve a certain 'end', i.e. certain benefits or higher order personal values and since the relation between attributes, benefits and consumer values may be different for each consumer, they constitute an excellent basis for segmentation. The advantage of the means-end chain theory over traditional bases for segmentation is therefore that it combines both the strengths of product-specific (attribute-based segmentation) and consumer-specific (value-based) bases of segmentation by explicitly linking attributes, benefits, and values at the segment level. This is clearly a distinctive advantage over separate approaches, such as conjoint segmentation where only the importance of attributes is used as a basis for segmentation, or VALS segmentation (see section 6.2.1.3) where only consumer values serve as a basis for segmentation. Recent examples where the means-end chain theory is used for market segmentation includes the research from Ter Hofstede et al. [262, 263].

The methodology that is most often used to reveal consumer's means-end structures is the so-called 'laddering' technique [228]. Basically, laddering is a qualitative interviewing technique that consists of three steps. In the first step, the objective is to identify attributes that are important to the consumer and to establish preferences within these attributes. For instance, I find it important that washing powder contains chemicals to make clothes whiter (product attribute). In the second step, the consumer is usually asked why he/she finds these attributes important. This way, benefits and consumer values can be identified. For instance, I find this important because then my kids have cleaner clothes (benefit). This way, I am a good parent (instrumental value) and this makes me feel good about myself (self-esteem: terminal value) [219]. Finally, in the last step, the results are analysed and a means-end structure is composed.

However, because laddering is time-consuming and thus only suitable for relatively small samples and involves highly trained interviewers, the technique is not preferred for large-scale samples that are typically required for market forecasts and segmentation, among others. This inspired Ter Hofstede et al. [262] to come up with an alternative methodology, called the association pattern technique (APT), which is much cheaper and faster to carry out and allows the researcher to collect data among a much larger sample of consumers. In the association pattern technique, two matrices are constructed, i.e. the attribute-consequence (benefits) matrix, and the consequence-value matrix. The consumer is then asked to indicate which attributes lead to which consequences, and in turn, which consequences lead to which values where the attributes, consequences and values are predetermined by the researcher. This way, both matrices consist of binary values, which can be modelled by a finite mixture model formulation [263]. The fact that attributes, benefits and values are pre-determined by the researcher and can be visually represented in tables to be scored by consumers makes the approach empirically tractable for large samples of consumers.

6.2.2 Segmentation Methods

The literature on segmentation methods is characterized by an enormous variety of methods and techniques. Basically, all techniques can be classified along two dimensions: 1) apriori versus post hoc and 2) predictive versus descriptive.

In *apriori* segmentation, the basis for segmentation is defined by the user/manager and segments are created according to a number of heuristics or business experience. In *post hoc* segmentation, statistical techniques are adopted to discover groups of customers that have similar values on a number of segmentation variables. Although careful selection of these segmentation variables is necessary, the clustering technique will identify which variables provide the best discrimination between the segments. Therefore, posthoc segmentation is often considered the better alternative.

Predictive segmentation methods analyse the relationship between a dependent and a set of independent variables. *Descriptive* methods, however, analyse the associations across a single set of segmentation variables, with no distinction between dependent and independent variables. The classification of clustering techniques along these two dimensions is illustrated in table 6.2 and will be elaborated upon in sections 6.2.2.1 to 6.2.2.4.

	A priori	**Post hoc**
Descriptive	Contingency tables, loglinear models	Hierarchical clustering, optimal clustering, latent class cluster models
Predictive	Cross-tabulation, regression, discriminant analysis	CART, CHAID, ANN, latent class regression models

Source: Wedel and Kamakura [291]

Table 6.2: Classification of clustering methods

6.2.2.1 Apriori descriptive methods

In apriori descriptive methods for clustering, the segmentation variables and the number of segments are chosen beforehand and displayed into a contingency table. For instance, supermarket shoppers may be segmented by shopping moment (week versus weekend shoppers) and by profession to find out whether there exists a relation between these two bases by calculating the Chi-squared statistic (for an example, see section 4.4.4) on the resulting contingency table.

In the case where multiple segmentation bases are used, higher order interactions are difficult to detect with contingency tables. The latter advocates the use of loglinear models (see section 3.2.3.2).

A well-known apriori descriptive method for clustering customers into distinct groups is the RFM (Recency-Frequency-Monetary value) framework. In this framework, customers are classified into groups according to their most recent purchase, their frequency of purchasing and the average amount spent per purchase. This way, response rates on mailing campaigns can be observed per segment and the best segments can be chosen for future marketing campaigns. An illustration of RFM-based segmentation on retail scanner data is provided later in this chapter in section 6.3.

6.2.2.2 Post hoc descriptive methods

Post hoc descriptive methods are probably the most popular methods for segmentation analysis. The purpose it to discover homogeneous segments of instances along a set of segmentation variables by means of analytical clustering methods.

Probably the oldest but still most frequently used techniques for clustering are based on measures of distance or similarity to group customers into segments on the basis of a number of segmentation variables. Depending on the type of the variable(s) being used, this distance or similarity measure may take different forms. Popular distance measures for metric variables are the correlation coefficient, Euclidean distance, Manhattan distance (also called city-block), Minkowski and Mahalanobis. Distance metrics for binary data include, amongst others, the simple matching coefficient and Jaccard's coefficient. In the case of variables of mixed types, Gower's distance measure is often used. In fact, the choice for a particular distance metric should be carefully evaluated in the context of theoretical arguments. Besides the selected distance measure, the clustering procedure itself may be of a *hierarchical* or a *non-hierarchical* nature [125].

Hierarchical clustering techniques in turn can be either *agglomerative* or *divisive*. Both techniques produce a so-called dendogram (see figure 3.2 in section 3.2.3.1) or cluster tree but they differ in terms of how this tree is built. Agglomerative techniques, such as single linkage, complete linkage, average linkage, centroid and Ward's initially produce as many clusters as there are instances. Subsequently, in each successive step aggregating those instances or clusters that are most similar from the previous step generates new clusters. The clustering ends when all instances are contained in a single cluster (for an example, see section 3.2.3.1). In contrast, divisive clustering, such as Howard-Harris, starts off with all instances in a single cluster and subsequently splits off clusters that are most dissimilar to the existing cluster solution. Divisive cluster methods are far less popular in segmentation research than agglomerative cluster methods.

Non-hierarchical cluster techniques, also called *optimal* clustering or k-means clustering, aims at finding an optimal partition of N instances into k segments on the basis of some criterion. Since an optimal solution has proven to be infeasible for most segmentation problems, a number of exchange-type algorithms have been developed. Such algorithms [153] move subjects around the segments and keep new segments thus obtained only when they improve the criterion of interest. Initial partitions can be obtained at random or by other methods, e.g. using the hierarchical cluster solution as the starting solution for the algorithm.

Recently, as a result of increased computer power, model-based clustering (also known as latent class cluster models or finite mixture models) [280] have gained increasing attention in the segmentation literature. This technique deserves our special interest since we will use it for clustering supermarket shoppers in the next chapter. Therefore, we will devote a separate section (7.1 to 7.3) to this technique. For now, it is important to remember that latent class cluster analysis is a statistical model-based clustering technique where the observed data is assumed to arise from a number of apriori unknown segments or latent classes that are mixed in unknown proportions. The objective of model-based clustering is then to 'unmix' the observations and to estimate the size of each segment and the parameters of the underlying density distributions within each segment.

Also from the data mining point of view, some interesting post hoc descriptive approaches to clustering have been developed. Specifically, the techniques of topographic maps and hypergraph partitioning are worth to be mentioned in this context since they do not rely on the notion of distance to cluster observations into groups.

Topographic maps were first introduced by Kohonen [162] as a special form of neural networks and were designed for multidimensional data reduction with topology-preserving properties. Similar to principal component analysis or multidimensional scaling, a SOM (self organizing map) constructs a low (usually two) dimensional mapping of the original higher-dimensional data. The objective is to reduce the dimensionality of the input data whilst preserving the structure of the input data as accurately as possible. In other words, neurons that are topological neighbours in the output space should also cover observations that are neighbours in the input space. Basically, the clustering proceeds as follows. Firstly, the two-dimensional output space (called a topographic map) consists of a number of neurons that are arranged in a topographically random way, i.e. with random vector weights. Subsequently, the observations (input data) are presented to the network in an iterative procedure. After the presentation of each input vector, the winning output neuron is identified by calculating the (Euclidean) distance between the input vector and the output vector. Subsequently, the winning output neuron's vector weights (and those of the neighbouring neurons) are updated at a particular, user-set, learning rate so that the winning neurons rapidly become prototypes or 'representatives' of a specific type of input data patterns (see also [276] for an extensive overview and improvement of this technique).

Hypergraph partitioning [126, 127] was already introduced in section 4.4.3. Briefly, a hypergraph $H=(V,E)$ consists of a set of vertices (V) and a set of hyperedges (E). In the case of association rules, the vertex set corresponds to the distinct items in the database and the hyperedges correspond to the relationships between the items as determined by the itemsets in which they appear. Furthermore, each hyperedge has associated with it a weight that corresponds to the strength of the hyperedge and is usually calculated as the average confidence of the rules that can be generated with the items in the hyperedge. Next, a hypergraph-partitioning algorithm is used to partition the hypergraph such that the weight of the hyperedges that are cut by the partitioning is minimized. In other words, the number of relations that are violated by partitioning the items into different clusters is minimized. This leads to a clustering of items into item groups. From these item groups, a clustering of the transactions (market baskets) can be obtained by assigning each transaction a cluster score. The more items a particular transaction has in common with a cluster of items, the higher the score on that cluster. The transaction is then finally assigned to the items cluster where it has the biggest score. Customers can be assigned to clusters in a similar way.

6.2.2.3 Apriori predictive methods

Apriori predictive clustering methods refer to methods where the type and number of segments are defined apriori, but where a set of independent variables is used to predict cluster membership. Basically, two approaches can be followed [293], i.e. the *forward* and the *backward* approach.

In the forward approach, general customer characteristics, such as demographics and psychographics are used to apriori form a number of clusters that are subsequently related to product-specific measures of consumer behaviour. For instance in our case, the supermarket retailer could segment the shoppers according to the ownership of a microwave and/or freezer, to subsequently test whether this affects the purchase frequency of pre-packed meals in the supermarket.

In the backward approach, product-specific measures of consumer behaviour are used to form apriori segments, to subsequently profile these segments according to general customer characteristics. For instance, the supermarket retailer would define a number of segments of ready-made meals in terms of the usage frequency, to subsequently profile heavy versus light users in terms of the ownership of a microwave or a freezer. In other words, the retailer would then test whether the ownership of a microwave and/or freezer are able to discriminate between light versus heavy users of the ready-made meals category.

Popular techniques to test this relationship are (logistic) regression, discriminant analysis and supervised data mining techniques, such as classification trees and neural networks.

6.2.2.4 Post hoc predictive methods

Post hoc predictive methods are used to identify customer segments on the basis of the estimated relationship between a dependent variable and a set of independent variables. The main difference between apriori and post hoc predictive methods is that in the former, the segments are formed on an apriori basis. In other words, the values of the dependent variable are grouped into classes by means of a business rule, a heuristic or something similar. However, in post hoc predictive segmentation, the selected algorithm carries out the grouping of the values on the dependent variable automatically.

6.2.3 Quality of Segmentation

In order to validate the quality of the obtained cluster solution in a market segmentation context, a number of criteria have been proposed in the literature [106, 165].

6.2.3.1 Identifiable clusters

Identifiability refers to the extent that segments can really be identified (separated) using real data.

6.2.3.2 Substantial clusters

Clusters should be substantial in size in order to be economically attractive. According to Henneking [135] this is particularly relevant for local retailers who are not part of a larger retail chain and who are therefore limited to advertise in local magazines only. Local retailers can not afford national or pan-regional advertising and therefore they can not advertise for multiple small segments. It is however our opinion that the definition of 'substantial' is probably different for the type of retail business. For instance, an online retailer (e-commerce store) is much more flexible in setting up a personal retail store based on the store visitor's profile. Indeed, an online retail store can be customized with almost no extra cost whereas this is almost impossible in a traditional store environment. The traditional retail store is probably most rigid of all since the store layout can not be changed for each different customer. Somewhere in between is probably the post-order business, where some level of customization is possible by printing different coupons and brochures per customer segment.

6.2.3.3 Accessible clusters

Accessibility refers to the extent to which the customers in the segments can be reached, e.g. by means of advertising or by a more direct approach. The question is whether the retailer knows where the customers in each segment are located. The use of frequent shopper cards in the retail business facilitates the accessibility of those clusters since the shopper usually provides address information in return for participation in the frequent shopper program. It then largely depends on the type of communication that the retailer wants to setup, which determines how easy it is to access the customers in those segments. For instance, if customers present their frequent shopper card at the checkout, their segment membership can be verified immediately and customized advertising messages can be printed on their cash receipts. For example, Catalina Marketing [73], a US-based retail consultancy company, sells point-of-purchase electronic couponing systems that can be implemented to print coupons for a particular category to be used for the next purchase, based on the customer's cluster membership or his current purchases made in other product categories.

6.2.3.4 Stable clusters

Cluster stability is expressed in terms of two criteria: structural stability and temporal stability. *Structural stability* refers to how well the cluster description remains unchanged for different cluster solutions. For instance, if the cluster solutions are very discontinuous over the range of parameter settings, the quality of the clustering can be questioned. However, if by adding more clusters to the cluster model, the cluster configuration does not change too drastic, we are more confident in the quality of the cluster solution. *Temporal stability* refers to the stability of a cluster over time. Particularly for smaller retailers, it is important that segments are relatively stable over time. Smaller retailers generally do not have the means to afford frequent segmentation analysis studies and consequently they allocate resources and marketing activities for a longer time. If segments change in the meantime, the allocation of resources is no longer optimal. Therefore, the segments should be stable at least for some period of time.

6.2.3.5 Responsive clusters

Responsiveness refers to the extent to which the different market segments respond uniquely to the marketing efforts directed at them. In other words, clusters are responsive when the customers in a particular cluster react in accordance with the expected response provoked by the retailer and when this reaction is different between the clusters. This is important since it is the ultimate objective of the segmentation, i.e. to create targeted marketing actions with improved overall success.

6.2.3.6 Actionable clusters

Actionability refers to the extent to which the identified market segmentation provides direction of marketing efforts. This not only means that segments must react differently to marketing efforts, but also that the required marketing efforts are consistent with the strengths and core competencies of the retailer. In other words, a segment is not actionable if it requires marketing actions that are not in line with the strategic marketing choices made by the retailer.

6.3 Segmentation Based on (R)FM

This section provides an illustration of RFM-based segmentation on retail scanner data.

6.3.1 Introduction

Segmentation based on recency-frequency-monetary value (RFM) is probably amongst the most popular in practice [156]. The idea behind it is rather simple and yet very powerful, namely that 'past behaviour is usually a good predictor for future behaviour'. In the case of RFM analysis, this behaviour is summarized by means of three variables: the recency of purchase (e.g. the time since the last purchase), the frequency of purchases (e.g. the number of purchases per year), and the monetary value per purchase (e.g. the average sales amount per purchase). RFM analysis is, however, an apriori segmentation method. This means that the clusters are determined in advance by the user. Therefore, the user needs to decide on two aspects: firstly, the level of detail of the analysis, which is determined by the amount of bins (intervals) for each variable, and secondly, the associated cutoff values for each bin. Usually, these decisions are made based on practical experience or on exploratory data analysis. Both decisions give rise to a matrix of which the number of cells are determined by the decision into how many bins each variable should be divided. In the case where each variable of the three RFM variables is divided into 3 bins, a matrix of 27 (3x3x3) cells is generated. In the context of mail-order companies, where RFM analysis is extremely popular [274], response rates are subsequently calculated for each cell in the matrix in order to determine which cells (e.g. those with the highest response rate) will be targeted for a particular marketing campaign. This way, mailings are sent only to customers in those segments that have the highest propensity to respond.

In the context of retail market basket analysis, RFM analysis can be useful to segment supermarket shoppers based on their recency of shopping (R), their shopping frequency (F) and the average monetary value spent per basket (M) in order to find out whether differences in the purchase behaviour between those segments can be discovered. In this context, differences in purchase behaviour will be expressed in terms of differences in frequent itemset combinations. More specifically, we are interested in finding out whether some product categories tend to be visited more frequently, or whether the interdependency between those product categories is much different between the defined segments. If so, the retailer may use this information to customize his offer towards those segments. For instance, if highly loyal and valuable shoppers tend to visit some product categories (say fresh vegetables and meat) more often than non-loyal or invaluable shoppers, then the retailer should make extra efforts to make sure that loyal customers are satisfied about the products and service provided in those categories. In the case where the interdependency between two product categories is much stronger in one segment than in the other, the retailer might also use this information to further examine the causal relationships between those product categories for purposes of pricing, product placement or promotions.

6.3.2 *Intervals and Cutoff Values*

Firstly, instead of using all three variables (R, F and M) we will only use frequency and monetary value to carry out the segmentation analysis. We will not use recency (R) since in the context of supermarket retailing, recency is of less importance given that shoppers usually visit a supermarket on a very regularly basis[32] (mostly weekly).

Secondly, we have to decide on the number of bins for each variable and the cutoff values for each bin. In fact, it was our first idea to divide each variable into three bins of equal frequency resulting into 9 cells in the FM matrix. However, this proved to be practically infeasible due to two reasons. Firstly, the number of retail baskets in each cell tends to become very small, especially for those cells with high values for each variable. Secondly, a comparative analysis of the purchase behaviour between the segments becomes problematic due to the fact that for each segment, hundreds or even thousands of frequent itemsets will be generated which makes comparison very cumbersome. Therefore, we decided to discretize the F and M variables into two bins and analyse only the two extreme segments, i.e high F and high M versus low F and low M.

In order to find reasonable cutoff values for this application, we rely on the distribution of the data for the number of visits over the period of data collection (F), and the average amount spent per basket (M). For both variables, we use the mean of the distribution as the cutoff value. The distribution of the number of visits over the period of data collection was already given in figure 2.3 in chapter 2. It turned out that for this dataset, the average number of visits over a period of 24 weeks was equal to 25. Therefore, we define high frequency customers for this application as those who visit the store at least 26 times in total.

---

[32] However, in the context of a mail-order company, recency is of greater importance since that the intervals in between two purchases are usually longer than in supermarket retailing.

Although this number could be debated, it is clearly much higher than the mode of the distribution and thus it reflects more or less the idea of high frequency shoppers. After all, these shoppers visit the store on average more than once per week.

Figure 2.1 in chapter 2 shows the distribution of the average amount (in old Belgian francs) spent per basket. The histogram shows that the average amount spent per customer and per basket equals 1276 Belgian francs. Therefore, we define high value customers as those who spent at least 1277 Belgian francs on average per visit. This number is clearly much higher than the modal amount spent per basket over the entire group of customers and therefore it reflects the idea of customers of high monetary value.

Table 6.3 summarizes the number of shoppers and their associated number of baskets (between brackets) for each segment in the FM matrix.

		Monetary value (M)		
		Low <=1276	High >1276	Total
Frequency (F)	Low <=25	**1117 (14673)**	908 (12295)	2025 (26968)
	High >25	765 (40828)	**361 (13833)**	1126 (54661)
	Total	1882 (55501)	1269 (26128)	3151 (81629)

Table 6.3: FM segments

As indicated before, the analysis will thus be concentrated on the two extreme segments (HFHM = high frequency and high monetary value, LFLM = low frequency and low monetary value) as indicated in bold in table 6.3. For each segment frequent itemsets will be generated and compared, as discussed in the following section. Furthermore, we use support and interest for each frequent set as a means of comparison. A ratio value above one (last column of table 6.4) indicates that the product categories contained in the frequent set are more strongly related in the group of high frequency and high monetary value customers than in the group of low frequency and low monetary value customers. A value below one indicates the reverse.

6.3.3 Comparison of Results

Frequent itemsets, with their respective support (definition 4.5) and interest (formula 4.1), were generated for each of the two segments. Table 6.4 shows the support and interest for those itemsets for which the difference in the two segments is big.

Itemset	Support			Interest		
	HFHM (1)	LFLM (2)	Ratio (1)/(2)	HFHM (3)	LFLM (4)	Ratio (3)/(4)
{fresh bread, fresh meat, yoghurt}	3.62	0.79	4.58	1.45	0.89	1.62
{fresh bread, coffee, fresh vegs & fruit}	2.30	0.56	4.11	1.47	1.00	1.47
{waters, fresh meat, fresh deli}	2.76	0.57	4.84	1.17	0.83	1.40
{waters, fresh meat, fresh cheese}	4.18	0.71	5.89	1.59	1.22	1.31
{fresh bread, milk, fresh meat}	4.45	1.58	2.82	1.38	1.07	1.29
{soft drinks, pastry}	1.35	0.54	2.5	1.10	0.86	1.28
{fresh bread, milk, fresh vegs & fruit}	4.08	1.28	3.19	1.38	1.09	1.27
{fresh bread, milk, margarine spread}	1.93	0.52	3.71	1.76	1.39	1.26
{fresh bread, milk, fresh meat, fresh vegs&fruit}	3.43	0.89	3.85	1.79	1.48	1.21
{candy bars, crisps}	3.54	0.59	6	1.97	1.76	1.12
{soft drinks, milk, waters}	3.86	0.99	3.90	3.14	4.69	0.67
{washing-up liquid, maintenance}	2.14	0.64	3.34	2.23	3.38	0.66
{canned vegetables, canned fish}	1.28	0.59	2.17	1.73	2.65	0.65
{frozen vegetables, frozen soups}	1.37	0.60	2.28	2.10	3.34	0.63
{candy bars, confectionery, dry biscuits}	2.96	0.62	4.77	3.77	6.01	0.63
{fruit juice, waters}	3.20	0.99	3.23	1.79	2.89	0.62
{maintenance, abrasives}	1.79	0.63	2.84	2.59	4.29	0.60
{milk, whipped cheese, yoghurt}	2.87	0.69	4.16	3.08	5.42	0.57
{refrigerated desserts, milk, yoghurt}	1.66	0.54	3.07	2.47	5.55	0.44
{crisps, soft drinks, nuts & appetizer biscuits}	1.41	0.52	2.71	5.15	14.31	0.36
{crisps, dry biscuits, nuts & appetizer biscuits}	1.66	0.62	2.68	5.39	15.58	0.34

Table 6.4: Differences in purchase behaviour for HFHM and LFLM segments

First of all, it can be observed from table 6.4 that the interdependency between 'fresh' product categories, such as fresh meat, fresh vegetables and fruit, fresh bread and waters is much stronger in the HFHM segment than in the

LFLM segment. This is maybe not surprising since the fact that most of the 'fresh' product categories contain perishable items and these items must be purchased on a regularly basis which may explain the relatively high store visit frequency for this group of customers. The fact that the interaction with 'fresh' product categories is much stronger in the HFHM segment, and since this is clearly an attractive group of customers for the retailer, should motivate the retailer to make sure that HFHM customers will be kept satisfied about the product offering and service in these product categories.

Secondly, with regard to the support of the frequent sets, it can be noticed that in general the support of any frequent set is bigger within the HFHM segment than in the LFLM segment. This can be explained by the fact that the average basket size within the HFHM segment is 11.14, whereas within the LFLM segment it is only 6.16, such that statistically speaking, the probability of two product categories occurring together in a basket in the HFHM segment is much bigger than in the LFLM segment. However, when looking at the ratio of the support of the frequent sets (i.e, ratio (1)/(2)), an interesting difference can be observed. In fact, for the frequent sets for which the interdependency is stronger in the LFLM segment (frequent sets below the dashed line in table 6.4), it turns out that, on average, the drop in support (LFLM versus HFHM) for those frequent sets is less pronounced than for the frequent sets above the dashed line. In fact, the average ratio (1)/(2) equals 4.15 above the dashed line and only 3.19 below the dashed line.

This leads to the following conclusion, not only the strength of the interdependency between 'fresh' product categories, but also their frequency of purchase, is stronger within the HFHM segment than within the LFLM segment. This can be concluded by the bigger ratio (3)/(4) and the smaller ratio (1)/(2) for those itemsets compared with the itemsets that do not contain 'fresh' product categories. One can therefore conclude that these 'fresh' product categories are of a strategic importance to the retailer in order to keep HFHM customers satisfied.

6.4 Segmentation Based on Basket Size

The size of the basket is another, yet natural way of segmenting customers.

6.4.1 Introduction

The reason why the size of a basket is often used as a means for segmentation is that it serves as an indicator for the type of shopping trip that the customer made [28, 148, 158, 224, 225]. For instance, in the 'Marsh Super Study' [225], stock-up shoppers are those customers who purchase more than 35 items (they account for only 16% of the customer population), routine shoppers buy 11 to 34 items and account for 41%, and fill-in shoppers buy 10 or fewer items and account for 43% of the customer population.

If differences in the purchase behaviour can be discovered for different basket sizes, then it might be appealing for the retailer to segment his customer population based on the average basket size and target those customers with customized product offers and promotional campaigns. For instance, suppose it is found that certain products or product categories (like fresh vegetables and fruit and bread) are extremely popular amongst stock-up shoppers, then a particular customer who is stocking-up once per week but who does not usually purchase bread in the store could be offered a reduction on bread for the next stock-up purchase in the store. In fact, the technology to print customized promotional offers on the customer's receipt is available today.

6.4.2 Intervals and Cutoff values

Here, we base our analysis on the same histogram (figure 2.1, chapter 2) of basket sizes. In that context, we previously defined fill-in baskets as those with 10 or fewer items. We will contrast these baskets against stock-up baskets, which we define here as those containing 20 items or more. This results in 49346 fill-in baskets versus 15229 stock-up baskets. In between are those baskets from size 11 to 19, which are not analysed here. So, we focus on a 'high-low' analysis where two datasets with two extreme and opposite basket sizes are used.

6.4.3 Comparison of Results

Finding a good way to compare the results from this market basket analysis for the two data sets is not easy. First of all, with a minimum support=1%, several thousands of frequent sets are generated per segment, which makes comparison very laborious. Furthermore, it is a problem to analyse frequent sets of a different size. Indeed, frequent sets containing many product categories will automatically have higher support in stock-up baskets compared to fill-in baskets since in the former, the average size of the baskets is much larger than in the latter. Therefore, we will concentrate the comparison of the results on frequent sets of size 2 and 3, for which the difference in 'interest' in both datasets is high, as illustrated by table 6.5. A ratio (last column) above one means a stronger interdependency between the items for stock-up purchases, whereas a ratio below one means a stronger interdependency for fill-in purchases.

The following conclusions can be drawn from examining the list of itemsets.

First of all, table 6.5 shows that, unless one or more 'fresh' product categories are involved (like fresh vegs & fruit, fresh bread, fresh meat, fresh cheese), the interest values for frequent sets are usually higher for fill-in baskets than for stock-up baskets. In fact, it seems that the interaction with the 'fresh' categories is much more pronounced within stock-up purchases than within fill-in purchases. Indeed, it is clear from looking at the interest values in table 6.5 that the interaction between most product categories, such as washing powders, washing-up liquids, softener, tobacco, chocolate, dry and fresh biscuits, etc. is much stronger for fill-in purchases than for stock-up purchases, except when there is at least one 'fresh' category involved.

Itemset	Support			Interest		
	Stock-up (1)	Fill-in (2)	Ratio (1)/(2)	Stock-up (3)	Fill-in (4)	Ratio (3)/(4)
{fresh bread, fresh sandw, fresh vegs&fruit}	1.99	0.11	18.09	1.37	0.63	2.16
{fresh bread, soft drinks, fresh deli}	2.24	0.11	20.36	1.06	0.51	2.06
{fresh bread, fresh meat, fresh deli}	4.86	0.91	5.34	1.24	0.83	1.49
{fresh bread, fresh sandwiches, fresh meat}	2.12	0.26	8.15	1.59	1.06	1.50
{pastry, fresh meat, fresh deli}	1.46	0.17	8.59	1.43	0.99	1.44
{pie/biscuit/cake, fresh meat, fresh vegs.fr}	1.66	0.20	8.30	1.26	0.90	1.39
{fresh bread, buns, fresh vegs&fruit}	1.67	0.23	7.26	1.59	1.21	1.31
{bake-off, fresh bread}	2.54	0.53	4.79	1.21	0.97	1.25
{fresh bread, fresh vegs&fruit, fresh cheese}	7.20	0.73	9.86	1.46	1.33	1.10
{dry biscuits, fresh biscuits}	21.04	1.14	18.46	1.25	2.53	0.49
{regular beers, heavy beers}	1.03	0.31	3.32	5.21	10.65	0.49
{tobacco paper, tobacco}	1.55	0.14	11.07	20.81	44.58	0.47
{pastas, sauces}	9.91	0.61	16.25	1.31	2.86	0.46
{frozen meat, frozen potatoes}	1.56	0.24	6.50	2.84	6.89	0.42
{frozen meals, frozen potatoes}	1.14	0.15	7.60	2.21	4.93	0.45
{frozen vegetables, frozen fish}	2.08	0.23	9.04	2.36	4.00	0.59
{washing-up liquid, maintenance}	4.31	0.28	15.39	1.66	4.19	0.40
{maintenance, liquid detergents}	3.21	0.19	16.89	1.63	4.41	0.37
{maintenance, abrasives}	3.66	0.28	13.07	1.88	5.84	0.32
{washing powder, softener}	2.55	0.17	15	2.45	8.17	0.30
{milk, whipped cheese, yoghurt}	6.74	0.12	56.16	1.50	5.04	0.30
{chocolate, dry biscuits, fresh biscuits}	9.15	0.15	61	1.68	5.92	0.28
{crisps, nuts & appetizer biscuits}	8.15	0.43	18.95	2.13	7.66	0.28
{soft drinks, waters, regular beers}	1.09	0.11	9.91	2.94	12.5	0.23
{dog food, cat food}	1.65	0.12	13.75	1.45	3.94	0.37
{baking margarine, margarine spread}	15.10	1.04	14.52	1.41	3.66	0.39

Table 6.5: Differences in purchase behaviour for big versus small baskets

Similar results were obtained for the FM analysis where the inclusion of a 'fresh' product category resulted in a higher interest value for the frequent set in the high frequency high monetary value segments, than in the low frequency low monetary value segment.

This is therefore a strong indication again that valuable customers, i.e. those that spend a lot and visit the store frequently, have a strong preference for (combinations of) 'fresh' products and that the retailer should pay special attention to these products to ensure that they satisfy the wants and needs of the shoppers.

Secondly, table 6.5 shows a stronger interaction between frozen food items, like frozen fish, frozen potatoes, frozen meat and frozen vegetables in fill-in baskets than in stock-up baskets, as indicated by the low ratio values (3)/(4) in the last column. In other words, although the probability of observing combinations of frozen food products is smaller in fill-in baskets than in stock-up baskets, the interaction between frozen food items tends to be much stronger in fill-in baskets than in stock-up baskets. The retailer could therefore promote frozen food products during weekdays (during which there are more fill-in purchases) instead of during weekends (during which there are more stock-up purchases) in order to boost sales of frozen food products and take advantage of the stronger interaction between them.

Thirdly, the support values for the frequent sets are much lower in the fill-in baskets than in the stock-up baskets. As already discussed in the previous section, this can be explained by the basket size, which in this application is even more different between fill-in and stock-up baskets.

6.5 Discussion

Despite the numereous applications where behaviour is used as a basis for segmentation, it is sometimes argued that in contrast to other bases for segmentation, behaviour has no theoretical justification in consumer theory. In our opinion, however, the justification of choosing behaviour as a basis for segmentation has more to do with the pragmatic methodology of segmentation, than with its theoretical underlyings. By this we mean that there are basically three ways of approaching segmentation from a methodological point of view (as already introduced in section 6.2.2.3, i.e. forward segmentation, backward segmentation and simultaneous segmentation.

In forward segmentation, segments are created based personal characteristics, preferences, values and lifestyle of custeromers, i.e. other variables than behaviour. Once that segments have been identified, one then hopes to find differences in consumer behaviour between the discovered segments. The advantage of this approach is that usually these segments can be easily reached since customers within the same group are similar in terms of, for instance, their age and residence.

In backward segmentation, segments are created based on the customer's purchase and usage behaviour after which these segments are profiled in terms of other variables than behaviour, like socio-demo and lifestyle data. The advantage of the backward segmentation approach is that homogeneous segments can be found in terms of the purchase and/or usage behaviour and that therefore it is easier to customize a product offer or communication policy towards these segments.

In the simultaneous approach, segments are created based on both personal customer characteristics and customer behaviour. The concomittant variable mixture approach, discussed in section 7.4.1 in the next chapter, is a perfect example of this approach. The advantage is that segments are created based on differences in purchase or usage behaviour and that other descriptive variables (like socio-demo and lifestyle data) are used to explain the differences in this behaviour simultaneously (i.e. during the formation of the clusters).

The behaviour-based segmentation approach discussed in this dissertation is thus clearly an example of the backward segmentation approach. However, our interest is primarily in finding good separated segments in terms of purchase behaviour and not really in profiling these segments afterwards. Nevertheless, the methods used in this dissertation do not exclude this second step.

CHAPTER 7
MODEL-BASED CLUSTERING

Instead of using basket size or recency-frequency-monetary value variables to segment customers '*apriori*', model-based clustering provides an '*a posteriori*' methodology for segmentation where segments can be automatically discovered from the data. Model-based clustering has gained a lot of attention during the last few years because it is a powerful methodology that provides a statistically sound basis for segmentation.

In this chapter, we will first introduce the general idea and principles of model-based clustering. After this, we will introduce a model-based clustering methodology based on multivariate Poisson mixtures. The idea is to cluster supermarket shoppers according to their purchase rates in a set of product categories. The methodology that we propose is therefore also a behaviour-based segmentation model, where the purchase rates in a number of product categories are used to cluster shoppers into groups. From a technical perspective, the objective is to introduce an innovative methodology for behaviour-based customer segmentation by means of a multivariate Poisson mixture model. More specifically, several variants of the mixture model will be proposed where the segmentation variables are defined as the customer purchase rates in a set of predefined product categories. The use of the Poisson distribution (instead of for instance the Normal distribution) is motivated based on the distributional characteristics of the purchase rates. Furthermore, the concept of product interdependence will again play an important role by explicitly accounting for interdependencies that exist between purchase rates in different product categories. In this respect, the treatment of the variance-covariance matrix for this kind of models is unique and is to our knowledge not yet presented in the literature before.

From a practical perspective, insights gained from cross-category analysis coupled with customer segmentation might be of interest both to retailers as to manufacturers and database marketeers. For retailers, the identification of segments with similar purchase rates in a set of product categories may help them to target different groups of customers in a different way. By setting-up a more personalized communication (either in or out of the store), customers can be presented more customized offers. Manufacturers of FMCG's, who market brands in related categories, could utilize these insights to rationalize marketing expenditures across two or more categories, e.g. fabric detergent and fabric softener and database marketeers are interested in discovering and converting cross-category dependencies to targeted cross-selling programs [30]. However, despite the significant industry interest in understanding multi-category purchase interdependencies, it is only recently that researchers have started examining purchases across multiple categories [232].

This research will thus result in a range of different models for segmentation. On the one end of the spectrum, there is the local independence Poisson mixture model (section 7.7.3), where no interdependencies between the product category purchase rates are assumed at all. On the other hand of the spectrum, there is the fully-saturated Poisson mixture model (section 7.7.1) where purchase rates between categories are allowed to be freely correlated. In between those two extreme models, there is the multivariate Poisson mixture model with common covariance (section 7.7.2), which is by far the most frequently used variant in the literature. The main contribution of this dissertation, however, will exist in the development of a theoretically more complete model in between the two extreme models detailed above, called the multivariate Poisson mixture model with restricted covariance structure (section 7.7.4). Indeed, knowledge about which purchase rates are correlated (i.e. knowledge about the existing product purchase interdependencies) in our opinion enables the formulation a much simpler model in between the two extremes by including into the variance/covariance matrix only those interaction terms that are statistically significant. The advantage of this restricted model is that it is the most parsimonious model including all relevant and no redundant interaction terms. As a result, it is flexible enough to accommodate most of variance in the data without adding more complexity to the segmentation model than necessary. Furthermore, it facilitates and speeds up the estimation of the multivariate model by reducing the number of parameters to be estimated as much as possible.

This chapter is based on work reported in [51, 57].

7.1 Introduction to Model-Based Clustering

The key idea [280, 291] in model-based clustering, also known as latent class clustering or finite mixture models, is that the observations (in our case customers) are assumed to originate from a mixture of density distributions for which the parameters of the distribution and the size and number of the segments are unknown. It is therefore the objective of model-based clustering to unmix the distributions and to find the optimal parameters of the distributions, and the number and size of the segments, given the underlying data.

The history of finite mixture models dates back more than 200 years to the work of Pearson [217]. Its breakthrough, however, came with the advent of high speed computers, turning the attention to likelihood estimation of the parameters in a mixture distribution [115, 195]. In particular, the formalization of the EM algorithm (expectation – maximization) by Dempster et al. [89] has given a new impetus to the research of finite mixture models. Since then, an extensive literature has been published on the topic, although the majority of publications date from 1995 onwards [197].

It is not within the scope of this chapter to provide an exhaustive overview of the model-based clustering approach, but rather to focus on the concepts and techniques that are most relevant for the development of the cluster model in this dissertation. Therefore, before proceeding to the formulation of the finite mixture model for clustering supermarket shoppers (see section 7.6), the next section will provide an overview of the general formulation of the finite mixture model and is mainly drawn from state-of-the-art books and review articles [96, 195, 197, 265, 289] in this domain.

7.2 Formal Description of Model-Based Clustering

In general, in model-based clustering, the observed data are assumed to arise from a number of apriori unknown segments[33] that are mixed in unknown proportions. The objective is then to 'unmix' the observations and to estimate the parameters of the underlying density distributions within each segment. The idea is that observations (in our case supermarket shoppers) belonging to the same class are similar with respect to the observed variables in the sense that their observed values are assumed to come from the same density distributions, whose parameters are unknown quantities to be estimated. The density distribution is used to estimate the probability of the observed values of the segmentation variable(s), conditional on knowing the mixture component from which those values were drawn.

The population of interest thus consists of k subpopulations and the density (or probability function) of the q-dimensional observation $\boldsymbol{y}$ from the j-th ($j=1,\ldots, k$) subpopulation is $f(\boldsymbol{y}|\theta_j)$ for some unknown vector of parameters θ_j. The interest lies in finding the values of the non-observable vector $\varphi = (\phi_1, \phi_2, \ldots, \phi_n)$ which contains the cluster labels for each observation ($i = 1, \ldots, n$) and $\phi_i = j$ if the i-th observation (e.g. household) belongs to the j-th subpopulation.

Since we do not observe the cluster labels, the unconditional density of the vector $\boldsymbol{y}$ is a mixture density of the form

$$f(y_i) = \sum_{j=1}^{k} p_j f(y_i \mid \theta_j) \tag{7.1}$$

where $0 < p_j < 1$, and $\sum_{j=1}^{k} p_j = 1$ are the mixing proportions. Note that the mixing proportion is the probability that a randomly selected observation belongs to the j-th cluster.

---

[33] Segments, components, latent classes or clusters are synonyms and will sometimes be used interchangably.

This is the classical mixture model (see [42, 197]). The purpose of model-based clustering is to estimate the parameters ($p_1, \ldots, p_{k-1}, \theta_1, \ldots, \theta_k$). Following the maximum likelihood (ML) estimation approach, this involves maximizing the loglikelihood:

$$L(y;\theta,p)=\sum_{i=1}^{n}\ln\left(\sum_{j=1}^{k}p_j f(y_i \mid \theta_j)\right) \tag{7.2}$$

which is not easy since there is often not a closed-form solution for calculating these parameters. Fortunately, due to the finite mixture representation, an expectation-maximization (EM) algorithm is applicable (see section 7.3.1 and 7.7.5).

In the case of a *multivariate* mixture model, it is often assumed that the observed variables are mutually independent within clusters in order to avoid computational difficulties [280]. Indeed, if no restrictions are imposed on the interdependence of variables, the model with multivariate probability density functions is applicable. In that case, the model-based clustering problem involves estimating a separate set of means, variances, and covariances for each mixture component, which quickly becomes computationally cumbersome.

In between the local independence model and the full covariance model, several types of restrictions with regard to the variance-covariance matrix can be imposed. In some cases, this may be necessary for practical reasons since the unrestricted model may be computationally infeasible. The reason is that the number of free parameters in the variance-covariance matrix for the full-covariance model increases rapidly with the number of mixture components and the number of indicator variables. Therefore, often more restricted models are defined by assuming certain pairs of y's to be mutually independent within mixture components by fixing some but not all covariances to zero. Also the equality of variance-covariance matrices across components is sometimes adopted as a restriction resulting in clusters that have the same form, but different locations [152, 173].

7.3 Model-Based Cluster Estimation

The purpose of model-based clustering, defined above, is to estimate the parameter vector Φ. The two main methods to estimate this parameter vector are maximum likelihood (ML) and maximum posterior (MAP) [280] of which the former is the most frequently used.

7.3.1 ML Estimation With the EM Algorithm

The purpose of model-based clustering is to estimate the parameters $(p_1, \ldots, p_{k-1}, \theta_1, \ldots, \theta_k)$. Following the maximum likelihood (ML) estimation approach, this involves maximizing the loglikelihood (7.2), as stated before. In other words, the objective is to find the optimal values for the parameter vector, say Φ_{opt}, such that the observations y_i $(i = 1, \ldots, n)$ are more likely to have come from $f(y_i|\Phi_{opt})$ than from $f(y_i|\Phi)$ for any other value of Φ.

In order to maximize this loglikelihood, most software tools either use Newton-Raphson [192] or expectation–maximization (EM) (see [89], [196]), or a combination of both. More recent techniques, such as MCMC (Markov Chain Monte Carlo) [229] and stochastic EM [95] are increasing in popularity. Furthermore, since EM is relatively slow and databases are increasing in size, recent research efforts concentrate on incremental EM algorithms for use on very large data sets [197]. Although the Newton-Raphson algorithm requires less iterations compared with the EM algorithm, convergence to a local optimum is not guaranteed [195]. Furthermore, because of its computational simplicity, the EM algorithm is the most widely used [264]. Later in this chapter (section 7.7.5), we will provide a detailed version of the EM for our mixture models. At this point, the EM can be described as an iterative algorithm that sequentially improves upon sets of starting values of the parameters, and enables simultaneous estimation of all model parameters (see [89], [133], [291]). More specifically, instead of maximizing the likelihood over the entire parameter space, the observed data y_i is augmented with the *unobserved* segment membership of subjects z_{ij}, which greatly simplifies the computation of the likelihood. More details about the EM computation can be found in [89, 195, 291]. Once an estimate for the optimal value of Φ has been found, the estimates of the posterior probability w_{ij}, i.e. the posterior probability for subject i to belong to component j, can be obtained for each observation vector y_i according to Bayes' rule. Indeed, after estimation we know the density distribution $f(y_i|\theta_j)$ within each mixture component j and we know the segment size p_j of each component such that we can calculate the posterior probability as

$$w_{ij} = \frac{p_j f\left(y_i \middle| \theta_j\right)}{\sum_{j=1}^{k} p_j f\left(y_i \middle| \theta_j\right)} \tag{7.3}$$

7.3.2 Determining the Number of Segments

In some applications of model-based clustering, there is sufficient *a priori* information about the number of components k in the mixture model to be specified with enough certainty. For instance, where the components correspond to externally existing groups. However, in many occasions, the number of components has to be inferred from the data, along with the parameters in the component densities [197]. Unfortunately, this important problem of finding the optimal number of components in a mixture model has not yet been completely solved [291]. From a theoretical perspective, Lindsay [175] proved that beyond a certain number of components, any new components are redundant. However, a more pragmatic perspective to determine the number of segments is based on the use of so-called information criteria to evaluate the quality of a cluster solution. The most well-known examples include AIC (Akaike information criterion) [17], $CAIC$ (Consistent Akaike information criterion) [46] and BIC (Bayes information criterion) [241]:

$$AIC = -2L_k + 2d_k \tag{7.4}$$

$$BIC = -2L_k + \ln(n)d_k \tag{7.5}$$

$$CAIC = -2L_k + [\ln(n)+1]d_k \tag{7.6}$$

Basically, information criteria are goodness of fit measures, which take into account model parsimony. The idea is that the increase of the likelihood of the mixture model (L_k) on a particular dataset of size n, is penalized by the increased number of parameters (d_k) needed to produce this increase of fit. The smaller the criterion, the better the model in comparison with another. However, it should be noted that several variants exist of these criteria, such as AIC3, JAAIC [47], information complexity criterion [45], minimum information ratio criterion [295], approximate weight of evidence (AWE) [26], such that they primarily serve as a guidance tool for the researcher to select the number of components.

Another way for deciding on the number of components in a mixture model is based on a hypothesis test, using the likelihood ratio as the test statistic. However, in order to assess the p-value of the likelihood ratio test, bootstrapping is required which is computationally much more demanding than information based criteria [16]. The idea is to draw T-1 random Monte Carlo samples of size N from a population having S components. The mixture model under consideration is then fit with S and with S+1 components to each of the generated samples and the likelihood ration statistic U is computed. The null hypothesis is then that the mixture model with S components fits as well as model with $S+1$ components. Now, if the value of U obtained from the observed data exceeds $T(1-\alpha)$ of the values of U obtained in the Monte Carlo samples, then the null hypothesis is rejected at the desired significance level α. Dillon and Kumar [96] advise the minimum value of T =20 when using a significance value of α=0.05.

Other methods for evaluating the quality of LC cluster models are based on the separation of the clusters, such as the entropy statistic $I(\mathrm{k})$, which measures how well the segmentation variables are able to predict class membership [197],

$$I(k) = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{k} w_{ij}\ln(w_{ij})}{n\ln(1/k)} \tag{7.7}$$

with the convention that $w_{ij}\ \ln(w_{ij}) = 0$ if $w_{ij} = 0$. In the case of a perfect classification, for each i there is only one $w_{ij} = 1$ and all the rest are 0. This implies a value for the criterion equal to 1. Thus, values near 1 show a good clustering.

7.3.3 Pros and Cons of the EM Algorithm

The EM algorithm is surely one of the most popular algorithms to estimate finite mixture models. This has to do with some of its appealing properties.

Firstly, the most important advantage of the EM algorithm is surely its convergence towards the optimum parameter values. This means that, given the current mixture model parameters, a single EM iteration provides new parameter estimates which are proven not to decrease the loglikelihood of the model [89, 196]. The convergence of the EM was proven by [199, 296]. Secondly, and particularly relevant for the models presented later in this chapter, the EM algorithm finds estimates for the parameters that are within the so-called 'admissible range'. This means that, for instance in the case of the Poisson distribution, the parameter values can not take negative values, which is ensured by the EM algorithm. Finally, the EM algorithm is quite easy programmable.

However, apart from these appealing properties of the EM algorithm, some cons have been identified as well.

First of all, the problem with EM-estimation is that the procedure may converge to a local but not a global optimum [195, 265]. It is generally accepted that the best way to prevent a local solution is to use multiple sets of starting values for the EM-algorithm and to observe the evolution of final likelihood for the different restarts of the EM-algorithm. Another solution is to use the partitioning of a k-means clustering as the initial starting values [195].

Secondly, the EM algorithm usually converges very slowly when compared to other iterative algorithms, such as Newton-Raphson. The reason is that EM converges linearly towards the optimum, whereas Newton-Raphson converges with quadratic speed towards the optimum [15]. Therefore, some commercial computer packages (e.g. Latent Gold [170]) use a combination of the EM and Newton-Raphson, i.e. they start with a number of EM iterations and when close enough to the final solution, they switch to Newton-Raphson. This is a way to combine the advantages of both algorithms, that is the stability of the EM even when far away from the optimum and the speed of Newton-Raphson when close to the optimum.

Thirdly, the convergence towards finding the globally optimal parameter values and its speed of convergence both heavily rely on the starting values for the parameters. Indeed, different initial parameter values may lead to different estimates and one can not be sure whether the global optimum has been achieved. This is another reasons for running the EM algorithm multiple times with different starting values.

Fourthly, an important problem, but rather ignored in the literature, is the stopping rule for the number of iterations. In fact, the EM is rather sensitive in the sense that different stopping rules can lead to different estimates [244, 245, 246]. According to Karlis [151], this is caused by the fact that at every iteration the loglikelihood increases by a very small amount and that at the same time the estimates can change heavily.

Finally, although the EM is a widely used algorithm and thus its general principles are well understood, in every problem one has to build the algorithm in a different way.

7.4 Comparison with Distance-Based Clustering

Model-based clustering has some advantages and disadvantages compared to traditional distance-based cluster methods.

7.4.1 Advantages of Model-Based Clustering

Firstly, when compared to model-based clustering, hierarchical clustering methods have large storage demands (i.e. all pairwise distances need to be stored in a distance matrix) and heavily rely on the distance measure and linkage method being used. K-means clustering, on the other hand, is computationally more feasible but its result heavily depends on the initial starting solution being chosen. Furthermore, k-means clustering is a heuristic methodology that works well only in a very small class of problems. In fact, it can be shown that k-means is a special case of a classification EM algorithm for finite multivariate normal mixtures with spherical components. Therefore, when dealing with data that are not Gaussian distributed, or where clusters are not spherical and there is a significant amount of correlation in the data, model-based clustering should be preferred over simple k-means clustering.

Secondly, recent advances in model-based clustering enable the inclusion of variables of mixed scale types (nominal, ordinal, continuous and count variables) in the same model [281, 282].

Thirdly, because the observations are assumed to be generated by a mixture of underlying probability distributions, a number of statistical tests can be used to check the validity of the cluster model. Indeed, the model-based clustering approach relies on specific assumptions related to the data and therefore allows for a statistical treatment of model selection and comparison.

Fourthly, since model-based clustering is a probabilistic clustering approach, the assignment of subjects to clusters is carried out in a probabilistic way (see Bayes rule in section 7.3.1). This makes it different from traditional clustering techniques where the subject is assigned to just one cluster. Therefore, model-based clustering is sometimes also mentioned in the context of fuzzy clustering, meaning that a subject may belong to different clusters according to some grade of membership. Furthermore, Bayes' rule (7.3) can be used to classify unseen observations into the identified clusters, since their values on the indicator variables can directly be used to compute their individual posterior class-membership probabilities.

Finally, a relatively recent evolution in model-based clustering, called *concomitant variable* mixture models, enables the simultaneous profiling of the derived segments with descriptor variables (also called covariates), such as socio-demographics [120, 149]. This eliminates the need for a second step of analysis with discriminant analysis or logistic regression to relate the cluster results to a set of descriptor variables.

7.4.2 Disadvantages of Model-Based Clustering

Model-based clustering also has some disadvantages compared to traditional distance-based clustering techniques.

First of all, especially when all parameters of the model are allowed to vary freely (in the case of the fully-saturated model), the computation of the cluster model can become very cumbersom. Our own experiments with model-based clsutering corroborate this finding, especially for large datasets such as in retailing. There are indeed severe restrictions both in terms of the number of subjects, the number of replications per subject, and the number of variables considered, which limits their utility in domains where data is abundant. Recent approaches aim at solving this problem, for example by developing incremental EM-algorithms [197]. Secondly, the current state of the algorithms still requires quite a lot of human intervention, especially in terms of the type of probability density distribution(s) to use, and how many and which dependencies between observations to incorporate.

7.5 Data Issues

7.5.1 Data Set

The dataset used for the empirical analysis of the models presented hereafter has been made available by Puneet Manchanda [183] from the University of Chicago Graduate School of Business. The data came from ACNielsen and describe the purchases made by 155 households in 4 product categories (i.e. cake mix, cake frosting, fabric detergent and fabric softener) over a period of 120 weeks from January 1993 to March 1995 and are from a large metropolitan market area in the western United States. The total number of market baskets, purchased by all the households in the sample, amounts to 17389. However, the analysis in this chapter will not be carried out on the raw market basket data with 0/1 purchases.

HouseholdID	C	F	D	S
1	0	1	1	0
1	1	1	0	0
1	0	0	1	0
2	0	1	0	0
2	1	1	0	1
2	1	1	1	1
3	0	0	1	1
3	0	1	1	1
3	0	0	0	1

$\Rightarrow$

HouseholdID	C	F	D	S
1	1	2	2	0
2	2	3	1	2
3	0	1	2	3

C = cakemix
F = frosting
D = detergent
S = softener

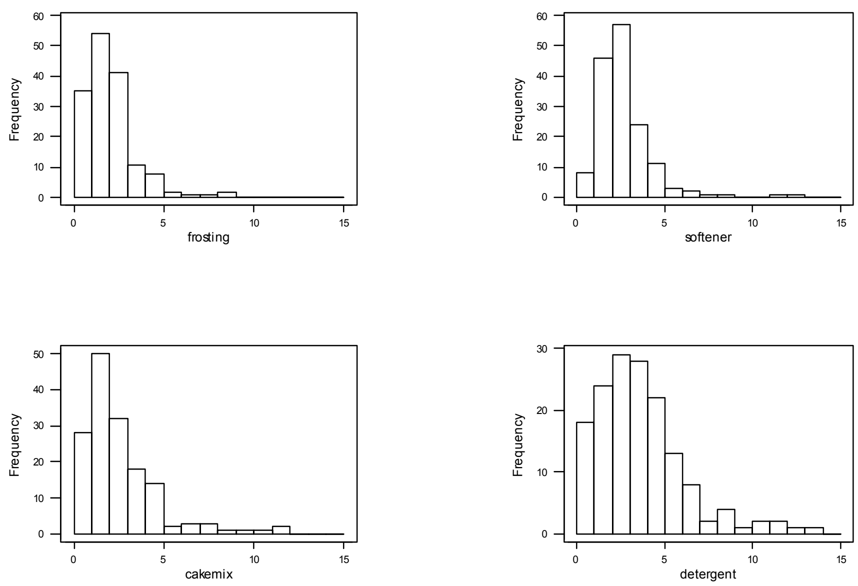
Figure 7.1: From 0/1 data to aggregated purchases per household

Our interest lies within the prediction of purchase rates per household and therefore the 0/1 purchases per household were aggregated into a vector of purchase rates for each of the 4 product categories, as shown in the example above (figure 7.1).

7.5.2 Exploratory Data Analysis

Figure 7.2 below shows the histogram of purchase rates for each variable together with the basic statistics, including mean and variance per variable. In fact, several interesting elements can be concluded from figure 7.2.

First of all, it can be seen from the histograms that the data are discrete integer values (i.e. count data) that can be modelled with a Poisson distribution. It is generally agreed upon in the literature that the Poisson distribution is well suited to model data of this kind (see section 7.6.2).



	Mean	Variance	Variance/mean
Cakemix	2.07742	4.46150	2.14762
Frosting	1.54839	2.18433	1.41071
Detergent	3.15484	6.52132	2.06709
Softener	2.20000	2.86234	1.30106

Figure 7.2: Distribution of purchase rates per variable

However, the basic statistics also show that the data is clearly overdispersed, i.e. the variance is clearly bigger than the mean and this is a problem when modelling the data with the Poisson distribution. Indeed, since the Poisson distribution is characterized only by a single parameter λ, the mean of the Poisson distribution is equal to its variance, which is not really accurate for these data. The solution to the problem of overdispersion is therefore to assume that the data are generated by a finite mixture of Poisson distributions, i.e. an unknown number of segments of households of unknown size with different unknown mean purchase rates (for more details, see section 7.6.2.2).

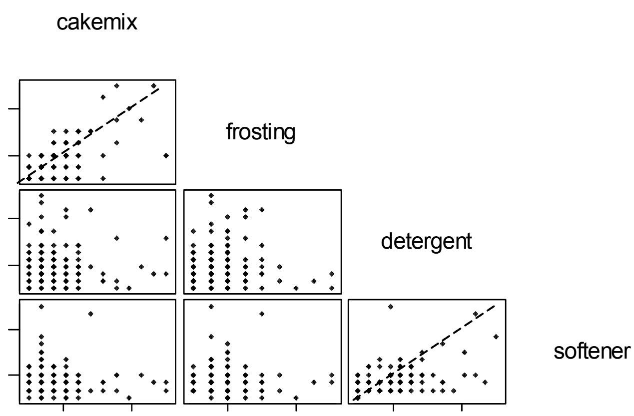


Figure 7.3: Bivariate scatter plots

Furthermore, when drawing bivariate scatter plots for all the variable combinations (see figure 7.3), it becomes visually clear that some interactions seem to exist between the purchase rates for different variables. In fact, a simple bivariate correlation analysis revealed two statistically significant interactions between cake mix and cake frosting (r=0.66) and between fabric detergent and fabric softener (r=0.48). However, since interactions may be even more complex (i.e. multivariate), loglinear analysis was carried out on these data to analyse the existence of potentially higher-order interactions between the given variables.

The contingency table below shows the frequency of occurrence of all possible purchase combinations of cake mix, cake frosting, fabric detergent and fabric softener for the total amount of 17389 purchase transactions.

C	F	D	S	Count
0	0	0	0	13975
1	0	0	0	415
0	1	0	0	219
0	0	1	0	1201
0	0	0	1	593
1	1	0	0	502
1	0	1	0	40
1	0	0	1	20
0	1	1	0	24
0	0	1	1	274
0	1	0	1	8
1	1	1	0	60
1	1	0	1	30
1	0	1	1	7
0	1	1	1	3
1	1	1	1	18

Table 7.1: Contingency table for four product categories

Table 7.1 shows that, out of the 17389 transactions, 13975 transactions contain none of the 4 product categories, whereas 274 transactions contain both fabric detergent and fabric softener but no cake mix nor cake frosting.

Performing a 2-fold (split-half[34]) loglinear analysis (see section 3.2.3.2) on these data shows that the saturated model can be significantly reduced to obtain a more parsimonious, unsaturated model containing less k-way interactions. Indeed, the likelihood ratio (LR) test shows that the most

[34] Each subset containing almost 50% of the observations drawn at random from the total dataset.

parsimonious model that fits the data well only consists of the main effects and two two-way interaction effects, i.e. between cakemix and frosting, and between detergent and softener.

Effects	**LR**	**df**	**p**
C+F+D+S+CF+CD+CS+FD+FS+DS+CFD+CFS+CDS+FDS+CFDS	0	0	.
C+F+D+S+CF+CD+CS+FD+FS+DS+CFD+CFS+CDS+FDS	1.76	1	0.19
C+F+D+S+CF+CD+CS+FD+FS+DS+CFD+CFS+CDS	1.76	2	0.41
C+F+D+S+CF+CD+CS+FD+FS+DS+CFS+CDS	1.78	3	0.62
C+F+D+S+CF+CD+CS+FD+FS+DS+CFS	2.47	4	0.65
C+F+D+S+CF+CD+CS+FD+FS+DS	3.95	5	0.56
C+F+D+S+CF+CD+CS+FS+DS	4.00	6	0.68
C+F+D+S+CF+CS+FS+DS	4.21	7	0.78
C+F+D+S+CF+CS+DS	4.63	8	0.80
C+F+D+S+CF+DS	**8.16**	**9**	**0.52**
C+F+D+S+CF	230.8	10	<0.001

Table 7.2: Loglinear analysis results for the first fold

Effects	**LR**	**df**	**p**
C+F+D+S+CF+CD+CS+FD+FS+DS+CFD+CFS+CDS+FDS+CFDS	0	0	.
C+F+D+S+CF+CD+CS+FD+FS+DS+CFD+CFS+CDS+FDS	0.01	1	0.94
C+F+D+S+CF+CD+CS+FD+FS+DS+CFD+CFS+CDS	0.01	2	0.99
C+F+D+S+CF+CD+CS+FD+FS+DS+CFD+CFS	0.02	3	1.0
C+F+D+S+CF+CD+CS+FD+FS+DS+CFS	0.05	4	1.0
C+F+D+S+CF+CD+CS+FD+FS+DS	1.01	5	0.96
C+F+D+S+CF+CD+CS+FD+DS	1.03	6	0.98
C+F+D+S+CF+CD+FD+DS	1.26	7	0.99
C+F+D+S+CF+FD+DS	1.77	8	0.99
C+F+D+S+CF+DS	**11.46**	**9**	**0.25**
C+F+D+S+CF	191.9	10	<0.001

Table 7.3: Loglinear analysis results for the second fold

This is illustrated by the two tables above (one table for each fold) that show the evolution of the LR test for the different nested models. Both tables show that the saturated model, shown in the first row, can be reduced significantly. Indeed, all 4-way, 3-way interactions and some 2-way interactions can be deleted, resulting in an unsaturated model of a much smaller size[35]. Furthermore, the loglinear analysis identifies the 2-way interactions between cakemix and frosting, and fabric detergent and softener as statistically significant. The only difference between the results of both folds is the order in which the terms are deleted from the loglinear model, dependent on the significance of their contribution.

The existence of these interactions between purchase rates for cakemix and frosting and for fabric detergent and softener motivates the use of a multivariate Poisson distribution instead of modelling the purchase rates separately (see section 7.7). The models introduced in this dissertation will therefore be multivariate mixtures of Poisson distributions, for the reasons stated above.

7.6 Models for Clustering Supermarket Shoppers

In this dissertation, it is thus the idea to introduce an innovative method based on model-based clustering to cluster supermarket shoppers based on the contents of their shopping baskets. In other words, customers will be clustered into distinct groups based on their similar purchase frequencies in a number of predefined product categories. As a result, the approach can be described as behaviour-based using a product-specific observable basis for segmentation (i.e. purchase frequencies) by means of a post-hoc descriptive method for segmentation (model-based clustering).

[35] The order in which insignificant terms are deleted is based on the chi-squared value, i.e. the term with the lowest chi-squared value during each iteration is deleted.

The reason for choosing model-based clustering instead of traditional distance-based cluster analysis, in this case, is motivated by some severe violations of the basic assumptions for distance-based cluster analysis for these data. Indeed, in traditional cluster analysis, it is assumed that the segmentation variables are orthogonal to each-other (i.e. no interdependency between them). This assumption is clearly not met for these data since loglinear analysis has shown that purchase rates are correlated between product categories as a result of product purchase interdependencies. Otherwise, a factor analysis is often carried out first to reduce the number of variables to a set of orthogonal feature vectors representing the underlying dimensions of the given variables. In contrast, the use of multivariate density distributions in model-based clustering exploits this correlation of observations and explicitly accounts for it whilst fitting the cluster model.

7.6.1 Previous Work on Mixture Models in a Supermarket Context

Cadez et al. [68] developed a mixture framework to model sales transactions instead of consumers. Their approach is, however, different from ours in a number of respects. Firstly, they assume individual transactions to arise from a mixture of multinomials, whereas in our approach we model a customer's transaction history (i.e. the aggregation of the counts over the collection of transactions per individual) by means of a multivariate Poisson mixture model. The use of the Poisson distribution in this context is motivated in section 7.6.2. Secondly, the focus of their contribution is not really on finding groups of customers in the data, but rather on making predictions and profiling each individual customer's behaviour. This is in contrast with our approach where the objective is to provide a methodology to discover groups of customers in the data having similar purchase rates in a number of product categories. To summarize, whereas our approach is more descriptive and on the segment level, their approach is more focussed on prediction and on the level of the individual customer. Furthermore, because Cadez et al. use the multinomial distribution, their approach offers a specific correlation structure between the probabilities, i.e. the covariance is always $-p_i * p_j$. Our method, however, explicitly accounts for interdependence effects and hereby it allows for greater flexibility, which will also lead to additional insights into the data, as discussed in section 7.9.1.3 of this chapter.

Ordonez et al. [212] used a mixture of Normal distributions to fit a sparse data set of binary vectors corresponding to the raw market baskets in a sales transaction database. However, they do not explicitly take correlations between product purchases into account since they assume diagonal covariance matrices.

Other examples include the use of the univariate Poisson mixture model [96] to find groups of customers with similar purchase rates of a particular candy product. The model identified light users (from 0 to 1.08 packs per week), heavy users (from 3.21 to 7.53 packs per week) and extremely heavy users (from 11.1 to 13.6 packs per week) of candy. A similar univariate Poisson mixture model was proposed by Brännäs and Rosenqvist [49] to model coffee purchase incidence data. However, the purpose of their model was to present a semi-parametric estimator for Poisson regression models with unobserved random heterogeneity. In fact, the individual Poisson parameter in their model depends both on observed covariates and on an unobserved component drawn from an unspecified mixing distribution. Their contribution mainly lies within the estimation of the shape of this unspecified distribution and the regression parameters. In this dissertation, we do not really deal with covariate information, although from a marketing point of view this would be very much welcome to explain the differences in the purchase rates between the different clusters (see section 7.10.1).

Russell and Kamakura [233] proposed a Poisson mixture model where the observed purchase quantities are assumed to be conditionally independent. Consequently, in contrast to our model specification, it does not include covariance terms into the model. Furthermore, their model allows for timevarying means (non-stationarity). Empirical validation of their model on purchase quantities in four paper goods categories (toilet paper, paper towels, facial tissue and paper napkins) revealed 7 segments.

Finally, Hoogendoorn and Sikkel [138] proposed a family of bivariate continuous Poisson mixtures, i.e. Poisson Gamma and Zero-inflated Poisson Gamma models to describe purchase incidence data. They motivate the use of the bivariate Poisson to model the interdependence between purchases rates in two subsequent sales periods (quarters). Different models were tested on purchase data from 450 households on three meat products (pork, beef and horse) for two sales periods (quarters).

To summarize, our model differs from existing approaches in basically three respects: 1) the use of the Poisson distribution instead of using Normal or multinomial distribution, 2) the true multivariate character of the model whereas most models treat the joint distribution as the product of the marginal univariate distributions, or greatly simplify the covariance structure, and 3) the systematic treatment of the covariance structure based on the underlying data. The next two sections motivate the use of the Poisson distribution and the multivariate version to model correlated purchase rates.

7.6.2 Modelling Purchase Rates with Poisson

Since we do not know exactly what drives each individual's purchase behaviour, the approach is based on the idea of modelling the purchase frequency as a Poisson-distributed random variable Y. In general, the Poisson random variable $Y_i(t)$ represents the number of occurrences of a rare event in a time interval of length t and is therefore very well suited for modelling the number of purchases within a product category over a certain period of time [13, 96, 290].

This means that we are given a number of supermarket shoppers ($i = 1, \ldots, n$) on which the random variable Y_i (i.e. purchase frequency within one particular product category) is measured over a certain period of time (t), e.g.

weeks or months. We assume the discrete random variable $Y_i(t)$ to be distributed Poisson, where $y_i = 0, 1, 2, \ldots$ and the rate parameter $\lambda t > 0$, i.e.

$$Po\left(Y_i(t) = y_i | \lambda\right) = \frac{(\lambda t)^{y_i} e^{-\lambda t}}{y_i!} \tag{7.8}$$

The mean and the variance of the Poisson distribution are $E(Y(t))= \lambda t$ and $\mathrm{var}(Y)=\lambda t$, respectively. The fact that the mean and the variance of the Poisson distribution are identical is however too restrictive in many applications where the variance of the data may exceed the mean [70]. This situation is called 'overdispersion' [190] and may be due to heterogeneity in the mean event rate of the Poisson parameter λ across the sample. Solutions to the problem of overdispersion therefore involve accommodating for the heterogeneity in the model, which can be done in a number of ways. Most well known are the *continuous* mixture specification and the *finite* mixture specification.

7.6.2.1 The continuous mixture specification

In the *continuous* mixture specification and for a unit time period, the Poisson parameter λ is assumed to be distributed across the population according to some underlying density distribution, such as for instance the Gamma distribution. This means that, at the individual level, the purchase rate is assumed to be Poisson distributed with rate parameter λ, and the rate parameter λ in turn is assumed to be Gamma distributed over the population with r the shape parameter and α the scale parameter, i.e.,

$$Po\left(Y_i = y_i | \lambda\right) = \frac{(\lambda)^{y_i} e^{-\lambda}}{y_i!} \tag{7.9}$$

$$g(\lambda) = \frac{\alpha^{r}\lambda^{r-1}e^{-\lambda\alpha}}{\Gamma(r)} \tag{7.10}$$

The distribution of purchase frequencies over the population is then given by:

$$P\left(Y_i = y_i\right) = \int_0^{\infty} P(Y_i = y_i | \lambda) g(\lambda) d\lambda \tag{7.11}$$

Substituting the Poisson and Gamma distribution produces:

$$P\left(Y_i = y_i\right) = \int_0^{\infty} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \frac{\alpha^{r}\lambda^{r-1}e^{-\alpha\lambda}}{\Gamma(r)} d\lambda \tag{7.12}$$

Bringing all possible terms that do not depend on λ out of the integral produces:

$$P\left(Y_i = y_i\right) = \frac{\alpha^{r}}{y_i!\Gamma(r)} \int_0^{\infty} \lambda^{y_i+r-1} e^{-\lambda(\alpha+1)} d\lambda \tag{7.13}$$

So, the problem comes down to solving the definite integral:

$$\int_0^{\infty} \lambda^{y_i+r-1} e^{-\lambda(\alpha+1)} d\lambda \tag{7.14}$$

Looking carefully at this integral, one can see that it strongly resembles the expression for the Gamma probability density distribution with shape parameter $y+r$ and scale parameter $\alpha+1$. Given that each probability density distribution integrates to '1', we can simplify (7.14) by multiplying it with:

$$\frac{\Gamma(y_i + r)(\alpha+1)^{y_i+r}}{(\alpha+1)^{y_i+r}\Gamma(y_i + r)} \tag{7.15}$$

and as a result:

$$P\left(Y_i = y_i\right) = \frac{\alpha^r \Gamma(r+y_i)}{y_i! \Gamma(r) (\alpha+1)^{r+y_i}} \underbrace{\int_0^{\infty} \frac{(\alpha+1)^{y_i+r} \lambda^{y_i+r-1} e^{-\lambda(\alpha+1)}}{\Gamma(y_i+r)}}_{=1} \tag{7.16}$$

$$= \frac{\alpha^r \Gamma(r+y_i)}{y_i! \Gamma(r) (\alpha+1)^{r+y_i}} \tag{7.17}$$

which finally turns out to be the Negative Binomial Distribution, known as the NBD model [207]. Although the NBD model is probably the most well known continuous mixture model in consumer research [103, 142], it is merely a member of a larger family of distributions in statistics, including the Poisson-Beta distribution, the Poisson-uniform distribution, the Poisson-inverse Gamma distribution, and many others [151]. In fact, another well-known member with applications in consumer research is the Poisson-Inverse Gaussian distribution [247].

7.6.2.2 The finite mixture specification

In this dissertation however, we will adopt the *finite* mixture specification which assumes that the underlying distribution of the Poisson parameter λ over the population can be approximated by a finite number of support points [290], which in the context of this study represent different segments or mixture components in the data. These support points and their respective probability masses can be estimated by a maximum likelihood approach. For instance, in the case of a k-segment single product category model, we assume k support points. In other words, we assume there are k groups of customers with their own size p_k and latent trait parameter $\lambda_k = \theta_k$. Consequently, the k-segment single product category model can be formulated as:

$$P\left[Y_i(t) = y_i\right] = \sum_{j=1}^{k} P\left[group_j\right] \times P\left[Y_i(t) = y_i \middle| group_j\right]$$

$$= \sum_{j=1}^{k} p_j \frac{\left(t\theta_j\right)^{y_i} \exp(-t\theta_j)}{y_i!} \tag{7.18}$$

The loglikelihood function is then defined as:

$$L\left(p, \forall \theta_j | data\right) = \sum_{i=1}^{n} \ln \left(\sum_{j=1}^{k} p_j \frac{\left(t\theta_j\right)^{y_i} \exp(-t\theta_j)}{y_i!} \right) \tag{7.19}$$

In this dissertation, however, we are interested in generalizing the finite mixture model towards multiple segments and multiple product categories. In other words, we are interested in clustering supermarket shoppers based on their purchase rate in a number (q) of product categories. Consequently, the basis for segmentation is defined as the consumer's purchase rate within *multiple* product categories. Therefore, in the next sections, the multivariate case will be developed in detail, both for uncorrelated and for correlated observations.

7.6.3 The Multivariate Poisson Distribution: General Treatment

Without loss of generality, we will restrict our exposition to the case with 4 variables. Define the sets $R_1 = \{1, 2, 3, 4\}$, $R_2 = \{12, 13, 14, 23, 24, 34\}$, $R_3 = \{123, 124, 134, 234\}$, and $R_4 = \{1234\}$, and let $R = \bigcup_{i=1}^{4} R_i$.

Now consider the latent variables X_j which follow Poisson distributions with parameters θ_j (denoted as $Po\,(x_j \mid \theta_j\,)$ with $j \in R$ respectively. Furthermore, we define the observed variables of interest Y_i, with $i = 1, 2, 3, 4$ as

$$Y_i = \sum_j X_j \tag{7.20}$$

where $j \in R$ and j contains the subscript i. For example, the general, fully-saturated covariance model for the case with 4 observed variables, where $R = \bigcup_{i=1}^{4} R_i$, is written as:

$$
\begin{aligned}
Y_1 &= X_1 + X_{12} + X_{13} + X_{14} + X_{123} + X_{124} + X_{134} + X_{1234} \\
Y_2 &= X_2 + X_{12} + X_{23} + X_{24} + X_{123} + X_{124} + X_{234} + X_{1234} \\
Y_3 &= X_3 + X_{13} + X_{23} + X_{34} + X_{123} + X_{134} + X_{234} + X_{1234} \\
Y_4 &= X_4 + X_{14} + X_{24} + X_{34} + X_{124} + X_{134} + X_{234} + X_{1234}
\end{aligned}
\tag{7.21}
$$

Looking carefully, one can see that the parameters θ_j ($j \in R_m$, m=2,3,4) correspond to m-way covariance similar to the m-way interaction terms and, thus, they impose structure to the data. Moreover, this enables to construct some interesting submodels by appropriately defining the set R. Namely,

- if $R = R_1$ then the we obtain an independence model (later referred to as the local independence (LI) model, see section 7.7.3)
- if $R = R_1 \bigcup R_4$ then we obtain a model with one common covariance term (later referred to as the common covariance (CC) model, see section 7.7.2).
- if we assume that $R = R_1 \bigcup \{12, 34\}$ then we allow only for some special 2-way covariances (later referred to as the restricted covariance (RC) model, see section 7.7.4).

Note that omitting the set of parameters θ_j ($j \in R_m$), is equivalent to setting θ_j=0. Thus, one may consider any submodel by simply assuming that the corresponding θ's are 0.

Now, denote the cardinality of R as J, which in fact for our 4-variate model equals J=15. Then, using the above notation and considering the most general model with all the covariance terms (though we will argue later that it imposes unnecessarily large structure), the joint probability density of the corresponding multivariate Poisson distribution is given as

$$P(\mathbf{y} \mid \boldsymbol{\theta}) = P(y_1, y_2, y_3, y_4 \mid \theta_j, j \in R) =$$
$$= \sum \dots \sum \prod_{j \in R}^{J} Po(x_j \mid \theta_j) \tag{7.22}$$

where the summation is extended over all the combinations of x_j such that $y_i \geq \sum_k x_k$, where $k \in R$ and k contains the subscript i. It will be illustrated in section 7.7.1, for the fully-saturated covariance model, that one needs 11 sums for the 4-variate case, which obviously implies a tremendously large computational burden. This difficulty in the calculation of the probability mass function has been the major obstacle in the use of the distribution in its most general form. Kano and Kawamura [150] described recursive schemes to reduce the computational burden, but the calculation remains computationally demanding for large dimensions.

This brings out the idea to create multivariate distributions with chosen covariances, i.e. not to include all the possible covariance terms but only to select covariance terms that are useful. Indeed, using all the m-fold covariance terms imposes too much structure while complicating the whole procedure without adding any further insight into the data. For this reason, after a preliminary examination, one may identify interesting covariance terms that may be included into the model and to exclude the others. This corresponds to fixing the value of the Poisson parameter, i.e. the corresponding λ's.

Based on this general treatment of the multivariate Poisson distribution and the relationship with more parsimonious submodels, we will in the next few sections provide a detailed treatment of each model.

7.7 Multivariate Poisson Mixture Models

In the subsequent section, we will elaborate on each multivariate Poisson mixture model in detail.

7.7.1 The Fully-Saturated MV Poisson Mixture Model

Without loss of generality, the methodological development of the fully saturated MVP mixture model will be illustrated by making use of the cakemix, cake frosting, fabric detergent and fabric softener data given earlier in section 7.5.

Suppose the objective is to cluster supermarket shoppers based on their purchase rates in a set of four product categories, i.e. cake mix I, cake frosting (F), fabric detergent (D) and fabric softener (S). Following the notation of Marshall and Olkin [186], and based on the discussion in section 7.6.3, a 4-variate Poisson distribution (Y_C , Y_F , Y_D , Y_S) with parameters (λ_C , λ_F , λ_D , λ_S , λ_{CF} , λ_{CD} , λ_{CS} , λ_{FD} , λ_{FS} , λ_{DS} , λ_{CFD} , λ_{CFS} , λ_{CDS} , λ_{FDS} , λ_{CFDS}) can then be constructed from a number of independent univariate Poisson distributions as follows:

$$\begin{aligned}
Y_C &= X_C + X_{CF} + X_{CD} + X_{CS} + X_{CFD} + X_{CFS} + X_{CDS} + X_{CFDS} \\
Y_F &= X_F + X_{CF} + X_{FD} + X_{FS} + X_{CFD} + X_{CFS} + X_{FDS} + X_{CFDS} \\
Y_D &= X_D + X_{CD} + X_{FD} + X_{DS} + X_{CFD} + X_{CDS} + X_{FDS} + X_{CFDS} \\
Y_S &= X_S + X_{CS} + X_{FS} + X_{DS} + X_{CFS} + X_{CDS} + X_{FDS} + X_{CFDS}
\end{aligned} \tag{7.23}$$

with all X's independent univariate Poisson distributions with their respective means λ_C , λ_F , λ_D , λ_S , λ_{CF} , λ_{CD} , λ_{CS} , λ_{FD} , λ_{FS} , λ_{DS} , λ_{CFD} , λ_{CFS} , λ_{CDS} , λ_{FDS} , λ_{CFDS}. Now, the question is how to obtain $P(Y_C = y_C , Y_F = y_F , Y_D = y_D , Y_S = y_S)$. The solution is based on the observation that $P(Y_C = y_C , Y_F = y_F , Y_D = y_D , Y_S = y_S)$ is the marginal distribution from $P(Y_C = y_C , Y_F = y_F , Y_D = y_D , Y_S = y_S , X_{CF} = x_{CF} , X_{CD} = x_{CD}$

$, X_{CS} = x_{CS}$, $X_{FD} = x_{FD}$, $X_{FS} = x_{FS}$, $X_{DS} = x_{DS}$, $X_{CFD} = x_{CFD}$, $X_{CFS} = x_{CFS}$, $X_{CDS} = x_{CDS}$, $X_{FDS} = x_{FDS}$, $X_{CFDS} = x_{CFDS}$) and can be obtained by summing out over all x 's, i.e.,

$$P(Y_C = y_C, Y_F = y_F, Y_D = y_D, Y_S = y_S) =$$

$$\sum_{x_{CF}=0}^{L_1} \sum_{x_{CD}=0}^{L_2} \sum_{x_{CS}=0}^{L_3} \sum_{x_{FD}=0}^{L_4} \sum_{x_{FS}=0}^{L_5} \sum_{x_{DS}=0}^{L_6} \sum_{x_{CFD}=0}^{L_7} \sum_{x_{CFS}=0}^{L_8} \sum_{x_{CDS}=0}^{L_9} \sum_{x_{FDS}=0}^{L_{10}} \sum_{x_{CFDS}=0}^{L_{11}} \tag{7.24}$$

$P(Y_C = y_C, Y_F = y_F, Y_D = y_D, Y_S = y_S, X_{CF} = x_{CF}, X_{CD} = x_{CD}, X_{CS} = x_{CS}, X_{FD} = x_{FD}, X_{FS} = x_{FS}, X_{DS} = x_{DS}, X_{CFD} = x_{CFD}, X_{CFS} = x_{CFS}, X_{CDS} = x_{CDS}, X_{FDS} = x_{FDS}, X_{CFDS} = x_{CFDS})$

The above expression (7.24) shows that the x's are summed out over all possible values of the respective x's. At this point, since the X's are Poisson distributed, it is known that the x's should take on zero or positive integer values. However, the upper bounds (L) of the different x's are unknown and will depend on the values of y_C , y_F , y_D , y_S, as illustrated below (7.29). Indeed, substituting the Y's for the X's in (7.24) results in:

$$\sum_{x_{CF}=0}^{L_1} \sum_{x_{CD}=0}^{L_2} \sum_{x_{CS}=0}^{L_3} \sum_{x_{FD}=0}^{L_4} \sum_{x_{FS}=0}^{L_5} \sum_{x_{DS}=0}^{L_6} \sum_{x_{CFD}=0}^{L_7} \sum_{x_{CFS}=0}^{L_8} \sum_{x_{CDS}=0}^{L_9} \sum_{x_{FDS}=0}^{L_{10}} \sum_{x_{CFDS}=0}^{L_{11}} \tag{7.25}$$

$P(X_c = y_c - x_{CF} - x_{CD} - x_{CS} - x_{CFD} - x_{CFS} - x_{CDS} - x_{CFDS}, X_F = y_F - x_{CF} - x_{FD} - x_{FS} - x_{CFD} - x_{CFS} - x_{FDS} - x_{CFDS}, X_D = y_D - x_{CD} - x_{FD} - x_{DS} - x_{CFD} - x_{CDS} - x_{FDS} - x_{CFDS}, X_S = y_S - x_{CS} - x_{FS} - x_{DS} - x_{CFS} - x_{CDS} - x_{FDS} - x_{CFDS}, X_{CF} = x_{CF}, X_{CD} = x_{CD}, X_{CS} = x_{CS}, X_{FD} = x_{FD}, X_{FS} = x_{FS}, X_{DS} = x_{DS}, X_{CFD} = x_{CFD}, X_{CFS} = x_{CFS}, X_{CDS} = x_{CDS}, X_{FDS} = x_{FDS}, X_{CFDS} = x_{CFDS})$

and since the X's are independent univariate Poisson variables, the joint distribution reduces to the product of the univariate distributions:

$$\sum_{x_{CF}=0}^{L_1} \sum_{x_{CD}=0}^{L_2} \sum_{x_{CS}=0}^{L_3} \sum_{x_{FD}=0}^{L_4} \sum_{x_{FS}=0}^{L_5} \sum_{x_{DS}=0}^{L_6} \sum_{x_{CFD}=0}^{L_7} \sum_{x_{CFS}=0}^{L_8} \sum_{x_{CDS}=0}^{L_9} \sum_{x_{FDS}=0}^{L_{10}} \sum_{x_{CFDS}=0}^{L_{11}} \tag{7.26}$$

$P(X_C = y_C - x_{CF} - x_{CD} - x_{CS} - x_{CFD} - x_{CFS} - x_{CDS} - x_{CFDS}) * \mathrm{P}(X_F = y_F - x_{CF} - x_{FD} - x_{FS} - x_{CFD} - x_{CFS} - x_{FDS} - x_{CFDS}) * P(X_D = y_D - x_{CD} - x_{FD} - x_{DS} - x_{CFD} - x_{CDS} - x_{FDS} - x_{CFDS}) * P(X_S = y_S - x_{CS} - x_{FS} - x_{DS} - x_{CFS} - x_{CDS} - x_{FDS} - x_{CFDS}) * P(X_{CF} = x_{CF}) * P(X_{CD} = x_{CD})$

$* P(X_{CS} = x_{CS}) * P(X_{FD} = x_{FD}) * P(X_{FS} = x_{FS}) * P(X_{DS} = x_{DS}) * P(X_{CFD} = x_{CFD}) * P$ $(X_{CFS} = x_{CFS}) * P(X_{CDS} = x_{CDS}) * P(X_{FDS} = x_{FDS}) * P(X_{CFDS} = x_{CFDS})$

Now, since the X's are univariate Poisson distributions, and since the Poisson distribution is only defined for positive integer values, the following four conditions on X_C , X_F , X_D and X_S must be satisfied (as stated earlier in section 7.6.3):

$$\begin{aligned}
&y_C - x_{CF} - x_{CD} - x_{CS} - x_{CFD} - x_{CFS} - x_{CDS} - x_{CFDS} >= 0 \\
&y_F - x_{CF} - x_{FD} - x_{FS} - x_{CFD} - x_{CFS} - x_{FDS} - x_{CFDS} >= 0 \\
&y_D - x_{CD} - x_{FD} - x_{DS} - x_{CFD} - x_{CDS} - x_{FDS} - x_{CFDS} >= 0 \\
&y_S - x_{CS} - x_{FS} - x_{DS} - x_{CFS} - x_{CDS} - x_{FDS} - x_{CFDS} >= 0
\end{aligned} \tag{7.27}$$

These conditions imply that all x's can not just take on any positive integer value, but in fact depend on the values of y_C , y_F , y_D and y_S. As a result, the distribution for $P(Y_C = y_C, Y_F = y_F, Y_D = y_D, Y_S = y_S)$ by summing out all the x's is:

$$\mathrm{P}(Y_C = y_C, Y_F = y_F, Y_D = y_D, Y_S = y_S) =$$

$$\sum_{x_{CF}=0}^{L_1} \sum_{x_{CD}=0}^{L_2} \sum_{x_{CS}=0}^{L_3} \cdots \sum_{x_{CFDS}=0}^{L_{11}} e^{-\lambda} \frac{\lambda_C^{y_C - x_{CF} - x_{CD} - x_{CS} - x_{CFD} - x_{CFS} - x_{CDS} - x_{CFDS}}}{(y_C - x_{CF} - x_{CD} - x_{CS} - x_{CFD} - x_{CFS} - x_{CDS} - x_{CFDS})!} *$$

$$\frac{\lambda_F^{y_F - x_{CF} - x_{FD} - x_{FS} - x_{CFD} - x_{CFS} - x_{FDS} - x_{CFDS}}}{(y_F - x_{CF} - x_{FD} - x_{FS} - x_{CFD} - x_{CFS} - x_{FDS} - x_{CFDS})!} *$$

$$\frac{\lambda_D^{y_D - x_{CD} - x_{FD} - x_{DS} - x_{CFD} - x_{CDS} - x_{FDS} - x_{CFDS}}}{(y_D - x_{CD} - x_{FD} - x_{DS} - x_{CFD} - x_{CDS} - x_{FDS} - x_{CFDS})!} *$$

$$\frac{\lambda_S^{y_S - x_{CS} - x_{FS} - x_{DS} - x_{CFS} - x_{CDS} - x_{FDS} - x_{CFDS}}}{(y_S - x_{CS} - x_{FS} - x_{DS} - x_{CFS} - x_{CDS} - x_{FDS} - x_{CFDS})!} *$$

$$\frac{\lambda_{CF}^{x_{CF}}}{x_{CF}!} * \frac{\lambda_{CD}^{x_{CD}}}{x_{CD}!} * \frac{\lambda_{CS}^{x_{CS}}}{x_{CS}!} * \frac{\lambda_{FD}^{x_{FD}}}{x_{FD}!} * \frac{\lambda_{FS}^{x_{FS}}}{x_{FS}!} * \frac{\lambda_{DS}^{x_{DS}}}{x_{DS}!} * \frac{\lambda_{CFD}^{x_{CFD}}}{x_{CFD}!} * \frac{\lambda_{CFS}^{x_{CFS}}}{x_{CFS}!} * \frac{\lambda_{CDS}^{x_{CDS}}}{x_{CDS}!} * \frac{\lambda_{FDS}^{x_{FDS}}}{x_{FDS}!} * \frac{\lambda_{CFDS}^{x_{CFDS}}}{x_{CFDS}!}$$

(7.28)

with $L_1 = min(y_C , y_F)$ $L_2 = min(y_C - x_{CF} , y_D)$ $L_3 = min(y_C - x_{CF} - x_{CD} , y_S)$

…

$L_{11} = min(y_C\text{-}x_{CF}\text{-}x_{CD}\text{-}x_{CS}\text{-}x_{CFD}\text{-}x_{CFS}\text{-}x_{CDS},\ y_F\text{-}x_{CF}\text{-}x_{FD}\text{-}x_{FS}\text{-}x_{CFD}\text{-}x_{CFS}\text{-}x_{FDS},\ y_3\text{-}x_{CD}\text{-}x_{FD}\text{-}x_{DS}\text{-}x_{CFD}\text{-}x_{CFS}\text{-}x_{FDS},\ y_S\text{-}x_{CS}\text{-}x_{FS}\text{-}x_{DS}\text{-}x_{CDS}\text{-}x_{CFS}\text{-}x_{FDS})$ (7.29)

and

$$\lambda = \lambda_C+\lambda_F+\lambda_D+\lambda_S+\lambda_{CF}+\lambda_{CD}+\lambda_{CS}+\lambda_{FD}+\lambda_{FS}+\lambda_{DS}+ \lambda_{CFD}+\lambda_{CFS}+\lambda_{CDS}+\lambda_{FDS}+\lambda_{CFDS} \quad (7.30)$$

To see why L_1 must equal $min(y_C, y_F)$, one must look at the first two of the four conditions on X_C , X_F , X_D and X_S specified in (7.27). Indeed, we know that all x's should be zero or a positive integer. Therefore, if all x's except from x_{CF} would be zero, then the maximum allowable value for x_{CF} can be $min(y_C , y_F)$ in order to satisfy the first two conditions. Analogously, the values for the other x's can be computed based on the preceding values[36], resulting in the admissible ranges for all L's.

The above formulated 4-variate Poisson model incorporates all possible interactions (i.e. 2-way, 3-way and 4-way) that can exist between the purchases of the 4 product categories considered. In other words, it accounts for all possible covariance between the purchase frequencies.

The mixture variant of the multivariate Poisson model simply extends the multivariate Poisson model by assuming k groups of customers having different parameter values for the λ's. Obviously, the specification of different groups in the data quickly increases the number of parameters to be optimized. In general, for a q-variate k-segments latent class model, the number of parameters to be optimized equals $(k\text{-}1)+k^*(2^q - 1)$, which increases linearly in the number of segments (k) and exponentially in the number of variables (q) considered.

---

[36] The interested reader is referred to Mahamunulu [181] for additional mathematical treatment on how those limits can be defined more rigorously. However, since this is not within the scope of our text, we will not further report on this.

7.7.2 The MV Poisson Mixture With Common Covariance Structure

As a result of the large number of parameters needed to estimate the fully-saturated multivariate Poisson (mixture) model (see section 7.7.1), an alternative approach has been proposed in the literature to simplify the model by representing variance/covariance by means of one common term [144, 173].

In this approach, following the exposition in section 7.6.3, the 4-variate Poisson distribution (Y_C , Y_F , Y_D , Y_S) with one common covariance term, later also referred to as the 'CC'-model, is defined as:

$$\begin{aligned} Y_C &= X_C + X_{CFDS} \\ Y_F &= X_F + X_{CFDS} \\ Y_D &= X_D + X_{CFDS} \\ Y_S &= X_S + X_{CFDS} \end{aligned} \tag{7.31}$$

with all X's independent univariate Poisson distributions with respective parameters $(\lambda_C , \lambda_F , \lambda_D , \lambda_S , \lambda_{CFDS})$. Though the covariance structure of this construction is limited compared to the more general definition of the multivariate Poisson, there are a manageable number of parameters to handle. In fact, a 4-variate Poisson distribution $P(Y_C = y_C ,Y_F = y_F ,Y_D = y_D ,Y_S = y_S)$ can now be obtained as the marginal distribution from $P(Y_C = y_C ,Y_F = y_F , Y_D = y_D , Y_S = y_S , X_{CFDS} = x_{CFDS})$ as follows :

$$P(Y_C = y_C ,Y_F = y_F ,Y_D = y_D ,Y_S = y_S) =$$

$$\sum_{x_{CFDS}=0}^{\min(y_C, y_F, y_D, y_S)} P(Y_C = y_C, Y_F = y_F, Y_D = y_D, Y_S = y_S, X_{CFDS} = x_{CFDS}) \tag{7.32}$$

Substituting the Y's for the X's in (7.32) results in:

$$P\,(Y_C = y_C\,, Y_F = y_F\,, Y_D = y_D\,, Y_S = y_S) =$$

$$\sum_{x_{CFDS}=0}^{\min(y_C, y_F, y_D, y_S)} P(X_C = y_C - x_{CFDS}, X_F = y_F - x_{CFDS}, X_D = y_D - x_{CFDS},$$

$$X_S = y_S - x_{CFDS}, X_{CFDS} = x_{CFDS}) \tag{7.33}$$

with all X's independent univariate Poisson distributions, thus:

$$P\,(Y_C = y_C\,, Y_F = y_F\,, Y_D = y_D\,, Y_S = y_S) =$$

$$\sum_{x_{CFDS}=0}^{\min(y_C, y_F, y_D, y_S)} P(X_C = y_C - x_{CFDS}) * P(X_F = y_F - x_{CFDS}) * P(X_D = y_D - x_{CFDS})$$

$$* P(X_S = y_S - x_{CFDS}) * P(X_{CFDS} = x_{CFDS})$$

$$=$$

$$\sum_{x_{CFDS}=0}^{\min(y_C, y_F, y_D, y_S)} e^{-\lambda} * \frac{\lambda_C^{y_C - x_{CFDS}}}{(y_C - x_{CFDS})!} * \frac{\lambda_F^{y_F - x_{CFDS}}}{(y_F - x_{CFDS})!} * \frac{\lambda_D^{y_D - x_{CFDS}}}{(y_D - x_{CFDS})!} * \frac{\lambda_S^{y_S - x_{CFDS}}}{(y_S - x_{CFDS})!} * \frac{\lambda_{CFDS}^{x_{CFDS}}}{x_{CFDS}!}$$

(7.34)

Analogous to the fully-saturated model presented in the previous section, the general k-segment q-variate Poisson mixture model requires the estimation of $(k\text{-}1)+k*(q+1)$ parameters, which increases linearly both with the number of variables and segments considered.

7.7.3 The MV Poisson Mixture With Local Independence

In the previous sections, we argued that in the case of 4 product categories, the joint probability of observing multiple outcomes $P(Y_{i1} = y_{i1}\,, Y_{i2} = y_{i2}\,, Y_{i3} = y_{i3}\,, Y_{i4} = y_{i4})$ for a customer 'i' is distributed according to the multivariate (MV) Poisson distribution. However, under the assumption of local independence of the purchase rates within each mixture component, this joint probability reduces to the product of the product category specific densities, i.e.

$$P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, Y_{i3} = y_{i3}, Y_{i4} = y_{i4}) =$$
$$P(Y_{i1} = y_{i1}) \,.\, P(Y_{i2} = y_{i2}) \,.\, P(Y_{i3} = y_{i3}) \,.\, P(Y_{i4} = y_{i4}) \qquad (7.35)$$

This means that the following representation is obtained for the Y_i 's:

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_2 \\ Y_3 &= X_3 \\ Y_4 &= X_4 \end{aligned} \qquad (7.36)$$

In this case, the likelihood function for the general k-segment mixture model for q product categories takes a very simple form:

$$L\left(\Theta; y_{il}\right) = \prod_{i=1}^{n} f\left(y_{il}|\Theta\right) = \prod_{i=1}^{n} \sum_{j=1}^{k} p_j \prod_{l=1}^{q} \frac{\left(\theta_{lj}\right)^{y_{il}} \exp(-\theta_{lj})}{y_{il}!} \qquad (7.37)$$

In fact, for the general k-segment mixture model for q product categories, we have k-1 different p's, and k different λ's per product category. For reasons of estimation, the log-likelihood is then expressed as:

$$LL\left(\forall p_s, \theta_{ij}|data\right) = \sum_{i=1}^{n} \ln\left(\sum_{j=1}^{k} p_j \prod_{l=1}^{q} \frac{\left(\theta_{lj}\right)^{y_{il}} \exp(-\theta_{lj})}{y_{il}!} \right) \qquad (7.38)$$

7.7.4 The MV Poisson Mixture with Restricted Covariance

The two multivariate Poisson mixture models presented in sections 7.7.1 and 7.7.2 represent two extreme approaches to modelling interdependent purchase rates. Indeed, the fully-saturated model including all possible purchase interdependencies on the one hand, and the model with common covariance structure with the most strongly limited covariance structure on the other hand. From a theoretical point of view, the fully-saturated model is preferable over the model with common covariance structure because the former captures more of the existing variance in the data. However, from the practical point of view, the model with common covariance structure is preferable over the fully-saturated model because it requires less parameters to be estimated.

The principle question is therefore whether a model somewhere in between the two presented extremes can be found that is both a) theoretically sound enough to capture most of the existing covariances, and b) is practically acceptable (feasible) in terms of the number of parameters to be estimated.

This section makes an attempt at introducing such a model. The underlying idea is to simplify the variance/covariance structure as much as possible by including only statistically significant k-fold interactions. For this 4-variate model, the statistical significance of the purchase interactions between cakemix (C), frosting (F), detergent (D) and softener (S) was already studied by means of loglinear analysis in section 7.5.2 and demonstrated that there are only two significant 2-fold interactions, i.e. between cakemix and frosting, and between fabric detergent and softener. The loglinear analysis is particularly relevant for the development of a simpler multivariate Poisson mixture model for clustering since it enables to discover which interaction terms in the variance/covariance matrix can be set equal to zero. Indeed, we make use of the latent variables $X = (X_C, X_F, X_D, X_S, X_{CF}, X_{DS})$, i.e. we use only two covariance terms where, for example, the term X_{DS} is the covariance term between detergent and softener. The vector of parameters is now $\boldsymbol{\theta} = (\lambda_C, \lambda_F, \lambda_D, \lambda_S, \lambda_{CF}, \lambda_{DS})$ and thus we have

$$\begin{aligned} Y_C &= X_C + X_{CF} \\ Y_F &= X_F + X_{CF} \\ Y_D &= X_D + X_{DS} \\ Y_S &= X_S + X_{DS} \end{aligned} \qquad (7.39)$$

Our definition of the model, in fact, assumes that the conditional probability function is the product of two independent bivariate Poisson distributions [161], one bivariate Poisson for cakemix and frosting, and another bivariate Poisson for fabric detergent and softener. In general, we denote the probability mass function of the bivariate Poisson (BP) distribution for cakemix and frosting as $BP(y_C, y_F; \lambda_C, \lambda_F, \lambda_{CF})$, where $\lambda_C, \lambda_F, \lambda_{CF}$ are the parameters and the probability mass function is given as

$$BP(y_C, y_F; \lambda_C, \lambda_F, \lambda_{CF}) = \frac{e^{-\lambda_C}\lambda_C^{y_C}}{y_C!}\frac{e^{-\lambda_F}\lambda_F^{y_F}}{y_F!}\sum_{i=0}^{\min(y_C, y_F)}\binom{y_C}{i}\binom{y_F}{i} i! \left(\frac{\lambda_{CF}}{\lambda_C\lambda_F}\right)^i \quad (7.40)$$

with $y_C, y_F = 0, 1, \ldots$

Thus the conditional probability function of an observation $Y = (Y_C, Y_F, Y_D, Y_S)$ is given as

$$\begin{aligned} P(y \mid \theta) &= P(y_C, y_F, y_D, y_S \mid \theta) \\ &= BP(y_C, y_F; \lambda_C, \lambda_F, \lambda_{CF}) BP(y_D, y_S; \lambda_D, \lambda_S, \lambda_{DS}) \end{aligned} \quad (7.41)$$

Note, however, that unconditionally the model is not two bivariate Poissons! Conditionally, the model is such, but unconditionally (and the problem is the unconditional problem since we do not know the cluster to which each observation belongs), it is not. The unconditional probability mass function is given under a mixture with k-components model by

$$P(y) = \sum_{j=1}^{k} p_j P(y_C, y_F, y_D, y_S \mid \theta_j) \quad (7.42)$$

Consequently, implicitly, the model assumes covariance between all the variables since it is imposed by the mixing distribution. In addition, variables Y_C and Y_F and Y_D and Y_S have increased covariance due to their intrinsic covariance induced by the model.

For a model with k components the number of parameters equals $7k$-1 (i.e. 6 lambda's per component plus the segment size) which, compared with the fully-saturated model that contains $(k-1)+k*(2^q-1)$ parameters, increases linearly instead of exponentially with the number of segments considered.

7.7.5 MV Poisson Mixture Estimation via EM

As already discussed in section 7.3, expectation-maximization (EM) algorithms are particularly useful to estimate the optimal values of the mixture model parameters. The EM algorithm is based on the missing data representation of the problem. Recall that, beyond the observed vector $\boldsymbol{Y_i}$ for the i-th observation, there are also the unobserved data (latent variables) represented by the vector $\boldsymbol{X_i}$. The EM algorithm proceeds by estimating the unobserved (missing) data by their conditional expectation at the E-step, and then at the M-step by maximizing the likelihood of the complete data. More specifically, for our multivariate Poisson models, at the E-step we need to obtain the quantities

$$s_{ijm} = E(y_{ijm} \mid \text{data}, \Theta) \ , \ i = 1,...,n \ , \ j = 1,...,J \ , m = 1,...,k \qquad (7.43)$$

where where i denotes the observation, j denotes the latent variable and m denotes the component number. Note that Θ denotes the total set of the parameters and that according to the complexity of the variance-covariance structure of the multivariate Poisson model, the number of latent variables *'J'* is different for each cluster model. Indeed, remember that *'J'* denotes the cardinality of the set *'R'* (see section 7.6.3), and this cardinality, which is determined by the number of latent variables *'X'* is different for each model. Indeed, as can be seen from (7.23), the fully-saturated Poisson model contains 15 latent variables, whereas the restricted covariance Poisson model (7.39) only has 6 latent variables. Therefore, the general form to calculate S_{ijm} equals

$$s_{ijm} = E(y_{ijm} \mid \text{data}, \Theta) = \frac{\sum_{x_j=0}^{T} x_j \left[\sum \dots \sum Po(x_j \mid \theta_j) \prod_{g \in R-\{j\}}^{J} Po(x_g \mid \theta_g) \right]}{f_m(y_i \mid \theta_m)} \quad (7.44)$$

where $T = min(y_{il})$ with $l \in R$ and l contains the subscript j. Furthermore, the number of summations in the nominator are determined by the number of latent variables in the cluster model. Similarly at the E-step we obtain the quantities

$$w_{ij} = \frac{p_j f_j(y_i \mid \theta_j)}{f(y_i)}, \quad i=1,..,n, \ j=1,...,k \quad (7.45)$$

which are simply the posterior probabilities that the i-th observation belongs to the j-th component. The M-step is relatively easier since it updates the parameters by simply calculating

$$p_j^{new} = \frac{\sum_{i=1}^{n} w_{ij}}{n}, j=1,...k \quad (7.46)$$

and

$$\theta_{jm}^{new} = \frac{\sum_{i=1}^{n} w_{ij} s_{ijm}}{\sum_{i=1}^{n} w_{ij}} \quad (7.47)$$

where θ_{jm}^{new} is the j-th parameter, $j=1,...J$ of the m-th component density, $m=1,\dots, k$. If some convergence criterion with respect to the evolution of the loglikelihood is satisfied, stop iterating, otherwise go back to the *E-step.*

The similarities with the standard EM algorithm for finite mixtures are obvious. The quantities w_{ij} at the termination of the algorithm are the posterior probabilities that the i-th observation belongs to the j-th cluster and thus they can be used to assign observations to the cluster with higher posterior probability.

With respect to computational issues, two issues need to be addressed. Firstly, the main complication of this EM algorithm is the fact that the expectations S_{ijm} (7.44) are very cumbersome to calculate due to the multitude of summations in the nominator. In fact, the number of summations needed is equal to those needed for the joint probability function. For instance, we need 11 summations (see probability function 7.24) for the fully-saturated model! In contrast, for the common covariance model, one only needs 1 sum (see probability function 7.32). The restricted covariance model that we propose in this dissertation needs 2 sums. Thus, the EM-algorithm presented above can be used for any of the presented models. Apart from the complexity of dealing with multiple sums, the only difference lies in not updating some of the parameters in the M-step. For instance, remember that for non-saturated models, like the common covariance model and the restricted covariance model, some of the θ_{jm} were set equal to 0 (see section 7.6.3). In that case, the EM-algorithm works by not updating these parameters at the M-step. In addition the form of the component densities $f_j(x|\theta_j)$ are simpler allowing for easier algebraic manipulations. Secondly, an important feature of our model with regard to scalability is that we may use frequency tables to simplify the calculations. The description of the EM algorithm above is given without using frequencies. However, the discrete nature of the data allows for using frequency tables instead of raw data. Thus, the sample size is not at all important for the computing time, since the original data can be collapsed into frequency tables. As a result, our model is scalable to very large databases. In fact, even with a very large database, the clustering is done without any additional effort. This is of particular interest in real applications, where usually the amount of observations is large.

To summarize, the scalability of the presented mixture model depends on basically two factors, i.e. the dimensions of the problem (which determines the number of observed variables and indirectly also the number of latent variables) and the variance-covariance structure considered (which determines the number of latent variables). It is well known that the speed of the EM algorithm depends on the 'missing' information. One could measure the missing information as the ratio of the observed information to the missing information. Thus, the more latent variables, the longer the computing time. The same is true as far as the number of dimensions is concerned. More dimensions lead to more latent variables. If the structure is not more complicated, the algorithm will perform relatively the same, but if the structure is more complicated, then we expect more effort. This is a strong indication that the structure imposed before the fit of the model must remain in moderate levels. In other words, if the number of dimensions is increased, but a simple covariance structure can be maintained, then the computational effort will remain relatively the same. However, if the number of dimensions significantly increases the number of free variance-covariance parameters in the covariance structure, then the computational effort will increase dramatically due to the above mentioned reasons.

7.8 Relevance for the Retailer

The practical relevance of the multivariate Poisson mixture model with limited variance/covariance structure for segmenting supermarket customers, or retail customers in general, obviously depends on its applicability on real transactional data. However, these data are typically huge, both in terms of the number of customers, as in terms of the number of product categories, or SKU's, being considered. Especially, the number of observed variables (e.g. product categories) represents an important challenge with respect to the current application of mixture models for segmentation, mainly because of the rapidly growing size of the variance/covariance structure.

Indeed, although our formulation of the mixture model with limited variance/covariance structure significantly reduces the number of free parameters to be estimated compared to the more general fully-saturated specification, the number of parameters increases quickly with the number of variables being considered for clustering. Furthermore, since the optimization function is highly nonlinear, finding the optimal values for the parameters is tricky. As a result, current implementations of mixture models for clustering are mostly univariate or bivariate and involve relatively simple applications. However, as computer resources and optimization algorithms become more powerful, one can expect more complex specifications and applications of mixture models to arise. Especially, recent developments in Bayesian estimation algorithms [270] and EM algorithms for large data sets [48, 197] provide interesting opportunities for estimating larger mixture models.

In the meantime, the application of multivariate Poisson mixture modelling for clustering customers in retailing is probably particularly relevant for category managers who manage a limited number of product categories and who want to discover groups of customers showing a different behaviour towards purchasing goods in those categories. For instance, one might be interested in finding out how purchases of baby food, baby care and baby toys are interrelated, or whether there exists a relationship between the purchase rates of fresh bread, sandwich filling and fresh cheese.

7.9 Empirical Analysis

In this section, the empirical results of the multivariate Poisson mixture model with restricted covariance structure (section 7.7.4) will be discussed and compared with the results for the local independence model (7.7.3) and the MVP mixture model with common covariance structure (7.7.2). Empirical results of the fully saturated multivariate Poisson mixture model will not be discussed since to our current knowledge, there does not exist any method to estimate the parameters of the fully-saturated multivariate Poisson mixture model in a reliable way. As indicated in section 7.7.5, the computation of the fully-saturated model involves a great number of summations and parameters to be estimated and this remains a difficulty. As a result, a comparison with the fully-saturated covariance model can not be made.

7.9.1 Results for the Different MVP Models

All three models, i.e. the local independence model (section 7.7.3), the common covariance model (7.7.2), and the model with restricted covariance structure (7.7.4) were at least implemented sequentially for 1 to 10 components ($k = 1,...,10$). For the model with restricted covariance structure, it was even implemented up till 16 components because the loglikelihood stopped increasing after $k>16$. Furthermore, in order to overcome the well-known drawback of the EM algorithm, i.e. the dependence on the initial starting values for the model parameters, 10 different sets of starting values were chosen at random. In fact, the mixing proportions (p) were uniform random numbers and rescaled so as to sum at 1, while the λ's were generated from a uniform distribution over the range of the data points. For each set of starting values, the algorithm was run for 100 iterations without caring about any convergence criterion. Then, from the solution with the largest likelihood, EM iterations were continued until a rather strict convergence criterion was satisfied, i.e. until the relative change of the loglikelihood between two successive iterations was smaller than 10^{-12}. This procedure was repeated 10 times for each value of k. As expected, problems with multiple maxima occurred for large values of k, while for smaller values of k the algorithm usually terminated at the same solution with small perturbations.

In terms of the number of clusters, we based our selection on the most well-known information criteria, i.e. Akaike Information Criterion (AIC), Consistent Akaike Information Criterion ($CAIC$) and the Bayesian Information Criterion (BIC) [197] (see formula 7.4 – 7.6). Note that in the literature, there are several equivalent variants of these criteria. Therefore, they only serve as a guide for the researcher to select the optimal number of components in the data.

Furthermore, we evaluate the quality of the clustering solution by means of the entropy statistic [197] (formula 7.7). The rule for allocating observations to clusters is the higher posterior probability rule. This means that an observation is allocated to the cluster for which the posterior probability, as measured by w_{ij}, is larger. Note that w_{ij} are readily available after the termination of the EM algorithm, as they constitute the E-step of the algorithm.

7.9.1.1 Results for the local independence model

Figure 7.4 shows the evolution of the loglikelihood for different components ($k = 1, \ldots, 10$) of the local independence MVP mixture model (see section 7.7.3). The figure shows that AIC selects 5 components whereas the $CAIC$ and the BIC select only 3 components. Note that the values for the AIC, $CAIC$ and BIC are rescaled in order to be comparable to the loglikelihood.

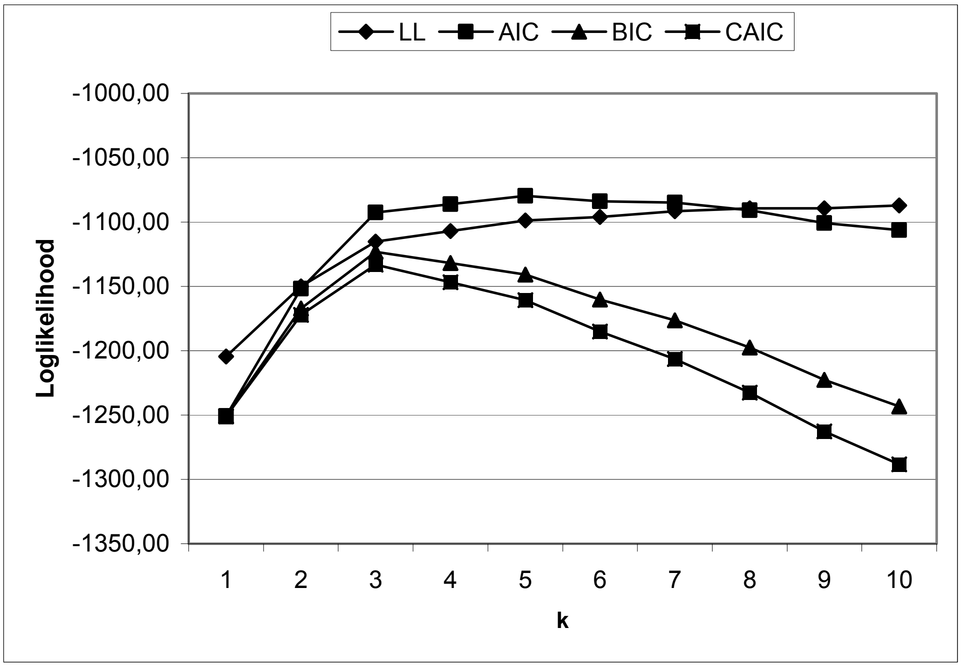


Figure 7.4: Loglikelihood, AIC, CAIC and BIC against the number of components for the local independence MVP mixture model

Additionally, the entropy statistic shows a good separation between the components for the 3 cluster solution $I(3)=0.83$ but a rather weak separation for the 5 cluster solution $I(5)=0.66$.

Figure 7.5 shows the optimal value of the mixing proportions for the range of models used (values of k from 2 to 10). It is worth to note that there does not seem to be a stable configuration, i.e. we can not observe a set of clusters that remains relatively stable over the different values of k. In fact, the mixing proportions tend to fluctuate over the different component solutions and additional analysis showed that the mixing proportions average out for larger component solutions ($k>10$).

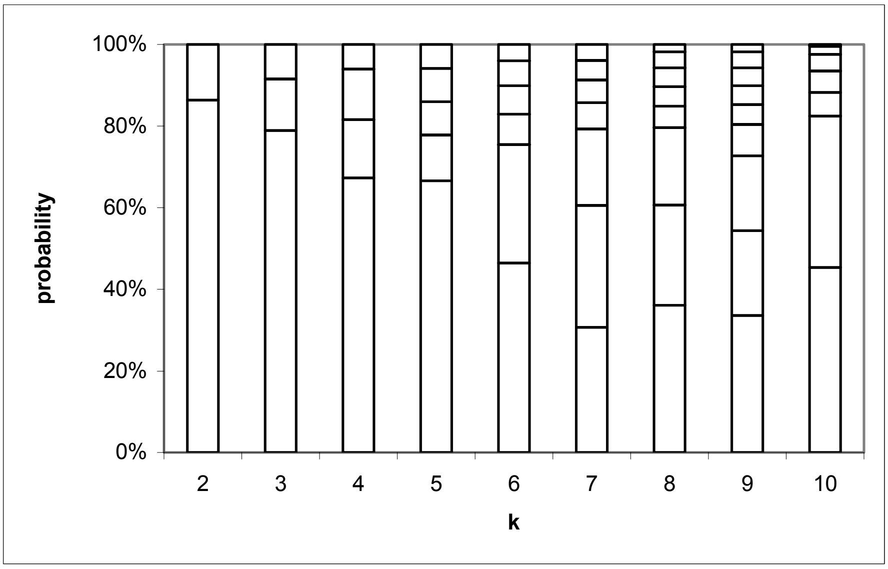


Figure 7.5: The mixing proportions for model solutions with k=2 to 10 components

In figures 7.6 and 7.7, we have plotted bubble-plots for pairs of the parameters. In fact, each graph depicts the joint mixing distribution for the selected pair. The plots depict both the 3 and the 5 components solution, with the 5 components solution represented as a bold circle. Furthermore, the size of the circle reflects the mixing proportion, the larger the size, the larger the mixing proportion.

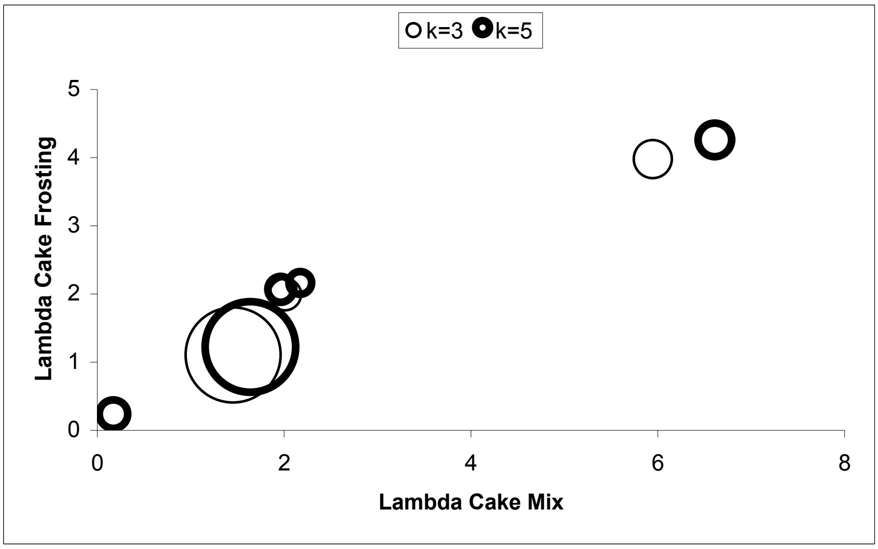


Figure 7.6: Bubble plot for λ_c against λ_F for the 3 and 5 component solution

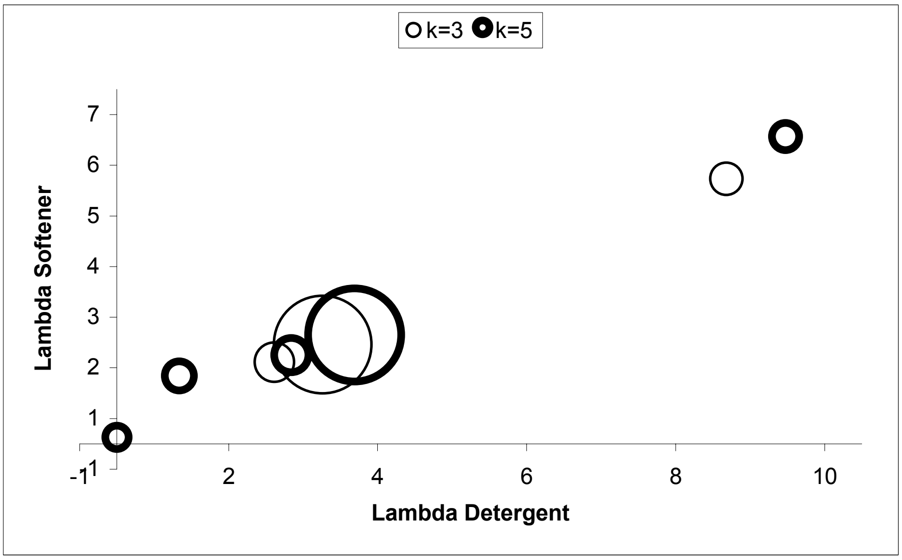


Figure 7.7: Bubble plot for λ_D against λ_S for the 3 and 5 component solution

It is clear from the graphs 7.6 and 7.7 that the two solutions differ only slightly and that indeed the 5 components solution contains two additional clusters separated from the 3 components solution, where the profile in terms of the estimated average purchase rates (λ_C, λ_F, λ_D, λ_S) of the existing 3 clusters in the 3 component solution remains almost unchanged in the 5 component solution.

Table 7.4 and table 7.5 contain the parameter estimates for the model with 3 components and 5 components respectively. One can see that all the components of the 3-cluster solution still exist in the 5-cluster solution, however with a slightly different profile in terms of their average purchase rates. Furthermore, their cluster size has decreased a little in favour of the creation of two additional clusters with rather extreme profiles, i.e. cluster 3 and cluster 5 in the 5-component solution. Nevertheless, there exist two big clusters, which together account for more than 80% of all the customers.

	Parameters				
Cluster	λ_C	λ_F	λ_D	λ_S	p
1	1.456	1.105	2.762	1.957	0.782
2	5.950	3.977	2.114	1.611	0.129
3	2.013	1.991	8.183	5.231	0.089

Table 7.4: Parameters for the 3-components local independence model

	Parameters				
Cluster	λ_C	λ_F	λ_D	λ_S	p
1	1.642	1.221	3.195	2.145	0.718
2	6.612	4.259	2.345	1.749	0.100
3	0.173	0.233	0.838	1.340	0.072
4	1.966	2.067	8.977	6.067	0.062
5	2.176	2.159	0.002	0.123	0.047

Table 7.5: Parameters for the 5-components local independence model

7.9.1.2 Results for the common covariance model

Figure 7.8 shows the evolution of the loglikelihood for different components (k = 1, …, 10) of the common covariance MVP mixture model (see section 7.7.2). Furthermore, the figure shows that *AIC* selects 5 components whereas the *CAIC* and the *BIC* select only 3 components. Note that the values for the *AIC*, *CAIC* and *BIC* are again rescaled in order to be comparable to the loglikelihood.

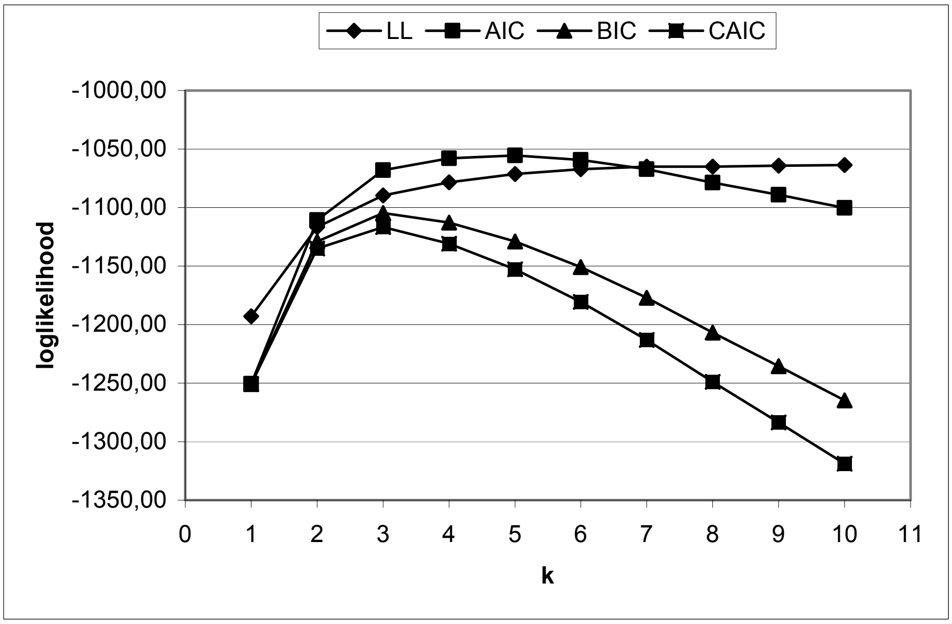


Figure 7.8: Loglikelihood, AIC, CAIC and BIC against the number of components for the common covariance MVP mixture model

Furthermore, the entropy statistics for the 3 component solution (selected by *CAIC* and *BIC*) $I(3)$=0.748 and for the 5 component solution (selected by *AIC*) $I(5)$=0.799 indicate a relatively good separation between the different components.

Figure 7.9 shows the optimal value of the mixing proportions for the entire range of models used (values of k from 2 to 10).

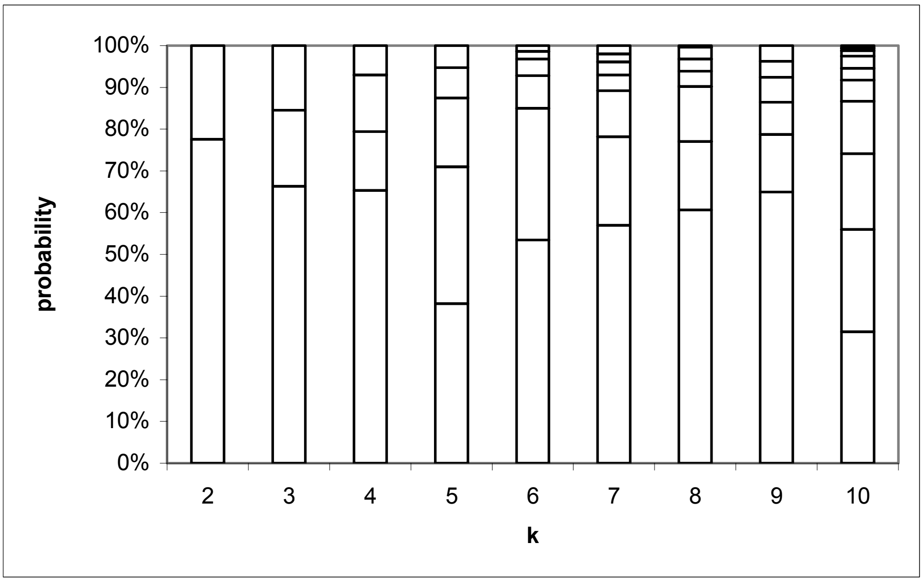


Figure 7.9: The mixing proportions for model solutions with k=2 to 10 components

Again, the graph does not show a stable cluster configuration, i.e. a clustering that remains relatively stable over the different component solutions. In other words, the cluster proportions tend to fluctuate and do not arrive at a relatively stable configuration.

In figures 7.10 and 7.11, we have plotted bubble-plots for pairs of the parameters. In fact, each graph depicts the joint mixing distribution for the selected pair. The plots depict both the 3 and the 5 components solution, with the 5 components solution represented as a bold circle. Furthermore, the size of the circle reflects the mixing proportion, the larger the size, the larger the mixing proportion. It is visually clear that the 5 component solution corresponds to a split of two clusters in the 3 component solution into two new clusters each, together with a slight shift in the position of cluster 3 in the 3 component solution compared to the 5 component solution.

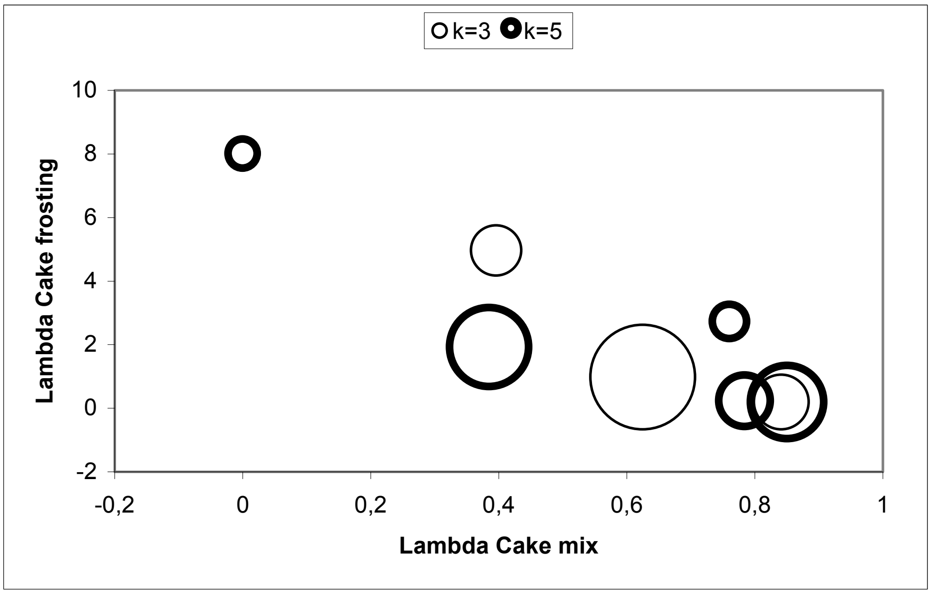


Figure 7.10: Bubble plot for λ_c against λ_F for the 3 and 5 component solution

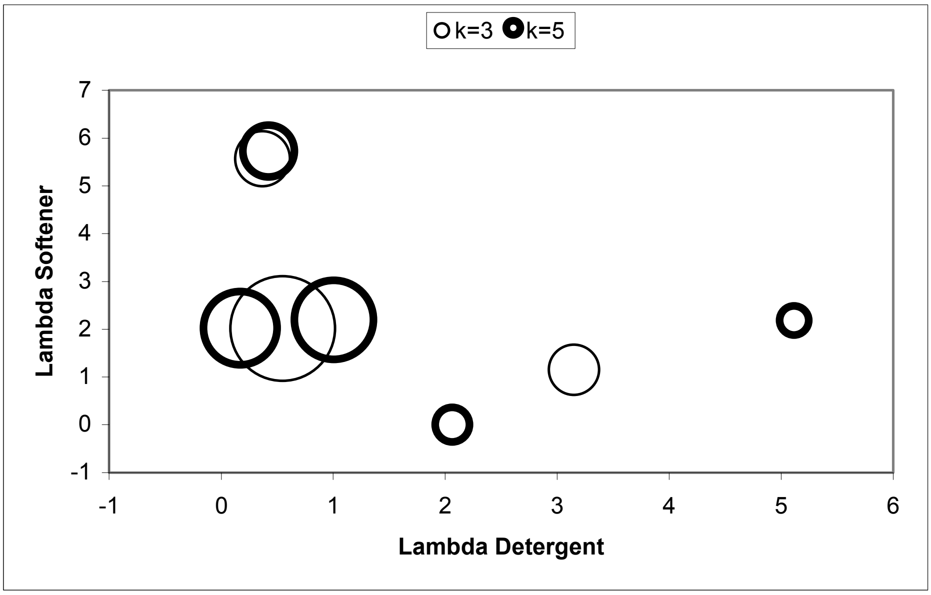


Figure 7.11: Bubble plot for λ_D against λ_S for the 3 and 5 component solution

Table 7.6 and table 7.7 contain the parameter estimates for the model with 3 components and 5 components respectively. One can see that the cluster solutions are clearly different in terms of their optimal parameter values. Furthermore, analysis of the cluster memberships for each observation showed that cluster 2 in the 3-component solution is split up in two new clusters in the 5-component solution (i.e. cluster 1 and 4). Furthermore, cluster 1 in the 3-component solution is split up in two new clusters in the 5-component solution (i.e. cluster 2 and 3). Finally, cluster 3 in the 3-component solution loses some observations and has a slightly different profile in the 5-component solution (see cluster 5 in the 5 component solution).

	Parameters					
Cluster	λ_C	λ_F	λ_D	λ_S	λ_{CFDS}	p
1	0.396	4.963	3.152	1.146	1.049	0.154
2	0.625	0.974	0.550	2.010	1.262	0.663
3	0.842	0.197	0.370	5.563	3.133	0.183

Table 7.6: Estimated parameters for the 3-components common covariance model

	Parameters					
Cluster	λ_C	λ_F	λ_D	λ_S	λ_{CFDS}	p
1	0.385	1.920	1.006	2.195	1.576	0.380
2	0.000	8.009	5.115	2.179	1.700	0.053
3	0.760	2.725	2.064	0.000	0.000	0.072
4	0.850	0.196	0.172	2.020	1.072	0.326
5	0.784	0.237	0.424	5.724	3.367	0.164

Table 7.7: Estimated parameters for the 5-components common covariance model

7.9.1.3 Results for the restricted covariance model

Figure 7.12 shows the evolution of the loglikelihood for different components (k = 1, …, 10) of the restricted covariance MVP mixture model (see section 7.7.4). Furthermore, figure 7.12 shows that the AIC criterion selects 6 components whereas the $CAIC$ and BIC criterion select 3 components. The depicted values are again rescaled so as to be comparable to the loglikelihood.

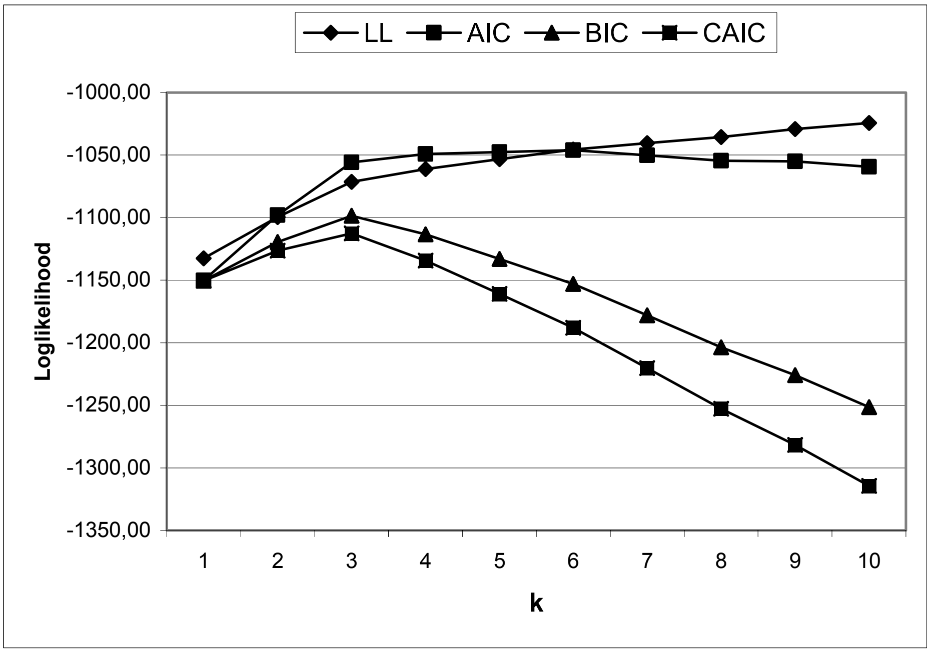


Figure 7.12: Loglikelihood, AIC, CAIC and BIC against the number of components for the restricted MVP mixture model

The entropy statistic for the 3 component solution $I(3)$=0.87 for the 6 component solution $I(6)$=0.81 indicate a very good separation between the clusters.

Figure 7.13 shows the optimal value of the mixing proportions for the entire range of models used (values of k from 2 to 10). It is apparent from figure 7.13 that usually the additional component corresponds to a split of an existing component in two parts, perhaps with some minor modification for the rest of the components, especially if they have estimates close to the component split.

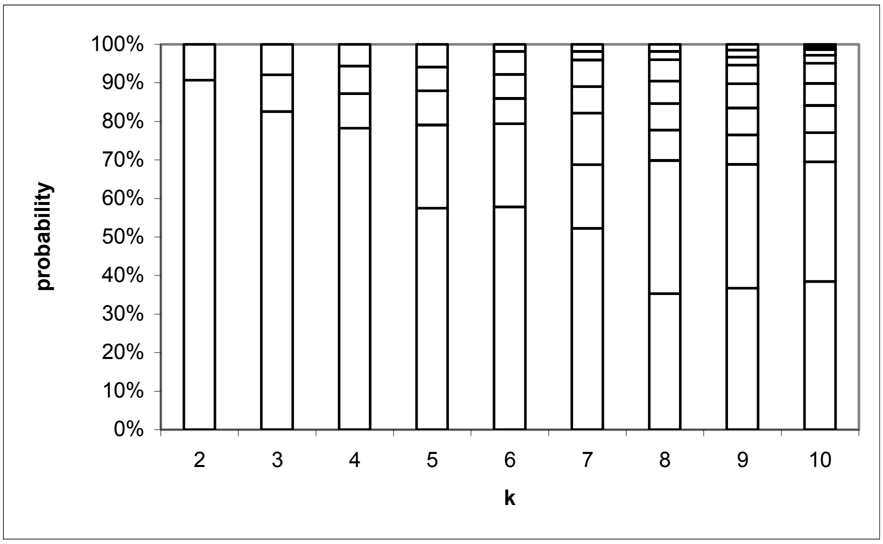


Figure 7.13: Mixing proportions for model solutions with k=2 to 10 components

This illustrates the stability of the model and the existence of two larger components, which together cover almost 80% of all observations (see also table 7.8 and 7.9). It is also quite interesting to see that the solution with 5 and 6 components differ only slightly. This is interesting from the retailer point of view for which the existence of a limited number of clusters is important. Indeed, if a large number of clusters would exist, it is impossible for the retailer to manage all segments separately, i.e. it would neither be cost-effective, nor practical to set-up different merchandising strategies for each (small) segment. Given that, in contrast to the earlier models, the loglikelihood remains to be increasing quite strongly for larger component solutions, we have opted here to select and discuss the 5 and 6 component solution, as indicated by the AIC.

In figure 7.14 and 7.15, we have plotted selected bubble-plots for pairs of the parameters. In fact, each graph depicts the joint mixing distribution for the selected pair. The plots depict both the 5 and the 6-cluster solution. The bold circles represent the 6-cluster solution and the thin circles the 5 cluster solution. The size of the circle reflects the mixing proportion, the larger the size the larger the mixing proportion. It is clear from the graph that the two solutions differ only slightly and that the 6-cluster solution just splits up one of the existing clusters as indicated by the arrows.

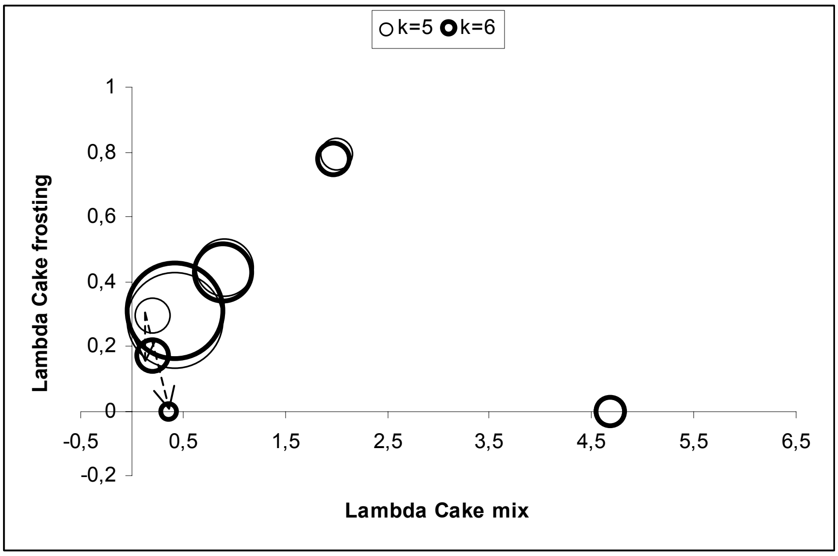


Figure 7.14: Bubble plot for λ_c against λ_F for the 5 and 6 components solution

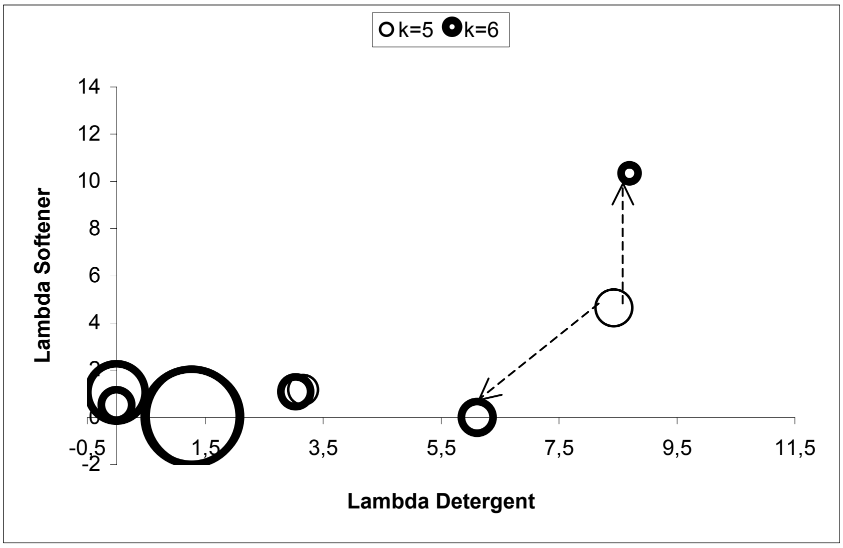


Figure 7.15: Bubble plot for λ_D against λ_S for the 5 and 6 components solution

Table 7.8 and table 7.9 contain the parameter estimates for the model with 5 components and 6 components respectively. One can see that all the components of the 5-cluster solution still exist in the 6-cluster solution, but an additional component appeared (number 2 in the 6-cluster solution) that seems to take observations from the old components 1 and 3 of the 5-cluster solution. In both solutions, there are 2 clusters of large size that are very similar, indicating the existence of two relatively stable clusters, which together account for almost 80% of all the customers.

	Parameters						
Cluster	λ_C	λ_F	λ_{CF}	λ_D	λ_S	λ_{DS}	P
1	0.207	0.295	1.507	8.431	4.639	0.000	0.088
2	0.427	0.279	1.093	1.347	0.031	1.955	0.575
3	0.908	0.441	0.555	0.000	1.030	0.977	0.216
4	2.000	0.792	4.292	0.000	0.524	1.187	0.062
5	4.668	0.000	1.223	3.166	1.161	0.702	0.059

Table 7.8: Estimated parameters for the 5-components model

	Parameters						
Cluster	λ_C	λ_F	λ_{CF}	λ_D	λ_S	λ_{DS}	P
1	0.205	0.171	1.523	6.116	0.000	2.061	0.066
2	0.356	0.000	2.063	8.698	10.33	0.000	0.019
3	0.424	0.311	1.061	1.275	0.026	2.083	0.578
4	0.897	0.425	0.587	0.000	1.047	0.972	0.215
5	1.975	0.776	4.287	0.000	0.521	1.192	0.062
6	4.684	0.000	1.219	3.040	1.085	0.782	0.059

Table 7.9: Estimated parameters for the 6-components model

Another interesting feature about the results in table 7.8 and 7.9 is the interpretation of the zero values. If the zero value corresponds to covariance parameters (i.e. λ_{CF}, λ_{DS}) then this implies that the two variables are not correlated at all for this component, i.e. the purchase rate of a product is independent from the purchase rate of the other product. The interpretation of a zero value for the other lambdas is a little more complicated. For instance, in table 7.9, take the lambda values for cakemix and cake frosting of the last component. In general, $\lambda_F = 0$ has the interpretation of the mean purchase rate of frosting after having removed the effect of cakemix. Recall that the marginal mean for frosting equals $\lambda_F + \lambda_{CF}$. Now, if λ_{CF} is large with respect to λ_F, then this implies that the purchases of product frosting are strongly correlated with cakemix.

Figure 7.16 depicts the clusters for the 5-clusters solution. The pairwise scatterplots present the 155 observations labeled with the cluster to which they belong (each cluster obtains a different label) and the values on both axes are the purchase rates for each variable. The figure shows that the clusters are quite close together with usually just one variable that differentiates the clusters from each other. Take for example cluster 2 and 5 in table 7.10: their difference is mainly due to the large difference for the variable cakemix.

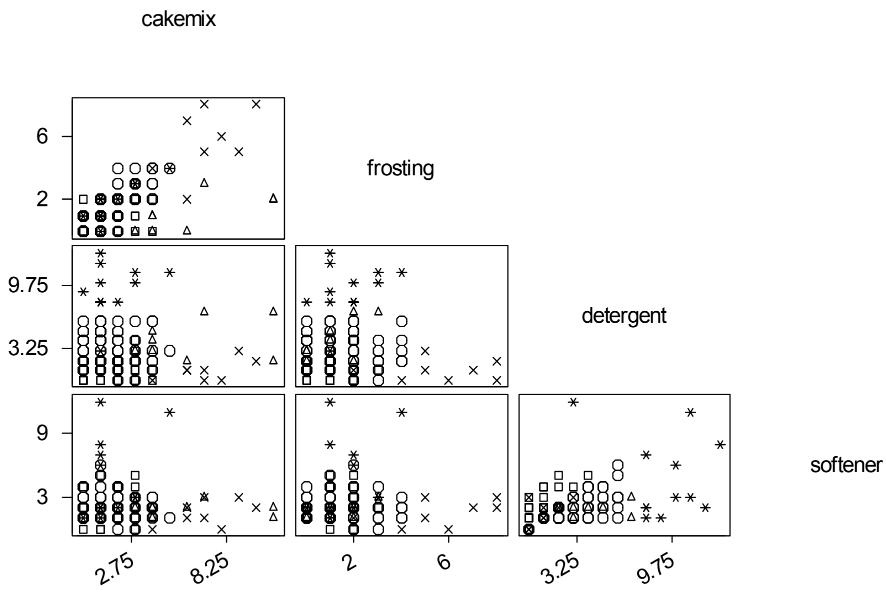


Figure 7.16: Pairwise clusters

Cluster	*Cakemix*	*Frosting*	*Detergent*	*Softener*	*Obs.*
1	1.667	1.750	9.250	4.833	12
2	1.505	1.419	3.290	2.022	93
3	1.618	0.971	0.912	2.000	34
4	7.125	5.625	1.000	1.500	8
5	6.250	1.125	4.125	1.875	8
Overall mean	2.077	1.548	3.155	2.200	155

Table 7.10: Cluster centres for the 5-component mixture model

In order to interpret the cluster differences with regard to the original data, table 7.10 contains the cluster centres for the 5-components solution. The last row contains the sample centroids, i.e. the mean purchase rate for each variable over the entire dataset and for the period of data collection.

Looking at the two major clusters (cluster 2 and 3) in table 7.8, it can be observed that they have a rather different profile. Especially with regard to fabric detergent and fabric softener, both clusters show indeed a rather different behaviour. Cluster 2 shows a very low average purchase rate of fabric softener ($\lambda_S = 0.031$) but a rather high covariance between fabric detergent and fabric softener ($\lambda_{DS} = 1.955$). This is shown in table 7.10: people in cluster 2 have an average purchase rate of fabric softener of 2.022, which is largely due to the covariance with fabric detergent ($\lambda_{DS} = 1.955$). Consequently, the own sales of fabric softener in cluster 2 are almost non-existent ($\lambda_S = 0.031$) but they occur mainly as a result of cross-selling with fabric detergent. In contrast, cluster 3 in table 7.8 shows a rather opposite profile. Cluster 3 shows no own purchases of fabric detergent at all ($\lambda_D = 0.000$), but again a relatively strong covariance with fabric softener ($\lambda_{DS} = 0.977$). This is again shown in table 7.10: people in cluster 3 have an average purchase rate of fabric detergent of 0.912, which is rather low compared to the total sample average, but this purchase rate is exclusively due to the covariance with fabric softener ($\lambda_{DS} = 0.977$).

These are important findings since they potentially have interesting implications for marketing decision making, e.g. for targeted merchandising strategies. For instance, cluster 2 in the 5-segment solution represents an important customer segment because it contains almost 60% of the observations and it is known that, in this cluster, the purchase rate of softener is rather low, compared to the sales of detergent, and that the covariance between the sales of softener and the sales of detergent is high.

Therefore, in order to take advantage of this strong interdependence effect, retail management could decide to make softener more salient, i.e. to bring softener more under the attention of customers shopping for detergent, for instance by putting detergent and softener in the same display. Moreover, placing highly interdependent products closer together in the store may also have positive affect at the shopping pleasure of time-pressured shoppers who typically do not want to waste time by looking for items in the store.

Furthermore, knowledge about correlated category usage patterns enables category managers (and manufacturers) to implement cross-category marketing strategies. For instance, Catalina Marketing [73] sells point-of-purchase electronic couponing systems that can be implemented to print coupons for a particular category, based on the purchases made in other categories[37].

Finally, cluster 1 in the 5-component solution shows another interesting, yet different profile compared to the two bigger clusters discussed before. Customers in this cluster purchase large quantities of fabric detergent ($\lambda_D = 8.431$) and fabric softener ($\lambda_S = 4.639$), however, the purchases are not correlated at all ($\lambda_{DS} = 0.000$). This means that although customers in cluster 1 purchase high amounts of detergent and softener, their purchase rates are not interdependent. Consequently, promotional campaigns (like price reduction or special display) on one of both products (say detergent) will probably not influence the sales of the other product (softener).

[37] To what extent such coupon action will benefit the retailer overall is difficult to say since other effects like cannibalization, stock-piling and competitive switching may determine the ultimate profitability of this action.

7.9.2 Comparison of Different Models

Looking at the empirical results of the different model formulations in the previous sections (7.9.1.1 to 7.9.1.3), the following conclusions can be drawn with regard to the fit and the quality of the different cluster solutions.

7.9.2.1 Fit of the different model specifications

With regard to the fit of the different models, it is clear from figure 7.17 that by enabling additional purchase correlations, the fit of the model, as indicated by the loglikelihood values, increases significantly.

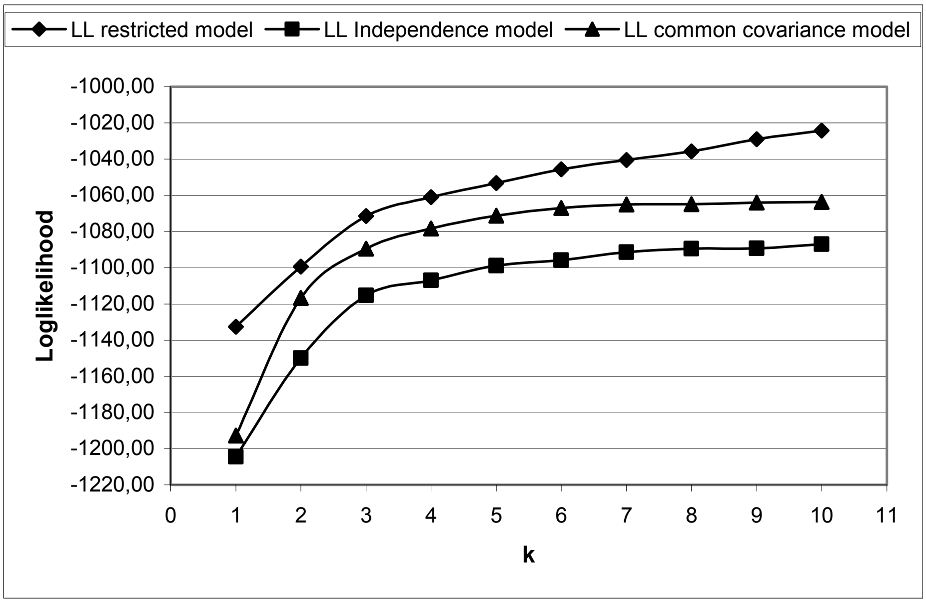


Figure 7.17: Loglikelihoods for local independence, common covariance and restricted covariance MVP mixture models for different component solutions

Figure 7.17 indeed shows that the loglikelihood of the restricted covariance model clearly dominates the loglikelihoods of the local independence and common covariance model over the entire range of component solutions ($k=1$ to 10). From the viewpoint of model fit, this partly justifies the use of the model with restricted covariance structure. The question whether the full covariance model would still produce a significantly better fit can not be answered since currently no reliable procedures exist to fit the full covariance model. Probably, a slight increase in the loglikelihood could can be expected but since the restricted covariance model contains all the significant purchase interactions, we expect this increase of fit not to be significant.

This raises the issue of testing the significance of difference in model fit between the different model specifications. At first sight, the likelihood ratio test (LRT) seems suitable to carry out this test. Under the null hypothesis (i.e. the fit of both models is equal), the LRT is asymptotically distributed as chi-square with degrees of freedom equal to the difference in the number of parameters if one model is nested in the other. Since, for instance, the local independence model is nested in the common covariance model by deleting the common interaction parameter, this therefore seems like a reasonable test. But the regularity conditions needed to use the LRT are not satisfied, because the parameters that allow to go from one model to the other take a value at the boundary of the parameter space. Recall that the parameters of any multivariate Poisson model are positive, so the value 0 is at the boundary. This makes the use of the LRT statistic impossible. The same problem arises when testing for model fit between different component solutions and is well documented in the literature [41, 197].

Another solution might be to construct some type of information criterion like AIC, BIC and CAIC to test the difference between the models. However, these information criteria compare point estimates and not the difference between entire curves so this does not seem to be applicable either. Therefore, the only way of comparing the different solutions is by visually inspecting figure 7.17 and since it shows that the restricted covariance model lies above the common covariance and local independence model along the entire range of components, it could be concluded that this model is clearly the better alternative.

7.9.2.2 Quality of clustering

Apart from the evaluation of the model fit, the multivariate Poisson mixture model with restricted covariance structure provides the most detailed information about purchase rates and their interdependence from the different empirically evaluated models. Indeed, in contrast to the local independence model, the model with restricted covariance structure provides additional insight into the purchase behaviour of individuals and uses the extra information about correlated purchases to increase the fit of the clustering solution. In the light of the typical quality criteria to evaluate cluster solutions, we will now focus on each of the quality criteria, introduced in section 6.2.3, for the restricted covariance model, although some conclusions may also apply for the local independence and common covariance model.

With regard to the *identifiability* of the multivariate Poisson mixture model with restricted covariance structure, it can be concluded from the entropy statistic (see section 7.9.1.1) that the segments are well separated. Indeed, the entropy is well above zero, and even close to one. Furthermore, the characteristics of the clusters, in terms of their optimal parameter values, are clearly different (see figure 7.14 and 7.15 and tables 7.8 and 7.9).

Furthermore, the cluster solution produces a limited set of clusters, with two of them of a *substantial* size. Indeed, in the restricted model two substantial clusters appear, containing almost 80% of all observations, which remain almost unchanged for larger component solutions. However, due to the possible effect of outliers, which leads to the creation of a number of very small clusters, they show up rather late.

The cluster solutions are *accessible* since the basis for segmentation has been chosen as the category purchase rates, which can easily be tracked and stored in a database, and which at the checkout offer opportunities to differentiate between customers of different segments. Indeed, the basis for segmentation is observable and product specific and thus provides excellent opportunities for targeted marketing campaigns, e.g. for printing customized coupons at the checkout. On the other hand, no covariate information is included into the model (e.g. socio-demographic and/or lifestyle data) such that the differences in the structure of each cluster can not be explained by means of covariate information (like in concomitant variable models). This may limit the applicability of the model in a retailing context because the model does not link the cluster solution to the loyalty card information about the customers.

The *structural stability* of the clusters is satisfactory. It is shown that usually the introduction of an additional component corresponds to a split of an existing component in two parts, perhaps with some minor modification for the rest components, especially if they have estimates close to the component split. The *temporal stability* of the clusters could not be examined since we did not possess purchase data for the same individuals at a later period in time. The existence of such data would enable to compare the cluster solution and cluster membership of the observations in order to evaluate its stability over time. Furthermore, the lambda parameters in the multivariate Poisson are specified as stationary parameters, which are assumed not to change over time.

The model does not provide insights into the *responsiveness* of the discovered clusters. For instance, cluster 1 in the 5 components solution of the restricted covariance model (see table 7.8) shows no interdependence between the purchases made in the categories fabric detergent and fabric softener. It could therefore be expected that promotions on fabric detergent would not have an effect on the purchases of fabric softener in that particular segment. However, the model does not support such interpretations. The reason is that we do not possess information about the (cross-)promotional elasticity of the products, which would be necessary to assess the responsiveness of particular promotional campaigns.

Finally, the *actionability* of the cluster solution depends on the strategic positioning of the retail firm. The extent to which the retailer wants to use the results of the model to devise customized marketing campaigns will partially depend on the information technology available at the retailer and his willingness to experiment. Currently, many (European) retailers are not very keen on customizing promotions towards their customers for the risk of wrong perception by the consumer. In fact, one supermarket retailer told us that he was afraid of not treating all consumers alike because of the risk that consumers might feel manipulated by the retailer (why does my neighbour get different promotions?) or may feel harmed in their personal sphere of life (privacy). However, these problems are common for any segmentation approach where the results are used to set up a customized communication program.

7.9.2.3 Overlap between cluster solutions

Besides comparing the fit of the different clustering solutions, we are also interested in evaluating how the additional complexity of the 4-variate Poisson mixture model with restricted covariance in comparison with the other simpler models introduced in the text leads to a different clustering solution. In other words, we want to examine how customers, who end up in the same cluster for one clustering method, say the common covariance model (CC) or the local independence model (LI), are spread over the clusters of the restricted covariance model (RC). This must enable to evaluate, for this particular dataset[38], whether the additional complexity of the restricted covariance model results in some additional insights into the data that can not be obtained by applying simpler models. More specifically, we will compare segment memberships for each observation and for each clustering method such that pairwise cross-tabulations can be made to visualize the scattering of the households over the different segments for the different clustering methods. However, since that the restricted covariance model is the main contribution in this chapter, we will compare all the other models against the restricted covariance model. In each case, we will use the number of clusters selected by the AIC criterion for comparison.

[38] Clearly, such evaluation is only valid with regard to this data set since for any other data set the interactions between the variables will be different.

Additionally, we will compare the results of the restricted covariance (RC) model with those obtained from a new model, not presented in this chapter before. In fact, the new model consists of the combination of two bivariate Poisson mixture models, i.e. one bivariate model between cakemix and frosting and one bivariate model between fabric detergent and softener. The idea is that the loglinear analysis showed an important interaction between cakemix and frosting on the one hand, and fabric detergent and softener on the other hand such that the additional complexity of the 4-variate model may not reveal any additional insights that can not be obtained by cross-tabulating the clustering results for the two bivariate models.

Local independence (LI) versus restricted covariance (RC)

The results of the local independence model, reported in section 7.9.1.1, reveal 5 clusters, whereas the results of the restricted covariance model, reported in section 7.9.1.3, report 5 (or 6) clusters. Therefore, a comparison between the clustering solutions for both models results in a cross-tabulation of size 5x5, where the clusters of the LI model are presented in the rows, and those of the RC model in the columns.

Table 7.11 shows that cluster 1 in the LI model mainly splits up into two clusters (cluster 2 and 3) in the RC model. The same is true for cluster 3 in the LI model, for which most of the observations also get split up into clusters 2 and 3 of the RC model.

When examining the parameter values for cluster 1 of the LI model (see table 7.5 in section 7.9.1.1), then it turns out that, in terms of the average purchase rates for the four products (λ_C=1.642, λ_F=1.221, λ_D=3.195, λ_S=2.145) these values correspond more or less with the average purchase rates of cluster 2 (λ_C=1.505, λ_F=1.419, λ_D=3.290, λ_S=2.022) and cluster 3 (λ_C=1.618, λ_F=0.971, λ_D=0.912, λ_S=2.000) of the RC model (see table 7.10 in section 7.9.1.3).

		RC model					
		1	2	3	4	5	Total
	1	4	81	25	0	5	115
	2	0	4	0	7	3	14
LI model	3	0	5	5	0	0	10
	4	8	1	0	0	0	9
	5	0	2	4	1	0	7
	Total	12	93	34	8	8	155

Table 7.11: Comparing the results of the local independence model versus the restricted covariance model

However, since the local independence model does not explicitly include covariances between the purchase rates of the 4 variables, it is not able to account for differences in the structure of the interdependencies that might exist between the product purchase rates. This is exactly why cluster 1 in the LI model gest split up into cluster 2 and 3 of the RC model. Indeed, when looking back to table 7.8 in section 7.9.1.3, it becomes clear that some differences exist in the main effects and the interactions between fabric detergent and softener. In fact, whereas the average purchase rate of softener in cluster 2 of the RC model is heavily dependent on the covariance with detergent, the reverse is true for cluster 3. Such interpretation can not be inferred from the results of the LI model since it does not explicitly account for such interactions within the formulation of the mixture model.

For cluster 3 of the LI model, it is clear that the average purchase rates for the 4 variables correspond most closely to clusters 2 and 3 of the RC model than with any other clusters in the RC model. However, why this cluster gets split up into two parts will probably again have to do with the reasons discussed before, i.e. the interdependence between the variables that are not accounted for in the LI model.

Common covariance (CC) versus restricted covariance (RC)

The first model that explicitly accounts for interactions between the variables is the common covariance model (see section 7.9.1.2). However, instead of modelling specific interactions between variables separately, this model accounts for interactions in a rather global way, i.e. by means of one common covariance term. It is therefore interesting to find out whether the specific bivariate interactions accounted for in the RC model reveal information that leads to a different clustering of the observations than the clustering produced by the common covariance model. A cross-tabulation of the 5-component solution for both models produces table 7.12.

Table 7.12 shows that clusters 1 and 4 of the CC model split up into clusters 2 and 3 of the RC model and that cluster 5 in the CC model splits up into clusters 1 and 2 of the RC model.

		RC model					
		1	2	3	4	5	Total
CC model	1	0	58	17	0	0	75
	2	5	0	0	0	0	5
	3	0	3	1	8	1	13
	4	0	18	15	0	7	40
	5	7	14	1	0	0	22
	Total	12	93	34	8	8	155

Table 7.12: Comparing the results of the common covariance model versus the restricted covariance model

However, when comparing the average purchase rates, the clusters still look quite different from each other. For instance, recall that for the common covariance model, the marginal mean purchase rate for a variable (say cakemix) equals $\lambda_C + \lambda_{CFDS}$ and can be calculated from table 7.7 Thus, for cluster 1 in the CC model, these marginal means (cakemix=1.961, frosting=3.496, detergent=2.582, softener=3.771) can be compared with the marginal means obtained for cluster 2 in the RC model (cakemix=1.505, frosting=1.419, detergent=3.290, softener=2.022) (see table 7.10). As one can see, although there is an important overlap between cluster 1 in the CC model with cluster 2 in the RC model, the mean purchase rates for each cluster are nevertheless quite different. The same is true for cluster 4. The reason is that the observations of cluster 2 in the RC model get distributed over multiple clusters (1, 3, 4 and 5) in the CC model such that the mean purchase rates for cluster 2 in the RC model is basically a weighted average of the mean purchase rates of cluster 1, 3, 4 and 5 of the CC model.

Combined bivariate (BI) versus restricted covariance (RC)

This comparison is based on two independent bivariate Poisson mixture models between cakemix and frosting on the one hand, and between fabric detergent and softener on the other hand. Details of both bivariate models are shown in appendix 9. Table 7.13 illustrates the cross-tabulation of the cluster memberships for the 155 households. The rows of the table are the combined clusters after grouping together the k=2 component solutions for each of the bivariate Poisson mixture models. The columns represent the k=5 cluster solution of the RC model.

		RC model					
		1	2	3	4	5	Total
Combined BI model	CF=1, DS=1	0	84	34	1	4	123
	CF=2, DS=1	0	0	0	7	2	9
	CF=1, DS=2	12	9	0	0	0	21
	CF=2, DS=2	0	0	0	0	2	2
	Total	12	93	34	8	8	155

Table 7.13: Comparing the results of the pairwise bivariate model against the restricted covariance model

From this cross-tabulation, it is clear that the observations in the first cluster of the combined BI model get split up over clusters 2 and 3 of the RC model. The same is true for the observations contained in cluster 3 of the combined BI model, which are spread out over clusters 1 and 2 of the RC model. A closer look at the parameters values of each cluster reveals the logic of this separation.

For the first cluster of the combined model (CF=1, DS=1), the separation is most probably made based on the values of λ_D, λ_S, and λ_{DS}. In fact, one can see from table 7.8 (see section 7.9.1.3) that for cluster 2 the main purchase rate of softener is very low (λ_S=0.031), such that the average purchase rate for softener in cluster 2 of the RC model is mainly due to the strong covariance with detergent (λ_{DS}=1.955) whereas in cluster 3 of the RC model, the reverse effect is true. Indeed, in cluster 3 the main purchase rate for detergent is very low (λ_D=0.00), such that the average purchase rate for detergent is mainly due to the strong purchase interaction with softener. In other words, the observations in the first cluster of the combined bivariate model get separated according to how the main effects of detergent and softener are related to the interaction between both products.

For the third cluster of the combined BI model, the separation is probably also made based on the values of λ_D, λ_S, and λ_{DS}. In fact, in the first cluster of the RC model (see table 7.8 in section 7.9.1.3), there is no interdependence between the purchases of detergent and softener (λ_{DS}=0.000), whereas in the second cluster, the interaction between detergent and softener is quite large (λ_{DS}=1.955).

These results tend to favour the RC model over the two bivariate models. However, when looking at the k=4 components solution of the bivariate detergent-softener model (see table A.9.4 in appendix 9), then it can be seen that this separation actually takes place in the k=4 components solution. Indeed, the table shows that cluster 1 of the bivariate detergent-softener model corresponds closely, both in profile and in size, to cluster 3 in the RC model (see table 7.8), and that cluster 4 of the bivariate model corresponds closely to cluster 2 of the RC model. This is, however, not true for the bivariate cakemix-frosting model where no clear correspondence between the clusters of the RC model can be found.

To summarize, one could say that the bivariate detergent-softener model can largely reproduce the detergent-softener part of the RC model, but that the same does not hold for the cakemix-frosting part. In fact, the clusters of the RC model show some differences between the main effects and interactions of cakemix and softener that are not identified by the bivariate model cakemix-frosting.

7.10 Limitations and Contributions

7.10.1 Limitations

Firstly, the models presented in this dissertation do not include marketing variables (covariates). This may limit their accessibility in a practical retail situation where often profiling information is used to target people with customized campaigns. However, when a detailed history of sales transactions about individuals is available, the presented methods are able to allocate people to the different segments, and covariate information may not be necessary. Nevertheless, covariates may help explain the differences between the behaviour in the clusters, which in this case is not possible. This issue is therefore high on the list of topics for future research.

Secondly, no data were available on price or promotional elasticity of the products included in this study. As a result, the suggested merchandising strategies should be adopted with care. Indeed, the price and promotion sensitivity of consumers is not known such that results of particular merchandising strategies, based on knowledge about purchase rates, can not be predicted with great certainty. Furthermore, the effect of promotions on the long-run purchase rates of consumers are not clear and depend on consumer purchase behaviour such as stock-piling, category expansion, cannibalism and competitive brand switching and store switching. For instance, a promotion on a particular package size of detergent brand may increase the short-term sales of that brand due to forward buying, at the expense of other package sizes of the same brand (cannibalism), and/or at the expense of other detergent brands (competitive brand switching) within the category, followed by a post-promotion dip. Furthermore, if the store is not able to attract consumers from competing stores (store switching), the long-term category expansion will be low and purchase rates may be left unaffected in the long run. Recent data [2] for a sample of Belgian supermarkets shows, however, that price promotions on detergent have a medium (i.e. neither low, nor high) impact on category expansion.

Thirdly, the segment specific purchase rates are treated as static parameters in the model, whereas in practice, they will probably change over time. This could result in customers switching from one cluster to another, or entire clusters to change profile over time, i.e. move from one position to another. This dynamic aspect has not been accounted for in this study. However, we do not consider this as being problematic in our study given the relatively short time horizon over which the data were collected.

Fourthly, loglinear analysis may identify two products purchase rates as being independent, as a result of existing opposite product interdependence effects in subgroups of the sample. For instance, in one subgroup, purchases can be positively correlated, whereas for another group they can be negatively correlated such that overall, no interdependence exists. In that case, in the current methodology, the respective covariance term will be set to zero in the covariance matrix and the model can no longer identify any interdependence effect between both products in each of the subgroups.

Fifthly, although it was shown in this dissertation how to significantly reduce the complexity of the multivariate Poisson mixture model, and how the model scales towards more observations (e.g. more customers or a longer transaction history), the scalability towards including more product categories requires more empirical study. Future research topics therefore include the evaluation of Markov Chain Monte Carlo (MCMC) methods to estimate the parameters for larger versions of the multivariate Poisson mixture model. Moreover, EM algorithms for the general multivariate Poisson mixture model are of interest, such as the construction of efficient recursive schemes for the evaluation of the probability mass function of a multivariate Poisson model. Such schemes can speed up the estimation of these models considerably.

Finally, but related to the data set used, some of the clusters (for instance cluster 4 and 5 in the restricted covariance model) obtained in the analysis contain very few observations and might be an indication of outliers in the data. By removing these observations prior to analysis the big clusters might show up sooner, i.e. after fewer components. In some applications, however, there is a high interest in discovering small but highly valuable clusters, such as in astronomy. Furthermore, the fact that the method separates these observations from the rest indicates that the method is able to identify such extreme observations and not let them spoil the other clusters.

7.10.2 Contributions

From a theoretical point of view, we have presented a multivariate Poisson mixture model to cluster supermarket shoppers based on their purchase rates in a number of product categories. However, instead of using the general formulation of the multivariate Poisson distribution, i.e. with fully-saturated variance/covariance matrix, it was shown that the number of free parameters can be reduced significantly by preliminary examination of the interdependencies between product category purchases. Knowledge about these interactions can be obtained in different ways. For small problems, which include just a few product categories, loglinear analysis of multi-way contingency tables containing the frequencies of product purchase co-occurrences can reveal statistically significant purchase associations. For larger problems, a data mining approach can be followed where significant product interactions can be discovered by means of association rule analysis. The result is a more parsimonious version of the multivariate Poisson mixture model that is easier and faster to estimate whilst it still accounts for most of the existing covariance in the underlying data. Furthermore, an EM algorithm was presented to estimate the parameters of the model.

From a practical point of view, the model was tested on a real supermarket dataset including the purchases of 155 households over a period of 26 weeks in 4 product categories (cakemix, frosting, fabric detergent and softener). The results of the model indicated that two big clusters, accounting for almost 80% of the observations, could be found with a distinct purchasing profile in terms of the purchase rates and purchase interactions between the product categories considered. Moreover, it was illustrated how the results can be used to devise merchandising strategies for the different clusters, based on their purchasing profile.

Finally, a comparison between the different suggested models was made in order to detect whether the more complex 'restricted covariance model' leads to additional insights into the data that can not be obtained by simpler models, such as the local independence model, the common covariance model, or a combination of two bivariate models between cakemix-frosting and fabric detergent-softener. Our comparison was based on a cross-tabulation of the cluster memberships in order to evaluate the overlap between different cluster solutions. However, in the end, the best way to compare the quality of two cluster solutions (say the two bivariate models against the restricted covariance model) would be to simulate data from the restricted covariance model and to fit two independent bivariate Poisson models on the data to see whether they are able to reveal the underlying structure.

CHAPTER 8
CONCLUSIONS AND FUTURE RESEARCH

8.1 Conclusions

The entrance of new technologies in supermarkets, such as barcode scanners and large database systems, has led to an abundance of data in the retailing sector. Although originally those systems were used to facilitate inventory management and speed up the checkout of customers, retailers have recently realized that inside these data there may be hidden useful information on consumer purchase behaviour. However, since these data are of secondary nature, i.e. they were initially not collected for decision-making, they are often noisy and dirty (contain outliers and inconsistencies). On the other hand, receipt data have a number of advantages that make them well suited for market basket analysis, i.e. they are quickly available, have a low acquisition cost, are collected at a very detailed level, and above all, reflect product purchase interdependencies. Previous research has demonstrated that product interdependence effects play an important role in marketing decision-making and that failing to consider such interdependencies may lead to disappointing results.

Our interest in this dissertation therefore went to two important problems in retailing, i.e. product selection and customer segmentation, where product interdependence effects may play an important role. Additionally, we were interested in finding out how receipt data and a recent technology for finding product co-occurrences in market basket data, i.e. association rules, are suited to support solutions to the problems of product selection and customer segmentation.

More specifically, we proposed an integer programming framework for product selection (PROFSET) and a multivariate Poisson mixture model for behaviour-based customer segmentation. Although very different with regard to the marketing problem that they tackle, in both models product or category co-occurrence information was explicitly account for.

For instance, in chapter 5, we were able to demonstrate that the PROFSET product selection model is able to take advantage of cross-selling effects to select products for a convenience store and for selecting products to position at attractive locations in a traditional store environment. However, since scanner data are typically very large, both in terms of the number of transactions as in the number of products/categories, finding such cross-selling effects within reasonable time is not straightforward (see section 3.2.3.1). We therefore proposed a relatively recent data mining technique, i.e. the discovery of frequent itemsets in association rule mining, to discover such (multi-way) cross-selling effects in an efficient way (see chapter 4). Despite the lack of product cost information and the existence of high market concentration in the data, we were able to demonstrate that cross-selling effects should be taken into account when trading-off products against each other. Indeed, from the marketing point of view, and especially in product categories where the category shares for the different brands are less concentrated towards a limited set of brands, we were able to demonstrate that by incorporating cross-selling effects between brands, some brands tend to become more important than others in terms of their overall profitability for the selected assortment. From the data mining point of view, we introduced a new framework to evaluate the interestingness of product associations. Indeed, most contributions to the field of interestingness of product associations have been focussed on statistical (interest, chi-squared, intensity of implication, …) or subjective (surprisingness, actionability, …) criteria. In contrast, we argued that that the business value of product associations is crucial in the treatment of the interestingness problem within a retailing context. Furthermore, we argued that although the discovery of association rules itself may reveal new insights into customer purchase behaviour, association rule mining should not be the endpoint but rather serve as the input for other modelling efforts, such as for product selection.

In chapter 6, we introduced the idea of behaviour-based customer segmentation within the context of market segmentation research. In this context, we provided a literature overview of both the most common variables and methods for segmentation. Furthermore, we provided two concrete illustrations of behaviour-based segmentation on a real dataset of supermarket sales transactions. In the first illustration, frequency and monetary value were used to segment supermarket shoppers into two groups and to investigate whether differences in the purchase behaviour of those groups could be found in terms of frequently co-occuring category purchases. In the second illustration, the size of the shopping basket (i.e. the number of distinct items) was used to segment supermarket shoppers into three groups. Both for fill-in baskets as for stocking-up baskets, it was investigated whether differences could be found in the purchase of frequently co-occuring product categories.

In chapter 7, we introduced a flexible multivariate Poisson mixture model based methodology for behaviour-based customer segmentation. From a statistical point of view, we argued that the general formulation of the multivariate Poisson mixture model with full variance-covariance structure (i.e. the fully-saturated model) often imposes more structure than needed in practice, and that as a result of this, the calculation of the probability function and the estimation of its parameters becomes computationally infeasible. Researchers have solved this by reducing the variance-covariance matrix down to one common covariance term. However, this specification, although much easier to compute, is often overly simplistic in practice. Therefore, we proposed an intermediate model that uses information from the examination of marginal interdependencies to construct a more parsimonious covariance structure. This way, the model accounts for most of the existing covariance in the data, whilst its parameters can still be computed within a reasonable amount of time. A theoretical development of the model is given, together with an EM algorithm to estimate its parameters. Estimation of the model on real data showed that distinct customer segments could be found, that the segments are well separated, and that the specification of a well-tuned variance-covariance matrix significantly increases the fit over the local independence and common covariance model.

From a marketing point of view, we showed that product purchase co-occurrence information enables the identification of customer segments with different purchase rates in a set of product categories, that these purchase rates can be interdependent and that strength of the interdependencies are segment specific. Furthermore, we argued that information of this kind could be used by category managers, for instance, to setup customized marketing campaigns and reorganize category layout.

Despite these interesting results, the use of product co-occurrence information for product selection and behaviour based customer segmentation is also subject to some limitations. In fact, one should be careful in predicting the outcome of marketing-mix decisions (such as pricing, promotion) purely on the basis of co-occurrence information. This is in general a limitation of all methods that measure co-occurrence (such as association rules, association coefficients and loglinear analysis). Indeed, in chapter 3 and 4, it was argued that co-occurrence does not reflect the reasons for co-occurrence and thus it is hard to predict what will be the outcome of particular marketing actions. In other words, different reasons (e.g. usage complements, location in the store) may lie on the basis for the observed co-occurrence and receipt data typically do not reflect such information. As a result, one should be careful in taking marketing actions to exploit such co-occurrences. For instance, assume that the reason for observing a high purchase co-occurrence between cheese and milk would be that they are located close to each other in the store. In that case, reducing the price of milk will not necessarily have a positive impact on the sales of cheese since the reason for their co-occurrence in shopping baskets is different, i.e. there is not necessarily a positive cross price elasticity between both products.

Some other limitations of association rules were also discussed in chapter 4, such as how to reduce the amount of association rules to the most interesting ones. For instance, one issue is that of redundancy. Indeed, since association rules are neither mutually exclusive, nor collectively exhaustive, multiple rules may cover the same instance, which leads to the problem of redundancy in an association rules ruleset. Therefore, Toivonen et al. proposed the RuleCover heuristic to select the least redundant set of rules by means of an iterative rule covering principle. The RuleCover heuristic, however, suffers from one important deficiency, i.e. the selection of rules is dependent on the ordering of the rules. In other words, the stepwise selection of a subsequent rule is dependent on which rules have been chosen during the previous steps. Thus, the heuristic uses only locally optimal information. Therefore, we presented the SetCover integer-programming model, which selects the least redundant set of rules by means global optimization. Experiments showed that the use of global information instead of local information enables to further reduce the amount of rules by a factor of 1 to 10%.

8.2 Future Research

Research is never finished, as illustrated by our future research agenda. Recently, product rationalization has been indicated as a problem of extremely high interest, both to practitioners and academics and we believe that the PROFSET model can be adapted to contribute to tackling this problem. Indeed, the rapid proliferation of brands within particular product categories not only places a heavy burden on the operational costs for the retailer, but also creates problems towards the consumer, i.e., it creates an overload of information and thus increases the difficulty of deciding which product(s) to choose. Yet, retailers are often reluctant to reduce the number of SKU's because they are afraid of the reactions of the consumer. Yet, several studies have demonstrated that the number of SKU's can be significantly reduced without a significant effect on the sales of the category, as long as people still find their favourite brands and the total number of facings devoted to the category remains unchanged. In that case, the question arises which of those brands to eliminate from the assortment. All too often, rather simple heuristics are used in practice, e.g. cutting the bottom third. Yet, cross-selling effects between products may again play an important role such that cutting away particular products will have a larger effect overall than can be expected from the loss of sales of the deleted items. We have ideas how to adapt PROFSET to take into account these cross-selling effects in order to take into account this information for better product deletion.

Secondly, the choice of a good support (and confidence) threshold has been a subject for debate as long as association rules exist. The debate is motivated by the finding that if the support is too high, interesting associations with low support may be missed, and if support is too low, the statistical significance of low frequency associations becomes an important problem. Therefore, a new framework based on Markov blankets [72] was recently proposed to mine for less frequent but highly significant associations. Briefly, a Markov blanket can be defined as follows. Suppose that, in the terminology of association rules (see chapter 4), X, Y and Z are sets of items ($X, Y, Z \subseteq I$), then a Markov blanket is a set of items/products Z such that an item Y is conditionally independent of another set of items X, given Z. In other words, for an association rule $X \Rightarrow Y$, Z is called a Markov blanket if Y is conditionally independent of X, given Z, or thus, $P(Y \mid X, Z) = P(Y \mid Z)$. The discovery of Markov blankets is interesting for the PROFSET optimization framework because it solves the problem of first mining for frequent itemsets and then deciding on how to distribute the margin of a transaction across a given collection of frequent itemsets, based on some measure for interdependency. Indeed, Markov blankets enable to carry out both steps at once and it is insensitive for the frequency threshold. It is therefore on our list of topics for future research to empirically investigate the contribution of Markov blankets to the problem of product selection.

Thirdly, more research is needed to investigate the impact of external/environmental circumstances or decisions on the association between products or categories. In fact, it was argued in chapter 5 that it is assumed that frequent itemsets can be used as a measure for the strength of the interdependency between retail products/categories. However, as far as we know, there exists no research on how aspects such as product placement, pricing and promotion may (positively or negatively) affect the strength of the association being measured between a set of products or categories. Therefore, in-store experiments are needed to find out whether these effects indeed play an important role and whether these effects are weaker or stronger for particular products or product categories. Cross-sectional data from different retail stores or in-store experiments could be helpful in this context.

Fourthly, because of the limitations of using purchase co-occurrence information for behaviour-based customer segmentation (i.e. we don't know the reasons for observed co-occurrence), we are currently investigating the inclusion of covariate information (socio-demographic and marketing mix) into the multivariate Poisson mixture model. This may help explain the purchase rates and the interactions between them.

Finally, a Bayesian version of the behaviour-based segmentation model is of interest. We expect that this would facilitate the estimation of the model since then one simply needs to simulate from a density without having to calculate its expectation. In fact, previous research has shown that the Bayesian methodology is scalable for large versions (i.e. many variables) of the multivariate Poisson model with common covariance [270]. It is therefore worthwhile to investigate how our model with restricted covariance structure can be scaled towards more dimensions with the use of Bayesian estimation.

References

[1] Abramson, C., Andrews, R.L., Currim, I.S., and Jones, M. (2000), Parameter bias from unobserved effects in the multinomial logit model of consumer choice, in: *Journal of Marketing Research*, 17 (November), pp. 410-426.

[2] Accuris, Hoe past u uw promoties aan om een maximale opbrengst op het winkelpunt te realiseren?, presented at: *Winnen op het Winkelpunt*, IIR conferentie, Brussels, May 23-24.

[3] ACNielsen (1992), *Category Management: Positioning Your Organization to Win*, NTC Business Books in association with Nielsen Marketing Research and the American Marketing Association.

[4] ACNielsen (2001), Het Voedingsuniversum in België 2001, *Company report*.

[5] Agarwal, R.C., Aggarwal, C.C., and Prasad, V.V.V. (2000), Depth first generation of long patterns, in: *Proceedings of the 6th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Boston (USA) pp. 108-118.

[6] Aggarwal, C., Sun, Z., and Yu, P.S. (1998), Online algorithms for finding profile association rules, in: *Proceedings of the 7th ACM CIKM International Conference on Information and Knowledge Management*, Bethesda (USA), pp. 86-95.

[7] Aggarwal, C.C., and Yu, P.S. (1998), A new framework for item set generation, in: *Proceedings of the ACM-PODS Symposium on Principles of Database Systems*, Seattle, Washington (USA), pp. 18-24.

[8] Agrawal, R., Imielinski, T., and Swami, A. (1993), Mining association rules between sets of items in large databases, in: *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data,* Washington DC (USA), pp. 207-216.

[9] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A. (1996), Fast discovery of association rules, in: *Advances in Knowledge Discovery and Data Mining*, AAAI Press, pp. 307-328.

[10] Agrawal, R., and Srikant, R. (2000), Privacy-preserving data mining, in: *Proceedings of the ACM SIGMOD 2000 Conference on Management of Data*, Dallas (USA), May 2000, pp. 439-450.

[11] Agrawal, R., and Srikant, R. (1995), Mining sequential patterns, in: *Proceedings of the 11th International Conference on Data Engineering (ICDE)*, Taipei (Taiwan), pp. 3-14.

[12] Agrawal, R., and Srikant, R. (1994), Fast algorithms for mining association rules, in: *Proceedings of the 20th International Conference on Very Large Databases (VLDB)*, Santiago (Chile), pp. 487-499.

[13] Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, Wiley Series in Probability and Statistics.

[14] Ainslie, A.S. (1998), Similarities and differences in brand purchase behaviour across categories, *PhD. Dissertation*, University of Chicago.

[15] Aitkin, M., and Aitkin, I. (1996), An hybrid EM/Gauss-Newton algorithm for maximum likelihood in mixture distributions, in: *Statistics and Computing*, Vol. 6, pp. 127-130.

[16] Aitkin, M., Anderson, D., and Hinde, J. (1981), Statistical modelling of data on teaching styles, in: *Journal of the Royal Statistical Society*, Series A, Vol. 144, pp. 419-461.

[17] Akaike, H. (1974), A new look at statistical model identification, in: *IEEE Transactions on Automatic Control*, AC-19, pp. 716-723.

[18] Ali, K., Manganaris, S., and Srikant, R. (1997), Partial Classification using Association Rules, in: *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, Newport Beach, California (USA), pp. 115-118.

[19] Allenby, G. M., and Rossi, P.E. (1999), Marketing models of heterogeneity, in: *Journal of Econometrics*, Vol. 89, pp. 57-78.

[20] Anderson, E.E., and Amato, H.N. (1974), A mathematical model for simultaneously determining the optimal brand-collection and display-area allocation, in: *Operations Research*, Vol. 22(1), pp. 13-21.

[21] Anand, S.S., Hughes, J.G., Bell, D.A., and Patrick A.R. (1997), Tackling the cross-sales problem using data mining, in: *Proceedings of the first Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Singapore, pp. 331-343.

[22] Angiulli, F., Ianni, G., and Palopoli, L. (2001), On the complexity of mining association rules, in: *Proceedings Nono Convegno Nazionale su Sistemi Evoluti di Basi di Dati (SEBD*), Venice (Italy), pp. 177-184.

[23] Asseal, H. (1984), *Consumer Behaviour and Marketing Action*, Boston, Kent Publishing Company.

[24] Babcock, C. (1994), Parallel Processing Mines Retail Data, in: *Computerworld*, 6, September 26.

[25] Baesens, B., Viaene, S., and Vanthienen, J. (2000), Post-processing of Association Rules, in: *Workshop notes of the workshop entitled 'Post Processing in Machine Learning and Data Mining' of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[26] Banfield, J., and Raftery, A. (1993), Model-based gaussian and non-gaussion clustering, in: *Biometrics*, Vol. 49, pp. 803-821.

[27] Bayardo, R.J., Agrawal, R., and Gunopulos, D. (1999), Constraint-based rule mining in large dense databases, in: *Data Mining and Knowledge Discovery*, Vol. 4(2/3), pp. 217-240.

[28] Bell, D.R., and Lattin, J.M. (1998), Shopping behavior and consumer preference for store price format: why 'large basket' shoppers prefer EDLP, in: *Marketing Science*, Vol. 17(1), pp. 66-88.

[29] Berekoven, L. (1995), *Erfolgreiches Einzelhandelsmarketing: Grundlagen und Entscheidungshilfen*, C.H. Beck'sche Verlagsbuchhandlung, München.

[30] Berry, J. (1994), Database marketing, in: *Business Week* (September 5), pp. 56-62.

[31] Berry, M.J.A. (1999), The Privacy Backlash, in: *Intelligent Enterprise Magazine- Decision Support*, Vol. 2(15).

[32] Berry, M.J.A., and Linoff, G. (1997), *Data Mining Techniques for Marketing, Sales and Customer Support*, Wiley, New York.

[33] Best, R.J. (1997), *Market-Based Management: Strategies for Growing Customer Value and Profitability*, Prentice Hall.

[34] Bixby, R.E., Fenelon, M., Gu, Z., Rothberg, E., and Wunderling, R. (1999), *MIP: theory and practice – closing the gap*, ILOG CPLEX Technical Report.

[35] Blattberg, R.C., and Neslin, S.A. (1990), *Sales Promotion: Concepts, Methods, and Strategies*, Englewood Cliffs, New York, Prentice Hall.

[36] Blischok, T. (1995), Every transaction tells a story, in: *Chain Store Age Executive with Shopping Center Age*, Vol. 71 (3), pp. 50-57.

[37] Bloemer, J.M.M., Brijs, T., Swinnen, G., and Vanhoof, K. (2002), Identifying latently dissatisfied customers and measures for dissatisfaction management, in: *International Journal of Bank Marketing*, Vol 20(2), pp. 27-37.

[38] Bloemer, J.M.M., Brijs, T., Swinnen, G., and Vanhoof, K. (2002), Comparing complete and partial classification for the identification of customers at risk, forthcoming in: *International Journal of Research in Marketing*.

[39] Bloom, P.N., Adler, R., and Milne, G.R. (1994), Identifying the Legal and Ethical Risks and Costs of Using New Information Technologies to Support Marketing Programs, in: *The Marketing Information Revolution*, Harvard Business School Press, Boston, Massachusetts, pp.289-305.

[40] Böcker, F. (1978), die Bestimmung der Kaufverbundenheit von Produkten, in: *Schriften zum Marketing*, Vol. 7.

[41] Böhning, D., Dietz, E., Schaub, R., Schlatttmann, P., and Lindsay, B.G. (1994), The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family, in: *Annals of the Institute of Statistics and Mathematics*, Vol. 46, pp. 373-388.

[42] Böhning, D. (1999), *Computer assisted analysis of mixtures & applications in meta-analysis, disease mapping & others*, CRC press.

[43] Borin, N., and Farris, P. (1990), An empirical comparison of direct product profit and existing measures of SKU productivity, in: *Journal of Retailing*, Vol. 66(3), pp. 297-314.

[44] Borin, N., Farris, P., and Freeland, J.R. (1994), A model for determining retail product category assortment and shelf space allocation, in: *Decision Sciences*, Vol. 25(3), pp. 359-384.

[45] Bozdogan, H. (1988), ICOMP: a new model selection criterion, in: H. Bock (ed.) *Classification and Related Methods of Data Analysis*, Amsterdam: Elsevier, pp. 599-608.

[46] Bozdogan, H. (1987), Model selection and akaike's information criterion (AIC): the general theory and its analytical extensions, in: *Psychometrika*, Vol. 52, pp. 345-370.

[47] Bozdogan, H. (1981), *Multi-Sample Cluster Analysis and Approaches to Validity Studies in Clustering Individuals*, Ph.D. thesis, Department of Mathematics, University of Illinois at Chicago.

[48] Bradley, P.S., Fayyad, U.M., Reina, C.A. (1998), Scaling EM clustering to large databases, in: *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*, New York (USA), pp. 9-15.

[49] Brännäs, K., and Rosenqvist, G. (1994), Semiparametric estimation of heterogeneous count data models, in: *European Journal of Operational Research*, Vol. 76, pp. 247-258.

[50] Briesch, R.A., Chintagunta, P.K., and Matzkin, R.I. (1997), Nonparametric and Semiparametric Models of Brand Choice Behavior, *working paper*.

[51] Brijs, T., Karlis, D., Swinnen, G., Vanhoof, K., and Wets, G. (2002), Tuning the multivariate Poisson mixture model for clustering supermarket shoppers, presented at: *Statistical Modelling and Inference for Complex Data Structures*, poster paper, Louvain-La-Neuve, May 21-23.

[52] Brijs, T., Swinnen, G., Vanhoof, K. and Wets, G. (2002), Building an Association Rules Framework to Improve Product Assortment Decisions, forthcoming in: *Knowledge Discovery and Data Mining*.

[53] Brijs, T., Swinnen, G., Vanhoof, K., and Wets, G. (1999), The Use of Association Rules for Product Assortment Decisions: A Case Study, in: *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining*, San Diego (USA), pp. 254-260.

[54] Brijs, T., Swinnen, G., Vanhoof, K., and Wets, G. (1999), Using Association Rules for Product Assortment Decisions in Automated Convenience Stores, in: *Proceedings of the 10th International Conference on Research in the Distributive Trades*, Stirling (Scotland), pp. 708-716.

[55] Brijs, T., Swinnen, G., Vanhoof, K., and Wets, G. (2000), A data mining framework for optimal product selection in conveniences stores, in: *Proceedings of the European Conference on Information Systems*, Vienna (Austria), pp. 1001-1008.

[56] Brijs, T., Swinnen, G., Vanhoof, K., and Wets, G. (2000), A data mining framework for optimal product selection in retail supermarket data: the generalized PROFSET model, in: *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining*, Boston MA (USA), pp. 300-304.

[57] Brijs, T., Swinnen, G., Vanhoof, K., and Wets, G. (2001), Using shopping baskets to cluster supermarket customers, in: *Proceedings of the 12th Annual Advanced Research Techniques Forum of the American Marketing Association*, Florida (USA).

[58] Brijs, T., Vanhoof, K., and Wets, G. (1999), Concept hierarchies and virtual items in retail market basket analysis, in: *Book of abstracts of the 6th International Conference on Recent Advances in Retailing and Services Science,* Las Croabas (Puerto Rico), pp. 10.

[59] Brijs, T., Vanhoof, K., and Wets, G. (2000), Reducing redundancy in characteristic rule discovery by using integer programming techniques, in: *Intelligent Data Analysis Journal*, Vol. 4(3&4), pp. 229-240.

[60] Brin, S., Motwani, R., and Silverstein, C. (1998), Beyond market baskets: generalizing association rules to dependence rules, in: *Data Mining and Knowledge Discovery*, Vol. 2(1), pp. 39-68.

[61] Brin, S., Motwani, R., Ullman, J.D., and Tsur, S. (1997), Dynamic itemset counting and implication rules for market basket data, in: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Tucson, Arizona (USA), pp. 255-264.

[62] Broniarczyk, S.M., Hoyer, W.D., and McAlister, L. (1998), Consumer's perceptions of the assortment offered in a grocery category: the impact of item reduction, in: *Journal of Marketing Research*, Vol. 35 (May 1998), pp. 166-176.

[63] Brown, C.L. (1996), Behavioral perspectives on market-basket phenomena, *working paper*, New York University.

[64] Bucklin, R.E., and Gupta, S. (1999), Commercial use of UPC scanner data: industry and academic perspectives, in: *Marketing Science*, Vol. 18(3), pp. 247-273.

[65] Bultez, A., and Naert, P. (1988), S.H.A.R.P.: shelf allocation for retailers' profit, in: *Marketing Science*, Vol. 7, pp. 211-231.

[66] Bultez, A., Gijsbrechts, E., Naert, P., and Vanden Abeele, P. (1989), Asymmetric cannibalism in retail assortments, in: *Journal of Retailing*, Vol. 65(2), pp. 153-192.

[67] Bultez, A., Julander, C-R, and Nisol, P. (1995), Structuring retail assortments according to in-store shopping, *Working paper*, Centre for Research on the Economic Efficiency of Retailing.

[68] Cadez, I.V., Smyth, P., and Mannila, H. (2001), Probabilistic modeling of transaction data with applications to profiling, visualization, and prediction, in: *Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining*, San Francisco CA (USA), pp. 37-46.

[69] Calders, T., and Goethals, B. (2002), Mining all non-derivable frequent itemsets, in: *Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Helsinki (Finland), pp. 74-85.

[70] Cameron, A.C., and Trivedi, P.K. (1986), Econometric models based on count data: comparisons and applications of some estimators and tests, in: *Journal of Applied Econometrics*, Vol. 1, pp. 29-55.

[71] Campo, K., Gijsbrechts, E., and Nisol, P. (2000), Towards understanding consumer response to stock-outs, in: *Journal of Retailing*, Vol. 76(2), pp. 219-242.

[72] Castelo, R., Feelders, A., and Siebes, A. (2001), MAMBO: Discovering association rules based on conditional independencies, in: F. Hoffmann et al. (eds.), *Advances in Intelligent Data Analysis*, Springer LNCS 2189, pp. 289-298.

[73] Catalina Marketing, Catalina CheckOut Coupon ®, Catalina Marketing Website, www.catalinamarketing.com.

[74] Cavoukian, A. (1998), Data Mining: Staking a Claim on Your Privacy, *report by the Information and Privacy Commissioner/Ontario*. (www.ipc.on.ca).

[75] Chain Store Age (1963), *Cifrino's space yield formula: a breakthrough for measuring product profit*, Vol. 39(11), p. 83.

[76] Chain Store Age (1965), *Shelf Allocation Breakthrough*, Vol. 41(6), pp. 77-88.

[77] Chamberlain, G. (1980), Analysis of Covariance with Qualitative Data, in: *Review of Economic Studies*, Vol. 47, pp. 225-238.

[78] Chintagunta, P.K., Jain, D.C., and Vilcassim, N.J. (1991), Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data, in: *Journal of Marketing Research*, 28 (November), pp. 417-428.

[79] Chintagunta, P.K., Kyriazidou, E., and Perktold, J. (1998), Panel data analysis of household brand choices, *research paper*.

[80] Churchill, G.A. Jr. (1994), *Marketing Research: Methodological Foundations*, 6th ed., Orlando, Florida, The Dryden Press.

[81] Clifton, C., and Marks, D. (1996), Security and privacy implications of data mining, in: *Proceedings of the ACM SIGMOD Workshop on Data Mining and Knowledge Discovery*, Montreal (Canada), pp. 15-19.

[82] Corstjens, M. L., and Corstjens J. (1995), *Store Wars: the Battle for Mindspace and Shelfspace*, Wiley.

[83] Corstjens, M., and Doyle, P. (1981), A model for optimizing retail space allocations, in: *Management Science*, Vol. 27, pp. 822-833.

[84] Corstjens, M. L., and Gautschi D.A. (1983), Formal Choice Models in Marketing, in: *Marketing Science*, Vol. 2(1), pp. 19-56.

[85] Cosmos: New hope for profits, in: *Chain Store Age*, February, pp. 34-35.

[86] Cummings, P., White, D., and Wisniowski, S. (1990), Strategic simplicity, in: *McKinsey Quarterly*, Vol. 3, pp. 80-90.

[87] Cutter, K., and Rowe, C. (1990), Scanning in the Supermarket for Better or Worse: A Case Study in Introducing Electronic Point of Sale, in: *Behavior and Information Technology*, Vol. 9, pp. 157-169.

[88] Danneels, E. (1996), Market segmentation: normative model versus business reality: an exploratory study of apparel retailing in Belgium, in: *European Journal of Marketing*, Vol. 30(6), pp. 36-51.

[89] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977), Maximum likelihood from incomplete data via the EM-algorithm, in: *Journal of the Royal Statistical Society*, Series B39, pp. 1-38.

[90] De Pelsmacker, P., and Van Kenhove, P. (1996), *Marktonderzoek: Methoden en Toepassingen*, Leuven, Garant, 2nd ed.

[91] DeSarbo, W. S., Ansari, A., Chintagunta, P., Himmelberg, C., Jedidi, K., Johnson, R., Kamakura, A. W., Lenk, P., Srinivasan, K., and Wedel, M. (1997), Representing heterogeneity in consumer response models, in: *Marketing Letters*, Vol. 8(3), pp. 335-348.

[92] De Schamphelaere, J., Van den Poel, D., and Van Kenhove, P. (2002), Direct and indirect effects of retail promotions on sales and profits in the do-it-yourself market, in: *Proceedings of the 31st European Marketing Academy (EMAC) conference*, May 28-31 (+ personal communications.)

[93] Dhalla, N.K., and Mahatoo, W.H. (1976), Expanding the scope of segmentation research, in: *Journal of Marketing*, Vol. 40, pp. 34-41.

[94] Dickson, P.R. (1982), Person-situation: segmentation's missing link, in: *Journal of Marketing*, Vol. 46, pp. 56-64.

[95] Diebolt, J., and Ip, E.H.S. (1996), Stochastic EM: method and application, in: *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall, pp. 259-274.

[96] Dillon, W.R., and Kumar, A. (1994), Latent structure and other mixture models in marketing: an integrative survey and overview, in: *Advanced Methods in Marketing Research*, ed. Bagozzi R., Cambridge, MA: Blackwell, pp. 295-351.

[97] Donovan, R.J., Rossiter, J.R., Marcoolyn, G., and Nesdale, A. (1994), Strore atmospherics and purchasing behavior, in: *Journal of Retailing*, Vol. 70, pp. 283-294.

[98] Doyle, P., and Gidengil, Z. (1977), A review of in-store experiments, in: *Journal of Retailing*, Vol. 53, pp. 47-62.

[99] Drèze, X., Hoch, S.J., and Purk, M.E. (1994), Shelf management and space elasticity, in: *Journal of Retailing*, Vol. 70(4), pp. 301-326.

[100] Duban, A.J. (1978), A failing retailer makes a poor customer, in: *International Association of Chain Stores Quarterly Review*, Vol. 27, pp. 7-10.

[101] DuMouchel, W., and Pregibon, D. (2001), Empirical bayes screening for multi-item associations, in: *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco CA (USA), pp. 67-76.

[102] van Es, D.P. (2000), Product selection based on cross-selling effects, *Technical Report 9634010*, University of Leiden (Holland).

[103] Fader, P.S., Hardie, B.G.S., and Huang, C-Y. (2001), An integrated trial/repeat model for new product sales, *research paper*.

[104] Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. (1996), From data mining to knowlegde discovery: an overview, in: *Advances in Knowledge Discovery and Data Mining*, AAAI Press / The MIT Press, pp. 1-34.

[105] Frank, R.E., and Massy, W.F. (1965), Market segmentation and the effectiveness of a brand's price and dealing policies, in: *Journal of Business*, Vol. 38, pp. 186-200.

[106] Frank, R.E., Massy, W.F., and Wind, Y. (1972), *Market Segmentation*, Englewood Cliffs, NJ: Prentice Hall.

[107] Foekens, E.W. (1995), Scanner Data Based Marketing Modelling: Empirical Applications, *Ph.D. dissertation*, Rijksuniversiteit Groningen.

[108] Forcht, K.A., and Cochran, K. (1999), Using data mining and datawarehousing techniques, in: *Industrial Management & Data Systems*, Vol. 5, pp. 189-196.

[109] Geerts, F., Goethals, B., and Van den Bussche, J. (2001), A tight upper bound on the number of candidate patterns, in: *Proceedings of the First IEEE International Conference on Data Mining*, San Jose (USA), pp. 155-162.

[110] Gensch, D.H. (1985), Empirically testing a disaggregate choice model for segments, in: *Journal of Marketing Research*, Vol. 22(4), pp. 462-467.

[111] GFK (2002), *Carrefour na GB*, presented at: Client Day 2002, Groot Bijgaerde, April 25.

[112] GFK (2002), *GFK Panel Services*, at: Winnen op het winkelpunt, IIR Seminar, May 23-24, Brussels.

[113] Goethals, B., and Van den Bussche, J. (2000), On supporting interactive association rule mining, in: *Proceedings of the 2nd International Conference on Data Warehousing and Knowledge Discovery*, London-Greenwich (GB), pp. 307-316.

[114] Gönül, F., and Srinivasan, K. (1993), Modeling multiple sources of heteregoneity in multinomial logit models: methodological and managerial issues, in: *Marketing Science*, Vol. 12(3), pp. 213-229.

[115] Goodman, L.A. (1974), Exploratory latent structure analysis using both identifiable and unidentifiable models, in: *Biometrika*, Vol. 61, pp. 215-231.

[116] Greenberg, M., and Schwartz, S.M. (1989), Successful needs/benefits segmentation: a user's guide, in: *Journal of Consumer Marketing*, Summer 1989, pp. 29-36.

[117] Guadagni, P.M., and Little, J.D. (1983), A Logit Model of Brand Choice Calibrated on Scanner Data, in: *Marketing Science*, Vol. 2(3), pp. 203-238.

[118] Guillaume, S., Guillet, F., and Philippe, J. (1998), Improving the discovery of association rules with intensity of implication, in : *Principles of Data Mining and Knowledge Discovery*, Volume 1510 of *Lecture Notes in Artificial Intelligence*, Springer, pp. 318-327.

[119] Gunter, B., and Furnham, A. (1992), *Consumer Profiles: An Introduction to Psychographics*, London: Routledge.

[120] Gupta, S., and Chintagunta, P. (1994), On using demographic variables to determine segment membership in logit mixture models, in: *Journal of Marketing Research*, Vol. 31, pp. 128-136.

[121] Gupta, S., Chintagunta, P., Kaul, A., and Wittink, D.R. (1996), Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data, in: *Journal of Marketing Research*, Vol. 33, November 1996, pp. 383-398.

[122] Gutman, J. (1982), A means-end model based on consumer categorization processes, in: *Journal of Marketing*, Vol. 46(Spring), pp. 60-72.

[123] Gutman, J., and Mills, M.K. (1991), Fashion life style, self-concept, shopping orientation, and store patronage: an integrative analysis, in: *Journal of Retailing & Distribution Management*, Vol. 19(3), pp. 64-86.

[124] Haeck, C. (1997), Shop 24: Een Exploratieve Studie van Alternatieve Locaties voor Volautomatische Winkels (in Dutch), *Undergraduate thesis*, Department of applied economics, Antwerp University, Belgium.

[125] Hair, J.F., Anderson, R.E., Tatham, R.L., and Black, W.C. (1998), *Multivariate Data Analysis*, 5th edition, Prentice Hall Intl.

[126] Han, E-H., Karypis, G., Kumar, V., and Mobasher, B. (1997), Clustering based on association rule hypergraphs, in: *Proceedings of the Workshop on Research Issues on Data Mining and Knowledge Discovery*, pp. 9-13.

[127] Han, E-H., Karypis, G. , Kumar, V., and Mobasher, B. (1998), Clustering in a high-dimensional space using hypergraph models, *Technical Report-97-063*, Department of Computer Science, University of Minnesota, Minneapolis.

[128] Han, J., and Fu, Y. (1999), Discovery of multiple level association rules from large databases, in: *IEEE Transactions on Knowledge and Data Engineering*, Vol. 11(5).

[129] Han, J., Pei, J., and Yin, Y. (2000), Mining frequent patterns without candidate generation, in: *Proceedings of the ACM SIGMOD Conference on Management of Data*, Dallas (USA), pp. 1-12.

[130] Hand, D.J., Mannila, H., and Smyth, P. (2001), *Principles of Data Mining*, The MIT Press.

[131] Hansen, P., and Heinsbroek, H. (1979), Product selection and space allocation in supermarkets, in: *European Journal of Operational Research*, Vol. 3, pp. 474-484.

[132] Harlam, B.A., and Lodish, L.M. (1995), Modeling Consumers' Choices of Multiple Items, in: *Journal of Marketing Research*, Vol. 32(November), pp. 404-418.

[133] Hasselblad, V. (1969), Estimation of finite mixtures from the exponential family, in: *Journal of the American Statistical Association*, Vol. 64, pp. 1459-1471.

[134] Heckman, J.J. (1981), The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time – Discrete Data Stochastic Process, in: *Structural Analysis of Discrete Data With Econometric Applications*, MA: MIT Press.

[135] Henneking, K-M. (1998), Marketingcontrolling im Einzelhandel, *Ph. D. dissertation*, Wiesbaden, Deutscher UniversitätsVerlag.

[136] Hernant, M., and Persson, P-G (1995), In search of a Deal Prone Consumer: A study of consumers' usage of sales promotions, introducing a new type of single source data. *Working paper nr. 11*, The Economic Research Institute, Stockholm School of Economics.

[137] Hilderman, R.J., and Hamilton, H.J. (1999), Knowledge discovery and interestingness measures: a survey, *Technical Report CS-99-04*, Department of Computer Science, University of Regina.

[138] Hoogendoorn, A.W., and Sikkel, D. (1999), Description of purchase incidence by multivariate heterogeneous Poisson processes, in: *Statistica Neerlandica*, Vol. 53(1), pp. 21-35.

[139] Houtsma, M., and Swami, A. (1995), Set-oriented data mining in relational databases, in: *Data and Knowledge Engineering*, Vol. 17, pp. 245-262.

[140] Howard, J.A., and Sheth, J.N. (1969), *The theory of buyer behavior*, New York: John Wiley & Sons.

[141] Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge, UK: Cambridge University Press.

[142] Huang, C-Y., and Hardie, B.G.S. (2000), Come often and stay longer, please: modelling a website's stickiness, in: *Proceedings of the 2000 International Conference of the Theories and Practices of Electronic Commerce*, Taipei.

[143] Jain, D.C., Vilcassim, N.J., and Chintagunta, P.K. (1994), A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data, in: *Journal of Business & Economic Statistics*, Vol. 12(3), pp. 317-328.

[144] Johnson, N.L., and Kotz, S. (1969), *Distributions in Statistics: Discrete Distributions*, Boston, Houghton Mifflin.

[145] Jones, J.M., and Landwehr, J.T. (1988), Removing heterogeneity bias from logit model estimation, in: *Marketing Science*, Vol. 7(Winter), pp. 41-59.

[146] Julander, C-R (1992), Basket Analysis: a New Way of Analysing Scanner Data, in: *International Journal of Retail & Distribution Management*, Vol. 20(7), pp. 10-18.

[147] Kahle, L.R., ed. (1983), *Social Values and Social Change: Adaptation to life in America*, New York: Praeger.

[148] Kahn, B.E., and Schmittlein, D.C. (1992), The relationship between purchases made on promotion and shopping trip behavior, in: *Marketing Letters*, Vol. 68, pp. 294-315.

[149] Kamakura, W.A., Wedel, M., and Agrawal, J. (1994), Concomitant variable latent class models for conjoint analysis, in: *International Journal of Research in Marketing*, Vol. 11, pp. 451-464.

[150] Kano, K., and Kawamura, K. (1991), On recurrence relations for the probability function of multivariate generalized Poisson distribution, in: *Communications in Statistics–Theory and Methods*, Vol. 20, pp. 165-178.

[151] Karlis, D. (1998), Estimation and testing problems in Poisson mixtures, *Ph.D. dissertation*, Department of Statistics, Athens University of Economics and Business, Greece, pp. 40-63.

[152] Karlis, D. (2001), An EM algorithm for multivariate Poisson distribution and related models, *Technical Report 112*, Department of Statistics, Athens University of Economics and Business, Greece.

[153] Kaufman, L., and Rousseeuw, P.J. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, NY.

[154] KBC (1999), Economisch Financiële Berichten, Veertiendaags Tijdschrift KBC, issue 20.

[155] Kendall, M., and Stuart, A. (1979), *The Advanced Theory of Statistics: Inference and Relationship*, Vol. 2, 4^{th} edition, Charles Griffin and Company Ltd., London.

[156] Kestnbaum, R.D. (1992), Quantitative database methods, in: *The Direct Marketing Handbook*, Nash E.L., pp. 588-597.

[157] Kim, S-M, Kim, J-D, Hong, J-H, Nam, D-W, Lee, D-H, and Lee, J-Y (2000), A system for association rule finding from an internet portal site, in: *Proceedings of the INFORMS-KORMS 2000 conference*, Seoul.

[158] Kim, B-D, and Park, K. (1997), Studying patterns of consumer's grocery shopping trip, in: *Journal of Retailing*, Vol. 73(4), pp. 501-517.

[159] Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., and Verkamo, A.I. (1994), Finding interesting rules from large sets of discovered association rules, in: *Proceedings of the 3rd International Conference on Information and Knowledge Management*, pp. 401-407.

[160] Knoke, D., and Burke, P.J. (1980), *Log-Linear Models*, Sage Publications, California, USA.

[161] Kocherlakota, S., and Kocherlakota, K. (1992), *Bivariate Discrete Distributions*, New York: Marcel Dekker.

[162] Kohonen, T. (1982), Self-organized formation of topologically correct feature maps, in: *Biological Cybernetics*, Vol. 43, pp. 59-69.

[163] Kok, J.N., and Kosters, W.A. (2000), Natural data mining techniques, in: *Bulletin of the European Association for Theoretical Computer Science*, Vol. 71, pp. 133-142.

[164] Kopp, R.J., Eng, R.J., and Tigert, D.J. (1989), A competitive structure and segmentation analysis of the Chicago fashin market, in: *Journal of Retailing*, Vol. 65(4), pp. 496-515.

[165] Kotler, P. (1988), *Marketing Management*, Englewood Cliffs, Prentice-Hall.

[166] Lagasse, L., Van Kenhove P., and van Waterschoot, W. (2000), *Management van het Distributiekanaal*, Antwerpen, Standaard Uitgeverij.

[167] Lakshmanan, L.V.S., Ng, R., Han, J., and Pang, A. (1999), Optimization of constrained frequent set queries with 2-variable constraints, in: *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, Philadelphia, Pennsylvania (USA) pp. 157-168.

[168] Lancaster, K. (1966), A new approach to consumer theory, in: *Journal of Political Economy*, Vol. 74 (April), pp. 132-157.

[169] Lancaster, K. (1971), *Consumer demand: a new approach*, New York: Columbia University Press.

[170] *Latent Gold*, Computer package for mixture modelling distributed by Statistical Innovations (www.statisticalinnovations.com).

[171] Lent, B., Swami, A., and Widom, J. (1997), Clustering association rules, in: *Proceedings of the International Conference on Data Engineering*, Birmingham (UK), pp. 220-231.

[172] Lewis, J.E. (1981), ACCUSPACE… The advanced-technology space management system for the 1980's and beyond, in: *Proceedings of the 25th Annual Executive Congress of the International Association of Chain Stores*, June 1, Hamburg, Germany.

[173] Li, C-S, Lu, J-C, Park, J., Kim, K., Brinkley P.A., and Peterson, J.P. (1999), Multivariate zero-inflated Poisson models and their applications, in: *American Statistical Association*, Vol. 41(1), pp. 29-38.

[174] Lilien, G.L., Kotler, P., and Moorthy, K.S. (1992), *Marketing Models*, Prentice Hall.

[175] Lindsay, B. (1995), *Mixture Models: Theory, Geometry and Applications*, Regional Conference Series in Probability and Statistics, Vol. 5, Institute of Mathematical Statistics and American Statistical Association.

[176] Little, J.D.C. (1990), Information Technology in Marketing, *working paper 1860-87*, Sloan School of Management, MIT.

[177] Liu, B., Hsu, W., and Ma, Y. (1999), Pruning and summarizing the discovered associations, in: *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining*, San Diego (USA), pp. 125-134.

[178] Lucas, G.H. Jr, Bush, R.P., and Gresham, L.G. (1994), *Retailing*, Houghton Mifflin Company.

[179] Luce, R. (1959), *Individual choice behavior*, New York, John Wiley and Sons.

[180] Lunn, T. (1986), Segmenting and constructing markets, in: *Consumer Market Research Handbook*, eds. Worcester, R., and Downham, J., ESOMAR publication series, McGraw-Hill, chapter 14.

[181] Mahamunulu, D.M. (1967), A note on regression in the multivariate Poisson distribution, in: *Journal of the American Statistical Association*, Vol. 62(317), pp. 251-258.

[182] Malsagne, R. (1972), La productivité de la surface de vente passe maintenant par l'ordinateur, in: *Travail et Methodes*, Vol. 274, pp. 3-8.

[183] Manchanda, P., Ansari, A., and Gupta, S. (1999), A Model for Multi-Category Purchase Incidence Decisions, in: *Marketing Science*, Vol 18(2), pp. 95-114.

[184] Mannila, H., Toivonen, H., and Verkamo, A.I. (1994), Efficient algorithms for discovering association rules, in: *Knowledge Discovery in Databases*, AAAI Press, pp. 181-192.

[185] Marketing News (2002), Group warns of grocery grab for data, in: *Marketing News*, April edition, pp.7.

[186] Marshall, A.W., and Olkin, I. (1985), A family of bivariate distributions generated by the bivariate Bernouilli distribution, in: *Journal of the American Statistical Association*, Vol. 80(390), pp. 332-338.

[187] McAlister, L. (1982), A dynamic attribute satiation model for choices made across time, in: *Journal of Consumer Research*, Vol. 9(3), pp. 141-150.

[188] McCann, J.M. (1974), Market segment response to the marketing decision variables, in: *Journal of Marketing Research*, Vol. 11, pp. 399-412.

[189] McCann, J.M., and Gallagher, J.P. (1990), Expert Systems for Scanner Data Environments, *International Series in Quantitative Marketing*, Kluwer Academic Publishers.

[190] McCullagh, P., and Nelder, J.A. (1989), *Generalized Linear Models*, 2nd edition, Chapman and Hall, London.

[191] McCurley Hortman, S., Allaway, A.W., Mason, J.B., and Rasp, J. (1990), Multisegment analysis of supermarket patronage, in: *Journal of Business Research*, Vol. 21, pp. 209-223.

[192] McHugh, R.B. (1956), Efficient estimation and local identification in latent class analysis, in: *Psychometrika*, Vol. 21, pp. 331-347.

[193] McKinsey General Foods Study (1963), *The economics of food distributors*, New York: General Foods.

[194] MCKINSEY (1974), *Evaluating feasibility of SPNS in the UK grocery industry,* Institute of Grocery Distribution, Watford.

[195] McLachlan, G.J., and Basford, K.E. (1988), *Mixture Models: Inference and Applications to Clustering*, Statistics: textbooks and monographs, Marcel Dekker, Inc.

[196] McLachlan, G., and Krishnan, T. (1997), *The EM algorithm and extensions*, Wiley: NY.

[197] McLachlan, G., and Peel, D. (2000), *Finite Mixture Models*, NY: Wiley Publications.

[198] Meeus, R. (2002), Supermarktketens zien groeiende markt in buurtwinkels, in: *De Morgen*, donderdag 29 augustus, p.19.

[199] Meilijson, I. (1989), A fast improvement of the EM on its own terms, in: *Journal of the Royal Statistical Society*, B51, pp. 127-138.

[200] Merkle, E. (1981), die Erfassung und Nutzung von Informationen über den Sortimentsverbund in Handelsbetrieben, in: *Schirften zum Marketing*, Vol. 11.

[201] Miller, K.E., and Granzin, K.L. (1979), Simultaneous loyalty and benefit segmentation of retail store customers, in: *Journal of Retailing*, Vol. 55(1), pp. 47-60.

[202] Mitchell, A. (1983), *The Nine American Lifestyles*, New York, Macmillan.

[203] Monshouwer, T., Oosterom, A., and Rovers, J. (1966), Het belang van weloverwogen assortimentsbeheer, in: *Het Levensmiddelenbedrijf*, December, pp. 385-393.

[204] Mooijman, R. (2002), Snelle boodschappen vormen groeiende markt, in: *De Morgen*, woensdag 24 juli, p. 18.

[205] Moore, D.S. (1986), *Tests of chi-squared type. Goodnes-of-fit techniques*, Marcel Dekker, New York, pp. 63-95.

[206] Morishita, S., and Sese, J. (2000), Traversing itemset lattices with statistical metric pruning, in: *Symposium on Principles of Databases Systems*, Dallas, Texas (USA), pp. 226-236.

[207] Morrison, D.G., and Schmittlein, D.C. (1988), Generalizing the NBD model for customer purchases: what are the implications and is it worth the effort? In: *Journal of Business and Economic Statistics*, Vol. 6(April), pp. 145-159.

[208] Moules, J. (1998), House of cards, in: *Information Strategy*, April 1998, pp. 48-49.

[209] Mulhern, F.J., and Leone, R.P. (1991), Implicit price bundling of retail products: a multiproduct approach to maximizing store profitability, in: *Journal of Marketing*, Vol. 55(October), pp. 63-76.

[210] Ng, R.T., Lakshmanan, L.V.S., and Han, J. (1998), Exploratory mining and pruning optimizations of constrained association rules, in: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Seattle, Washington (USA), pp. 13-24.

[211] Olson, J., and Reynolds, R. (1983), Understanding consumer's cognitive structures: implications for advertising and consumer psychology, in: *Advertising and Consumer Psychology*, Lexington Books, Lexington, pp. 77-90.

[212] Ordonez, C., Omiecinski, E., and Ezquerra, N. (2001), A fast algorithm to cluster high dimensional basket data, in: *Proceedings of the IEEE International Conference on Data Mining*, San Jose CA (USA), pp. 633-636.

[213] Özden, B., Ramaswamy, S., and Silberschatz, A. (1998), Cyclic association rules, in: *Proceedings of the Conference on Data Engineering*, Orlando, Florida (USA), pp. 412-421.

[214] Pareto, V. (1909), *Manuel d'economie politique*, Paris.

[215] Park, C.W., Easwer, I., and Smith, D.C. (1989), The effects of situational factors on in-store grocery behavior: the role of store environment and time available for shopping, in: *Journal of Consumer Research*, Vol. 15, pp. 422-433.

[216] Park, J.S., Chen, M-S, and Yu, P.S. (1995), An efficient hash based algorithm for mining association rules, in: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, San Jose CA (USA), pp. 175-186.

[217] Pearson, K. (1894), Contributions to the mathematical theory of evolution, in: *Philosophical Transactions*, Vol. A185, pp. 71-110.

[218] Pessemier, E. (1980), Retail assortments - some theoretical and applied problems, *Technical Report*, Marketing Science Institute Research Program.

[219] Peter, J.P., Olson, J.C., and Grunert, K.G. (1999), *Consumer Behaviour and Marketing Strategy*, McGraw-Hill Publishing Company, London.

[220] Piatetsky-Shapiro, G., and Matheus, C.J. (1994), The interestingness of deviations, in: *Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases*, Seattle, Washington (USA), pp. 25-36.

[221] Plat, F., van Mens, M., and Merkx, M. (1997), *Klantenkaarten en Klantprogramma's*, Interactie Marketing, Kluwer BedrijfsInformatie, Deventer, 190 pg.

[222] Polpinij, J., Jaruskulchai, C., Haddawy, P. (2001), A probabilistic analysis of factors affecting consumer purchase of insurance policies, in: *Proceedings of the Second International Conference on Intelligent Technologies*, Bankok (Thailand), pp. 323-327.

[223] POPAI (2001), Consumer Buying Habits Study.

[224] Popkowski Leszczyc, P.T.L., and Timmermans, H.J.P. (1992), Store switching behavior, in: *Marketing Letters*, Vol. 8(2), pp. 193-204.

[225] Progressive Grocer (1992), *How Consumers Shop*, December.

[226] Ramaswamy, S., Mahajan, S., and Silberschatz, A. (1998), On the discovery of interesting patterns in association rules, in: *Proceedings of the 24th Conference on Very Large Databases (VLDB)*, New York (USA), pp. 368-379.

[227] Reutterer, T. (1998), Competitive market structure and segmentation analysis with self-organizing feature maps, in: *Proceedings of the 27th EMAC Conference*, Stockholm (SE), pp. 85-115.

[228] Reynolds, T., and Gutman, J. (1988), Laddering theory, method, analysis, and interpretation, in: *Journal of Advertising Research*, Vol. 28(1), pp. 11-31.

[229] Robert, C.P. (1996), Mixture of distributions: inference and estimation, in: *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall, pp. 441-464.

[230] Rossi, P.E., and Allenby, G.M. (1993), A Bayesian approach to estimating household parameters, in: *Journal of Marketing Research*, Vol. 30 (May), pp. 171-182.

[231] Rudolph, T.C. (1993), *Positionierungs- und Profilierungsstrategien im Europäischen Einzelhandel*, Band 10 der FAH-Schriftenreihe "Marketing Management", Ph.d. dissertation.

[232] Russell, G., Bell, D., Bodapati, A., Brown, C.L., Chiang, J., Gaeth, G., Gupta, S., and Manchanda, P. (1997), Perspectives on multiple category choice, in: *Marketing Letters*, Vol. 8(3), pp. 297-305.

[233] Russell, G., and Kamakura, W. (1997), Modeling multiple category brand preference with household basket data, in: *Journal of Retailing*, Vol. 73(4), pp. 439-461.

[234] Russell, G., Petersen, A. (2000), Analysis of cross category dependence in market basket selection, in: *Journal of Retailing*, Vol. 76(3), pp. 367-392.

[235] Savasere, A., Omiecienski, E., and Navathe, S. (1995), An efficient algorithm for mining association rules in large databases, in: *Proceedings of the 21st International Conference on Very Large Databases*, Zurich (Switzerland), pp. 432-444.

[236] Savasere, A., Omiecienski, E., and Navathe, S. (1998), Mining for strong negative associations in a large database of customer transactions, in: *Proceedings of the International Conference on Data Engineering*, Orlando, Florida (USA), pp.494-502.

[237] Saygin, Y., Verykios, V.S., and Clifton, C. (2001), Using unknowns to prevent discovery of association rules, in: *SIGMOD Record*, Vol. 30(4), pp. 45-54.

[238] Schultz, H. (1938), *The theory and measurement of demand*, Chicago.

[239] Schwartz, S.H. (1992), Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries, in: *Advances in Experimental Social Psychology*, pp. 1-65.

[240] Schwartz, S.H., and Bilsky, W. (1987), Toward a universal structure of human values, in: *Journal of Personality and Social Psychology*, Vol. 3, pp. 550-562.

[241] Schwarz, G. (1978), Estimating the dimensions of a model, in: *The Annals of Statistics*, Vol. 6, pp. 461-464.

[242] Seetharaman, P.B., Ainslie, A., and Chintagunta, P.K. (1999), Investigating household state dependence effects across categories, in: *Journal of Marketing Research*, Vol. 36 (November), pp. 488-500.

[243] Segal, M.N., and Giacobbe, R.W. (1994), Market segmentation and competitive analysis for supermarket retailing, in: *International Journal of Retail & Distribution Management*, Vol. 22(1), pp. 38-48.

[244] Seidel, W., Mosler, K., and Alker, M. (2000a), A cautionary note on likelihood ratio tests in mixture models, in: *Annals of the Institute of Statistical Mathematics*, Vol. 52, pp. 481-487.

[245] Seidel, W., Mosler, K., and Alker, M. (2000b), Likelihood ratio tests based on subglobal optimisation: a power comparison in exponential mixture models, in: *Statistische Hefte*, Vol. 41, pp. 85-98.

[246] Seidel, W., Sevcikova, H., and Alker, M. (2000c), On the power of different versions of the likelihood ratio test for homogeneity in an exponential mixture model, in: *Diskussionsbeitrage zur Statistik und Quantitativen Ökonomik*, 92-2000, Universität der Bundeswehr Hamburg.

[247] Sichel, H.S. (1982), Repeat-buying and the generalized inverse Gaussian-Poisson distribution, in: *Applied Statistics*, Vol. 31, pp. 193-204.

[248] Silberschatz, A., and Tuzhilin, A. (1996), What makes patterns interesting in knowledge discovery systems, in: *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8(6), pp. 970-974.

[249] Simkin, L. (1996), Tackling barriers to effective implementation of modelling in retail marketing applications, in: *The International Review of Retail, Distribution and Consumer Research*, Vol. 6(3), pp. 225-241.

[250] Smith, W. (1956), Product differentiation and market segmentation as alternative marketing strategies, in: *Journal of Marketing*, Vol. 21, pp. 3-8.

[251] SOPRES (1998), The MOSAIC typology, company document.

[252] Srikant, R., and Agrawal, R. (1996), Mining sequential patterns: generalizations and performance improvements, in: *Proceedings of the 5th International Conference on Extending Database Technology*, Avignon (FR), pp. 3-17.

[253] Srikant, R., and Agrawal, R. (1996), Mining quantitative association rules in large relational tables, in: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Montreal (CA), pp. 1-12.

[254] Srikant, R., and Agrawal, R. (1995), Mining Generalized Association Rules, in: *Proceedings of the 21st International Conference on Very Large Databases*, Zurich (Switzerland), pp. 407-419.

[255] Srikant, R., Vu, Q., and Agrawal, R. (1997), Mining association rules with item constraints, in: *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, Newport Beach CA (USA), pp. 67-73.

[256] Suzuki, E., and Kodratoff, Y. (1998), Discovery of surprising exception rules based on intensity of implication, in : *Principles of Data Mining and Knowledge Discovery*, Volume 1510 of *Lecture Notes in Artificial Intelligence*, Springer, pp. 10-18.

[257] Swinnen, G. (1983), Decisions on Product-mix Changes in Supermarket Chains, *Ph.D. Dissertation*, UFSIA.

[258] Swinyard, W.R. (1993), The effects of mood, involvement and quality of store experience on shopping intentions, in: *Journal of Consumer Research*, Vol. 20, pp. 271-280.

[259] Tan, P-N, and Kumar, V. (2000), Interestingness measures for association patterns: a perspective, *Technical Report TR00-036*, Department of Computer Science, University of Minnesota.

[260] Tan, P-N, Kumar, V., and Srivastava, J. (2000), Indirect association: mining higher order dependencies in data, *Technical Report TR00-037*, Department of Computer Science, University of Minnesota.

[261] Tellis, G.J. (1988), Advertising exposure, loyalty, and brand purchase: a two-stage model of choice, in: *Journal of Marketing Research*, Vol. 15 (May), pp. 134-144.

[262] Ter Hofstede, F., Audenaert, A., Steenkamp, J.B.E.M., and Wedel, M. (1998), An investigation into the assocation pattern technique as a quantitative approach to measuring means-end chains, in: *International Journal of Research in Marketing*, Vol. 15, pp. 37-50.

[263] Ter Hofstede, F., Steenkamp, J.B.E.M., and Wedel, M. (1999), International market segmentation based on consumer-product relations, in: *Journal of Marketing Research*, Vol. 36(February), pp. 1-17.

[264] Titterington, D.M. (1990), Some recent research in the analysis of mixture distributions, in: *Statistics*, Vol. 4, pp. 619-641.

[265] Titterington, D.M., Smith, A.F.M., and Makov, U.E. (1985), *Statistical Analysis of Finite Mixture Distributions*, Wiley Series in Probability and Mathematical Statistics.

[266] Toivonen, H. (1996), Sampling large databases for association rules, in: *Proceedings of the 22nd International Conference on Very Large Databases*, Bombay (India), pp. 432-444.

[267] Toivonen, H., Klemettinen, M., Ronkainen, P., Hätönen, K., and Mannila, H. (1995), Pruning and Grouping Discovered Association Rules, in: *Workshop Notes of the ECML-95 Workshop on Statistics, Machine Learning, and Knowledge Discovery in Databases*, Heraklion (Greece), pp. 47-52.

[268] Tordjman, A. (1994), European retailing: convergences, differences and perspectives, in: *International Journal of Retail & Distribution Management*, Vol. 22(5), pp. 3-19.

[269] Triffin, R. (1940), *Monopolistic Competition and General Equilibrium Theory*, Cambridge, Mass, 4e ed.

[270] Tsionas, E.G. (1999), Bayesian analysis of the multivariate Poisson distribution, in: *Communications in Statistics – Theory and Methods*, Vol. 28(2), pp. 431-451.

[271] Twedt, D.W. (1967), Some practical applications of 'heavy-half' theory, in: *Market Segmentation: Concepts and applications*, New York, Holt/Rinehart and Winston, pp. 265-271.

[272] Uncles, M. (1994), Do you or your customers need a loyalty card scheme?, in: *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 3, pp. 335-350.

[273] Urban, T. (1998), An inventory-theoretic approach to product assortment and shelf-space allocation, in: *Journal of Retailing*, Vol. 74, pp. 15-35.

[274] Van Den Poel, D. (1999), Response Modeling for Database Marketing using Binary Classification, *Ph.D. dissertation 129*, Catholic University of Leuven.

[275] Van der Ster, W., and van Wissen, P. (1993), *Marketing & detailhandel*, Wolters-Noordhoff.

[276] Van Hulle, M. M. (2000), *Faithful Representations and Topographic Maps: From Distortion- to Information-Based Self-Organization*, John Wiley & Sons.

[277] Van Kenhove, P., Van Ossel, G., and Desrumaux, P. (1996), Koopmomenten sturen winkelgedrag, in: *Tijdschrift voor Marketing*, juni 1996, pp. 23-25.

[278] Van Kenhove, P., Van Waterschoot, W., and De Wulf, K. (1999), The impact of task definition on store attribute saliences and store choice, in: *Journal of Retailing*, Spring 1999, pp. 125-137.

[279] Van Vugt, Th. (1995), Liever een koffiezetapparaat dan een weekendje Londen, in: *Adformatie*, Vol. 39.

[280] Vermunt, J.K., and Magidson, J. (2000), Latent class cluster analysis, *research paper*.

[281] Vermunt, J.K., and Magidson, J. (2000), Latent class cluster analysis, chapter 3 in J.A. Hagenaars and A.L. McCutcheon (eds.), *Applied Latent Class Analysis*, Cambridge University Press.

[282] Vermunt, J.K., and Magidson, J. (2000), *Latent GOLD 2.0 User's Guide*, Belmont, MA: Statistical Innovations, Inc.

[283] Viveros, M.S., Nearhos, J.P., and Rothman M.J. (1996), Applying Data Mining Techniques to a Health Insurance Information System, in: *Proceedings of the 22nd VLDB Conference*, Bombay (India), pp. 286-294.

[284] von Neumann, J., and Morgenstern, O. (1947), *The theory of games and economic behavior*, Princeton, N.J., Princeton University Press.

[285] Vriens, M., and Ter Hofstede, F. (2000), Linking attributes, benefits, and consumer values, in: *Marketing Research*, Vol. 12(3), pp. 4-10.

[286] Walters, R.G. (1991), Assessing the impact of retail price promotions on product substitution, complementary purchase, and interstore sales displacement, in: *Journal of Marketing*, Vol. 55(April), pp. 17-28.

[287] Wang, K., Tay, S.H.W., and Liu, B. (1998), Interestingness-based interval merger for numeric association rules, in: *Proceedings of the 4th International Conference on Knowledge Discovery & Data Mining*, New York (USA), pp. 121-127.

[288] Webster, F.E. jr. (1965), The deal-prone consumer, in: *Journal of Marketing Research*, Vol. 2 (May 1965), pp. 186-189.

[289] Wedel, M., and DeSarbo, W.S. (1994), A review of latent class regression models and their applications, in: *Advanced Methods in Marketing Research*, ed. Bagozzi R., Cambridge, MA: Blackwell, pp. 353-388.

[290] Wedel, M., DeSarbo, W.S., Bult, J.R., and Ramaswamy, V. (1993), A latent class poisson regression model for heterogeneous count data, in: *Journal of Applied Econometrics*, Vol. 8, pp. 397-411.

[291] Wedel, M., and Kamakura, W.A. (2000), *Market segmentation: conceptual and methodological foundations*, 2nd ed., Kluwer, Boston.

[292] Wedel, M., Kamakura, W.A., and Böckenholt, U. (2000), Marketing data, models and decisions, in: *International Journal of Research in Marketing*, Vol. 17(2-3), pp. 203-208.

[293] Wilkie, W.L., and Cohen, J.B. (1977), An overview of market segmentation: behavioral concepts and research approaches, *Marketing Science Institute Working Paper*.

[294] Wind, Y. (1978), Issues and advances in segmentation research, in: *Journal of Marketing Research*, Vol. 15 (August), pp. 317-337.

[295] Windham, M.P., and Cutler, A. (1992), Information ratios for validating mixture analyses, in: *Journal of the American Statistical Association*, Vol. 87(420), pp. 1188-1192.

[296] Wu, C.F.J. (1983), On the convergence of the EM Algorithm, in: *Annals of Statistics*, Vol. 11, pp. 95-103.

[297] Zaki, M.J. (2000), Generating non-redundant association rules, in: *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston MA (USA), pp. 34-43.

[298] Zaki, M.J., Parthasarathy, S., Ogihara, M., and Li, W. (1997), New algorithms for fast discovery of association rules, *Technical Report* no. TR651.

[299] Zeithaml, V.A. (1988), Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence, in: *Journal of Marketing*, Vol. 52(July), p. 2-22.

APPENDIX 1

Belgian retail store types

(source ACNielsen Belgium [4])

F1 Stores: Large distribution (short list)

Colruyt N.V

Delhaize "LE LION"

Group Carrefour Belgium : Super GB, Maxi GB and Bigg's

Group Louis Delhaize: Cora & Match

Group Mestdagh: Super M and Champion

F2 Stores: Average-size integrated distribution (short list)

Aldi

Delhaize "Le Lion": Delhaize 2 and Delhaize City

Lidl

Group Louis Delhaize: Profi

Laurus : Battard, Central Cash (all EDA stores closed in 2000)

F2NI Stores: Average-size non-integrated distribution

Alvo

Delhaize "Le Lion": AD Delhaize and Superettes Delhaize Group

Carrefour Belgium: Unic, Nopri, Super GB Partner, Super GB Contact,

Intermarché

Samgo

Laurus: SPAR, Unidis supermarkets

Distrigroup 21: Cash Fresh

Lambrechts: CARAT and SPAR

Other independent supermarkets above 400 m²

F3 Stores:

All self-service stores below 400m² (e.g. Supra, Prima, Louis Delhaize not mentioned above), plus stores with traditional service amongst which night shops.

APPENDIX 2

Using 'interest' to find retail merchandise lines

Instead of using association coefficients as a measure to express similarity between product categories, the interest measure (section 4.6.4) from association rule mining could also be used as an alternative. Therefore, we were interested in finding out whether the 'interest' value produces similar results as association coefficients to discover retail merchandise lines from shopping patterns. Therefore, all possible 2-itemsets and their interest values for the 100 best selling product categories were generated from the same supermarket dataset. The results below show the 10-cluster solution by using Ward's hierarchical clustering (table A.2.1), and two graphs (figure A.2.1 and A.2.2) with the multi-dimensional scaling results (stress=0.034). When comparing these results with those presented in section 3.2.3.1, then it is clear that both ways of measuring interdependency between product categories produce similar results. However, in order to statistically test this hypothesis, we set up a Fisher's exact test by cross-tabulating the clustering results for both solutions (table A.2.2), where the cluster solution based on association coefficients is depicted in the rows and the cluster solution based on interest values is depicted in the columns. The Fisher's exact test [13] is applicable when the statistical conditions for using the standard chi-squared statistic are violated. This happens when there are cells in the contingency table that have expected values below 1, and more than 20% of the cells have expected values below 5, which is clearly the case in this situation.

However, since the Fisher's exact test takes too much time to calculate for this contingency table, we calculated a Monte Carlo estimate for the exact test using one million samples. The test shows no statistical difference on the 1% significance level between the two solutions, which indicates that both methods for measuring assocations (i.e. by using association coefficients or by using 'interest') produces the same results.

Cluster	Product categories	Size
1	Red wine, white wine	2
2	Light beer, heavy beer	2
3	Frozen ready-made meals, frozen meat products, frozen fish, frozen vegetables, frozen soups, icecream, frozen potato products	7
4	Sanitary towels, shampoo, toothpaste, brosserie, shower gel, makeup, beauty	7
5	Paperware, cutlery, electricity, stationery, candles	5
6	Maintenance products, toilet paper & kitchen roll, washing powder, dishwashing, maintenance tools, softener, abrasives, liquid detergent	8
7	Canned meat, spices, speculaas, tea, stockings, canned fish, rice, dried products, oils, gingerbread, frozen pizza, low calory, dry cookies, confectionery, soft drinks, milk, chocolate, sauces, crisps, yoghurt, fresh cookies, canned vegetables, margarine spread, waters, whipped cheese, baking margarine, coffee, soups, pasta, eggs, flour products, desserts, candy bars, grain products, sugar, fruit juice, canned fruit, cheese spread, filter & waste bags, baby food, mayonnaise, biscuit, nuts & appetizer biscuits, hard cheese, broth	46
8	Buns, bake-off products, fresh bread, pastry, pie & biscuit & cake	5
9	Soft cheese, butter, packed bread, cream, salads, prepacked meat	6
10	Catfood, cigarettes, vegetables, fresh vegetables & fruit, fresh meat, bread, sandwich filling, newspaper & magazines, fresh cheese, dogfood, sliced vegetables & fruit, tobacco	12

Table A.2.1: 10-cluster solution using 'interest' as a measure of association

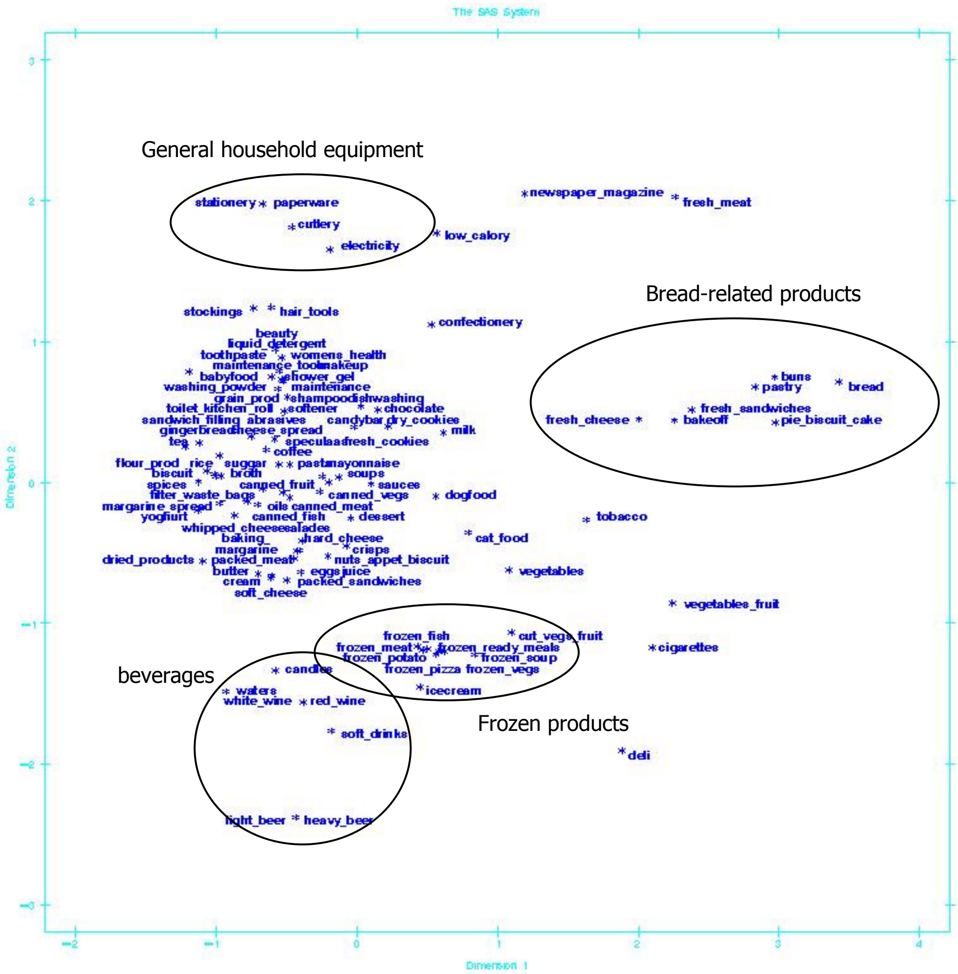


Figure A.2.1: Multi-dimensional scaling results using 'interest' as a measure of association

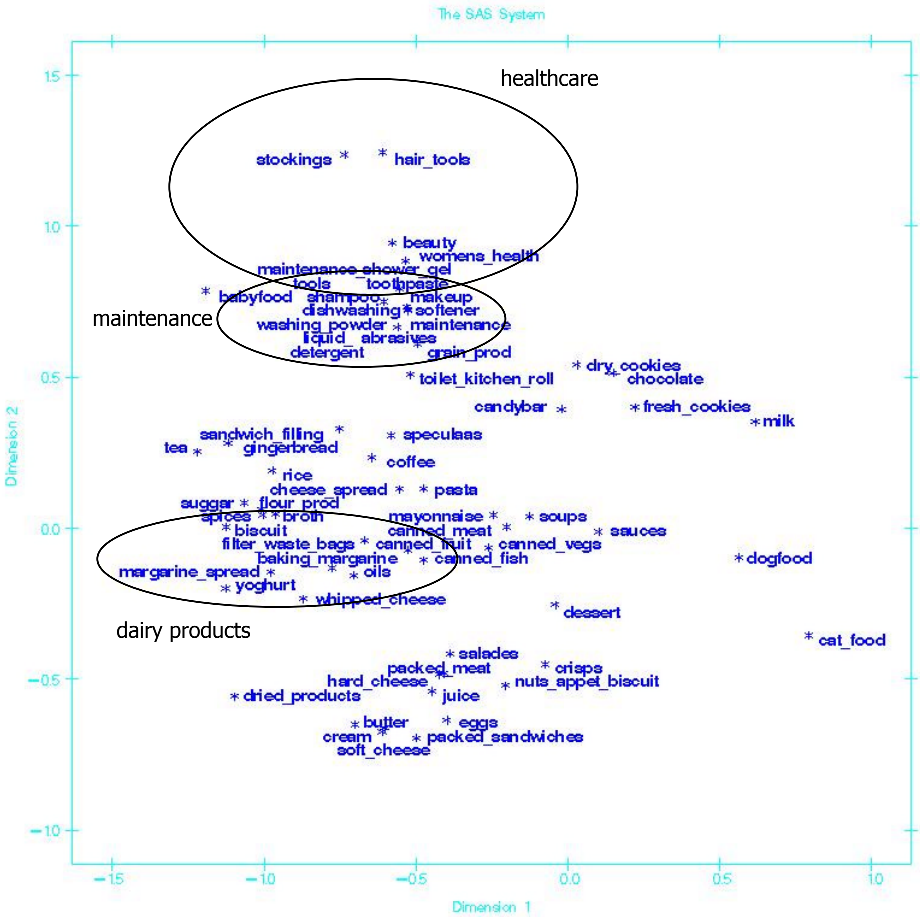


Figure A.2.2: Enlargement of the cluttered points in figure A.2.1.

cluster	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	2
2	0	0	0	0	0	0	0	5	0	2
3	2	0	0	0	0	0	1	0	0	0
4	0	2	0	0	0	0	2	0	0	0
5	0	0	0	0	0	8	0	0	0	0
6	0	0	0	0	0	0	7	0	0	0
7	0	0	7	0	0	0	1	0	0	3
8	0	0	0	0	0	0	35	0	6	0
9	0	0	0	0	5	0	0	0	0	5
10	0	0	0	7	0	0	0	0	0	0

Table A.2.2: Contingency table comparing cluster solutions based on association coefficients (rows) and interest (columns)

APPENDIX 3

Minimum, maximum and selected number of products per category for model 1

CATEGORY NAME	MIN	MAX	SELECTED
AFTER-SHAVE	0	1	0
DISH WASHING	1	2	2
LIGHTERS AND MATCHES	0	1	0
APPETIZER DRINKS	1	3	1
BABY CARE	0	1	0
BATH FOAM	0	1	0
BAKE-OFF PRODUCTS	1	2	2
BAKING MARGARINE	1	2	2
BALSAM	1	3	1
BEAUTY	0	2	0
SANDWICH FILLING	2	4	2
REFRIGERATED READY MEALS	1	3	1
BREAKFAST BISCUITS	1	2	1
BITTER	0	1	0
BROTH	0	2	2
BREAD	2	3	3
BREAD PREPACKED	1	2	1
SANDWICHES PREPACKED	1	2	1
FRESH SANDWICHES	2	3	2
CACAO	1	1	1
CANDY BARS	4	5	4
CHAMPAGNE AND SPARKLING WINE	1	2	1
CRISPS	2	4	4
CHOCOLATE	2	3	2

CATEGORY NAME	MIN	MAX	SELECTED
SWEETS	3	5	3
CANNED VEGETABLES	2	3	2
PASTA	2	4	2
DEODORANT	1	2	1
REFRIGERATED DESSERTS	2	3	2
SHOWER GEL	1	2	1
DRY PREPARED MEALS	0	1	0
DRY BISCUITS	3	5	3
FRAGRANCES	0	1	0
HANDCRAFT BUTTER	1	1	1
EGGS	1	1	1
ELECTRICITY	0	3	1
FARMACEUTICALS	1	2	1
WASTE BAGS DEEP FREEZE BAGS	1	2	1
FILTERS	0	1	0
FRENCH CHEESE	0	1	0
SOFT DRINKS	3	5	5
FRUIT	0	1	0
FRUIT JUICES	2	4	2
PASTRY	1	2	1
DRY PRODUCTS	0	1	0
HAIR GEL	1	2	1
SMOKED/MARINADED FISH	0	1	0
CUT VEGETABLES/FRUIT	0	1	0
GRAIN PRODUCTS	2	3	3
HAIR SPRAY	1	2	1
SETTING LOTION	0	1	0
HARD CHEESE	1	2	1
DUTCH CHEESE	0	1	0

CATEGORY NAME	MIN	MAX	SELECTED
DOG FOOD	1	2	1
INSECTICIDES	0	1	0
CANDLES	0	1	1
CAT FOOD	1	2	1
BABY FOOD	1	2	1
BABY DRINKS	0	1	0
COFFEE	2	3	3
COFFEE BUNS	1	2	1
CANNED FRUIT	1	3	1
CANNED MEALS	0	1	0
CANNED FISH	1	2	1
CANNED MEAT	1	2	2
SPICES	0	2	0
REGULAR BEERS	3	5	3
DIAPERS	1	3	3
WOMEN CARE	2	4	2
MAKE UP	0	2	0
MAYONAISE	2	3	2
FLOUR PRODUCTS	2	3	2
MILK	2	3	3
NUTS AND APPETIZER BISCUITS	3	4	3
OILS	1	3	1
GENERAL MAINTENANCE PRODUCTS	0	2	0
ORIENTAL PRODUCTS	0	1	0
FORCED PRODUCTS	0	2	0
SMOKED/MARINATED FISH	0	1	0
INSTANT COFFEE	1	2	1
PERFUMES	0	1	0
GINGER BREAD	0	1	0

CATEGORY NAME	MIN	MAX	SELECTED
WHIPPED CHEESE	1	3	1
CLEANING PRODUCTS	1	3	1
RICE	2	4	2
CREAM	1	2	1
REFRIGERATED SALADES	0	1	0
SAUCES	1	4	2
SHAVING GEAR	0	2	0
SHAVING SOAP	0	2	0
SHOE POLISH	0	2	0
ABRASIVES	0	2	0
SHAMPOO	1	2	1
SYRUPS	0	1	0
CHEESE SPREAD	1	2	1
MARGARINE SPREAD	2	3	3
SOUPS	2	3	2
SPECULAAS	0	1	1
VAPO STICK	0	1	0
SUGAR	1	2	2
TABLE BEERS	0	1	0
TOOTH BRUSH	1	2	1
TOOTH PASTE	1	2	1
TEA	2	3	2
TOILET SOAP	1	2	1
FRESH ANIMAL FOOD	0	1	0
FRESH BISCUITS	1	3	1
FRESH SOUPS	1	2	1
FRESH FISH	1	1	1
FATS	1	2	1
PIE / BISCUIT / CAKE	1	2	1

CATEGORY NAME	MIN	MAX	SELECTED
LIQUID DETERGENT	2	4	3
READY MADE MEALS	0	1	0
BIRD FOOD	0	1	0
PREPACKED MEAT	1	3	1
WASHING POWDER	2	4	2
WASHING SOFTENER	2	4	2
WATERS	2	3	3
TOILET PAPER AND KITCHEN CLOTH	1	2	2
RED WINE	1	2	1
ROSE WINE	0	1	0
WHITE WINE	1	2	1
YOGURT	2	3	3
SOFT CHEESE	2	3	2
SUN LOTION	0	2	0
HEAVY BEERS	2	4	4

APPENDIX 4

Overview of selected products for model 1

Description	Category	Cat. Position	Overall Position
LVQR (La Vache Qui Rit)	Cheese spread	1	79
Becel essential	Margarine spread	**4**	47
Butter Cerrygold	Handcraft butter	1	205
Grated Emmental private label	Hard cheese	1	254
Stassano cream	Cream	1	578
Yakult	Milk	3	25
Glacé	Pastry	1	110
White bun	Fresh sandwiches	2	1088
Always Ultra Long	Women care	2	75
Always Ultra Normal	Women care	1	30
Appelsientje	Fruit juice	1	68
Backerbsen	Flour products	**3**	303
Baking margarine Becel	Baking margarine	2	65
Baking margarine Solo	Baking margarine	1	5
Batteries AA Duracel	Electricity	1	55
Frying oil Becel	Oils	1	63
Margarine spread Becel	Margarine spread	1	22
Margarine spread Bertoli	Margarine spread	3	36
Prepacked bread Bioform	Bread prepacked	1	257
Bo French bread	Bake-off products	1	44
Bo Kaiser bread	Bake-off products	2	78
Bolognese sauce Manna	Sauces	1	45
Calgon	Cleaning products	1	70
Center wafers LU	Dry biscuits	**5**	134
Chat. Bel Air Bordeaux Red wine	Red wine	1	293
Cha Cha LU	Dry biscuits	1	35
Mushrooms Parador	Canned vegetables	2	141

Paprika crisps Croky	Crisps	4	168
Paprika crisps Smiths	Crisps	2	88
Salty crisps Croky	Crisps	3	124
Salty crisps Smiths	Crisps	1	86
Choco crispies Kellogs	Grain products	3	107
Choco Nutella	Sandwich filling	1	38
Double Lait chocolate C.d'Or	Chocolate	2	71
Milk chocolate C.d'Or	Chocolate	1	51
Chocomousse Jacky	Refrig. desserts	1	135
Coca Cola Regular	Soft drinks	1	1
Coca Cola Light	Soft drinks	2	3
Coral Intens	Liquid detergent	3	84
Crack-a-Nut paprika Duyvis	Nuts/appetizer bisc.	2	375
Cristal Alken	Regular Beers	2	27
Cury sauce Knorr	Sauces	2	97
Actimel Danone	Yogurt	2	42
Dash Futur	Liquid detergent	2	74
Dash Scoops	Washing powder	1	6
Decafe vacuum Douwe Eghberts	Coffee	2	7
Deo spray Dove	Deodorant	1	637
Dessert vacuum Douwe Eghberts	Coffee	1	2
Dixan doses	Washing powder	2	41
Shower cream Dove	Shower gel	1	528
Downy Active Frish	Washing softener	2	266
Dreft compact liquid	Liquid detergent	1	17
Dreft household liquid	Dish washing	2	49
Dreft Ultra dishwashing liquid	Dish washing	1	132
Duvel	Heavy Beers	2	48
Duyvis peanuts	Nuts/appetizer bisc.	1	231
Elastic Hansaplast	Farmaceuticals	1	789
Elnett Hairspray normal	Hair spray	1	250
Fanta Orange	Soft drink	3	21
Coffee Filters Mellita	Filters	1	46

Frolic cow	Dog food	1	195
Frosties Kellogs	Grain products	1	56
Grey shrimps Pieters	Fresh fish	1	447
Grey bread	Bread	1	9
Chocolate confetti Meurisse	Sandwich filling	**4**	176
Sugar lumps Tienen	Sugar	**2**	76
Semi-skimmed milk Inza	Milk	2	11
Semi-skimmed milk private label	Milk	1	10
Imperial Salmon Fancy pink	Canned fish	1	214
Jupiler	Regular Beers	1	8
Candles	Candles	1	14
Kinder Surprise chocolate	Sweets	**4**	346
Chicken broth Knorr	Broth	1	85
Currant cake	Coffee buns	1	122
Leffe blond	Heavy Beers	4	99
Leffe brown	Heavy Beers	3	81
Lenor Ultra Alps Freshness	Washing softener	1	239
Leo pack	Candy bars	1	32
Lipton Ice Tea	Soft drink	4	23
Lipton Ice Tea Light	Soft drink	5	31
M&M's pack	Candy bars	4	101
Mars pack	Candy bars	3	83
Mascarpone cheese	Soft cheese	1	169
Mayonaise Lemon D.L.	Mayonaise	**3**	311
Mayonaise Egg D.L.	Mayonaise	1	18
Flour sugar Graeffe	Sugar	**3**	96
Mentos mint	Sweets	2	321
Mildou vacuum Douwe Eghberts	Coffee	3	69
Minute Maid Orange juice	Fruit juice	2	130
Miracoli Spaghetti Tomato	Pasta	1	61
Multi-grain bread	Bread	3	28
NA shampoo Elseve	Balsam	1	260
Nescafe select extra	Instant coffee	1	182

Nesquick	Cacao	1	216
New wave styling gel	Hair gel	1	896
North cherry peeled Hak	Canned fruit	1	258
Ozo frying fat	Fat	1	162
Palm	Heavy Beers	1	33
Pampers baby dry max	Diapers	1	20
Pampers premium cot.like junior	Diapers	3	57
Pampers premium cot.like maxi	Diapers	2	26
Pancakes cash-fresh	Refrig. Ready meals	1	261
Breadcrumbs Anco	Breakfast biscuits	**3**	384
Pariguette Baguette	Sandwiches prep.	1	173
Petit Gervais Strawberry	Whipped cheese	1	39
Philadelphia Light	Soft cheese	2	385
Pickwick teabags	Tea	1	100
Red port	Appetizer drinks	1	16
Rice Bosto	Rice	1	237
Rice pudding Manna	Rice	2	774
Rice pudding Jacky	Refrig. Desserts	2	219
Rice pie	Pie/biscuit/cake	1	334
Rocher	Sweets	1	294
Samson Bubbles	Champ./spark.wine	**6**	1100
Sandwiches	Fresh sandwiches	1	12
Shampoo Pantene normal	Shampoo	1	703
Soave classico pasqua	White wine	1	326
Spaghetti extra fine Anco	Pasta	2	377
Special K Kellogs	Grain products	2	72
Speculaas Lotus	Speculaas	1	98
Bacon Natura Herta	Prepacked meat	1	139
Stella Artois	Regular Beers	3	234
Tooth brush aquafr flex cont	Toothbrush	**2**	1409
Tea time Delacre	Dry biscuits	2	53
Tea Y-label Lipton	Tea	2	145
Toilet soap Sunlight	Toilet soap	1	561

Tomato concentrate Elvea	Canned vegetables	1	138
Tom. soup with meat balls priv.l	Fresh Soups	1	1449
Tomato ketchup Heinz	Sauces	3	116
Toothpaste Signal Ultra protect.	Toothpaste	1	429
TUC mixed	Nuts/appetizer bisc	3	524
TV sausages Zwan	Canned meat	1	19
Twix pack	Candy bar	2	82
Unox tomato cream soup	Soup	1	382
Unox tomato soup with meat bal	Soup	2	438
Fresh Eggs	Eggs	1	24
Vitabis	Baby food	1	87
Meat broth Knorr	Broth	2	94
Wafels Eigenbak	Fresh biscuits	1	60
Water Still Evian	Water	**4**	67
Water Spa sparkling	Water	2	13
Water Spa still	Water	1	4
Toilet paper private label	Toilet paper	2	121
Toilet paper white isis	Toilet paper	1	62
Vienna sausages Zwan	Canned meat	2	37
Whiskas cocktail	Cat food	1	919
White bread	Bread	2	15
Yogurt light panache Danone	Yogurt	3	43
Yogurt fruit Vitalinea	Yogurt	1	40
Self-raising flour Anco	Flour products	**5**	541

APPENDIX 5

Overview of selected products for model 2

Description	Category	Overall Position
La Vache qui rit	Cheese spread	79
Alpro Soya Bake and Fry	Margarine spread	190
Becel essential	Margarine spread	47
Fruit yogurt Lelie	Yogurt	59
Yakult	Milk	25
Glace	pastry	110
Alpro soya minarine	Margarine spread	54
Always Ultra Long	Women care	75
Always Ultra Normal	Women care	30
Appelsientje	Fruit juice	68
Baby tissues Pampers	Baby care	102
Baking margarine Becel	Baking margarine	65
Baking margarine Bertoli	Baking margarine	90
Baking margarine Fama	Baking margarine	151
Baking margarine Solo	Baking margarine	5
Batteries AA Duracel	Electricity	55
Frying oil Becel	Oils	63
Margarine spread healthy Becel	Margarine spread	22
Margarine spread Bertoli	Margarine spread	36
Bifi sausage	Canned meat	103
Biotex blue	Washing powder	118
Bo bread	Bake-off products	92
Bo French bread	Bake-off products	44
Bo Kaiser bread	Bake-off products	78
Bolognese sauce Manna	Sauces	45
Bounty pack	Candy bars	215
Calgon	Cleaning products	70
Canderel tablets	Sugar	64

Cecemel chocolate milk	Milk	131
Center wafers LU	Dry biscuits	134
Cha cha LU	Dry biscuits	35
Mushrooms Parador	Canned vegetables	141
Paprika crisps Croky	Crisps	168
Paprika crisps Smiths	Crisps	88
Salty crisps Croky	Crisps	124
Salty crisps Smiths	Crisps	86
Choco crispies Kellogs	Grain products	107
Choco Nutella	Sandwich filling	38
Double Lait chocolate C.d'Or	Chocolate	71
Milk chocolate C.d'Or	Chocolate	51
Chocomousse Jacky	Refrig. Desserts	135
Chocos Kellogs	Grain products	133
Coca cola regular	Soft drinks	1
Coca cola light	Soft drinks	3
Coral intens	Liquid detergent	84
Cristal Alken	Regular Beers	27
Cury sauce Knorr	Sauces	97
Actimel Danone	Yogurt	42
Extra light cheese Danone	Whipped cheese	112
Light Yogurt Danone	Yogurt	58
Dash Futur	Liquid detergent	74
Dash scoops	Washing powder	6
Decafe vacuum Douwe Eghberts	Coffee	7
Dessert vacuum Douwe Eghberts	Coffee	2
Dixan doses	Washing powder	41
Dreft compact liquid	Liquid detergent	17
Dreft household liquid	Dish washing	49
Dreft Ultra dishwashing liquid	Dish washing	132
Duvel	Heavy Beers	48
Effi Minarine	Margarine spread	105
Fanta orange	Soft drinks	21

Coffee filters Melitta	Filters	46
Fristi	Milk	93
Frosties Kellogs	Grain products	56
Grey bread	Bread	9
Chocolate confetti Kwatta	Sandwich filling	123
Chocolate confetti Meurisse	Sandwich filling	176
Sugar lumps Tienen	Sugar	76
Herring fillets Korenbloem	Refrigerated salads	140
Semi-skimmed milk Inza	Milk	11
Semi-skimmed milk private label	Milk	10
Jupiler	Regular Beers	8
Candles	Candles	14
Chicken broth Knorr	Broth	85
Coffee arome Douwe Eghberts	Coffee	119
Crystallized sugar Grand pont	Sugar	148
Leffe blond	Heavy Beers	99
Leffe brown	Heavy Beers	81
Leo pack	Candy bars	32
Lipton ice tea	Soft drinks	23
Lipton ice tea light	Soft drinks	31
M&M's pack	Candy bars	101
Skimmed milk Inza	Milk	126
Mais eggs	Eggs	106
Mars pack	Candy bars	83
Mayonaise egg D.L.	Mayonaise	18
Flour sugar Graeffe	Sugar	96
Mildou vacuum Douwe Eghberts	Coffee	69
Milky Way pack	Candy bars	194
Minelma	Margarine spread	104
Minute Maid orange juice	Fruit juice	130
Miracoli spaghetti tomato	Pasta	61
Multi-grain bread	Bread	28
Ozo frying fat	Fat	162

Palm	Heavy Beers	33
Pampers baby dry junior	Diapers	77
Pampers by dry maxi	Diapers	20
Pampers premium cot.like junior	Diapers	57
Pampers premium cot.like maxi	Diapers	26
Pepsi max	Soft drink	50
Pepsi regular	Soft drink	115
Petit Gervais strawberry	Whipped cheese	39
Pick up Bahlsen	Dry biscuits	125
Pickwick tea bags	Tea	100
Planta	Margarine spread	34
Pokemon energy wafel	Dry biscuits	109
Red port	Appetizer drinks	16
Grinded coffee Fort	Coffee	108
Roda	Margarine spread	181
Rye-bread	Bread	164
Raisin bread	Bread	120
Sandwiches	Fresh sandwiches	12
Free-range eggs	Eggs	80
Napkins	Waste/deep freeze bag	200
Sherry dry	Appetizer drinks	111
Orange juice private label	Fruit juice	154
Snickers pack	Candy bars	179
Spa lemonade Lemon	Soft drink	198
Spa lemonade orange	Soft drink	156
Special K Kellogs	Grain products	72
Spekulaas Lotus	Spekulaas	98
Sprite	Soft drink	52
Tea time Delacre	Dry biscuits	53
Tomato concentrate Elvea	Canned vegetables	138
Tomato ketchup Heinz	Sauces	116
TV sausages Zwan	Canned meat	19
Twix pack	Candy bars	82

Fresh eggs	Eggs	24
Vitabis	Baby food	87
Vitelma healthy spread	Margarine spread	89
Meat broth Knorr	Broth	94
Whole-meal bread	Bread	117
Whole milk Inza	Milk	29
Wafels Eigenbak	Fresh biscuits	60
Water still Contrex	Water	114
Water still Evian	Water	67
Water sparkling private label	Water	113
Water still private label	Water	66
Water sparkling Spa	Water	13
Water still Spa	Water	4
Water still Vittel	Water	73
Toilet paper private label	Toilet paper	121
Toilet paper white Isis	Toilet paper	62
Vienna sausages Zwan	Canned meat	37
White bread	Bread	15
Yogurt Light Panache	Yogurt	43
Yogurt fruit Vitalinea	Yogurt	40
Bran bread	Bread	91

APPENDIX 6

Overview of selected products for model 3

Description	Category	Overall position
La Vache qui rit	Cheese spread	79
Becel essential	Margarine spread	47
Becel pro-activ	Margarine spread	143
Yogurt danone light	Yogurt	153
Yogurt Lelie fruit	Yogurt	59
Video Walt Disney: Tarzan	Electricity	157
Yakult	Milk	25
Toilet paper Scottex	Toilet paper	174
Glace	Pastry	110
Sausage-roll	Pastry	167
Alpro soya minarine	Margarine spread	54
Always Normal	Women care	152
Always Ultra long	Women care	75
Always Ultra Normal	Women care	30
Antikal	Cleaning products	153
Appelsientje	Fruit juice	68
Baby tissues pampers	Baby care	102
Baking margarine Becel	Baking margarine	65
Baking margarine Bertoli	Baking margarine	90
Baking margarine Fama	Baking margarine	151
Baking margarine Solo	Baking margarine	5
Batteries cigarette 1.5 V Duracell	Electricity	127
Batteries AA Duracel	Electricity	55
Frying oil Becel	Oils	63
Margarine spread healthy Becel	Margarine spread	22
Margarine spread Bertoli	Margarine spread	36
Biotex blue	Washing powder	118
Bo bread	Bake-off products	92

Bo French bread	Bake-off products	44
Bo Kaiser bread	Bake-off products	78
Bolognese sauce Manna	Sauces	45
Bonux scoops	Washin powder	149
Calgon	Cleaning products	70
Canderel tablets	Sugar	64
Center wafers	Dry biscuits	134
Cha Cha LU	Dry biscuits	35
Mushrooms Parador	Canned vegetables	141
Paprika crisps Smiths	Crisps	88
Salty crisps Croky	Crisps	124
Salty crisps Smiths	Crisps	86
Choco Boerinneke	Sandwich filling	159
Choco crispies Kellogs	Breakfast cereals	107
Choco Nutella	Sandwich filling	38
Double Lait chocolate C.d'Or	Chocolate	71
Milk chocolate c.d'Or	Chocolate	51
Milk chocolate St-Nicolas	Chocolate	161
Milk chocolate Zero	Chocolate	236
Coca Cola Regular	Soft drinks	1
Coca Cola Light	Soft drinks	3
Strawberry jam Materne	Sandwich filling	218
Coral Intens	Liquid detergent	84
Cristal Alken	Regular Beers	27
Cury sauce Knorr	Sauces	97
Daycreme Nivea	Beauty	150
Actimel Danone	Yogurt	42
Extra light cheese Danone	Whipped cheese	112
Light yogurt Danone	Yogurt	58
Dash Compact	Washing powder	171
Dash Futur	Liquid detergent	74
Dash Scoops	Washing powder	6
Decafe vacuum Douwe Eghberts	Coffee	7

Dessert vacuum private label	Coffee	129
Dessert vacuum Douwe Eghberts	Coffee	2
Dixan doses	Washing powder	41
Dixan gel	Liquid detergent	147
Dreft compact liquid	Liquid detergent	17
Dreft household liquid	Dish washing	49
Dreft Ultra dishwashing liquid	Dish washing	132
Duvel	Heavy Beers	48
Fanta Orange	Soft drink	21
Coffee Filters Mellita	Filters	46
Fristi	Milk	93
Frying oil VDM	Oils	128
Frosties Kellogs	Breakfast cereals	56
Grany bilberry biscuit LU	Dry biscuits	172
Grey bread	Bread	9
Chocolate confetti Kwatta	Sandwich filling	123
Sugar lumps Tienen	Sugar	76
Herring filets Korenbloem	Refrigerated salads	140
Semi-skimmed milk Inza	Milk	11
Semi-skimmed milk private label	Milk	10
Jupiler	Regular Beers	8
Candles	Candles	14
Chicken broth Knorr	Broth	85
Coffee Arome Douwe Eghberts	Coffee	119
Currant cake	Ginger bread	122
Leffe blond	Heavy Beers	99
Leffe brown	Heavy Beers	81
Leo pack	Candy bars	32
Lipton Ice Tea	Soft drink	23
Lipton Ice Tea light	Soft drink	31
M&M's pack	Candy bars	101
Mais eggs	Eggs	106
Mars pack	Candy bars	83

Mascarpone cheese	Soft cheese	169
Mayonaise Egg D.L.	Mayonaise	18
Flour sugar Graeffe	Sugar	96
Mildou vacuum Douwe Eghberts	Coffee	69
Minelma	Margarine spread	104
Minute Maid Orange juice	Fruit juice	130
Miracoli Spaghetti Tomato	Pasta	61
Multi-grain bread	Bread	28
Palm	Heavy Beers	33
Pampers baby dry junior	Diapers	77
Pampers baby dry maxi	Diapers	20
Pampers baby dry midi	Diapers	160
Pampers premium cot.like junior	Diapers	57
Pampers premium cot.like maxi	Diapers	26
Pariguette Baguette	Sandwiches prepared	173
Pepsi Max	Soft drink	50
Pepsi Regular	Soft drink	115
Petit Gervais Strawberry	Whipped cheese	39
Pick up Bahlsen	Dry biscuits	125
Pickwick teabags	Tea	100
Planta	Margarine spread	34
Pokemon energy wafel LU	Dry biscuits	109
Red port	Appetizer drinks	16
White port	Appetizer drinks	142
Grinded Coffee Fort	Coffee	108
Sandwiches	Fresh sandwiches	12
Free-range eggs	Eggs	80
Sherry dry	Appetizer drinks	111
Soup-greens private label	Cut vegetables/fruit	136
Special K Kellogs	Breakfast cereals	72
Bacon Natura Herta	Prepacked meat	139
Sprite	Soft drink	52
Tea time Delacre	Dry biscuits	53

Tea Y-label Lipton	Tea	145
TV sausages Zwan	Canned meat	19
Applejuice Varesa	Fruit juice	137
Fresh Eggs	Eggs	24
Vitabis	Baby food	87
Whole-meal bread	Bread	117
Whole milk Inza	Milk	29
Wafels eigenbak	Fresh biscuits	60
Washing powder Coral	Washing powder	170
Water still Contrex	Water	114
Water still Evian	Water	67
Water still private label	Water	66
Water sprankling Spa	Water	13
Water still Spa	Water	4
Water still Vittel	Water	73
Toilet paper private label	Toilet paper	121
Toilet paper white Isis	Toilet paper	62
Vienna sausages Zwan	Canned meat	37
Westmalle Tripel	Heavy Beers	166
White bread	Bread	15
Yogurt light panache Danone	Yogurt	43
Yogurt fruit Vitalinea	Yogurt	40
Bran bread	Bread	91

APPENDIX 7

Overview of selected products for model 4

Description	Category	Overall position
La Vache qui rit	Cheese spread	79
Becel essential	Margarine spread	47
Becel pro-activ	Margarine spread	143
Yogurt danone light	Yogurt	**153**
Yogurt Lelie fruit	Yogurt	59
Video Walt Disney: Tarzan	Electricity	**157**
Yakult	Milk	25
Toilet paper Scottex	Toilet paper	**174**
Apple bun	Pastry	**287**
Glace	Pastry	110
Alpro soya minarine	Margarine spread	54
Always Normal	Women care	**152**
Always Ultra long	Women care	75
Always Ultra Normal	Women care	30
Antikal	Cleaning products	**153**
Appelsientje	Fruit juice	68
Baby tissues pampers	Baby care	102
Baking margarine Becel	Baking margarine	65
Baking margarine Solo	Baking margarine	5
Batteries cigarette 1.5 V Duracell	Electricity	127
Batteries AA Duracel	Electricity	55
Frying oil Becel	Oils	63
Margarine spread healthy Becel	Margarine spread	22
Margarine spread Bertoli	Margarine spread	36
Bifi sausage	Canned meat	103
Biotex blue	Washing powder	118
Bo French bread	Bake-off products	44
Bo Kaiser bread	Bake-off products	78

Bolognese sauce Manna	Sauces	45
Bonux scoops	Washing powder	149
Calgon	Cleaning products	70
Canderel tablets	Sugar	64
Cecemel chocolate milk	Milk	131
Center wafers LU	Dry biscuits	134
Cha Cha LU	Dry biscuits	35
Mushrooms Parador	Canned vegetables	141
Salty crisps Croky	Crisps	124
Salty crisps Smiths	Crisps	86
Choco crispies Kellogs	Breakfast cereals	107
Choco kr.&milk bar Kellogs	Grain products	**270**
Choco Nutella	Sandwich filling	38
Double Lait chocolate C.d'Or	Chocolate	71
Milk chocolate c.d'Or	Chocolate	51
Chocomousse Jacky	Refrig. Desserts	135
Chocos Kellogs	Breakfast cereals	133
Coca Cola Regular	Soft drinks	1
Coca Cola Light	Soft drinks	3
Coral Intens	Liquid detergent	84
Cristal Alken	Regular Beers	27
Cury sauce Knorr	Sauces	97
Daycreme Nivea	Beauty	150
Actimel Danone	Yogurt	42
Extra light cheese Danone	Cheese	112
Light yogurt Danone	Yogurt	58
Dash Compact	Washing powder	**171**
Dash Futur	Liquid detergent	74
Dash Scoops	Washing powder	6
Decafe vacuum Douwe Eghberts	Coffee	7
Dessert vacuum private label	Coffee	129
Dessert vacuum Douwe Eghberts	Coffee	2
Dixan doses	Washing powder	41

Dixan gel	Liquid detergent	147
Dreft compact liquid	Liquid detergent	17
Dreft household liquid	Dish washing	49
Dreft Ultra dishwashing liquid	Dish washing	132
Duvel	Heavy Beers	48
Effi Minarine	Margarine spread	105
Fanta Orange	Soft drink	21
Coffee Filters Mellita	Filters	46
Fristi	Milk	93
Frying oil VDM	Oils	128
Frosties Kellogs	Breakfast cereals	56
Grany bilberry biscuit LU	Dry biscuits	**172**
Grey bread	Bread	9
Semi-skimmed milk Inza	Milk	11
Semi-skimmed milk private label	Milk	10
Jupiler	Regular Beers	8
Candles	Candles	14
Chicken broth Knorr	Broth	85
Coffee Arome Douwe Eghberts	Coffee	119
Currant cake	Ginger bread	122
Leffe blond	Heavy Beers	99
Leffe brown	Heavy Beers	81
Leo pack	Candy bars	32
Lipton Ice Tea	Soft drink	23
Lipton Ice Tea light	Soft drink	31
M&M's pack	Candy bars	101
Mais eggs	Eggs	106
Mars pack	Candy bars	83
Mascarpone cheese	Soft cheese	**169**
Mayonaise Egg D.L.	Mayonaise	18
Flour sugar Graeffe	Sugar	96
Mildou vacuum Douwe Eghberts	Coffee	69
Minelma	Margarine spread	104

Minute Maid Orange juice	Fruit juice	130
Miracoli Spaghetti Tomato	Pasta	61
Multi-grain bread	Bread	28
Ozo frying fat	Fat	**162**
Palm	Heavy Beers	33
Pampers baby dry junior	Diapers	77
Pampers baby dry maxi	Diapers	20
Pampers baby dry midi	Diapers	**160**
Pampers premium cot.like junior	Diapers	57
Pampers premium cot.like maxi	Diapers	26
Pariguette Baguette	Sandwiches prepared	**173**
Pepsi Max	Soft drink	50
Pepsi Regular	Soft drink	115
Petit Gervais Strawberry	Whipped cheese	39
Pick up Bahlsen	Dry biscuits	125
Pickwick teabags	Tea	100
Planta	Margarine spread	34
Pokemon energy wafel LU	Dry biscuits	109
Red port	Appetizer drinks	16
Grinded Coffee Fort	Coffee	108
Sandwiches	Fresh sandwiches	12
Free-range eggs	Eggs	80
Sherry dry	Appetizer drinks	111
Orange juice private label	Fruit juice	**154**
Special K Kellogs	Breakfast cereals	72
Bacon Natura Herta	Prepacked meat	139
Sprite	Soft drink	52
Tea time Delacre	Dry biscuits	53
Tea Y-label Lipton	Tea	145
Tomate ketchup Heinz	Sauces	116
TV sausages Zwan	Canned meat	19
Twix pack	Candy bars	82
Applejuice Varesa	Fruit juice	137

Fresh Eggs	Eggs	24
Vitabis	Baby food	87
Vitelma healthy spread	Margarine spread	89
Meat broth Knorr	Broth	94
Whole-meal bread	Bread	117
Whole milk Inza	Milk	29
Wafels eigenbak	Fresh biscuits	60
Washing powder Coral	Washing powder	**170**
Water still Chaudfontaine	Water	**345**
Water still Contrex	Water	114
Water still Evian	Water	67
Water still private label	Water	66
Water sprankling Spa	Water	13
Water still Spa	Water	4
Water still Vittel	Water	73
Toilet paper private label	Toilet paper	121
Toilet paper white Isis	Toilet paper	62
Vienna sausages Zwan	Canned meat	37
Westmalle Tripel	Heavy Beers	**166**
White bread	Bread	15
Yogurt light panache Danone	Yogurt	43
Yogurt fruit Vitalinea	Yogurt	40
Bran bread	Bread	91

APPENDIX 8

Selected products for model 2 based on fill-in baskets

Description	**Category**	**Overall Position**
La Vache qui rit	Cheese spread	93
Becel essential	Margarine spread	66
Fruit yogurt Lelie	Yogurt	46
Video Walt Disney: Tarzan	Electricity	70
Yakult	Milk	27
Toilet paper Scottex	Toilet paper/kitchen roll	151
Apple bun	Pastry	68
Chocolate bun	Pastry	98
Glace	Pastry	56
Sausage roll	Pastry	96
Dubble sausage roll	Pastry	36
Alpro soya minarine	Margarine spread	71
Always Ultra Long	Women care	62
Always Ultra Normal	Women care	45
Apple compote Materne	Canned fruit	203
Apple muslin Materne	Canned fruit	253
Appelsientje	Fruit juice	155
Baby tissues Pampers	Baby care	139
Baking margarine Becel	Baking margarine	86
Baking margarine Bertoli	Baking margarine	156
Baking margarine Solo	Baking margarine	7
Batteries cigarette Duracel	Electricity	116
Batteries AA Duracel	Electricity	48
Frying oil Becel	Oils	77
Margarine spread healthy Becel	Margarine spread	31
Margarine spread Bertoli	Margarine spread	37
Biscuit 3 fruit	Pie/biscuit/cake	113
Bo bread	Bake-off products	64
Bo French bread	Bake-off products	23

Bo Kaiser bread	Bake-off products	54
Bolognese sauce Manna	Sauces	50
Calgon	Cleaning products	75
Canderel tablets	Sugar	92
Cha cha LU	Dry biscuits	65
Paprika crisps Smiths	Crisps	182
Salty crisps Smiths	Crisps	158
Choco Nutella	Sandwich filling	79
Chocolate milk Inza	Milk	97
Double Lait chocolate C.d'Or	Chocolate	90
Milk chocolate C.d'Or	Chocolate	73
Coca cola regular	Soft drinks	1
Coca cola light	Soft drinks	2
Coral intens	Liquid detergent	129
Cristal Alken	Regular Beers	13
Cury sauce Knorr	Sauces	128
Daycreme nivea	Beauty	118
Actimel Danone	Yogurt	63
Extra light cheese Danone	Whipped cheese	134
Light Yogurt Danone	Yogurt	89
Dash Futur	Liquid detergent	109
Dash scoops	Washing powder	12
Decafe vacuum Douwe Eghberts	Coffee	15
Dessert vacuum private label	Coffee	83
Dessert vacuum Douwe Eghberts	Coffee	3
Dixan doses	Washing powder	60
Dreft compact liquid	Liquid detergent	26
Dreft household liquid	Dish washing	58
Dreft Ultra dishwashing liquid	Dish washing	117
Duvel	Heavy Beers	22
Effi Minarine	Margarine spread	167
Fanta orange	Soft drinks	28
Coffee filters Melitta	Filters	53

Frying oil VDM	Oils	130
Frosties Kellogs	Grain products	87
Fruit basket	Fruit	100
Grimbergen dubbel	Heavy Beers	152
Grey bread	Bread	5
Chocolate confetti Kwatta	Sandwich filling	130
Sugar lumps Tienen	Sugar	120
Herring fillets Korenbloem	Refrigerated salads	165
Semi-skimmed milk Inza	Milk	17
Semi-skimmed milk private label	Milk	16
Jonge Bols brandy	Spirits	76
Jupiler	Regular Beers	6
Candles	Candles	10
Kasteelbier	Heavy beers	122
Chicken broth Knorr	Broth	123
Coffee buns	Pastry	82
Currant bun	Pastry	80
Crystallized sugar Grand pont	Sugar	127
Leffe blond	Heavy Beers	72
Leffe brown	Heavy Beers	47
Leo pack	Candy bars	52
Lipton ice tea	Soft drinks	30
Lipton ice tea light	Soft drinks	59
M&M's pack	Candy bars	146
Mais eggs	Eggs	125
Mayonaise egg D.L.	Mayonaise	21
Flour sugar Graeffe	Sugar	110
Mildou vacuum Douwe Eghberts	Coffee	104
Minelma	Margarine spread	163
Miracoli spaghetti tomato	Pasta	115
Multi-grain bread	Bread	11
Nescafe select extra	Coffee	150
Ozo frying fat	Fat	126

Palm	Heavy Beers	18
Pampers baby dry maxi	Diapers	19
Pampers baby dry midi	Diapers	101
Pampers premium cot.like junior	Diapers	32
Pampers premium cot.like maxi	Diapers	39
Pepsi max	Soft drink	24
Pepsi regular	Soft drink	132
Petit Gervais strawberry	Whipped cheese	99
Pickwick tea bags	Tea	103
Planta	Margarine spread	34
Pokemon energy wafel	Dry biscuits	131
Red port	Appetizer drinks	20
Grinded coffee Fort	Coffee	57
Red bull	Soft drinks	102
Rye-bread	Bread	74
Raisin bread	Bread	112
Sandwiches	Fresh sandwiches	9
Free-range eggs	Eggs	119
Sherry dry	Appetizer drinks	94
Soave Classico Pasqua	White wine	138
Soup vegetables private label	Cut vegetables/fruit	140
Spa lemonade Lemon	Soft drink	137
Spa lemonade orange	Soft drink	108
Special K Kellogs	Grain products	121
Spekulaas Lotus	Spekulaas	147
Sprite	Soft drink	61
Stella Artois	Regular beer	88
Tea time Delacre	Dry biscuits	29
Tea Y-label Lipton	Tea	148
TV sausages Zwan	Canned meat	38
Apple juice Varesa	Fruit juice	42
Fresh eggs	Eggs	25
Vitabis	Baby food	106

Vitelma healthy spread	Margarine spread	145
Meat broth Knorr	Broth	124
Whole-meal bread	Bread	51
Whole milk Inza	Milk	35
Wafels Eigenbak	Fresh biscuits	67
Water still Chaudfontaine	Water	143
Water still Contrex	Water	142
Water still Evian	Water	95
Water sparkling private label	Water	107
Water still private label	Water	41
Water sparkling Spa	Water	14
Water still Spa	Water	4
Water still Vittel	Water	85
Toilet paper white Isis	Toilet paper	111
Vienna sausages Zwan	Canned meat	55
Westmalle dubbel	Heavy beer	91
Westmalle tripel	Heavy beer	43
White bread	Bread	8
Yogurt Light Panache	Yogurt	84
Yogurt fruit Vitalinea	Yogurt	81
Bran bread	Bread	49
Zip fire starter blocks	maintenance	149

APPENDIX 9

Results for the two bivariate Poisson mixture models

The first bivariate Poisson mixture model describes the bivariate interaction between cakemix and frosting. The computational results with respect to the likelihoods, the information criteria and the optimal parameter values for the 2 (BIC/CAIC) and 3 (AIC) component solutions are given in figure A.9.1 and tables A.9.1 and A.9.2 below.

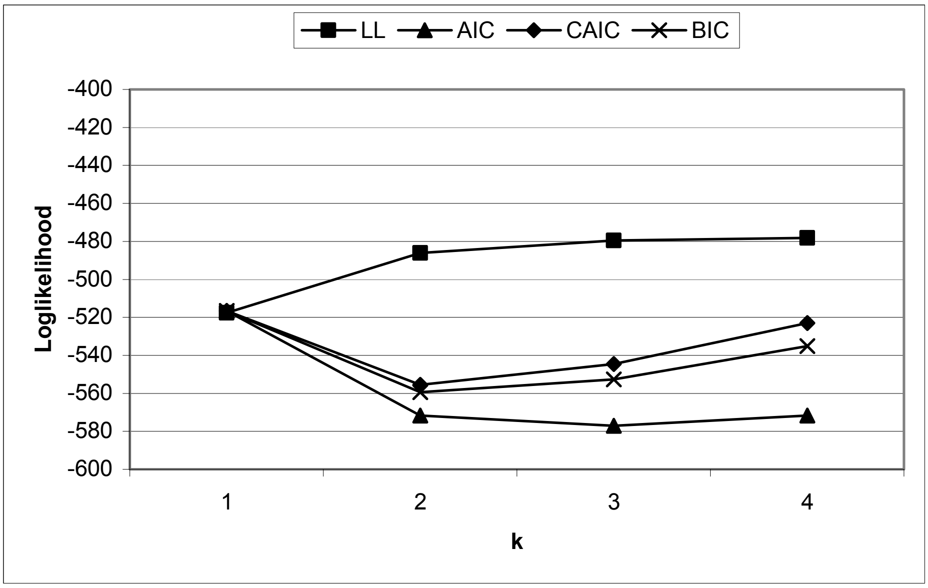


Figure A.9.1: Loglikelihood, AIC, CAIC and BIC against the number of components for the bivariate cakemix-frosting Poisson mixture model

	Parameters			
Cluster	λ_C	λ_F	λ_{CF}	p
1	0.614	0.337	0.985	0.914
2	6.155	2.952	1.009	0.086

Table A.9.1: Estimated parameters for the 2-components bivariate cakemix-frosting mixture model

	Parameters			
Cluster	λ_C	λ_F	λ_{CF}	p
1	0.402	0.287	1.132	0.819
2	7.862	4.656	0.000	0.068
3	2.548	0.630	0.000	0.113

Table A.9.2: Estimated parameters for the 3-components bivariate cakemix-frosting mixture model

The second bivariate Poisson mixture model describes the bivariate interaction between fabric detergent and softener. The computational results with respect to the likelihoods, the information criteria and the optimal parameter values for the 2 (CAIC/BIC) and 4 (AIC) component solutions are given in figure A.9.2 and tables A.9.3 and A.9.4 below.

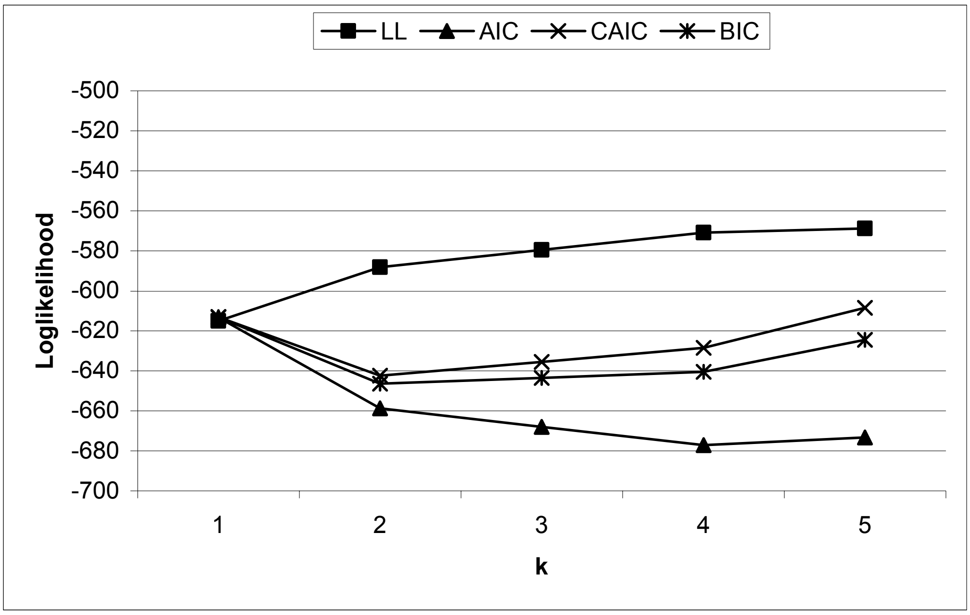


Figure A.9.2: Loglikelihood, AIC, CAIC and BIC against the number of components for the bivariate detergent-softener Poisson mixture model

	Parameters			
Cluster	λ_D	λ_S	λ_{DS}	p
1	1.406	0.836	1.095	0.874
2	7.489	3.864	0.200	0.126

Table A.9.3: Estimated parameters for the 2-components bivariate detergent-softener mixture model

	Parameters			
Cluster	λ_D	λ_S	λ_{DS}	p
1	0.000	1.150	0.797	0.214
2	5.760	0.669	1.374	0.103
3	8.824	9.367	0.000	0.024
4	1.133	0.089	1.952	0.659

Table A.9.4: Estimated parameters for the 4-components bivariate detergent-softener mixture model

When allocating the 155 households over the different clusters for each bivariate solution with k=2 components, the following cross-tabulation is obtained (see table A.9.5).

		Cakemix-frosting		
		1	2	Total
Detergent - softener	1	123	9	132
	2	21	2	23
	Total	144	11	155

Table A.9.5: Cross tabulation of clustering results for two bivariate mixture models for k=2 components.